Proceedings of ACM SIGGRAPH Applied Perception in Graphics and Visualization (APGV 2010), pages 149–156.

Depth Judgment Measures and Occluding Surfaces in Near-Field Augmented Reality

Gurjot Singh* J. Edward Swan II* J. Adam Jones* Stephen R. Ellis†
*Mississippi State University †NASA Ames Research Center

Abstract

In this paper we describe an apparatus and experiment that measured depth judgments in augmented reality at near-field distances of 34 to 50 centimeters. The experiment compared perceptual matching, a closed-loop task for measuring depth judgments, with blind reaching, a visually open-loop task for measuring depth judgments. The experiment also studied the effect of a highly salient occluding surface appearing behind, coincident with, and in front of a virtual object. The apparatus and closed-loop matching task were based on previous work by Ellis and Menges. The experiment found maximum average depth judgment errors of 5.5 cm, and found that the blind reaching judgments were less accurate than the perceptual matching judgments. The experiment found that the presence of a highly-salient occluding surface has a complicated effect on depth judgments, but does not lead to systematically larger or smaller errors.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality; I.4.8 [Scene Analysis]: Depth Cues; H.5.1 [Information Systems]: Multimedia Information Systems—Artificial, Augmented, and Virtual Realities H.1.2 [Information Systems]: User/Machine Systems—Human Factors

Keywords: depth perception, augmented reality, optical see-through display, x-ray vision

1 Introduction

Augmented reality (AR) is an important and promising technology, with many compelling applications [e.g., Azuma et al. 2001; Feiner 2002]. Over the past several years there has been much technical development in *video see-through AR*, where digital images are captured by a camera, and virtual objects are digitally composited into the video stream. However, the perceptual issues involved in video see-through AR are equivalent to a standard computer display. This work focuses on perceptual issues that arise from *optical see-through AR*, where an optical element combines virtual objects with an optical view of the real world. In optical see-through AR (called "AR" for the rest of this paper), unique perceptual issues arise because the human visual system simultaneously sees both real-world and superimposed virtual objects. One such perceptual issue, addressed in this paper, is how the depth of a virtual object is perceived relative to surrounding real-world objects.

Recent analyses of depth perception have focused on how human behaviors afford vision; among these are Cutting and Vishton [1995], who usefully classified the perceptual cues to depth and layout according to distances that relate to human action. They divided perceptual space into three distinct regions, which we term *near-field* (out to ~1.5 meters), *medium-field* (~1.5 to ~30 meters), and *far-field* (beyond ~30 meters) distances. The near-field is the distance within which the hands can easily manipulate or reach for objects. Many compelling near-field AR applications have been proposed; these include medical applications [e.g., Sielhorst et al. 2008], maintenance tasks [e.g., Henderson and Feiner 2009], and manufacturing [e.g., Curtis et al. 1998]. Sielhorst et al. [2008] have published a comprehensive review of the use of AR in medical applications; in this review they also discuss perceptual issues, and they identify correct depth perception as the most important unsolved problem: "While many problems of early systems have already been addressed, the issue of a correct depth visualization remains unsolved."

To date, depth judgments at medium-field distances have been widely studied in virtual reality [e.g., Creem-Regehr et al. 2005; Interrante et al. 2008; Loomis and Knapp 2003], and a few studies have also been conducted in augmented reality [e.g., Jones et al. 2008; Swan II et al. 2007]. Most of this work has measured depth judgments with *action-based tasks*, where the depth judgment is based on a physical action performed by the observer. The most common such task has been *blind walking* [Loomis and Knapp 2003]. Blind walking can be considered an open-loop task, in that while the task is being performed, the observer receives no visual feedback regarding their location.

However, to the best of our knowledge, to date the published work examining near-field AR depth judgments can be found in only four papers [Ellis and Menges 1998; McCandless et al. 2000; Rolland et al. 1995; Rolland et al. 2002]. The first of these papers is Rolland et al. [1995]; they examined depth judgments of real and virtual objects at distances of 80 to 120 cm, using a forced-choice task. They found that the depth of virtual objects was overestimated at the tested distances. Rolland et al. [2002] then ran additional experiments with an improved AR display, which further examined the 80 cm distance, and compared forced-choice and perceptual matching tasks. They found improved depth accuracy and no consistent depth judgment biases. Ellis and Menges [1998] summarize a series of four AR depth judgment experiments, which used a visually closed-loop perceptual matching task to examine near-field distances of 40 to 100 cm, and studied the effects of an occluding object (the "x-ray vision" condition), convergence, accommodation, observer age, and monocular, biocular, and stereo AR displays. They found that monocular viewing degraded the depth judgment, and that an occluding object caused a change in vergence angle which resulted in depth judgments being biased towards the observer. In particular, they found that when the occluder was placed at the same distance as the virtual object, the incorrect occlusion cues caused proximal vergence, in which the apparent nearness of the virtual object drove the change in vergence angle. Finally, McCandless et al. [2000] used the same experimental setup and task to additionally study motion parallax and AR system latency in monocular viewing conditions; they found that depth judgment errors increased systematically with increasing distance and latency.

* Email: Gurjot Singh <gs266@msstate.edu>, J. Edward Swan II <swan@acm.org>, J. Adam Jones <jaj33@msstate.edu>
† Email: Stephen R. Ellis < stephen.r.ellis@nasa.gov>

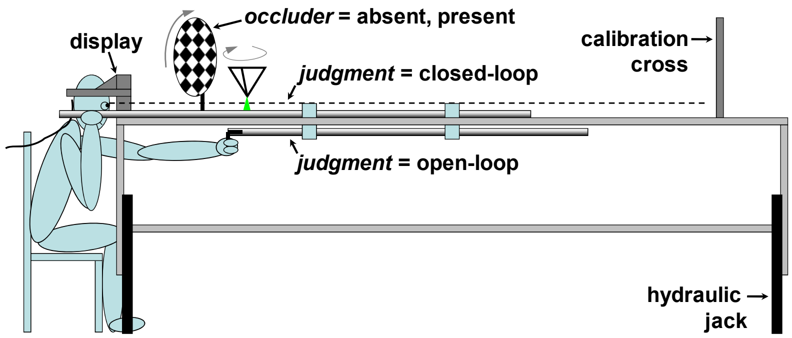

Figure 1: *Side-view diagram of the experimental table.*

2 Experimental Goals

In this paper, we describe an experimental apparatus to measure near-field AR depth judgments, and we present an experiment that partially replicates and extends the results first reported by Ellis and Menges [1998]. We pursued the following goals in this work:

(1) We sought to largely replicate the apparatus described by Ellis and Menges [1998], because to date that apparatus is one of only a very few to have studied near-field AR depth judgments, and we wanted to build on these previous results.

(2) In addition, compared to measuring medium-field AR depth judgments with techniques like blind walking [e.g., Jones et al. 2008], where effect sizes are on the order of 20 to 30 cm, here we expected effect sizes on the order of 10 to 5 cm or even less. Therefore, the apparatus had to measure depth judgments with a high degree of accuracy.

(3) Ellis and Menges [1998] studied visually closed-loop depth judgments by perceptually depth matching a virtual target with a mechanical pointer. We wanted to replicate this closed-loop depth judgment, and compare it to an open-loop depth judgment. We chose to implement *blind reaching*, where an observer reaches without sight to indicate a distance [e.g., Bingham et al. 2000; Mon-Williams and Bingham 2007; Mon-Williams and Tresilian 1999, 2000; Tresilian et al. 1999]. Blind reaching is the near-field equivalent of blind walking: it is an action-based task that does not involve visual feedback. Note, however, that in blind reaching observers do have proprioceptive feedback about their hand's location, so it is not a fully open-loop task; it is more precisely a *visually* open-loop task. Nevertheless, we use the term "open-loop" to refer to this task throughout the rest of this paper.

Action-based reaching tasks have been widely studied; Mon-Williams and Bingham [2007] and Bingham et al. [2000] provide surveys. This work comprises a wide variety of reaching tasks, which differ according to the presence or absence of visual or haptic feedback, whether the reach is in the air or involves sliding or manipulating a pointer, and whether the judged depth is indicated by a pointer held in the hand or by the participant's finger. The task that we have implemented here is most similar to the one reported by Mon-Williams and Tresilian [1999, 2000]. In their task, a participant positioned an unseen index finger at the same depth as the target; their apparatus ensured that the participant's finger was laterally displaced from the target by only a few centimeters. With this task, Mon-Williams and Tresilian [1999] found accurate distance judgments for real-world targets in full-cue conditions at distances of 16.7 to 50 cm. In our blind reaching task, a participant manipulates a slider until their unseen thumb is at the same depth as the target, and our apparatus also ensures that the participant's thumb is only laterally displaced from the target by a few centimeters.

(4) Finally, we wanted to test the relationship between *dark vergence* and the perceived depth of a virtual object. Dark vergence is the vergence distance of the eyes in the complete absence of light, when the ocular muscles that control vergence shift into a resting state. Dark vergence is unique for each person, and it has been found to predict the depth of visually impoverished objects [Gogel and Tietz 1973].

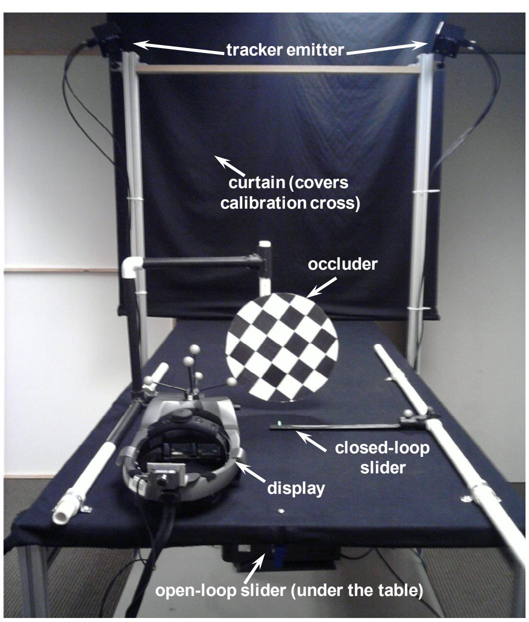

Figure 2: *Annotated photograph of the experimental table.*

3 Experimental Setup and Task

We developed a table apparatus for measuring near-field depth judgments; our table largely replicates the apparatus described by Ellis and Menges [1998], with modifications to support both open-loop and closed-loop depth judgments. Figure 1 shows a side-view diagram of the table and depicts the depth judgment tasks, while figure 2 shows an annotated photograph of our table.

The Table: The top of our table is a custom-ordered optical breadboard, 244 cm long by 92 cm wide by 3.8 cm thick; the top and bottom of the breadboard have a grid of 25 mm bolt holes for mounting equipment. As shown in figure 2, in order to prevent the grid from giving a strong perspective depth cue, we attached a matte black cloth to the top of the table, and mounted our equipment through the cloth. We created a frame for holding the table top, using custom-designed legs and attachments from 80/20 Inc. Our frame had six legs, with four legs at the corners and two legs in the middle. We mounted a second, particle-board table top, salvaged from an old office table, onto horizontal cross-beams 48 cm off of the ground; we use this second table to store equipment such as the tracking system computer. We obtained a hydraulic jack system from Dyna-Lift, Inc., and mounted six hydraulic jacks to the table legs, and bolted the compressor to the bottom of the table top. With the jacks, we can raise and lower the entire table apparatus such that the table top is between 104 and 134 cm above the ground. As shown in figure 1, we adjusted the table height so that sitting participants could comfortably rest the bottom of the AR display on the top of the table. In this position, when the display has been calibrated, participants' pupils are very close to 3.5 cm above the table surface.

Tracking: For tracking, we used an ARTtrack system by Advanced Realtime Tracking, GmbH. This system consists of two infrared emitters / cameras and retro-reflective spheres. We mounted the tracker cameras to the two middle legs of our table; these legs extend 104 cm above the table surface (see figure 2). By mounting the cameras to the table itself, the tracking does not have to be recalibrated when the table top is raised or lowered.

Display: Our augmented reality display was an nVisor ST model by NVIS, Inc.; this model has a screen resolution of 1280 × 1024 pixels, a 60° diagonal field-of-view, and 100% overlap between the screens. The optical elements are infinity collimated, and each optical unit is independently left / right adjustable.

Depth Judgment Tasks: We replicated the closed-loop matching task described by Ellis and Menges [1998] by sliding a length of white plastic PVC pipe through two collars that are attached to the right-hand side of table surface (see figures 1 and 2). Mounted to the pipe is an arm that extends to the middle of the table; at the tip of the arm is a green LED. As shown in figure 1, the participant uses their right hand to position the LED in depth by sliding the pipe. We mounted a retro-reflective sphere to the arm, and we use the tracker to automatically encode the participant's closed-loop depth judgment.

In addition, we developed a method for measuring open-loop blind reaching depth judgments. Also as shown in figures 1 and 2, we slid another length of PVC pipe into two collars mounted to the center of the bottom of the table. At the end of the pipe is a right angle bracket; we ask participants to grab this bracket with their thumb pointing upwards, and then slide the pipe until their thumb is directly underneath items seen on top of the table. As shown in figure 1, participants cannot see their thumb, so this is a blind reaching task. At the far end of the pipe we have attached a laser level that shoots a fan-shaped laser beam onto a meter stick that is also mounted under the table. We encode open-loop depth judgments by reading the value on the meter stick and entering it into our control software.

Occluder: We also replicated the occluder described by Ellis and Menges [1998]. We glued paper printed with a 5 cm square checkerboard pattern onto a circular foam disc with a diameter of 29 cm. We mounted the center of the disc onto a small motor that rotates the disc at 2 revolutions per minute. As shown in figures 1 and 2, we attached the disc and motor assembly to a pipe that runs through two collars mounted on the left-hand side of the table. With this mounting, our occluder can be positioned either in or out of the participant's field of view. When the occluder is in the field of view, it can be positioned at a range of distances from the participant. Following the goals of Ellis and Menges [1998], this occluder is as salient as possible: the black and white checkerboard pattern contains many strong accommodative cues, and the slow rotation further enhances the salience.

Virtual Target: For each task, participants judged the depth of a virtual target object. The target was a white, wireframe pyramid, rotating at 4 revolutions per minute, with the apex of the pyramid facing downwards (see figure 1). Having the point of the pyramid facing down facilitates precise matching with the LED pointer in the closed-loop task, and with the thumb in the open-loop task. The pyramid nominally has a base of 10 cm and a height of 5 cm. For each trial, we randomly scaled the pyramid from 70% to 130% of its actual size, in order to remove retinal size as a distance cue. The pyramid was seen against a black curtain that hung 220 cm from the participant (see figure 2).

Table 1: *Independent and Dependent Variables*

INDEPENDENT VARIABLES		
participant	16	(random variable)
judgment	2	closed-loop, open-loop
occluder	2	absent, present
distance	5	34, 38, 42, 46, 50 cm
repetition	3	1, 2, 3
DEPENDENT VARIABLES		
judged distance		in cm
error		*judged distance – distance*

When designing our table apparatus we strove to replicate the design of Ellis and Menges [1998] as closely possible, in as many details as possible. The primary difference is that the position of our table is considerably higher than it was for Ellis and Menges; in their experiment the table was at a normal height for an office table, and the target floated at eye height. We found that the higher table position was necessary to implement the open-loop task: pilot testing indicated that the participant's thumb needed to be just under the table, and the virtual target needed to be floating just over the table. In this configuration, the participant's thumb is only laterally displaced from the target by a few centimeters; Mon-Williams and Tresilian [1999] came to a similar conclusion and discuss additional reasons why the seen target and unseen pointing finger need to be close together. In addition, the top of the participant's shoulder needs to be below the bottom of the table surface in order for the experiment to measure the maximum amount of reach.

4 Variables and Design

Table 1 describes the experimental variables and design.

4.1 Independent Variables

Participant: We recruited 18 participants from a population of university students, faculty, and staff. The participants ranged in age from 19 to 30; the mean age was 21.6, and 7 were female and 11 male. Four participants were paid $12 an hour, and the rest received course credit. Most of the participants did not have problems learning the task or completing the experiment. However, we did not analyze the data for two participants: the first participant did not seem to take the experiment seriously, and the second participant realized part-way through the experiment that they were performing the depth judgments incorrectly. As indicated in table 1, we analyzed the data from the remaining 16 participants.

Judgment: Participants performed two kinds of depth judgments, *closed-loop* and *open-loop*. For a closed-loop judgment, participants used the pipe mounted on the top of the table to slide the green LED pointer until it was at the same depth as the apex of the rotating pyramid. For an open-loop judgment, participants grabbed the angle bracket on the end of the pipe mounted underneath the table with their thumb pointing upwards, and then slid the pipe until they believed that their unseen thumb was directly below the apex of the rotating pyramid.

Occluder: For each depth judgment, the occluder was either *absent* or *present*. When it was absent, the occluder did not rotate, and was positioned so that it was out of the participant's field of view. When the occluder was present, we placed it 42 cm from the participant, which matches the middle of the 5 tested distances. In this position the occluder subtends 34.6° of visual angle. The occluder rotated, and was positioned so that the apex of the rotating pyramid was visible just beyond the edge of the disc; the portion of the pyramid visible below the disc subtended 1.4° of visual angle.

Distance: The virtual target appeared at 5 different distances from the participant: 34, 38, 42, 46, and 50 cm. Pilot testing indicated that even relatively short-armed adults could reach 55 cm. At 34 cm the largest possible target subtends 17.4° of visual angle, while at 50 cm the smallest possible target subtends 4.4°.

Repetition: Participants saw 3 repetitions of each combination of the other dependent variables.

4.2 Dependent Variables

As shown in table 1, the primary dependent variable was *judged distance*, which we measured using either the closed-loop or the open-loop distance judgment. We also calculated *error* = *judged distance* – *distance*. An error near 0 cm indicates an accurately judged distance; an error > 0 cm indicates an overestimated distance; and an error < 0 cm indicates an underestimated distance.

4.3 Experimental Design

We used a factorial nesting of independent variables in a within-subjects, repeated-measures experimental design. The variables varied in the order that they are listed in table 1: *judgment* varied the slowest; within each judgment participants saw each *occluder* condition. The presentation order of each *judgment* ⊗ *occluder* block was controlled by the same 4 × 4 Latin square that is presented in Jones et al. [2008]. Within each *judgment* ⊗ *occluder* block, our control program generated a list of 5 (*distance*) × 3 (*repetition*) = 15 distances. The program then randomly permuted the presentation order of the distances, with the restriction that the same distance could not be presented twice in a row. We collected a total of 16 (*participant*) × 2 (*judgment*) × 2 (*occluder*) ×5 (*distance*) × 3 (*repetition*) = 960 data points.

With respect to the *judgment* ⊗ *occluder* blocks, this design can perfectly counterbalance multiples of 4 participants; the 16 participants analyzed here were perfectly counterbalanced in this manner. In addition, as discussed in more detail in Jones et al. [2008], this design not only counterbalances the presentation order of 4 conditions, it also counterbalances first-order carryover effects (which condition succeeds and precedes each other condition). Since publishing Jones et al. [2008], we have learned that this kind of Latin square was first described by Williams [1949].

4.4 Procedure

Each participant first filled out a standard informed consent form, a simulator sickness survey, and then a brief demographic survey. They next took a stereo acuity test, which all participants passed.

After this, the participant trained for and then used an apparatus that measured their *dark vergence*. As discussed above, we measured dark vergence because we hypothesized that it might describe some portion of the variance in the depth judgment data. The dark vergence apparatus is of our own design, and is based on a design described by Miller [1987]. However, to date we have not found any clear evidence to support our hypothesis, and there is not space in this paper to fully describe this apparatus and the related procedures. Therefore, here we will only report that participants next spent about 20 minutes training for and using our dark vergence apparatus.

After this, we described the two different depth judgment tasks to the participant, and they practiced each one several times. The participant did not yet wear the display; we held a physical model of the pyramid over the table while the participant practiced each matching task. We told the participant that the pyramid would randomly change size with each trial, and that they should not rely upon the size when making their depth judgment.

We next fitted the display on the participant's head, and calibrated the display using the technique described in Jones et al. [2008]. We originally developed this technique for a standing participant looking down a long hallway; the technique's parameters are the participant's eye height, inter-pupillary distance, and the distance from the participant's eyes to a calibration cross mounted at the participant's eye height at the end of the hallway. We modified the technique to work with our current setup, using an eye height of 3.5 cm above the table surface, and a calibration cross also 3.5 cm above the table surface; we drew this cross on a black cardboard surface that we mounted 220 cm from the participant (figure 1). This calibration method is described in detail in Jones et al. [2008]; after completing it (1) the participant is looking through the optical center of each of the display's eyepieces, (2) translational tracker errors related to the way the display fits on the participant's head are corrected, and (3) rotational tracker errors also related to the display's fit are corrected. After completing the calibration, we covered the calibration cross with a black curtain (figure 2).

The participant next completed four *judgment* ⊗ *occluder* blocks of 15 trials each. We allowed participants to take a break at any point during the experiment, and we mentioned the possibility of a break between each block. Most participants took one break; after each break we re-calibrated the display.

At the beginning of each block of trials, we adjusted the sliding pipe so that the LED pointer (closed-loop judgments) or angle bracket pointer (open-loop judgments) was as close to the participant as possible. For the first trial, the participant slid the pointer from this position to indicate their depth judgment. We trained participants to put their hands in their lap when they were satisfied with their depth judgment. When we saw them do this, we blanked the display, recorded the judgment, and then triggered the next trial. Therefore, for all trials except the first, the participant adjusted the pointer to the next depth from the previous depth position. Because our counterbalancing ensured that identical distances never appeared in subsequent trials, it was always necessary to move the pointer to make a correct judgment, and thus the pointer was moved from a wide variety of previous positions. We encouraged participants to be as accurate as possible, and did not set any time limit for the trials.

After completing the trials, participants filled out the simulator sickness exit survey, and we then debriefed them regarding their strategies and thoughts about the experiment.

5 Results and Discussion

We analyzed judged distance results from N = 960 data values. One data value was missing because of a tracker error, and examining histograms revealed 5 data values with *error* > +20 cm that we determined to be outliers. Using the procedure recommend by Barnett and Lewis [1994], we replaced these 6 values with the median of the remaining values in the *judgment* ⊗ *oc-*

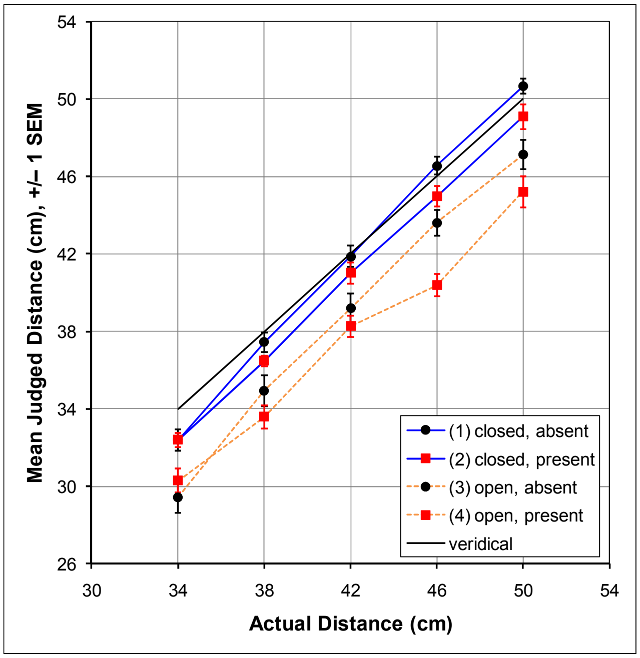

Figure 3: *The mean judged distances versus the actual distances for all of the data (N = 960). The numbers (1)–(4) indicate the four main conditions.*

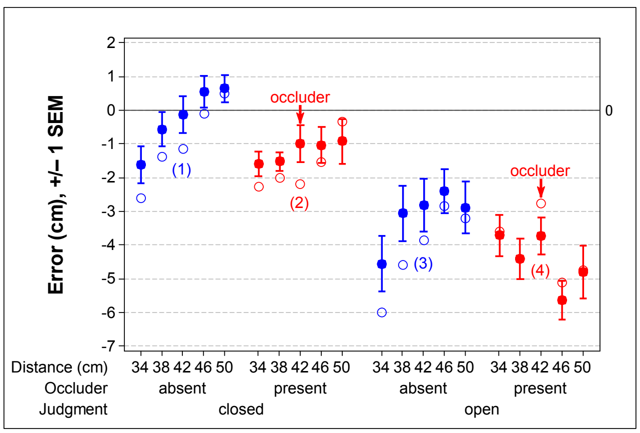

Figure 4: *The mean error for the overall depth judgments (N = 960); means are denoted with filled circles (●) and medians with hollow circles (○). As indicated, when the occluder was present it was located at 42 cm. The numbers (1)–(4) indicate the four main conditions.*

cluder ⊗ distance experimental cell. However, after completing the ANOVA analyses described below, we re-ran the same ANOVA models with the 5 outlying data points included. This ANOVA model has exactly the same pattern of results as what is described below, along with an additional main effect; we conclude that this additional effect is spurious because it is caused by the outlying data points. In addition, the histograms suggested a different distribution for the open-loop and closed-loop data, with the closed-loop results exhibiting more skew. To test for this potential violation of the assumptions underlying ANOVA modeling, we re-ran all of the ANOVA models discussed below with log-transformed data. These models also had the same pattern of results, but generally revealed greater experimental power: the *F* values were generally larger, which would be expected from better meeting the ANOVA assumptions. Nevertheless, here we report ANOVA analysis of the non-transformed error results.

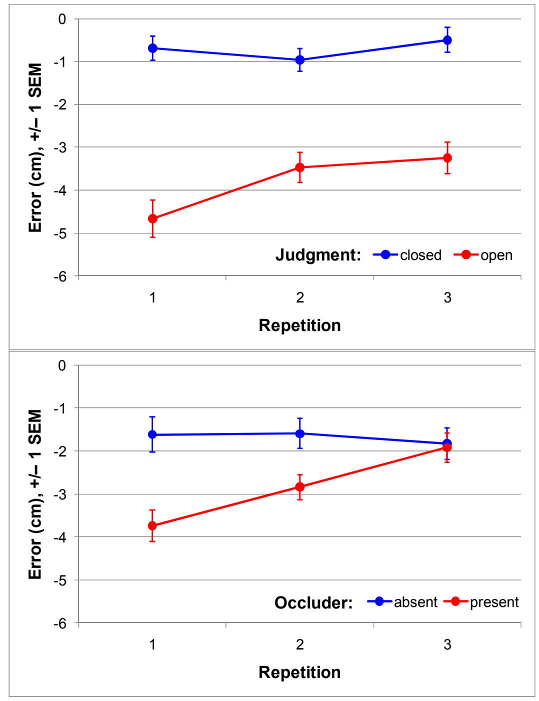

Figure 5: *Interactions with repetition for all depth judgments (N = 960). (upper) Judgment by repetition interaction. (lower) Occluder by repetition interaction.*

5.1 Results Over All Data

Figure 3 shows the main results as judged distance versus actual distance; figure 4 shows this same data plotted as error. We conducted an ANOVA analysis of this main experimental design. The main result is a different pattern of means by distance for each of the four main *judgment* ⊗ *occluder* conditions. This pattern would show up statistically as a three-way interaction between judgment, occluder, and distance. Although this three way interaction is not significant here, the following related main effect and two-way interaction are significant. First, there is a main effect of judgment on error ($F(1,15) = 21.0$, $p < .001$); open-loop responses are more underestimated than closed-loop responses. Second, there is an occluder by distance interaction on error ($F(4,60) = 4.5$, $p = .003$); this is reflected in the different shapes of the patterns by distance between the occluder absent and present conditions.

We also found significant calibration effects; figure 5 shows interactions with repetition. We found a judgment by repetition interaction on error ($F(2,30) = 3.4$, $p = .045$); closed-loop judgments had a relatively constant error of –0.71 cm, while open-loop judgments showed a calibration effect of 1.2 cm between repetition 1 and 2. In addition, we found an occluder by repetition interaction on error ($F(2,30) = 5.4$, $p = .010$); when the occluder was absent there was a relatively constant error of –1.68 cm, while when the occluder was present the error decreased linearly from –3.74 cm to near equivalency with the occluder = absent judgments.

These interactions by repetition indicated rapid calibration by the participants. We explored this calibration in more detail by running another ANOVA model that directly examined the 15 trials in each block. Figure 6 shows these results; we found a significant main effect of trial ($F(14,210) = 2.2$, $p = .010$), a significant judgment by trial interaction ($F(14,210) = 2.0$, $p = .022$), and a significant occluder by trial interaction ($F(14,210) = 2.1$, p =

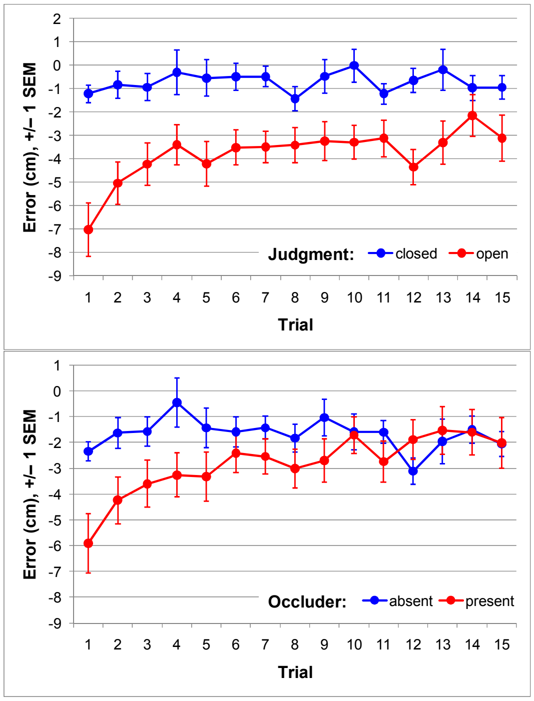

Figure 6: *Interactions with trial for all depth judgments (N = 960). (upper) Judgment by trial interaction. (lower) Occluder by trial interaction.*

.012). Figure 6 indicates rapid calibration over the first 5 trials; in both interactions, by trial 6 participants had more or less reached a steady state for the lower (red) responses (judgment = open-loop and occluder = present).

Calibration effects have been widely studied in open-loop reaching tasks. In situations where participants receive no visual or haptic feedback, reaching accuracy degrades over time [Bingham et al. 2000; Mon-Williams and Bingham 2007]. However, when participants receive visual or haptic feedback, they rapidly calibrate their reach according to this feedback. Mon-Williams and Bingham [2007] explore these calibration effects by systematically distorting this feedback; they find that participants rapidly recalibrate their reaching over a number of trials that is similar to what we have found here. This suggests that our participants are calibrating to the relatively novel perceptual situation of the open-loop reaching judgments as well as viewing the target in the presence of the occluder.

5.2 Results Over Stable Data

Based on these calibration effects, we next analyzed the $N = 640$ data points left over when repetition = 1 is removed from the data. Figures 5 and 6 suggest that the data from these two repetitions is a more stable representation of the participants' depth judgments. This data is plotted in figures 7 and 8.

Both figures demonstrate a three-way interaction between judgment, occluder, and distance on error ($F(4,60) = 22.1$, $p = .004$); included in this interaction are main effects of judgment ($F(1,15) = 13.6$, $p = .002$) and distance ($F(4,60) = 3.4$, $p = .014$), and an occluder by distance interaction ($F(4,60) = 4.6$, $p = .002$). This is a complex interaction; note that there is a different pattern of means by distance for each of the four main conditions. For the (1) closed-loop judgment, absent occluder condition, the judged distance shows a linear increase with increasing distance, beginning with error = –1.89 cm and progressing to error = +0.77 cm. When (2) the occluder is present with closed-loop judgments, the

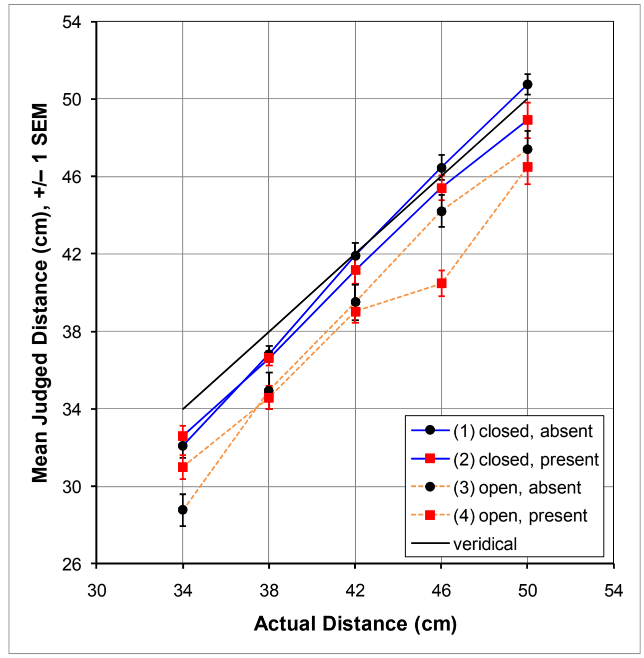

Figure 7: *The judged distances versus the actual distances for the stable judgments, where repetition = 2, 3 (N = 640). The numbers (1)–(4) indicate the four main conditions.*

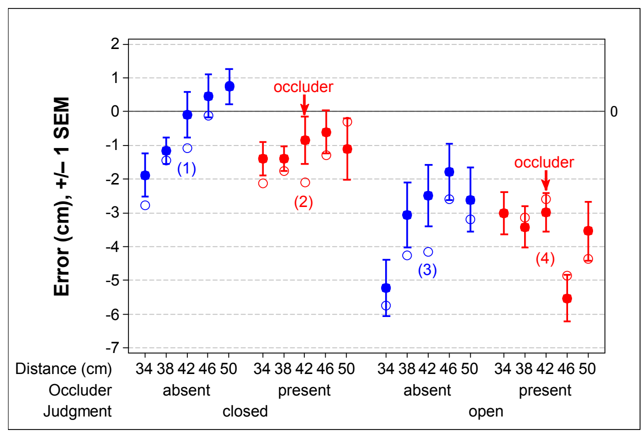

Figure 8: *The error for the stable judgments, where repetition = 2, 3 (N = 640); means are denoted with filled circles (●) and medians with hollow circles (○). As indicated, when the occluder was present it was located at 42 cm. The numbers (1)–(4) indicate the four main conditions.*

error remains relatively constant at –1.07 cm. In this case the presence of the occluder disrupts the liner pattern, probably by biasing convergence towards the occluding distance of 42 cm. For the (3) open-loop judgment, absent occluder condition, the judged distance again shows a general pattern of increasing with increasing distance. When (4) the occluder is present, four of the five means cluster around an average error of –3.23 cm, but when the target was at 46 cm the distance to it was judged to be considerably closer to the participant, at error = –5.52 cm. At this distance the target was 4 cm behind the occluder; here it is likely that, similar to Ellis and Menges [1998], the incorrect occlusion cues which suggest that the target is in front of the occluder cause a change in convergence, which results in the underestimated distance judgment. However, note that when the target was at 50 cm (8 cm behind the occluder) this effect is no longer operating, and the error is similar to the other open-loop depth judgments.

This result is consistent with an observation made by both us and Ellis and Menges [1998]: when a virtual object is initially located in front of a physical object, and the physical object is slowly moved towards an observer, at first the virtual object appears to be "pushed" closer to the observer by the physical object. At some point, however, the virtual object suddenly appears to "fall back" behind the physical object, which imparts a strong sense of transparency to the physical object. This effect is easy to see in an AR system using one's hand.

Finally, note that there is a general trend of greater underestimation for the open-loop judgments (error = –5.52 to –1.78 cm) relative to the closed-loop judgments (error = –1.89 to +0.77 cm). And within each judgment type, although the overall mean is similar whether the occluder is absent or present, the pattern of means by distance varies considerably.

6 Conclusions and Future Work

We have successfully developed an apparatus and related calibration and measuring techniques for collecting near-field depth judgments, using both closed-loop and open-loop tasks. Furthermore, our closed-loop results largely replicate the relevant results reported by Ellis and Menges [1998]. In addition, we directly compared closed-loop perceptual matching to visually open-loop blind reaching in the same experimental context. We found that blind reaching is significantly more underestimated than perceptual matching, with an average additional error of 3.1 cm for distances of 50 cm or less. This suggests that AR-presented virtual objects are perceived as being somewhat closer than they are perceptually matched with closed-loop tasks. We also found that the presence of a highly-salient occluding surface has a complicated effect on depth judgments, but it does not lead to systematically larger or smaller errors.

Finally, a practical finding is that this experiment determines an overall error bound for a range of reaching distances out to 50 cm. For the closed-loop task, we found a maximum mean localization error of 1.9 cm, and for the open-loop task a maximum mean error of 5.5 cm. It may be possible to implement some near-field AR applications within these error bounds.

In the future, we intend to further study the effect of calibration on depth judgment accuracy. In most near-field AR applications, the users of these systems would interact with co-registered virtual and real objects for long periods of time, and through this interaction the users may improve their initially-incorrect perception of the distance to virtual objects. Therefore, we imagine an experiment that uses a *pretest, intervention, posttest* design, where the pretest and posttest measure depth judgments using our apparatus, and the intervention is a period of time performing a task that involves manipulating virtual objects in the context of real objects, such as, for example, placing virtual blocks at specific locations on a real-world pattern. We hypothesize that the intervention would allow better calibration and hence improve the accuracy of the depth judgments.

In addition, in this experiment we used an occluder that is as salient as possible. In real AR applications it is likely that occluding surfaces would naturally be, or could be designed to be, much less salient. For example, in a medical AR application supporting endoscopic surgery, where the doctor uses AR to "see through" the patient's skin, we could imagine the skin being covered by a plain white sheet (or black, if that color was found to be more helpful). We intend to perform future studies that vary occluder salience, where the occluder is monochromatic and does not rotate.

Acknowledgements

This work was funded by the National Science Foundation, under award IIS-0713609. We are grateful for the assistance of Lorraine Lin, and we also acknowledge Andreas Werner of ART GmbH for a very helpful visit.

References

AZUMA, R., BAILLOT, Y., BEHRINGER, R., FEINER, S., JULIER, S. J., & MACINTYRE, B. [2001]. Recent Advances in Augmented Reality. *IEEE Computer Graphics & Applications, 21*(6), 34–47.

BARNETT, V., & LEWIS, T. [1994]. *Outliers in Statistical Data* (3rd ed.): John Wiley & Sons.

BINGHAM, G. P., ZAAL, F., ROBIN, D., & SHULL, J. A. [2000]. Distortions in Definite Distance and Shape Perception as Measured by Reaching Without and With Haptic Feedback. *Journal of Experimental Psychology: Human Perception and Performance, 26*(4), 1436–1460.

CREEM-REGEHR, S. H., WILLEMSEN, P., GOOCH, A. A., & THOMPSON, W. B. [2005]. The Influence of Restricted Viewing Conditions on Egocentric Distance Perception: Implications for Real and Virtual Environments. *Perception, 34*(2), 191–204.

CURTIS, D., MIZELL, D. W., GRUENBAUM, P., & JANIN, A. [1998]. Several Devils in the Details: Making an AR Application Work in the Airplane Factory. In *Augmented Reality: Placing Artificial Objects in Real Scenes; Proc. International Workshop on Augmented Reality (IWAR 98)* (pp. 47–60). Natick, Mass., USA: A.K. Peters.

CUTTING, J. E., & VISHTON, P. M. [1995]. Perceiving Layout and Knowing Distances: The Integration, Relative Potency, and Contextual Use of Different Information about Depth. In W. Epstein & S. Rogers (Eds.), *Handbook of Perception and Cognition: Perception of Space and Motion* (Vol. 5, pp. 69–117). San Diego, CA, USA: Academic Press.

ELLIS, S. R., & MENGES, B. M. [1998]. Localization of Virtual Objects in the Near Visual Field. *Human Factors, 40*(3), 415–431.

FEINER, S. [2002]. Augmented Reality: A New Way of Seeing. *Scientific American, 286*(4), 48–55.

GOGEL, W. C., & TIETZ, J. D. [1973]. Absolute Motion Parallax and the Specific Distance Tendency. *Perception & Psychophysics, 13*(2), 284–292.

HENDERSON, S. J., & FEINER, S. [2009]. Evaluating the Benefits of Augmented Reality for Task Localization in Maintenance of an Armored Personnel Carrier Turret. *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR) 2009*, October 19–22, Orlando, FL, USA, pp. 135–144.

INTERRANTE, V., RIES, B., LINDQUIST, J., KAEDING, M., & ANDERSON, L. [2008]. Elucidating Factors that Can Facilitate Veridical Spatial Perception in Immersive Virtual Environments. *Presence: Teleoperators and Virtual Environments, 17*(2), 176–198.

JONES, J. A., SWAN II, J. E., SINGH, G., KOLSTAD, E., & ELLIS, S. R. [2008]. The Effects of Virtual Reality, Augmented Reality, and Motion Parallax on Egocentric Depth Perception. *Proceedings of Proceedings of the Symposium on Applied*

Perception in Graphics and Visualization (APGV 2008), August 9–10, 2008, Los Angeles, CA, USA, pp. 9–14.

LOOMIS, J. M., & KNAPP, J. M. [2003]. Visual Perception of Egocentric Distance in Real and Virtual Environments. In L. J. Hettinger & J. W. Haas (Eds.), *Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues* (pp. 21–46). Mahwah, NJ, USA: Lawrence Erlbaum Associates.

MCCANDLESS, J. W., ELLIS, S. R., & ADELSTEIN, B. D. [2000]. Localization of a Time-Delayed, Monocular Virtual Object Superimposed on a Real Environment. *Presence: Teleoperators and Virtual Environments, 9*(1), 15–24.

MILLER, R. J. [1987]. Nonius Alignment Apparatus for Measuring Vergence. *American Journal of Optometry and Physiological Optics, 64*(6), 458–466.

MON-WILLIAMS, M., & BINGHAM, G. P. [2007]. Calibrating Reach Distance to Visual Targets. *Journal of Experimental Psychology: Human Perception and Performance, 33*(3), 645–656.

MON-WILLIAMS, M., & TRESILIAN, J. R. [1999]. Some Recent Studies on the Extraretinal Contribution to Distance Perception. *Perception, 28*, 167–181.

MON-WILLIAMS, M., & TRESILIAN, J. R. [2000]. Ordinal Depth Information from Accommodation? *Ergonomics, 43*(3), 391–404.

ROLLAND, J. P., GIBSON, W., & ARIELY, D. [1995]. Towards Quantifying Depth and Size Perception in Virtual Environments. *Presence: Teleoperators and Virtual Environments, 4*(1), 24–49.

ROLLAND, J. P., MEYER, C., ARTHUR, K., & RINALDUCCI, E. [2002]. Method of Adjustment versus Method of Constant Stimuli in the Quantification of Accuracy and Precision of Rendered Depth in Helmet-Mounted Displays. *Presence: Teleoperators and Virtual Environments, 11*(6), 610–625.

SIELHORST, T., FEUERSTEIN, M., & NAVAB, N. [2008]. Advanced Medical Displays: A Literature Review of Augmented Reality. *Journal of Display Technology, 4*(4), 451–467.

SWAN II, J. E., JONES, A., KOLSTAD, E., LIVINGSTON, M. A., & SMALLMAN, H. S. [2007]. Egocentric Depth Judgments in Optical, See-Through Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics (TVCG), 13*(3), 429–442.

TRESILIAN, J. R., MON-WILLIAMS, M., & KELLY, B. M. [1999]. Increasing Confidence in Vergence as a Cue to Distance. *Proceedings of the Royal Society of London; B: Biological Sciences, 266*, 39–44.

WILLIAMS, E. [1949]. Experimental Designs Balanced for the Estimation of Residual Effects of Treatments. *Australian Journal of Scientific Research, Series A, 2*, 149–168.