Studies of depth judgments in a see-through head-mounted display

Jannick P. Rolland[a], and Kevin Arthur[b]

[a] Center for Research and Education in Optics and Lasers (CREOL),
Departments of Electrical and Computer Engineering, and Computer Science,
University of Central Florida, Orlando, FL 32766-2700
[b]Department of Computer Science, University of North Carolina at Chapel Hill, NC 27599-3175

ABSTRACT

The goals of this research are to measure accuracy and precision of depth perception of 3D objects in see-through head-mounted displays for near-field visualization (i.e. within arm length) and to set benchmarks for current technology. In this paper we present results on accuracy and precision of perceived depth for stimuli of various forms (i.e. of varying shapes and sizes) presented either in a side by side or in a top-bottom configuration. A finding of these experiment is the unexpected relatively high human subject variability measured which we postulate to be correlated with aspects of the methodology used. Further investigations on methodology related to these types of experiments are being conducted. Finally, we report a new benchmark for the precision of perceived depth of 7 mm.

Keywords: quantification, depth, size, shape perception, virtual environments, benchmark, head-mounted displays

1. INTRODUCTION

The assessment of depth performance in head-mounted displays (HMDs) is critical to various visualization applications such as engineering, medical, and scientific applications[1]. The aims of this research are to enhance our understanding of human perception in such displays as well as evaluate the impact of engineering factors on perceived depth so that the design of HMDs and the presentation of virtual objects can be optimized. This research addresses two fundamental questions[2]: how accurate is the percept of depth in HMDs and how variable is the percept?

The image presentation and content for this research is of a different nature from investigations where the images being displayed are collimated for infinity viewing of symbology information[3-4]. This research is concerned with assessing depth perception of 3D virtual objects presented in the near field of vision, under 2 m away. Near field of vision is also referred to as *personal space* by Cutting and Vishton (1995) who make a distinction between *personal space* which extends to 2 meters, *action space,* which is just beyond personal space, and *vista space* which is beyond about 30 meters[5]. Experiments on the assessment of depth perception in HMDs for the near field of vision were also conducted by Ellis and colleagues[6-7]. Their emphasis has been on comapring binocular versus monocular, and biocular viewing conditions. As a consequence of a study of depth perception in personal space, results of the experiments will be especially relevant to applications involving visualization or manipulation within arms reach.

We report on two depth-perception experiments : a main and a control experiment, conducted using a bench prototype, optical see-through HMD shown in Figure 1. In both experiments, the method of constant stimuli was used and human subjects observed objects of various forms located at about 0.8 meter away. A stimulus form was characterized by a shape and a size factor. In the main experiment, three virtual shapes (i.e. a cube, a cylinder, and an octahedron) of various sizes were investigated. In the control experiment, two medium size octahedra were compared in a side by side or in a top-bottom presentation. All stimuli were displayed at a depth close to 0.8

meter, and the virtual (or optical) image planes were also located at this depth. In all experiments, the optical baseline of the viewer was set for each user to match the user's inter-pupillary distance.

There are many sources of information to depth from static to dynamic and monocular to binocular, and it is of ultimate interest to identify which sources will most effectively render the desired percept of depth. Cutting and Vishton note that typically six sources of information are most relevant in personal space[5]: occlusion, stereopsis, head-motion parallax, relative size, convergence and accommodation. We report here on experiments using: static, stylized, and untextured stimuli; stereopsis, convergence and accommodation as main sources of information to depth; along with the combination of shape and shading as a secondary source of information to depth[16].

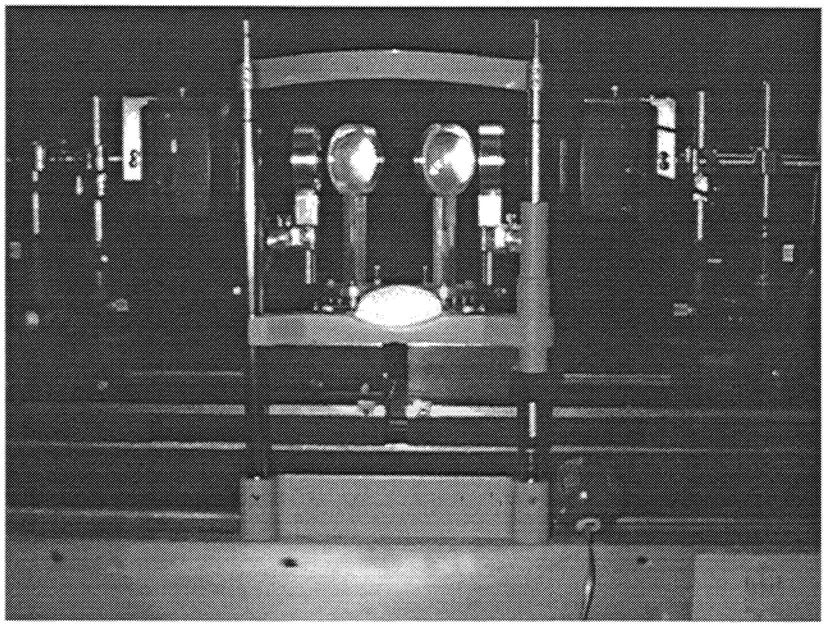

Figure 1. *Experimental setup. Shown is a chin rest for the human subject, the optical viewer, two LCD displays, and a two-button input device in the form of a ball for easy handling.*

The objects were displayed without any grounding support or any contextual background, as if they were floating in space. While the play of light on forms also yields shadows in most instances[8-9], shadows were not included as a source of information to depth in the experiments presented here. We share the view of Cutting and Vishton (1995) that shadows provide a source of information to shape, rather than a source of information to depth per se, at least for static objects[5]. This paper aims at the exploration of the role of object form which includes shape and mean size, on depth perception.

2. METHODS

Apparatus: We presented the stimuli using a third generation prototype of an optical see-through head-mounted display[10]. Two miniature LCD displays were used of 25.5x33.9 degrees field of view each. The number of addressable lines were 429x586, which yielded an effective resolution of 3.6 minutes of arc at the eye point. Depixelization screens from Microsharp Technology were added to the LCDs to blur the boundaries between pixels while minimizing the induced overall blur of the image.

Calibration: Careful calibration of head-mounted displays is typically of fundamental importance for assessment of depth perception. In the case of this study, high-accuracy calibration with respect to the world coordinate system was perhaps not as critical as in the general use of an augmented reality setup because virtual objects were compared only to each other. No comparison of real and virtual objects was conducted in the reported experiments. Nevertheless, a careful calibration was done on the system. The procedure used a laser beam alignment technique we developed to align the mirror, the lens, and the display.

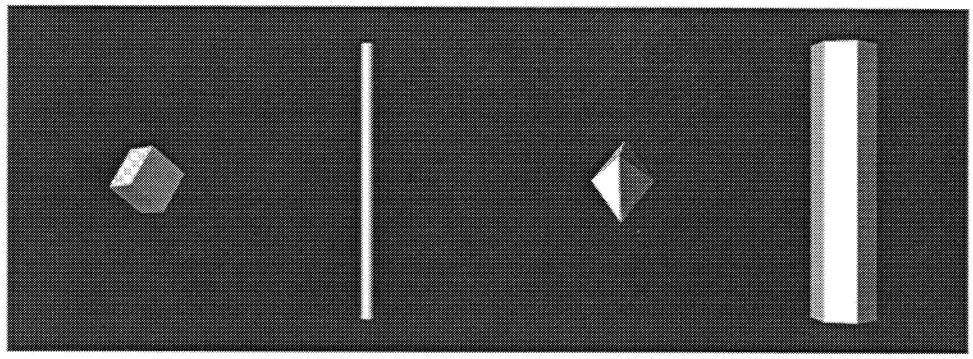

Figure 2. *Stimuli used in the investigation of the role of form on perceived depth. The smallest cube (40 mm on one side) and octahedron (size adjusted to be equal to the cube in volume) are shown here. Other sizes were 60 mm and 80 mm for the cube and the equivalents for the octahedron. The smooth cylinder and the larger faceted cylinder are also shown.*

Subjects : Five human subjects were selected to participate in the experiments. Their selection was based on a visual acuity of 20/20 (corrected or uncorrected) on the eye chart, and on their performance on the Howard-Dolman test[11].

Stimuli and task : The stimuli were simple geometrical forms such as a cube, a smooth cylinder, a faceted cylinder, and an octahedron (two five-sided pyramids back to back and joined at the base) as shown in Figure 2. The subject's task was to judge the depth of the objects' centers. The octahedron might have been the simplest stimulus with respect to finding its center because the tips of the two pyramids were on the axis that passed through the center of the octahedron. Thus, we postulated that a subject could more easily judge the distance of the object's center in depth by using the tips of the pyramids as cues. The center of the cylinder might also have been fairly easy to estimate based on the cylinder's projected shape. The cube, however, was presented at an odd angle (45 degree rotation around the longitudinal axis (z) and 45 degree rotation around the vertical axis (y)) , and thus its center might have been more difficult to judge since its projected shape was not symmetric.

Three cube sizes were chosen – small (40 mm per side), medium (60 mm per side), and large (80 mm per side). The cubes roughly subtended 4, 6, and 8 degrees visual angle, respectively. The octahedron shapes were constructed with the base side length equal to the total height (the height of the two pyramids back to back), and three sizes were chosen so that the volumes of the three octahedra were equal to the volumes of the three cubes. As a result, the small cube and the small octahedron gave at least the subjective impression of occupying roughly the same volume in space. This impression also generalized to the two other sizes.

The small cylinder was smooth-shaded and had a diameter of 13 mm (equivalent to 1 degree visual angle) and a height 235 mm (or 17 degree visual angle) as used in a previous experiment[10]. A larger, faceted cylinder (i.e. a hexagonal cross-section) of the same height was created with approximately four times the width of the small cylinder. The larger cylinder was given structure with added facets to create secondary cues to depth. The cylinder's maximum diameter was 58 mm; its average diameter was approximately 52 mm (that is the diameter of a circle covering the same area as the hexagon with distance 58 mm from the center to each corner).

The computer graphics were generated by Pixel Planes 5, a massive parallel graphics engine developed at the University of North Carolina at Chapel Hill under the direction of Henry Fuchs and John Poulton [12]. Left- and right-eye virtual images were warped to compensate for the system's optical distortion. Compensation for distortion was applied using a look-up table[13-14].

This study focuses on the study of stereopsis and convergence, however the stimuli were optically presented so that accommodating at roughly the converged distance would yield least image blur. The stimuli did not occlude each other, the head of the observer was fixed in a static chinrest, and we aimed at removing relative size as a source of information to depth by having the subjects compare objects of different shapes. We assumed here that relative size is most directly used between objects of similar shapes. It is most likely the case, however, that relative size also operates as a source of information to depth for objects of different shapes occupying roughly the same volume in space (e.g. a cube and an octahedron of the same volume). In an attempt to eliminate judgments based on relative size, we varied the size of one object randomly from trial to tril by 15% and we asked the human subjects to ignore size information telling them that it was random and unrelated to relative distance.

Presentation : Virtual stimuli of different shapes were presented two at a time (e.g. a cube and an octahedron) either side by side in the main experiment or in a top-bottom configuration in a control experiment. Unlike the field of view on a CRT display, whose borders limit the extent of the computer graphics, the field of view in a virtual-reality setting is limited by the imaging optics rather than the dimensions of the displays. For the user, the view is similar to that experienced when looking through a window into another 3D environment. The edges of the displays are not visible, but the field of view is limited by the optical components. Since the display edges are too close to the eyes to be in focus, each display edge has become, in effect, a transparent edge.

The spacing between the objects was computed so that the lateral distance between the two object centers was proportional to a linear measure of the total size of the objects. For the cylinders, we considered the measure of size to be the diameter. For the cube and the octahedron, we considered the size to be the diameter of a sphere with volume equal to that of the cube or the octahedron. Diameters for small, medium, and large spheres were 49.6 mm, 74.4 mm, and 99.3 mm. The calculation for object separation was based on the assumption that the spacing between a small cube and a small octahedron was 110 mm, center to center (± 55 mm from the center of the binocular field of view). The other distances were computed so as to keep the ratio "distance/average-size" constant and equal to 2.218, where 2.218 was computed as 110 / [1/2 (49.6 +49.6)]. The cubes, octahedra, and cylinders were presented as shown in Figure 2. The octahedron was tilted down around an horizontal axis by pi/12.

Relative depth perception was assessed using a two-alternative forced-choice method of constant stimuli. Ten values of depths for the object on the right in the main experiment (or on top for the control experiment) were presented around the nominal depth of the object on the left (or at the bottom) in step sizes of 6 or 7 mm. Human subjects were asked to judge whether the object on the right (or above) was in front or in the back of the object on the left (or below). Responses were entered on a two button hand-held device shown in Figure 1..

The size of the moving object located on the right in the main experiment (and on the top in the control experiment) was varied by ±15% around its mean value. This scaling aimed at preventing human subjects from judging depth essentially based on relative size estimation of the moving stimuli around its mean displayed size. We found that human subjects can learn the mean size of an object among a subset of displayed sizes in only ten or twenty trials. Such a scaling is a fairly standard procedure in psychophysics for controlling the strategy used in making a decision. We shall come back to this point in the discussion section. Here, the desired strategy is judging the depth of an object relative to another object, irrespective of size values.

Data Analysis : Data were analyzed using Probit analysis. For each psychometric function, the point of subjective equality (50% point on the psychometric curve) as well as the slope of the fitted curve (measured from the 16% and 84% points on the psychometric function) were estimated. The departure of the point of subjective equality (PSE) from the nominal value of 0.8 meter measures the accuracy of perceived depth, referred to as Delta-PSE (i.e. Delta-PSE= Measured Value – Nominal Value). The slope of the curve, referred to as discrimination threshold, measures the precision of perceived depth.

Discrimination threshold values reported here indicate upper-bound values. The step size of 6 or 7 mm was then kept constant across conditions. We postulate that threshold values estimated to be below 5 mm by the Probit analysis were not accurately estimated. In the data reported here, we then set to 5 mm all thresholds estimated to be below 5 mm. Thus, reported thresholds values indicate upper-bound values. Results from individual human subjects indicated that a smaller step size could be used in future experiments if more precise values of the thresholds were sought.

3. RESULTS AND DISCUSSION

Detailed results for the comparison of the cylinders with the cubes and the cylinders with the octahedra are shown in Figures 3 and 4. Figure 5 shows a summary for all forms compared. Figure 6 shows typical data for two individual human subjects under four of the conditions tested. Figure 7 shows the results for the control experiment. In all figures, "ALL" means the average over four (Figures 3-5) or five human subjects (Figure 7). Each data point on the graphs corresponds to an average across human subjects of 700 trials or three psychometric functions of 300 (1) and 200 (2) trials per human subject. Errors bars in Figures 3-5 correspond to across subject variability. In Figure 6 and 7, error bars correspond to within subject variability.

In Figure 3 and 4, results are plotted separately for each pair of stimuli (e.g. a cube and a cylinder). The two figures on the left represent accuracy of perceived depth measured as the perceived depth minus the nominal depth of 0.8 meter and referred to as Delta-PSE. Given two objects (e.g. a cube fixed in space on the left and a cylinder moving in depth on the right), the independent variable is the scale (i.e. the mean size) of the object on the right, which is represented on the x-axis. The two figures on the right represent the precision of perceived depth referred to as discrimination thresholds.

Results of the experiment show that an average upper bound for discrimination thresholds is about 7 mm regardless of the shape or size of the stimuli. The lower the discrimination threshold, the higher the resolution in depth, thus the higher the performance. These discrimination thresholds were found to be significantly smaller than previously reported in a similar experiment (a value of 15 mm was then reported) as a consequence of using higher resolution displays, depixelization screens, and anti-aliasing computation[10]. This value of 7 mm constitutes a new benchmark for future comparison of similar systems.

With respect to the accuracy of perceived depth, our hypothesis was that the task would be easier, therefore the accuracy would be higher, when using an octahedron and a cylinder instead of a cube and a cylinder because of the symmetry properties of the stimuli in the former condition. Moreover, it was unclear whether using a faceted larger cylinder would be beneficial compared to using a smooth small cylinder.

Results of the experiment clearly show that the large, faceted cylinder yields generally higher bias in perceived depth than the smooth, small cylinder. This is shown by an overall upper shift of the data points in Figures 5a and 5c, as well as in Figures 3a and 4a. The shift indicates that while the facets on the cylinder may subjectively give more 3D percept to the cylinder as suggested by the human subjects, the human subjects do in fact judge the 3D location of the small smooth cylinder more accurately than that of the larger, faceted cylinder.

Concerning the benefit of using an octahedron versus a cube to assess depth perception, results seem to indicate a slight preference for the medium size octahedron in terms of minimizing perceived bias for a pair of stimuli, however further investigation is required to evaluate a possible gain once human subjects variability has been lessened. In any case, we postulate that the gain in using a symmetrical stimuli is perhaps less than originally predicted.

Human subject's variability for the same conditions was examined by collecting three psychometric functions per human subject per condition. The design was balanced in the sense that all the conditions were run before they were repeated, and within a run, the conditions were randomly presented. The variability that found to be higher than expected, especially within subject's variability. Typical data for two individual human subjects are reported in Figure 6 for a few conditions. Because the task of the observers involved little if any cognitive processing, we did not expect the subject's responses to improve or degrade systematically through time. Scrutiny of the results though time (not plotted here) indeed show that depth errors for all pairs of stimuli varied uncorrelated within individuals. Variability over time for equivalent trials was random.

We postulate that at least part of the observed variability is a direct consequence of varying the size of one of the two displayed objects by ± 15%. This is supported by Davis and colleagues' as well as Cutting and Vishton's findings that relative size of objects is a strong cue to depth perception within a viewing distance of 1 meter[2,5]. The variation in size imposed on the objects was applied to encourage a strategy for depth judgment. We are currently investigating whether other methods such as the double staircase method with 2 alternative-forced-choice, for example, yields less variability and repeatable judgments.

We completed this study of form by looking at the depth judgments of two identical shape objects. We chose the medium-size octahedron in this case. We predicted that the point of subjective equality for the object being judged relative to the other should equal the nominal depth of the other object. We also repeated the experiment with the two octahedra displayed on top of each other (vertical or 90 degrees) instead of side by side (horizontal or 0 degrees). In this case, the side by side stimuli were simply rotated 90 degrees around the center point between the two objects to arrive at a vertical configuration.

Results, presented in Figure 7, show that judgments for the vertical condition were in average more accurate than for the horizontal condition. We considered that this finding could result from a handedness tendency; however, it is unlikely because one of the human subjects (i.e. KE) was left-handed, and the observed bias for subject KE is in the same direction as the measured bias for most right-handed human subjects (an exception was subject VA who was right-handed). We postulate that the observed bias is most likely a result of a tendency to fixate non-symmetrically with respect to each object, perhaps combined with some eye dominance tendency. This hypothesis is the subject of a study of its own on the impact of controlled fixations at various locations with respect to the two objects, versus free fixation in the virtual environment. In follow-up experiments, we propose to further investigate the impact of controlled fixation versus free fixation on depth-judgment performance in a virtual environment because it is of critical importance to future studies of depth perception as well as to the use of the technology for visualization of real-world applications.

Figure 7b also indicates that discrimination thresholds for the vertical condition are about equivalent to those for the horizontal condition. This control experiment thus suggests that a vertical configuration would be beneficial in future experiments to minimize systematic perceptual bias and judgment variability, at least for spatially confined stimuli such as the cube and the octahedron. Indeed, if a cylinder were used, it is important to note that when displayed horizontally, human subjects' ability to judge depth from stereo disparity would be highly reduced. Zerbolio and Walker (1989) showed that average errors in depth judgments were roughly the same for two rods displayed horizontally whether monocular or binocular vision was used. Errors on the order of 100 mm were reported[15].

4. CONCLUSION

We presented two experiments on the quantification of depth perception in virtual environments. Using the method of constant stimuli, human subjects judged the relative depth of pairs of virtual objects of three shapes and various sizes. We measured accuracy and precision of perceived depth as a function of the shape and the size of the stimuli. Precision of perceived depth was measured to be less than or equal to 7 mm, a factor of two improvement from a previous study[10]. This finding sets a new benchmark for the evaluation of comparable systems. Results also indicate that the use of a small and smooth cylinder is beneficial compared to a faceted cylinder contrary to subjective expectations. Using a symmetrical stimuli such as an octahedron instead of a cube yields a slight benefit in the accuracy assessment of perceived depth but the increase in performance is smaller than predicted. Performance is in fact currently limited by subjects's variability which we postulate to be at least in part an artifact of aspects of the methodology. The general issue of the choice of a methodology for depth perception assessment in head-mounted displays is far from being resolved and constitutes a remaining challenging issue.

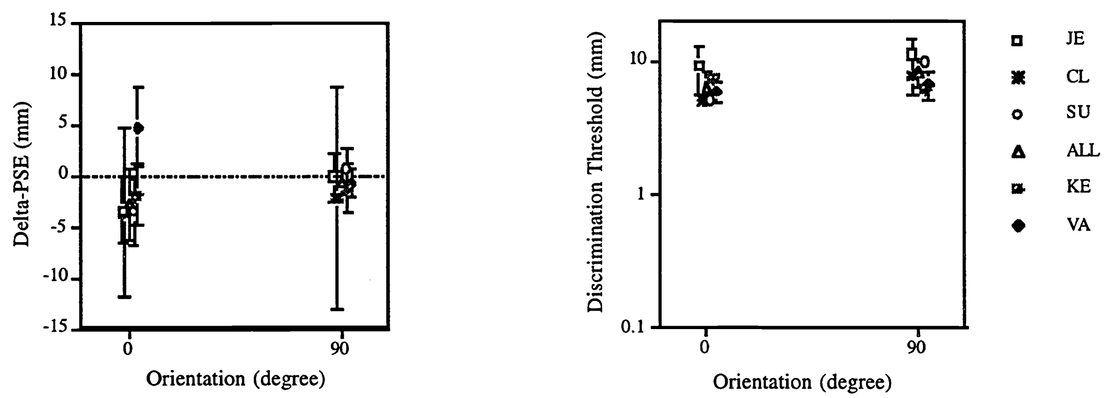

Figure 7. *Control Experiment: Accuracy (left graph) and precision (right graph) of perceived depth for two medium size octahedra (equivalent in volume to a 60 mm cube) displayed side by side (orientation of 0 degrees) and in a top- and bottom-configuration (90 degree orientation). Results are presented for five human subjects. Averaged performance over the five human subjects is also presented. Error bars for individual human subjects represent variability within subject. Error bars for the average condition (referred as "ALL") represent variability between human subjects.*

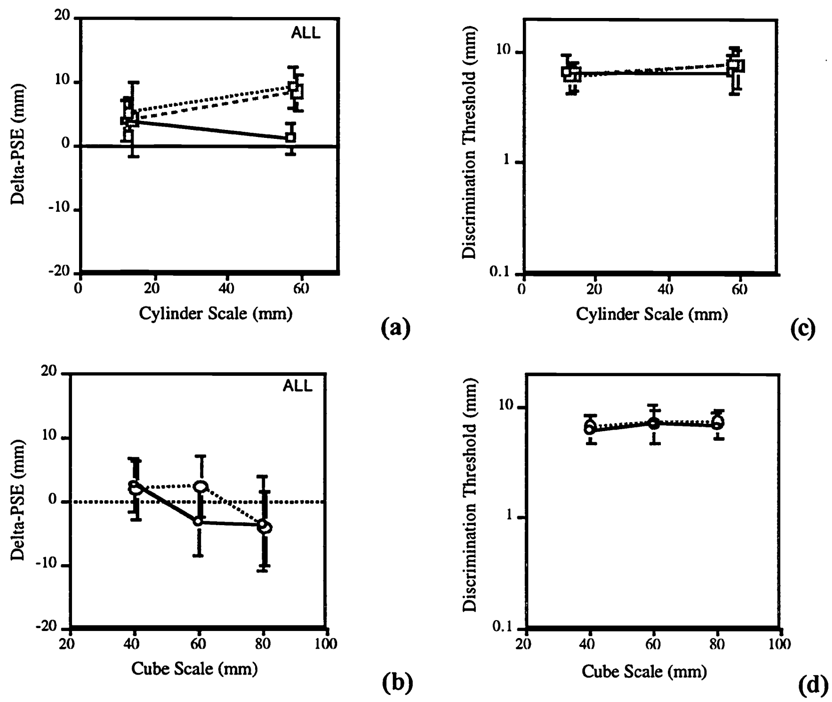

Figure 3: *(Left) Accuracy of relative perceived depth when judging the depth of: (a) a cube relative to a cylinder; and (b) a cylinder relative to a cube. (Right) Precision of perceived depth for the same conditions.*

"ALL" denotes that the data are summarized here for four human subjects.

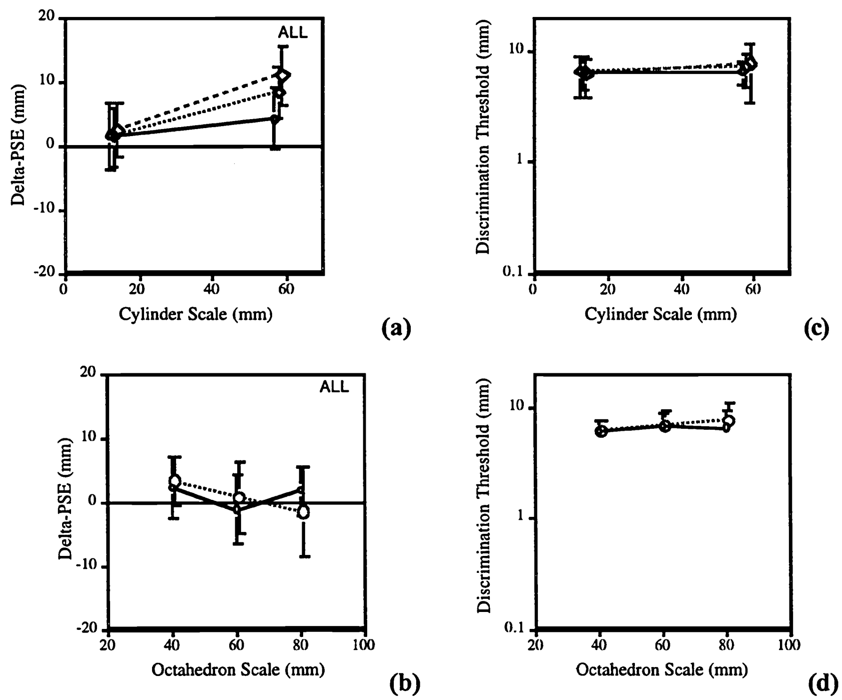

Figure 4: *(Left) Accuracy of relative perceived depth when judging the depth of: (a) an octahedron relative to a cylinder; and (b) a cylinder relative to an octahedron.*

(Right) Precision of perceived depth for the same conditions.

13 mm cylinder
58 mm cylinder

"40" mm octahedron
"60" mm octahedron
"80" mm octahedron

40 mm cube
60 mm cube
80 mm cube

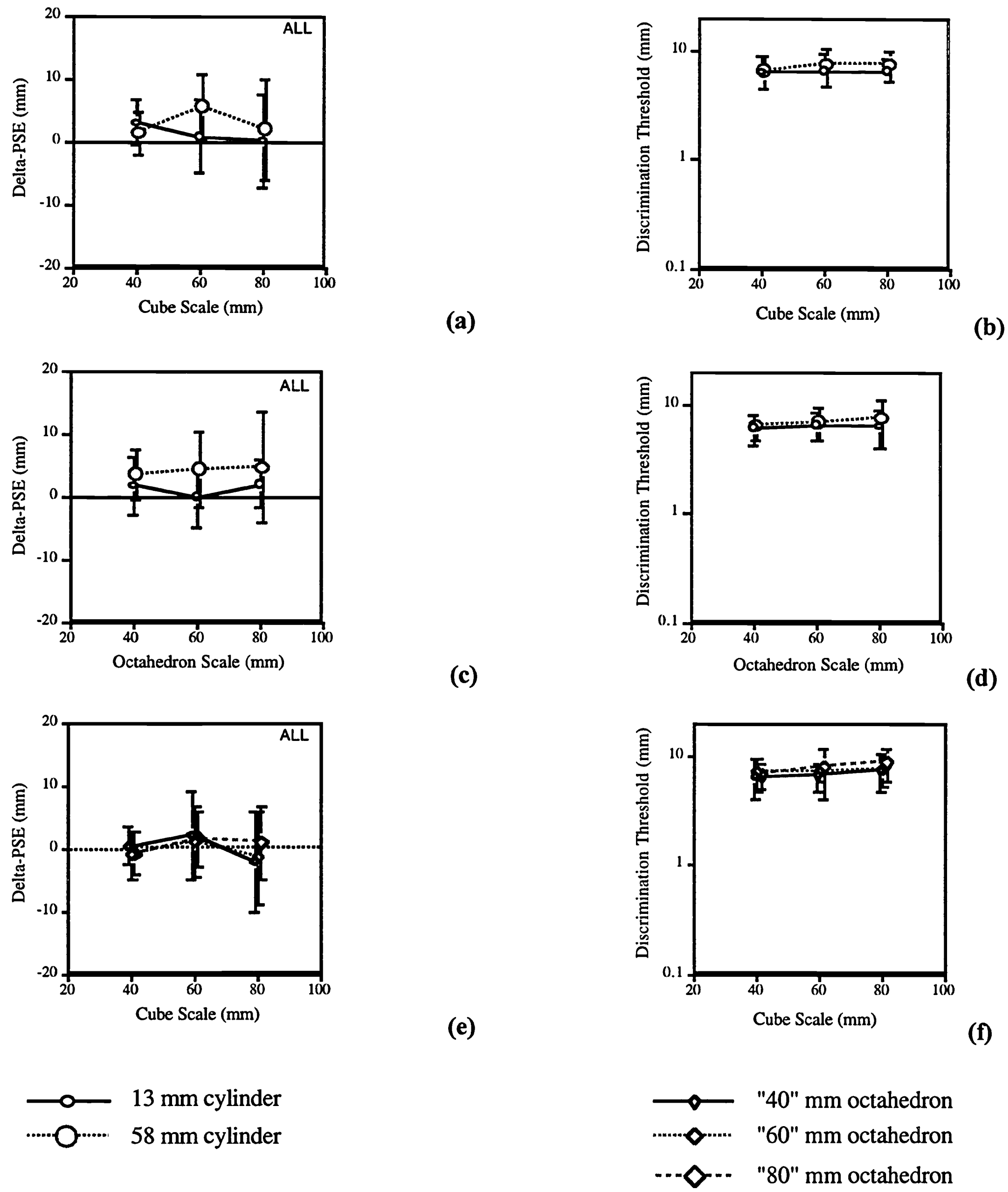

Figure 5. *(Left) Average accuracy of perceived depth across mirror pairs of virtual stimuli (e.g. a cube with a cylinder and a cylinder with a cube) when judging the depth of one with respect to the other. (Right) Precision of perceived depth for the same conditions.*

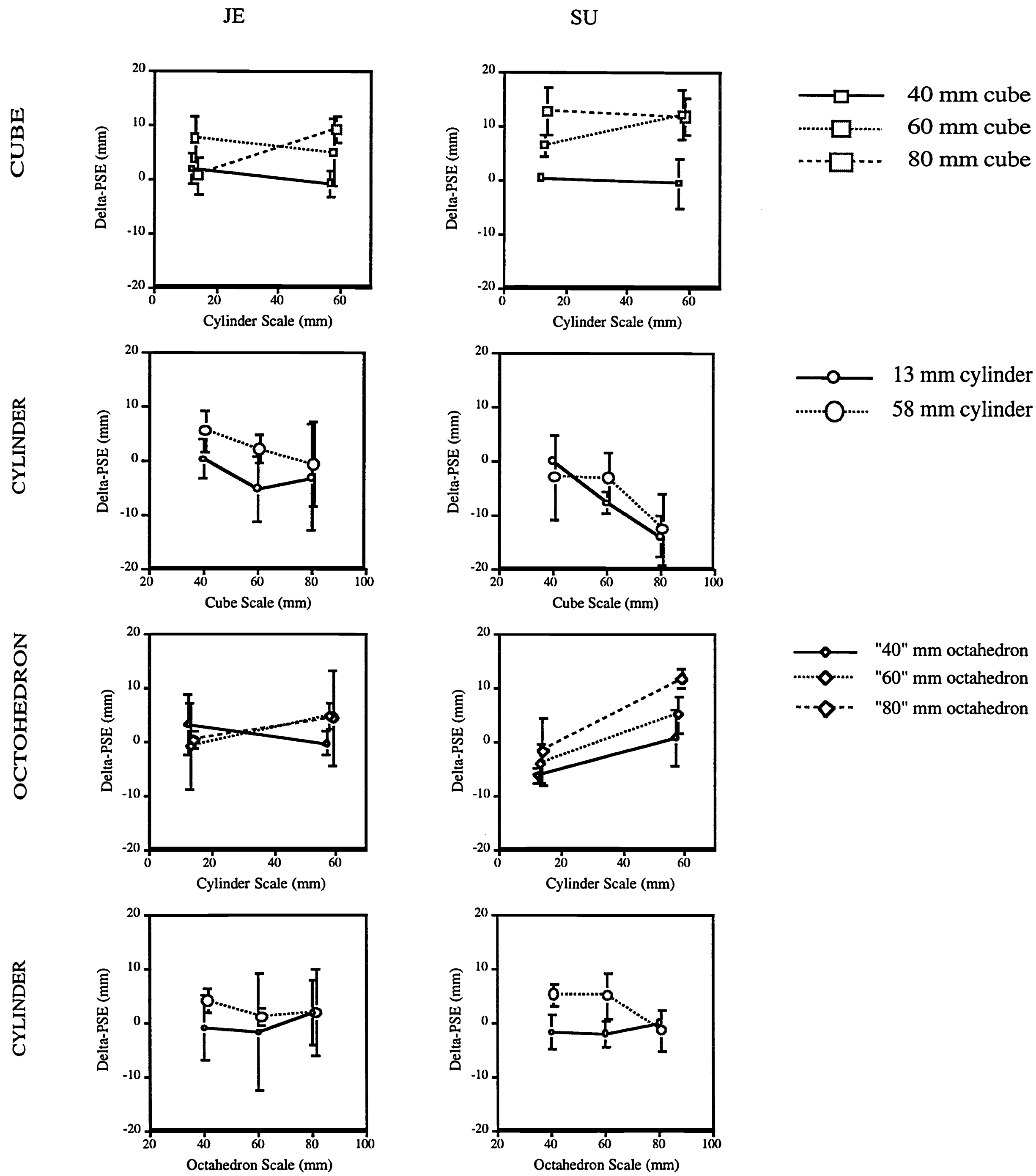

Figure 6 : *Accuracy of perceived depth for two human subjects (JE and SU) across four pairs of stimuli. Within-subject variability is shown.*

ACKNOWLEDGMENTS

We thank Robert S. Kennedy and Edward Rinalducci from the University of Central Florida (UCF) as well as Denise Varner from the Southwest Research Institute for their stimulative discussions on methodology. We also thank the Computer Graphics team at the University of North Carolina at Chapel Hill for their invaluable support while conducting this research. We are grateful to both Terry Allard and Lisa Achille from the Office of Naval Research for their overall support and their invigorating discussions about the research as a whole. We finally thank Microsharp Technology for donating depixelization screens for the LCD displays. This research was supported by the Office of Naval Research under grant N00014-94-1-0503.

REFERENCES

1. Bajura, M., Fuchs H., and Ohbuchi R. "Merging virtual objects with the real world". *Proc. of SIGGRAPH 92, in Computer Graphics* 26(2), 203-210, 1992.
2. Surdick, R.T., Davis, E.T., King, R.A., Corso, G.M., Shapiro, A., Hodges, L., and Elliot, K., "Relevant cues for the visual perception of depth: is where you see it where it is?", *Proceedings of the Human Factors and Ergonomic*s Society 38th Annual Meeting, 1994.
3. Roscoe, S.N., Olzak, L.A., and Randle, R.J. "Ground-referenced visual orientation with imaging displays: monocular versus binocular accommodation and judgments of relative size", *Proc. of the AGARD Conf. on Visual Presentation of Cockpit Information Including Special Devices for Particular Conditions of Flying (pp. A5.1-A5.9). N*euilly-sur-Seine. France: NATO, 1976.
4. Palmer, E., and Cronn, F.W., "Touchdown performance with computer graphics night visual attachment", *In Proceedings of the AIAA Visual and Motion Simulation Conference, 1-6 New York: American Institute o*f Aeronautics and Astronautics, 1973.
5. Cutting, J.E., and Vishton, P.M., "Perceiving Layout and Knowing distanes: the integration , relative potency, and contextual use of different information about depth", *In the Perception of Space and Motion, Ed. W. Epstein and S. Rogers, Academic Press*, 69-117, 1995.
6. Ellis, S.R., and Bucher, U.J., "Distance perception of stereoscopically presented virtual objects optically superimposed on physical objects in a head-mounted see-through display", *Proc. of the Human Factors and Ergonomic Society, Nashville*, 1994.
7. Ellis, S.R., and Menges, B.M., "Judged distance to virtual objects in the near visual field", *Proc. of the Human Factors and Ergonomic Society, San Diego*, 1995.
8. Bülthoff, H.H., and Mallot, H.A., "Interpreting depth modules: stereo and shading", *Journal of the Optical Society of America A,* 35, 357-37, 1988.
9. Yonas, A., and Granrud, C.E., "The development of sensitivity to kinetic, binocular, and pictorial depth information in human infants", *In D.J. Ingle, M. Jeannerod, and D.N. Lee (Eds). Brain mechanisms and spatial vision. Dordrecht, the Netherlands: Mertinus Nijhoff,* 113-145, 1985.
10. Rolland, J. P., Ariely, D., and Gibson, W., "Towards quantifying depth and size perception in virtual environments", *Presence* , 4(1), 24-49, 1995.
11. Howard, H.J.,"A test for the judgment of distance", *American Journal of ophthalmology,* 2, 656-675, 1919.
12. Fuchs, H., Poulton, J., Eyles, J., Greer, T., Goldfeather, J., Ellsworth, D., Molnar, S., and Israel, L., "Pixel-Planes 5: A heterogeneous multiprocessor graphics system using processor-enhanced memories," Computer Graphics (Proc. of SIGGRAPH '89), 23 (3), pp79-88, 1989.
13. Robinett, W., and Rolland, J.P.,"A computational model for the stereoscopic optics of a head mounted display", *Presence, 1* (1), 45-62, 1992.
14. Rolland, J.P., and Hopkins, T., "A method of computational correction for optical distortion in head-mounted displays", *Technical Report TR-93-045, Dept. of Computer Science, University of* North Carolina at Chapel Hill, 1993.
15. Zerbolio, D.J., and Walker, J.T., "Factorial design: binocular and monocular depth perception in vertical and horizontal stimuli", *Teaching of Psychology, 16(2,* 1989.
16. Boring, E.G. *Sensation and perception in the history of experimental psychology.* New York: Appleton-Century-Crofts, 1942.