Direct Interaction with Virtual Objects

Long Nguyen, Larry Keeler
NAWCTSD, Orlando FL
long.nguyen@navy.mil
laurence.keeler@navy.mil

Linda Malone
University of Central Florida
Orlando, FL
lmalone@mail.ucf.edu

Jannick Rolland
University of Rochester
Rochester, NY
rolland@optics.rochester.edu

ABSTRACT

Virtual Environment (VE) applications must afford users with mechanisms for perception and manipulation of virtual objects in order to be effective. Since VEs can only provide a subset of the cues experienced in the real world it is important to understand the impact of those cues supporting VEs. This study explores direct object interaction in personal space, in order to quantify achieved accuracy and performance time in VEs, and to provide insight into which factors contribute to these measures.

The study utilized a stereoscopic Head Mounted Projection Display (HMPD) with inherent high accuracy for the presentation of those visual cues most important in personal space, and addressed the full continuum of VE types (Immersive Virtual Environment, Mixed Reality, and Reality) and sensory modalities (visual, audio, and touch) for comprehensive evaluations. Two full-factorial across-subjects experiments were conducted and the results used to provide key insights into the effect of each type of environment and modality on accurate and timely interaction with virtual objects.

The mean depth perception error in personal space was less than four millimeters whether the stimuli presented were real, virtual, or mixed; the mean error for the simple task, button tapping, was less than four millimeters whether the buttons were real or virtual; and the mean task completion time was less than one second. The high accuracy and rapid task performance observed was attributed to the presentation accuracy of the visual cues, including occlusion, stereoscopy, accommodation, and convergence. With performance already near an optimal level with visual cues presented alone, adding proprioceptive, audio, and haptic cues did not significantly improve results.

ABOUT THE AUTHORS

Long Nguyen is an Engineer at the Naval Air Warfare Center Training Systems Division (NAWCTSD). His primary duties involve research, development, and procurement of simulation-based training devices. He holds a Ph.D. in Industrial Engineering, specializing in Interactive Simulation and Training.

Linda Malone is a Professor in the Industrial Engineering Department at UCF. Her primary research interests include Response Surface Analysis and Quality. She was an associate editor of the Journal of Statistical Computation and Simulation for over 25 years and was a founding co-editor of the STATS Magazine. She was awarded the honor of Fellow of the American Statistical Association.

Jannick Rolland is the Brian J. Thompson Professor of Optical Engineering and the Associate Director of the R.E. Hopkins Center at the University of Rochester. She has produced 18 patents, 6 book chapters, over 80 peer-reviewed publications. She is a Fellow of the OSA, the SPIE, and a senior member of IEEE.

Larry Keeler is an Engineer at NAWCTSD. His primary duties involve research, development and procurement of training systems and related support systems**.** He holds a Bachelor of Science in Electrical Engineering and a Ph.D. in Experimental Psychology, specializing in Learning.

Direct Interaction with Virtual Objects

Long Nguyen, Larry Keeler
NAWCTSD, Orlando, FL
long.nguyen@navy.mil
laurence.keeler@navy.mil

Linda Malone
University of Central Florida
Orlando, FL
lmalone@mail.ucf.edu

Jannick Rolland
University of Rochester
Rochester, NY
rolland@optics.rochester.edu

INTRODUCTION

Interacting with a Virtual Environment (VE) requires the user to perceive the relative position and orientation of virtual objects. For applications requiring interaction in personal space, the user may also need to accurately judge the position of virtual objects relative to that of real objects, for example, a virtual button and the user's real hand. This is difficult since VEs can only provide a subset of the cues experienced in the real world. Complicating matters further, technological limitations of currently available visual display apparatus may distort the VE.

An experimental test bed was designed (Nguyen, Mead, Fidopiastis, & Rolland, 2004) to provide the highest attainable accuracy for salient visual cues in personal space (Cutting & Vishton, 1995) using Head Mounted Projection Display (HMPD, Hua, Girardot, Gao, & Rolland, 2000; Rolland, Biocca, Hamza-Lup, Ha, & Martin, 2005) and retro-reflective screen technology. The test bed accommodated the full continuum of VE types and multiple sensory modalities for comprehensive comparison studies. A research scheme was designed to expand on previous experiments (Nguyen, Mead, Malone, Williams, Fidopiastis, & Rolland, 2005) that measured VE depth perception and interaction, and to explore and quantify fundamental accuracy and performance time for direct object interaction in VE.

Research Question I: Depth Perception in a VE

Given correct binocular visual cues and kinesthetic proprioception, what is the accuracy of perceived depth for virtual and real objects? This primary research question is parsed into detailed questions to provide further insight into the contribution of different VE types and the cue of kinesthetic proprioception to depth perception. Does a Mixed Reality (MR) environment enhance accuracy of depth perception over the purely Immersive Virtual Environment (IVE)? Does the addition of kinesthetic proprioception enhance the precision of depth perception? Finally, what is the accuracy of depth perception for Reality (RE) and is it significantly different than that for the MR or IVE with or without synchronized proprioceptive cues?

Research Question II: Direct Interaction in a VE

This question addresses interaction with virtual and real objects. Can one manipulate virtual controls as accurately and rapidly as one does real counterparts? This primary research question is parsed into several detailed questions to assess the contribution of each sensory modality on task performance. First, given the correct synchronized visual and kinesthetic proprioceptive cues in a VE, what are the performance time and accuracy for the task of interacting with simple controls, such as virtual cockpit buttons in a flight simulator? Second, do these performance measures improve with the addition of an auditory cue? Similarly, how does an additional, simple touch haptic cue impact the performance? How does performance compare in different environments, e.g., MR vs. RE?

METHOD

Two experiments were designed to address the research questions above. These included depth perception and interaction tasks and explored the entire Reality-Virtuality continuum (R-V, Milgram & Kishino, 1994; Milgram & Colquhoun, 1999) – Reality (RE), Mixed Reality (MR), and Virtuality (or Immersive Virtual Environment, IVE) – and all salient sensory modalities.

Apparatus

The primary apparatus used for collecting experimental data included VE equipment for the main data collection procedure and eye examination instruments for pre- and post- experiment procedures.

The visual display for collecting experimental data was an HMPD with high brightness Organic Light Emitting Diode (OLED) displays provided by eMagin Corporation and the Optical Diagnostic and Applications (ODA) Lab at UCF (See Figure 1).

Figure 1. ODA Lab's HMPD

The HMPD had adjustments for head size and position and for inter-pupilary-distance (IPD) to match those of each test participant and weighed less than 500g. Since the HMPD could be worn with prescription glasses, focus adjustment in lieu of the eyeglasses was unnecessary. HPMD technology, uniquely, provides high fidelity for a combination of cues that are crucial for VE depth perception and interaction in personal space (Nguyen, et al., 2004).

The computer simulation software used was DiSTI's GL Studio, which provided a user-friendly Graphical User Interface (GUI) for creating 3-D cockpit models. The software application supported generation of accurate binocular views and provided for programmable adjustment of field of view, aspect ratio, IPD, and perspective to match the visual display to those of the experiment participants.

An infrared detector with audio signal feedback was used to detect the breaking of the infrared beam, thereby indicating the hand touching the virtual object. A Sony Digital Hi-Definition camcorder and an HDTV were used for monitoring and, as a secondary method, for capturing distance measurements during the experiment.

The Snellen Eye Chart, the Dolman Depth Perception Box, and the Stereo Fly Test were used to measure each participant's static acuity, depth perception, and ability to perceive 3-D via stereoscopy, respectively in order to screen out participants who had less than normal vision. The Simulation Sickness Questionnaires (SSQs; Kennedy, Lane, Berbaum, & Lilienthal, 1993), as a safety measure, was used to ascertain any signs of adverse aftereffects (Kennedy, Dunlap, Jones, & Stanney, 1996; Drexler, Kennedy, & Malone, 2009).

Experiment Facility and Participants

A military installation with a pool of well disciplined personnel provided a near ideal experimental population. This included screened military participants who had the perfect or corrected eyesight required for the study. The experiment was carried out at the Naval Air Station, Jacksonville using a participant pool comprised mostly of military students and instructors. Participants were given ample opportunities to provide feedback and repeat data points where unintended inputs were made.

Experiment I: VE Depth Perception

The first experiment provided a complete set of empirical data on depth perception spanning all three environment types, and included the sensory cue of proprioception in addition to high-fidelity visual cues of stereoscopy, accommodation, and convergence afforded by the HMPD. The Method of Adjustment defined by Rolland (Rolland, Meyer, Arthur, & Rinalducci, 2002) was used for assessing depth perception. The stimuli were an octahedron and a cylinder (Figure 2) subtending four degrees and one degree in widths, respectively.

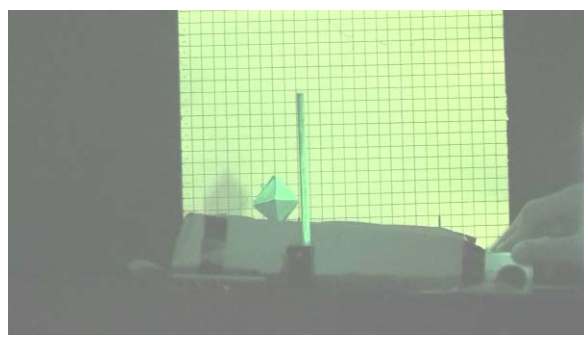

Figure 2. Stimuli – Cylinder and Octahedron

VE Depth Perception Experiment - Hypotheses

Using results from previous VE depth perception research (Ellis & Menges, 1998; Rolland et al., 2002; Nguyen, et al., 2005) the following hypotheses (H_a, alternate hypotheses) were formulated:

Given correct binocular visual and kinesthetic proprioceptive cues, the user can perceive depth with sub-centimeter accuracy between a real and a virtual object, between two real objects, and between two virtual objects.

Hypothesis I(a): With accurate visual cues for virtual objects, mean depth perception errors are small, less than 10 mm.

$$H_a: \text{Error} < 10\text{mm} \quad (1)$$

The addition of kinesthetic proprioception provides an extra cue, which could potentially improve accuracy of depth perception.

Hypothesis I(b): Proprioception significantly improves depth perception, i.e., the error is less given the added cue.

$$H_a: \text{Error}_{\text{proprioception}} < \text{Error}_{\text{no-proprioception}} \quad (2)$$

VE Depth Perception Experiment - Stimuli and Task

The task for the participant was to adjust a moveable object until it was perceived to be the same distance from the participant as a nearby fixed object using Roland's Method of Adjustment (Rolland et al., 2002). The fixed object was a cylinder, and the moveable object an octahedron, translated in one dimension, back and forth, by the participant. Both stimuli were virtual objects for the IVE condition and both were real objects for the RE condition. A real cylinder and a virtual octahedron were used for the MR condition. The fixed object, the cylinder, was 0.65 m (approximately the extreme reach) in front of the participant. For each trial, the moveable object, the octahedron, was initially positioned randomly to the left/right of, and closer/farther than the fixed object.

VE Depth Perception Experiment - Independent Variables and Dependent Measures

The Independent variables were the virtual environment type and the kinesthetic proprioceptive cue. The Virtual environment type had three levels: V/V, for virtual object next to virtual object, represented IVE. R/V, for real object next to virtual object, represented MR. R/R, for real object next to real object, represented RE. The proprioceptive factor had two levels: presence or absence of the cue. For the presence condition, the participant's hand moved inch for inch with the octahedron. For the absence condition, the participants pressed arrow buttons on the keyboard to translate the octahedron and did not get any relevant, direct proprioceptive feedback.

Error of depth perception was the primary quantity of interest for this experiment, and three dependent measures, bias, accuracy, and precision were made for each participant. (These measures are defined below for this study.)

For each trial, the signed error was determined by measuring the final distance in depth (horizontally in the sagittal plane) between the displaced octahedron and the fixed cylinder. It was arbitrarily defined as positive error distance if the octahedron was closer to the participant than the cylinder and negative error distance if farther away.

For each participant, these signed error distances were averaged over the repeated trials, to calculate the first dependent measure, bias. Bias can be positive or negative, since the error can be positive or negative. It is a measure of how far, on average, a participant overshoots, if positive, or undershoots, if negative, in perceiving depth.

Accuracy was the second dependent measure. For this study, it was defined as the average of the absolute values of the error distances. It was calculated for each participant by taking the absolute value of the error distance for each trial and averaging these values over all repeated trials. Accuracy can only be positive since it is the average of only positive values. It is a measure of average error of depth perception for each participant.

Precision was the final dependent measure. For this study, it was defined as the standard deviation of the signed (positive or negative) error distances. For each participant, the standard deviation of the error distances for the repeated trials were calculated. Although the error distances could be positive or negative, precision could only take on positive values since it was determined by calculation of standard deviation. Precision is a measure of the variability of depth perception for each participant.

VE Depth Perception Experiment - Procedures

The participant adjusted the position of the octahedron until its depth was believed to match the close by, stationary cylinder. The experiment quantified error of depth perception for all three types of Virtual Environments – RE, MR, and IVE. Half of the participants performed the task with kinesthetic proprioceptive cue and the other half performed the task without the feedback. With the kinesthetic proprioceptive cue, the participant moved the octahedron by translating a computer mouse which was calibrated, inch for inch, with the movement of the octahedron. Without the proprioceptive cue, the participant indirectly moved the object using computer keyboard buttons. Each participant repeated the adjustment task twenty times. The HMPD and head were fixed. The chin was rested on a fixed reference point.

Experiment II: VE Object Interaction

The VE Object Interaction Experiment was also a full-factorial, across-subjects design. The experimental test bed was set up in a night vision goggle training laboratory at the Naval Air Station in Jacksonville, Florida and included mainly the HMPD, retro-reflective fabric, the chin rest, and the stimuli (Figure 3).

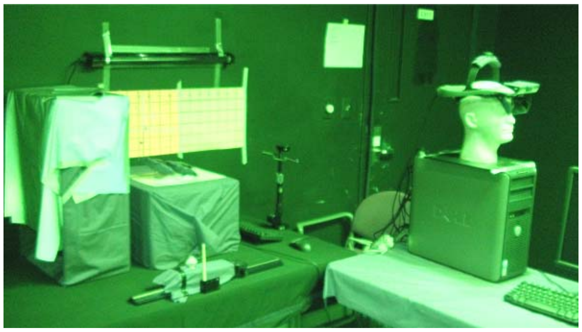

Figure 3. Experimental Set Up

VE Object Interaction Experiment - Hypotheses
The following hypotheses were drawn for the research questions discussed earlier.

Hypothesis II(a): Given accurate visual cues, including stereoscopy, accommodation, convergence, and occlusion, synchronized with proprioceptive cue, interaction with a simple object is highly accurate, in the sub-centimeter (millimeters) range.

$$H_a\text{: Error} < 10 \text{ mm} \quad (3)$$

Hypothesis II(b): Performance time and spatial error for RE is significantly less than that for MR.

$$H_a\text{: Perf Time}_{RE} < \text{Perf Time}_{MR} \quad (4)$$

$$H_a\text{: Error}_{RE} < \text{Error}_{MR} \quad (5)$$

Hypothesis II(c): The addition of an audio cue significantly improves task performance time.

$$H_a\text{: Perf Time}_{audio} < \text{Perf Time}_{no\text{-}audi} \quad (6)$$

Hypothesis II(d): The addition of a haptic cue significantly improves accuracy of task performance.

$$H_a\text{: Error}_{haptic} < \text{Error}_{no\text{-}haptic} \quad (7)$$

VE Object Interaction - Task and Stimuli
The task was activation of a keyboard button. Real and virtual buttons similar to those found in flight simulation cockpits were used as stimuli. Real buttons were presented using an optical mirage dish that provided true fidelity for all visual cues. Virtual buttons were modeled and rendered using the computer software application, GL Studio, and displayed by the stereoscopic HMPD and a retro-reflective screen. Haptic feedback was provided by placing a clear, essentially invisible, plastic panel at the distance that the optical mirage and the computer generated buttons were displayed. Audio feedback was provided by an optical sensor and a speaker set to provide a “ding” sound when the infrared beam (at the distance of the displayed button) was broken.

VE Object Interaction - Independent Variable and Dependent Measures
Independent variables were: virtual environment type, haptic cue, and audio cue. Virtual environment types included MR and RE. The MR condition was represented with the real hand manipulating the virtual (computer-generated, HMPD displayed) button. The RE environment type was represented with the real hand manipulating the real (optical mirage) button. The haptic cue had two levels, presence and absence (of the transparent panel). Likewise, the audio cue had two levels, presence and absence (of the auditory tone).

Dependent measures were spatial error (bias, accuracy, and precision) and performance time. Bias, accuracy, and precision were defined (as before) as the average, average of absolute values, and standard deviation of the error distances respectively. The error distance was measured from the center of the finger tip to the center of the stimuli, i.e., the buttons. Performance time was the duration necessary for participants to tap the virtual or real button.

VE Object Interaction Experiment - Procedures
For this experiment, the participant’s head was fixed. The hand and finger were not fixed and the finger tip’s position was measured relative to the fixed head and the fixed stimuli. Similar to the traditional Fitt’s tapping task conducted by Arsenault and Ware (2000), each participant tapped two buttons, alternating from one to the other, 32 times. Half of the participants tapped buttons that were real (optical mirages). The other half tapped buttons that were virtual (computer-generated and HMPD displayed). Half received haptic cue feedback and half did not. Likewise, half received audio feedback and half did not.

RESULTS

VE Depth Perception Experiment: Results

Experiment I measured depth perception and compared the results across VE types and proprioceptive conditions. VE types included R/R, R/V, and V/V which represented RE, MR, and IVE, respectively. The proprioceptive conditions included presence and absence of the cue.

The experiment had six treatments: Each treatment had seven participants randomly assigned. Each participant, for each experiment, contributed 20 data points from which the following measures were calculated.

1) Depth Perception Bias: average of the errors for the 20 trials
2) Depth Perception Accuracy: average of the absolute values of the errors for the 20 trials
3) Depth Perception Precision: standard deviation of the errors for the 20 trials

These bias, accuracy, and precision values were the dependent measures used for statistical analysis.

VE Depth Perception Bias
Depth perception bias was a measure of how far, on average, the participant overshot or undershot when attempting to align the moveable octahedron to line up with the fixed reference depth, the cylinder. The ANOVA F-Test provided statistics for the main effects and first order interaction. The test did not show statistical significance for either of the two terms, environment type or proprioception, nor their interaction (environment type by proprioception).

VE Depth Perception Accuracy
The ANOVA and confidence interval for the main effects were computed to determine if the three types of environment exhibited a significant difference in accuracy. The analysis indicated a statistical difference in the RE type. The 95% confidence interval plot indicated that the RE condition was statistically different from the MR and IVE conditions. The RE condition, where both of the stimuli, the octahedron and the cylinder, were real, showed high accuracy, 1.7 mm (Standard Deviation, SD = 0.9 mm). Mean accuracy for the MR and IVE conditions were 3.0 mm (SD = 1.2 mm) and 3.9 mm (SD = 1.7 mm), respectively.

VE Depth Perception Precision
ANOVA and confidence intervals for the main effects indicated a significant difference in precision only in the RE condition. Precision for the MR and IVE conditions were not significantly different. Participants clearly performed with better precision for the RE condition, with a mean precision of 1.7 mm (SD = 1.1 mm). For the MR and IVE conditions, participants averaged 3.9 mm (SD = 1.3 mm) and 3.6 mm (SD = 1.7 mm), respectively.

VE Depth Perception Observations
Hypothesis I(a), H_a: Error < 10 mm, was supported by the empirical data and statistical analysis above. Whether bias, accuracy, or precision was used as the measure of spatial error, the results clearly showed that mean error for all types of environments, whether the proprioceptive cue was provided or not, was less than 4 mm with a standard deviation of approximately 2 mm (1.6 mm for accuracy and precision and 2.2 mm for bias).

Hypothesis I(b), H_a: $Error_{proprioception} < Error_{no\text{-}proprioception}$, was not supported by the data. The results failed to show a significant difference attributable to the addition of the proprioceptive cue. While this hypothesis was not supported, it is noted that the spatial error measured turned out to be relatively small, less than 4 mm (SD = 2.2 mm) in all conditions (experimental treatments). With such small error to begin with, further improvement attributable to the additional sensory modality of proprioception was difficult to find and was not found.

The empirical data also showed that the difference in depth perception in RE compared with MR and IVE was statistically significant. Depth perception in RE was more accurate and more precise than that in MR and IVE. Although statistically significant, this difference was small, less than 3.0 mm (SD = 1.2 mm), and not of practical significance.

The spatial errors observed were relatively small, comparable to experimental apparatus tolerances. Of the three measures for spatial error, precision is the most reliable measure. Precision is a measure of standard deviation and is affected only by the precision of the measurement equipment, not its absolute accuracy, which was likely worse.

VE Object Interaction Experiment: Results

The second experiment measured spatial error and completion time associated with object interaction. Control variables included environment type, haptic condition, and audio condition. Environment types included RE and MR. Haptic conditions included presence and absence of a simple touch cue. Audio conditions included presence and absence of a simple sound cue. The experiment had eight treatments. Each treatment had five participants randomly assigned from the group used in the previous experiment. Each participant contributed 32 data points from which the dependent measures – bias, accuracy, precision, and performance time – were calculated for statistical analysis.

VE Object Interaction Depth Bias

Bias provided a measure that indicated whether each participant, on average, overshot (negative value) or undershot (positive value) when interacting with the virtual or real button. It was calculated for each participant by averaging the error distances, negative if farther and positive if closer, between the tip of the finger and the center of the button. ANOVA and factorial fit for all main effects and first order interaction terms indicated no statistically significant main effect. It revealed one statistically significant and one marginally significant interaction term, Environment by Audio and Environment by Touch, respectively. However, the error was only about one millimeter, within measurement instrument accuracy, and thus was not of practical significance. Overall, average bias was less than 1 mm regardless of condition - Virtual/Real, Audio Present/Absent, and Haptics Present/Absent.

VE Object Interaction Depth Accuracy

Accuracy was calculated from the average of the absolute values of the 32 error distances computed for each participant. ANOVA and 95% confidence intervals showed no significant effect from any main factor or interaction terms. Mean accuracy was 3.1 mm (SD = 1.2 mm) for interaction with virtual buttons and 2.2 mm (SD = 1.3 mm) for interaction with real buttons.

VE Object Interaction Depth Precision

Similar to the previous measures, no main factor or interaction term was statistically significant. Mean precision was 3.95 mm (SD = 1.2 mm) for interaction with virtual buttons and 3.3 mm (SD = 1.3 mm) for interaction with real buttons.

VE Object Interaction Performance Time

Reaction or performance time was the duration taken to tap the buttons 32 times. ANOVA revealed only one statistically significant term, audio, which was a main factor. Surprisingly, the effect of adding audio was counter-intuitive. The addition of audio cuing increased mean completion time from 24 seconds (SD = 8 seconds) to 30.3 seconds (SD = 10.6 seconds).

VE Object Interaction Observations

Hypothesis II(a), H_a: Error < 10 mm, was supported by the data collected. Mean error for interaction with real or virtual buttons was, surprisingly, less than 4 mm with standard deviation of 1.6 mm or less whether bias, accuracy, or precision was used as the measure of spatial error. This held true for all experimental treatments, with real or virtual buttons and with or without haptic and audio cues.

Hypothesis II(b), H_a: $Error_{RE} < Error_{MR}$ and H_a: $Time_{RE} < Time_{MR}$, were not completely supported by the empirical data. Spatial error, in terms of bias, accuracy, and precision were slightly less for interaction with real buttons compared to interactions with virtual buttons, but were not statistically significant. Although statistical significance was not observed, it is noted that spatial error measured for all eight treatments were unexpectedly found to be near optimal, less than 4 mm, not leaving much room for enhancement by the addition of touch or audio cues. For performance time, there was no evidence of any difference at all between interaction with real and virtual buttons. Again, while no significant difference was found, performance time was surprisingly near optimal, about one second per tap, for all treatments. The optimal result was likely due to the high fidelity of the visual cues provided for all treatments, MR or RE.

Hypothesis II(c), H_a: $\text{Perf-Time}_{audio} < \text{Perf-Time}_{no\text{-}audio}$, was also not supported by the data collected. On the contrary, the opposite was found to be true. Audio is a temporal cue and its addition was predicted to enhance performance time. However, the results showed significant degradation in performance time with audio cue added. This was likely due to the experimental apparatus. The audio cues were not provided immediately when the buttons were touched. The audio feedback system, an Enforcer® Alert System Model E-931CS22RC, had a noticeable delay of about 200 ms. Some participants hesitated and paced themselves waiting for the audio cue when tapping the buttons, especially when the haptic cue was absent.

Hypothesis II(d), H_a:$Error_{haptic} < Error_{no\text{-}haptic}$, was also not entirely supported. For interaction with virtual buttons, the addition of the touch haptic cue, on average, appeared to reduce the error, but by less than a millimeter. On the other hand, for interaction with real buttons, the cue appeared to increase the error, on average, by about the same amount. This relatively small difference is not of practical significance and could be due to an artifact of rounding errors during data collection and calculation or the experiment setup and apparatus.

Unexpected, subjective observations were also noted. Several participants expressed that they could not help but focus on the retro-reflective screen instead of the virtual object. Some took several tries over several minutes to wean themselves from focusing on the real screen to concentrating on the virtual buttons. Two expressed similarity of their experience with autostereograms (e.g., 'Magic Eye' pictures), which generally require several minutes of acclimation before the stereo image is perceived.

Even though the room was kept completely dark so that the participant could not notice any real objects, the retro-reflective screen was lit, necessarily so, by the HMPD. The hard, polished retro-reflective screen with high reflectivity had small engravings and minor scratches that were clearly visible. Having an unintended real object in the scene could have an effect on the results. For this experiment, the real screen may have made focusing on the virtual objects more difficult initially and some participants took longer to get adjusted to concentrating on the virtual buttons.

SUMMARY

A research scheme was developed to fully explore and quantify fundamental accuracy and performance time for direct object interaction in VE personal space. Two experiments were conducted to measure depth perception and task performance and to gain insight on which factors contribute to these measures, positively or negatively. Some performance differences were observed, but the results clearly showed that mean error was surprisingly low, less than 4 mm (SD = 2.2 or less) and mean performance time was within 1 second (SD = 0.2 second) for the simple task of tapping a button using HMPD technology, regardless of the VE type (RE, MR, or IVE) or whether other sensory cues (haptic or auditory) were provided.

Spatial Error for Depth Perception and Object Interaction

Using a display device that offered key visual cues in personal space, this study provided empirical evidence asserting that perception and interaction in VE is highly effective, with a mean spatial error of 4 mm or less, given correct, accurate binocular visual cues – accommodation, convergence, stereoscopy, and occlusion – instantaneously synchronized with proprioceptive cues. The results were upheld regardless of visual environment type and whether additional sensory modalities were presented or not.

The results are summarized in Tables 1 and 2. These show low mean spatial error (bias, accuracy, and precision < 4 mm) in all environment types (RE, MR, and IVE) and for both tasks (depth perception and object interaction). For depth perception, the differences between RE and MR and between RE and IVE were statistically significant. The differences between MR and IVE were not.

Table 1. VE Depth Perception Mean Error

	Depth Perception Mean Error				
	Bias (mm)	Accuracy (mm)		Precision (mm)	
RE	0.42	1.07		1.71	
MR	-0.63	2.95		3.89	
IVE	-2.06	3.89		3.61	
MR - RE	-1.05	1.88	p < 0.001	2.18	p < 0.001
IVE - RE	-2.48	2.82	p < 0.001	1.9	p < 0.001
MR - IVE	1.43	-0.94		0.28	

Table 2. VE Interaction Mean Error

	Object Manipulation Mean Error		
Buttons	Bias (mm)	Accuracy (mm)	Precision (mm)
Real	0.40	2.25	3.25
Virtual	0.60	3.10	3.95
Difference	0.20	0.85	0.70

Performance Time for Object Interaction

Performance time varied widely. Participants were asked to tap the buttons at a normal and comfortable pace. This pace varied widely even in RE. Regardless, the data showed a mean performance time of about one second per button tap. These results were similar for virtual or real buttons and with or without haptic or audio feedback. These findings were consistent with Schiefele's (2000) findings, where mean performance time for manipulating real objects was measured at 1.5 seconds. The results, however, were, on average, shorter than Schiefele's findings for virtual object manipulation of 3.5 seconds. The main difference in the two studies was the visual display, its fidelity, and its synchronization with proprioception.

The data also showed a slight difference with audio cue provided, but in the opposite direction as predicted. This degradation in performance was likely due to the participants' reaction to the delay in the audio feedback system, which was not previously considered and was realized only after the empirical data was collected and analyzed. The difference was statistically significant for the haptic-absent condition, as shown in Table 3 for completion time of the 32 taps.

Table 3. VE Interaction Mean Performance Time

Audio	*Performance Time (sec)*
Absent	33.6
Present	21.50
Difference	12.10 (p = 0.026)

(for Haptic-Absent Condition)

Additional Findings

The Simulation Sickness Questionnaire (SSQ) showed little side effects. Mean SSQ scores were less than five, indicating negligible symptoms (Kennedy et al., 1993). Two of the 42 participants indicated slight eye strain from trying to focus on the virtual buttons which were displayed at 0.4 m directly in front of the participant. The two indicated that the symptoms immediately disappeared when they stopped focusing on the stimuli. Much care was given to alignment of the visual display, matching the optics parameter with the software and graphics, and adjustment for the user, which may have contributed to the lack of side effects.

The presence of real objects appeared to have noticeable effects, both positive and negative. Some participants expressed difficulty focusing on the buttons because they found themselves focusing on the screen about 0.3 m farther away. They also identified similarities with autostereograms (two-dimensional images that create visual illusions of three-dimensional scenes), which generally take a minute or two of viewing before being resolved into the perceived stereo image. Some participants asserted that being able to see the real hand helped them see the virtual buttons more easily. Before lifting the hand into the scene, they reported seeing double images of the virtual buttons, indicating that they were focusing on the screen until the hand came into the scene near the buttons. Some asked if the light could be turned on dimly as it helped them see the virtual buttons better. Dialing up the room light reduced the relative intensity or contrast of the virtual buttons. At the same time, it illuminated nearby experimental apparatus and the hand, real objects for the eyes to focus on that were close to the same distance as the virtual objects. This highlights the positive effect of having nearby real objects next to virtual objects. It appears that participants tended to focus on real objects, which may have helped with awareness of and concentration on virtual objects nearby, but may have had negative effects with perception of virtual objects that were elsewhere (Ellis & Menges, 1997, 1998).

Although not statistically significant, the results showed a slight difference in bias, which was a measure of overshoot when adjusting the octahedron. On average, the virtual octahedron, used in the MR and IVE conditions, was adjusted about 2 mm (SD = 2.2 mm) farther away from the participant compared with the real octahedron used in the RE condition. The difference in bias could be due to the retro-reflective screen, which was another real object and an unintended real image in the experiment. The image screen was about 20 mm, farther away from the participant than the fixed stimuli, the cylinder. Ellis and Menges (1997, 1998) showed that depth of a virtual object could be perceived further than actual position, if a real object further in depth was also in the scene.

All participants expressed confidence in their ability to clearly perceive the virtual objects and know their exact positions. Many compared the experiment to the more familiar eye exams and commented afterwards that they had "passed the test"; even though individual results were not provided to the participants.

Occlusion or lack thereof, had a strong effect. After the experiment, some participants were asked to move a real reference rod right next to a virtual calibration line to verify alignment. It was noted that when the rod was placed right on the line, the participant tended to keep pushing the rod further away. Some expressed that the "line followed the stick". One potential explanation was occlusion. The participant may have expected the virtual line to eventually occlude the reference rod (which was not possible) and therefore kept on sliding it out. Another explanation would point back to the vividness or richness of the real object affecting the perception of the virtual object as discussed previously.

RECOMMEDATIONS

It is evident that virtual object interaction with correct visual cues presented, including synchronization with proprioception, is highly accurate (4 mm) and highly effective. Performance in MR is comparable to RE, with less than three millimeter difference, whether mean bias, accuracy, or precision is used as the measure of spatial error. Likewise, task completion times for simple button pushes are similar for RE and MR, approximately one second. This performance level suits many VE applications adequately. Moreover, this study was performed with simple geometry models. Higher accuracy may be achieved if the VE takes advantage of richer models or if other visual cues, such as size and motion perspective are included. This performance level exceeds the accuracy requirements for many applications including cockpit simulations, where controls are set at about 12 mm apart, the minimum separation recommended for ergonomic designs (Boff & Lincoln, 1986). The HMPD and MR technology has matured to a level where more complex and practical experimentation can be conducted for application specific environments.

This study showed that in a high-fidelity visual environment, the addition of an inaccurate cue, specifically a delayed sound cue, degrades performance compared with absence of the cue. For VE designs involving button pushes, having no audio feedback may provide a more effective VE than having an unsynchronized one. It also indicates that additional studies are required to examine and control for the impact of unintentional objects in the field of view when using VE.

ACKNOWLEDGEMENTS

The authors thank NAWCTSD 4.6 organization, UCF's Industrial Engineering Department and ODA Lab, eMagin Corp., DiSTI Corp, Design Interactive Inc., and RSK Assessments Inc. for their support in this study. Special thanks to Mike Muehlbauer, CDR Kevin Rasch, CDR Allen Crisp, and LCDR Michael Sypnewski at Naval Air Station, Jacksonville for providing the ideal NVG lab and participant pool for the experiments.

REFERENCES

Arsenault, R. & Ware, C. (2000). Eye-hand co-ordination with force feedback. *CHI Letters, 2,* 408-414.

Boff, K. R. & Lincoln, J. E. (1986). *Engineering data compendium: Human perception and performance*. Wright Patterson Air Force Base, OH: Harry G. Armstrong Aerospace Medical Research Laboratory.

Cutting, J. E. & Vishton, P. M. (1995). Perceiving layout and knowing distances: The interaction, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 69-117). San Diego, CA: Academic Press.

Drexler, J., Kennedy, R., & Malone, L. (2009). Virtual environment sickness and implications for training. In D. Nicholson, D. Schmorrow, & J. Cohn (Eds.), *Handbook of virtual environments for training and education Vol. 2*, (pp. 219-239). Westport, CT: Praeger.

Ellis, S. R. & Menges, B. M. (1997). Judgments of the distance to nearby virtual objects: Interaction of viewing conditions and accommodative demand. *Presence-Teleoperators and Virtual Environments, 6*, 452-460.

Ellis, S. R. & Menges, B. M. (1998). Localization of virtual objects in the near visual field. *Human Factors, 40,* 415-431.

Hua, H., Girardot, A., Gao, C., & Rolland, J. P. (2000). Engineering of head-mounted projective displays. *Applied Optics, 39,* 3814-3824.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire (SSQ): An enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology, 3(3),* 203-220.

Kennedy, R. S., Dunlap, W. P., Jones, M. ., & Stanney, K. M. (1996). *Screening users of virtual reality systems for after-effects such as motion sickness and balance problems* (Final Rep. No. NSF1-96-4). Arlington, VA: National Science Foundation.

Milgram, P. & Colquhoun, H. (1999). A taxonomy of real and virtual world display integration. In Y.Ohta & H. Tamura (Eds.), *Mixed reality - Merging real and virtual worlds* (pp. 5-30). Tokyo, Japan: Ohmsha.

Milgram, P. & Kishino, F. (1994). A taxonomy of mixed reality visual-displays. *IEEE Transactions on Information and Systems, E77D*, 1321-1329.

Nguyen, L., Mead, A., Fidopiastis, C., & Rolland, J. (2004). Enhancing Virtual Environment Interaction Using a Head Mounted Projection Display (HMPD). *Proceedings of I/ITSEC*, Orlando, FL, Nov 2004.

Nguyen, L., Mead, A., Malone, L., Williams, K., Fidopiastis, C., & Rolland, J. (2005). Are Correct Visual Cues Enough for Direct Manipulationof Virtual Objects? *Proceedings of the Human Computer Interaction International (HCII) Conference*, New Orleans, LA, Aug 2005.

Rolland, J.P., Biocca, F., Hamza-Lup, F., Ha, Y., & Martins, R. (2005). Development of head-mounted projection displays for distributed, collaborative augmented reality applications. *Presence: SI Immersive Projection Technology, 14(5),* 528-549 (2005).

Rolland, J. P., Meyer, C., Arthur, K., & Rinalducci, E. (2002). Method of adjustments versus method of constant stimuli in the quantification of accuracy and precision of rendered depth in head-mounted displays. *Presence-Teleoperators and Virtual Environments, 11*, 610-625.

Schiefele, J. (2000). Realization and evaluation of virtual cockpit simulation and virtual flight simulation. Thesis, Technical University Darmstadt.