

**J. P. Rolland**

jannick@odalab.ucf.edu

**C. Meyer**

ODA-Lab, School of  
Optics/CREOL  
4000 Central Florida Blvd.  
University of Central Florida  
Orlando, FL 32766-2700

**K. Arthur**

Department of Computer Science  
University of North Carolina  
Chapel Hill, NC 27599

**E. Rinalducci**

Department of Psychology  
University of Central Florida  
Orlando, FL 32766

# Method of Adjustments versus Method of Constant Stimuli in the Quantification of Accuracy and Precision of Rendered Depth in Head-Mounted Displays

---

**Abstract**

The utilization of head-mounted displays (HMDs) in high-end applications such as medical, engineering, and scientific visualization necessitates that the position of objects be rendered accurately and precisely. Accuracy and precision of rendered depth for near-field visualization were measured in a custom-designed bench prototype HMD. Experimental results were compared to theoretical predictions established from a computational model for rendering and presenting virtual images by Robinett and Rolland (1992). Such a theoretical model provided the necessary graphics transformations required so that rendered virtual objects be perceived at the rendered depth in binocular HMDs. Three object shapes of various sizes were investigated under two methodologies: the method of constant stimuli modified for random size presentation and the method of adjustments. Results show a 2 mm and an 8 mm performance for the accuracy and the precision of rendered depth in HMDs, respectively. Results of the assessment of rendered depth in HMDs for near-field visualization support employing the method of adjustments over the method of constant stimuli whether or not the method of constant stimuli is modified for random size presentation.

**I Introduction**

The basic principle underlying perceiving three dimensions using various stereoscopes, including head-mounted displays (HMDs), calls for the presentation of slightly disparate images to the eyes of the user (Wheatstone, 1838). HMDs differ from common stereoscopes in their ability to update the displayed images at interactive speed (Sutherland, 1968). It is well accepted that most stereoscopic devices provide a striking impression of depth.

With the development of applications such as medical, engineering, and scientific visualization, a striking impression of depth is not sufficient. Users typically need to perform tasks that include the fine manipulation of objects at arm length. Such tasks pose challenges for the engineering of the display device (Rolland & Fuchs, 2000; State et al., 1996). What is important is that depth can be rendered with accuracy and precision of approximately 1 mm. Rendered depth is especially critical for tasks in which errors may have high cost

associated with them, such as in some of the medical applications (Biocca & Rolland, 1998).

We previously reported (Rolland, Ariely, & Gibson, 1995) an investigation to address two fundamental questions: how accurate is rendered depth in HMDs and how variable is the depth percept? Results of the experiments showed various forms of perceptual bias and variability in the data. In an effort to design systems that can be employed effectively in high-end applications, we are investigating engineering issues related to HMDs and are employing psychophysical methods to assess the improved technology in a closed loop. In this paper, we specifically focus on the methodology to make measurements and we seek to establish new benchmark performance on an improved HMD device, given a computational model for rendering and displaying the stereoscopic images.

The experimental studies presented in this paper are based on the computational model of rendered depth and images presentation by Robinett and Rolland (1992), in which various possible errors for rendered depth of virtual objects were discussed. The computational model includes the specification of all required graphics transformations (including optical distortion) from the virtual world to the left and right eyepoints in a binocular HMD. We applied this model to a custom HMD technology and predict accuracy and precision of rendered depth obtainable in the system. We validated the model by measuring human performance in a set of psychophysical investigations. The human eye served as a measuring device to assess the technology under its naturally working conditions. Thus, perception is simply used in the assessment of the technology, and visual pathways and cognitive properties are not subjects of this investigation.

## 2 Previous Work and Motivation for the Research Presented

The need to see depth accurately in 3D visualization devices and thus to quantify depth perception in HMDs has been the focus of multiple studies starting in the 1960s when HMD technology was used to provide

additional information to pilots flying airplanes and flight simulators. Pilots conducted studies of size and depth perception to explain observed misperceptions. In these experiments, pilots made landings by reference to panel-mounted periscope screens in airplanes and to virtual *collimated* computer graphics images in flight simulators. (Collimated images refer to virtual monocular images formed at optical infinity, or distances  $\geq 6$  m.) Pilots consistently misjudged the runway as being smaller and farther away than it was, and consequently tended to overshoot their landings (Roscoe, Olzak, & Randle 1976; Palmer & Cronn 1973; Randle, Roscoe, & Pettit, 1980). Most of these misperceptions have been associated with the collimation of the virtual images.

Ellis and Bucher (1994), Ellis, Bucher, and Menges (1995), and Ellis and Merges (1997) have been investigating various types of perceptual bias of real and virtual objects in a haploscopic display using the method of adjustments. In such experiments, the virtual stimulus was a small pyramid pointing downward. Subjects were asked to locate the top of the pyramid typically with a light-emitting diode positioned at the end of a stick physically movable in depth. The combination of real and virtual objects allows establishing whether the virtual world is registered with respect to the real world and how objects are perceived in some combination of objects where they may or may not overlap. Equivalent studies with only real or virtual objects bring further control into explaining perceptual bias observed in mixing real and virtual objects.

Utsumi, Milgram, Takemuta, and Kishino (1995) have investigated the interplay of edge sharpness and binocular disparity of stereoscopically presented virtual objects on depth perception. The method of constant stimuli was adopted. They found large individual differences in importance given to the blur of the objects' edges. While for some subjects blur had almost no effect, for others the effect superseded that of stereopsis. Identifying factors that induce large individual differences, either correlated or random, is of primary importance to designing more effective systems and assessing their performance.

Surdick et al. (1994) have conducted psychophysical experiments to investigate the effectiveness of various

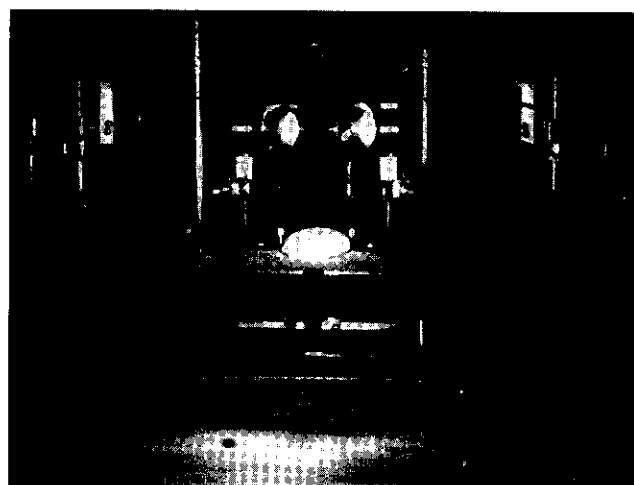
depth cues in a modified Wheatstone virtual display. They also investigated the effect of the viewing distance on cue effectiveness. Their stimulus was a square presented using various cues in isolation or in combination. The method of constant stimuli was used. The just noticeable difference (JND), which is related to the precision in measured rendered depth, was estimated from the collected data and was used as a measure of the effectiveness of the various cues present in the display. Understanding which cues are most effective under various conditions is essential to the optimization of display information.

In an earlier paper, we employed a 40 mm cube and a 13 mm dia. cylinder presented relative to each other at either 0.8 m or 1.2 m (Rolland et al., 1995). Subjects were asked to judge relative depth of the two objects whose depth had to be estimated from the location of their physical centers. The objects could be virtual, real, or a combination of the two. We had found large variability in all measurements involving virtual objects. We established two main possible causes of data variability in this study: first, conflicts of accommodation and convergence caused by presenting the virtual monocular images collimated (whereas judgments of relative depth were performed in the near field) (Robinett & Rolland, 1992; Wann, Rushton, & Mon-Williams, 1995; Marran & Schor, 1997) and, second, aspects of the methodology detailed in section 3. In the research presented here, we describe how we eliminated possible conflicts of accommodation and convergence, and we address issues related to methodology. Specifically, we present performance measures using two methodologies, and we demonstrate how the choice of the stimuli affects the measures of accuracy and precision of rendered depth.

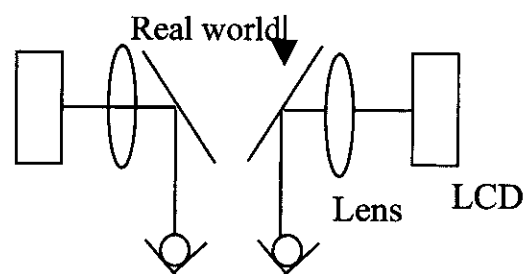
### 3 Methods

#### 3.1 Apparatus

A third-generation prototype of a conventional optical see-through, as opposed to projection-based see-through, HMD served as the experimental setup (Hua, Girardot, Gao, & Rolland, 2000). In this conventional, yet new instrument shown in figures 1(a) and



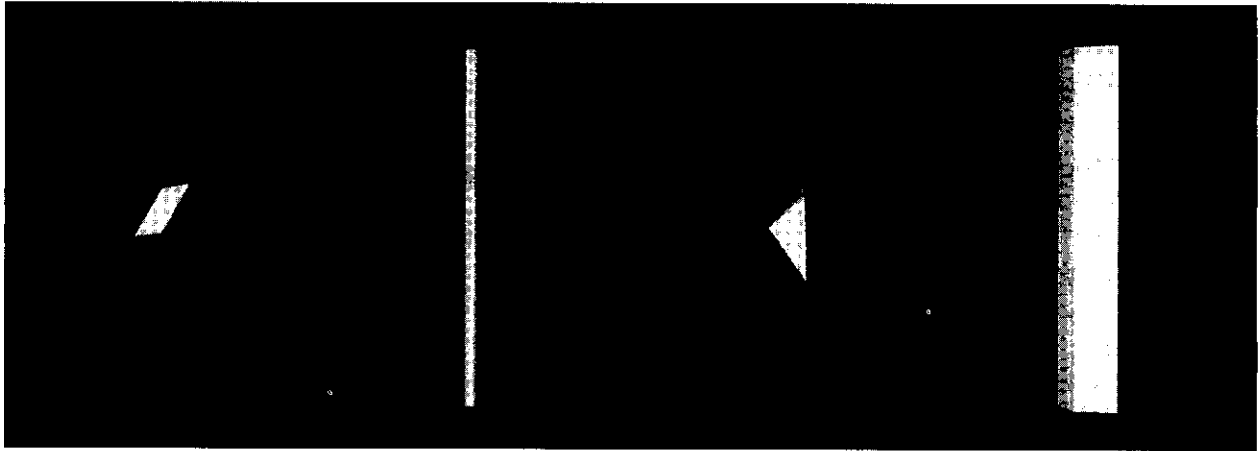
(a)



(b)

**Figure 1.** Experimental setup. (a) A chin rest for the human subject, the optical viewer, two depixelized LCD displays, and a two-button input device in the form of a ball for easy handling. The two-button device served in entering responses in the method of constant stimuli. A continuous dial located to the right also served to locate objects in depth in the method of adjustments. (b) A schematic of the optical layout.

1(b), micro adjustments permit precise alignment and setting of the various components. The overall geometry was modified from a previous V-shape geometry to a horizontal geometry to facilitate alignment (Rolland et al., 1995). New miniature higher resolution and contrast displays were also integrated in the system. The displays were off-the-shelf, 2.2 in. diagonal Casio active-matrix LCDs (model TV-7700B). The combination of a



**Figure 2.** Stimuli employed in the experiments. 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.

faceted cylinder and the larger faceted cylinder to determine if one would lead to higher performance than the other. For the larger faceted cylinder, the potential additional depth cue created by increased shading as a result of the facets is trading off its increased size.

The cube sizes were small (40 mm per side), medium (60 mm per side), and large (80 mm per side). The cubes roughly subtended 4, 6, and 8 deg. visual angle, respectively. In terms of stimuli shapes, we included an octahedron stimulus because it resembled the cube, yet we hypothesized that its center may be more easily estimated from external landmarks, which in this case correspond to the endpoints of the octahedron. We also varied the size of the octahedron to investigate the effect of its size on accuracy and precision of perceived depth. 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.

All stimuli are shown in figure 2, and their presentation is detailed in subsection 3.4. It is important to note

that in this investigation we are limited by a display with 3.5 arc-minutes of resolution, and therefore the results as they pertain to the specific stimuli and the specific cues to depths available for each stimuli may be valid only for displays with similar resolution. With a higher-resolution display, the effect of resolution on the results may be investigated as well. The computer graphics were generated by a Silicon Graphics Onyx Engine. Left- and right-eye virtual images were warped to compensate for the system's optical distortion (Rolland & Hopkins, 1993).

### 3.4 Presentation of the Stimuli

We eliminated conflicts of accommodation and convergence by placing the monocular optical images at the depth of 0.8 m where the reference 3D virtual objects were also presented. The elimination was accomplished by moving the miniature displays with respect to the focal point of the imaging optics. The correct location was verified by ensuring that there was no parallax between a real grid placed at 0.8 m and a virtual grid rendered at 0.8 m. Based on a stringent 1 arc-minute visual acuity criteria and a standard 4 mm dia. pupil under the viewing conditions, the depth of focus of the eye around 0.8 m is 49 mm and 44 mm behind and in

**Table 1.** Lateral Distance between Objects. The Bracketed Values were Conditions Not Used in the Sets of Reported Experiments because Similar Shapes were Not Compared to Each Other. Spacings are Reported as  $a/b$  where  $a$  is in mm and  $b$  is in degree of Visual Angle

	Small Cylinder	Faceted Cylinder	Small Cube or Octahedron	Medium Cube or Octahedron	Large Cube or Octahedron
Small cylinder	[28.8]/[2.0]	[78.7]/[5.6]	69.4/[5.0]	96.9/6.9	124.4/8.9
Faceted cylinder	[78.7]/[5.6]	[128.6]/[9.2]	119.3/8.5	146.8/10.5	174.3/12.4
Small cube or octahedron	69.4/5.0	119.3/8.5	110.0/7.9	137.5/9.9	165.0/11.8
Medium cube or octahedron	96.9/6.9	146.8/10.5	137.5/9.9	165.0/11.8	192.5/13.7
Large cube or octahedron	124.4/8.9	174.3/12.4	165.0/11.8	192.5/13.7	220.0/15.7

front, respectively (Rolland, Krueger, & Goon, 2000). All objects to be assessed were presented within these depths of focus values. Therefore, accommodation did not conflict with convergence. Stereopsis and convergence coupled to accommodation were the main sources of information to depth in the current study. Shape and shading served as secondary sources of information to depth (Reading, 1983).

Two stimuli of different shapes were presented at a time (such as a cube and an octahedron) side by side. 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, 74.4, and 99.3 mm, respectively. 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 ( $\pm 55$  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)]. Spacing for object pairs is given in table 1. The cube was presented at a specific angle (45 deg. rotation around the longitudinal axis ( $z$ ) and 45 deg. rotation around the vertical axis ( $y$ )). The octahedra were tilted down around a horizontal axis by  $\pi/12$ . The subject's task was to judge the relative depth of the objects' centers.

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 edges of the miniature displays. For the user, the view is similar to that experienced when looking through a window into another 3D environment. The edges of the miniature displays are optically located outside of the field of view of the optics and thus are not visible. The field of view is thus limited by the optical components, and, because those are not in focus, no sharp edges are perceived.

## 4 Measurement Methods

### 4.1 Method of Constant Stimuli with Random Size Presentation

In the first set of experiments, relative depth perception was assessed using a two-alternative forced-

choice method of constant stimuli (Guilford, 1954; Gescheider, 1976; Engen, 1972). Ten values of depths for the object on the right were presented around the nominal depth of the object on the left in step sizes of 6 or 7 mm. Human subjects were asked to judge whether the object on the right was in front or in back of the object on the left. Responses were entered on a two-button handheld device. No time limit was given on any trial. Participants completed 100 trials (ten depth values and ten presentations of each depth value) for each condition. Participants' variability for the same conditions was also examined by collecting three psychometric functions per subject per condition. The design was balanced in the sense that all the conditions were run before they were repeated, and the conditions were randomly presented within a run.

During pilot experiments of depth perception based on the method of constant stimuli, we established that human subjects could learn the mean size of an object among a subset of displayed sizes in only ten or twenty trials. Subjects could then substitute their depth judgments by size judgments and results would simply be the outcome of size assessment around a mean value imposed by the methodology. Therefore, for the method of constant stimuli to be useful in assessing depth perception, randomization of size must be applied. We thus randomly varied the size of the displayed object by  $\pm 15\%$  around its mean value. Such randomizing of a parameter is commonly done in psychophysics for prohibiting subjects to rely on certain cues. We refer to the method as the method of constant stimuli with random size presentation.

#### 4.2 Method of Adjustments

In the second set of experiments, relative depth perception was assessed using the method of adjustments (Guilford, 1954; Gescheider, 1976; Engen, 1972). Based on results of the experiments using the method of constant stimuli, the large cylinder stimulus was not tested in the method of adjustment experiment. Therefore, each experiment run consisted of thirty conditions: six cube-cylinder conditions, six octahedron-cylinder conditions, and eighteen cube-octahedron con-

ditions. In each trial, one of the two virtual stimuli was presented at a random depth between 0.64 m and 0.96 m. The other object was presented at the nominal depth of 0.8 m. Participants employed a continuous dial to move one object back and forth during each trial, and they moved the adjustable object so that the centers of the two objects were estimated to be at the same depth. The participant pushed a button on the handheld device to finalize the depth judgment and to initiate a new trial. No time limit was given on any trial. Participants completed twenty trials for each condition, and two participants ran the experiment three times to test for repeatability. It is important to note that the intrinsic issue with stimuli size that led to add randomization of size to the method of constant stimuli does not surface in the method of adjustments because the presentations of depth are not set to fixed values in advance throughout the experiment. Rather, in the method of adjustments, participants make a decision on depth perception after they adjust one of the stimuli themselves to a depth value. Therefore, there was no need within the method of adjustments to introduce any randomization of the size. Thus, a fair comparison was to assess the method of constant stimuli modified for random size presentation against the conventional method of adjustments, given that the end goal is to assess technology with a method that provides the least bias and variability in the data.

## 5 Data Analysis

### 5.1 Method of Constant Stimuli

Data were analyzed using probit analysis. For each psychometric function, the point of subjective equality (defined as the 50% point on the psychometric curve) and 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 m measured the accuracy of rendered depth, referred to as *Delta-PSE* ( $\text{Delta-PSE} = \text{Measured Value} - \text{Nominal Value}$ ). The slope of the curve, referred to as the *dis-*

*crimination threshold*, measured the precision of rendered depth.

Discrimination threshold values reported here indicate upper-bound values because the step size for increment and decrement in depth was chosen for each subject so that anchor points were collected for all psychometric functions, across stimuli shapes and sizes. Anchor points refer to points on the psychometric function where a subject is correct 100% of the time. A step size of 6 or 7 mm was set and kept constant across conditions. Although a smaller step size could have been used for certain conditions to measure precision of measured rendered depth more accurately, we chose not to change the step size so that it would not be a confounding factor in measuring the accuracy of rendered depth across various 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 truly 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.

## 5.2 Method of Adjustments

The PSE values for each subject were calculated by averaging the adjusted depth values over the twenty trials for each condition. Delta-PSE values were calculated from the difference of the PSE values and the nominal value, as in the method of constant stimuli. Discrimination threshold values were calculated by taking the standard deviation of the Delta-PSE values for each set of twenty trials.

## 6 Expected Findings

Based on the computational model of Robinett and Rolland (1992), objects presented in front or behind a reference object located at a given depth should be measured to be in front or behind accordingly. This prediction applies to any two virtual objects presented in

the experiments. Thus, 0 mm  $\pm$  1 mm, or 2 mm Delta-PSE served as the expected performance for accuracy of measured rendered depth in the display. It is important to note that the  $\pm$ 1 mm is the variability in Delta-PSE measured with a real cube and a real cylinder as reported by Rolland et al. (1995). In the graphs plotting Delta-PSE as a function of several variables, performance should lie on the dashed horizontal line passing through zero, or within 1 mm above or below.

The theoretical limitation of precision of rendered depth is set by visual stereoacuity that we shall denote  $\eta$ . Given a fixation plane located at a distance  $L$ , stereoacuity sets the smallest resolvable depth  $\Delta l$  around that plane, given by

$$|\Delta l_{\pm}| = \frac{|\eta| \cdot L^2}{\text{IOD} \mp |\eta| \cdot L}, \quad (1)$$

where IOD is the interocular distance. A comprehensive exploitation of this formula as well as equation (2) in the form of graphs may be found in Rolland et al. (2000). Stereoacuity varies widely between individuals. Values between 2 arc-sec and 130 arc-sec have been reported in the literature (Reading, 1983; Lit & Finn, 1976). Given an average IOD of 65 mm, a 30 arc-sec stereoacuity yields depth resolution  $\Delta l$  of 1.4 mm at 0.8 m.

However, the precision of rendered depth in HMDs is most often limited by the display resolution. Based on a 2.2 in. diagonal display and  $429 \times 586$  addressable lines, the equivalent pixel size is 77  $\mu$ m. The precision of rendered depth when imposed by the pixel resolution  $p$  can be shown to be related to  $p$  by

$$\Delta l = \frac{pL}{\text{IOD} - \frac{2 \left(1 - \frac{L}{f}\right) \cdot p}{f}}, \quad (2)$$

where  $f$  is the optics focal length (86 mm for the current system) (Rolland et al., 2000). Based on equation (2), the measured precision of rendered depth imposed by the display resolution is estimated to be 15 mm. This prediction is consistent with the theoretical finding reported by Rolland et al. (2000) as well, in which the

authors showed that the display pixel size should be in the order of  $6\ \mu\text{m}$  when combined with a 90 mm focal length optics (that is, a factor of ten smaller than it currently is) for precision of rendered depth to reach 1.5 mm and be limited by the human visual system.

## 7 Experimental Results and Discussion

Results using the methods of constant stimuli and adjustments are summarized in figures 3 and 4, and 5 and 6, respectively. In the figures, "ALL" means the average over four human subjects. Error bars in figures 3 and 5 correspond to across-subject variability. In figures 4 and 6, error bars correspond to within-subject variability.

### 7.1 Results Using the Method of Constant Stimuli Modified with Random Size Presentation

Each data point on the graphs corresponds to an average across human subjects of 700 trials or three psychometric functions, one with 300 trials and two with 200 trials per human subject. In figure 3, the pairs of mirror stimuli were averaged together, reducing the data to six plots: three on the left for the accuracy and three on the right for the precision. Figures 3(a) and 3(b) compare a cube and a cylinder, figures 3(c) and (d) compare an octahedron and a cylinder, and figures 3(e) and (f) compare a cube and an octahedron. Figure 4 shows typical data for two individual human subjects under four of the conditions tested.

Results show that an average upper bound for discrimination thresholds is about 7 mm across all the stimuli as shown in figures 3(b), 3(d), and 3(f). The lower the discrimination thresholds the higher the resolution in depth, thus the higher the performance. This value was found to be twice as small as predicted before depixelization of the display. Thus, we have further established that depixelization of the displays can lead to a factor-of-two gain in depth resolution in near-field visualization. Based on this finding, we infer that, if pixel sizes in the order of  $6\ \mu\text{m}$  were required to avoid limit-

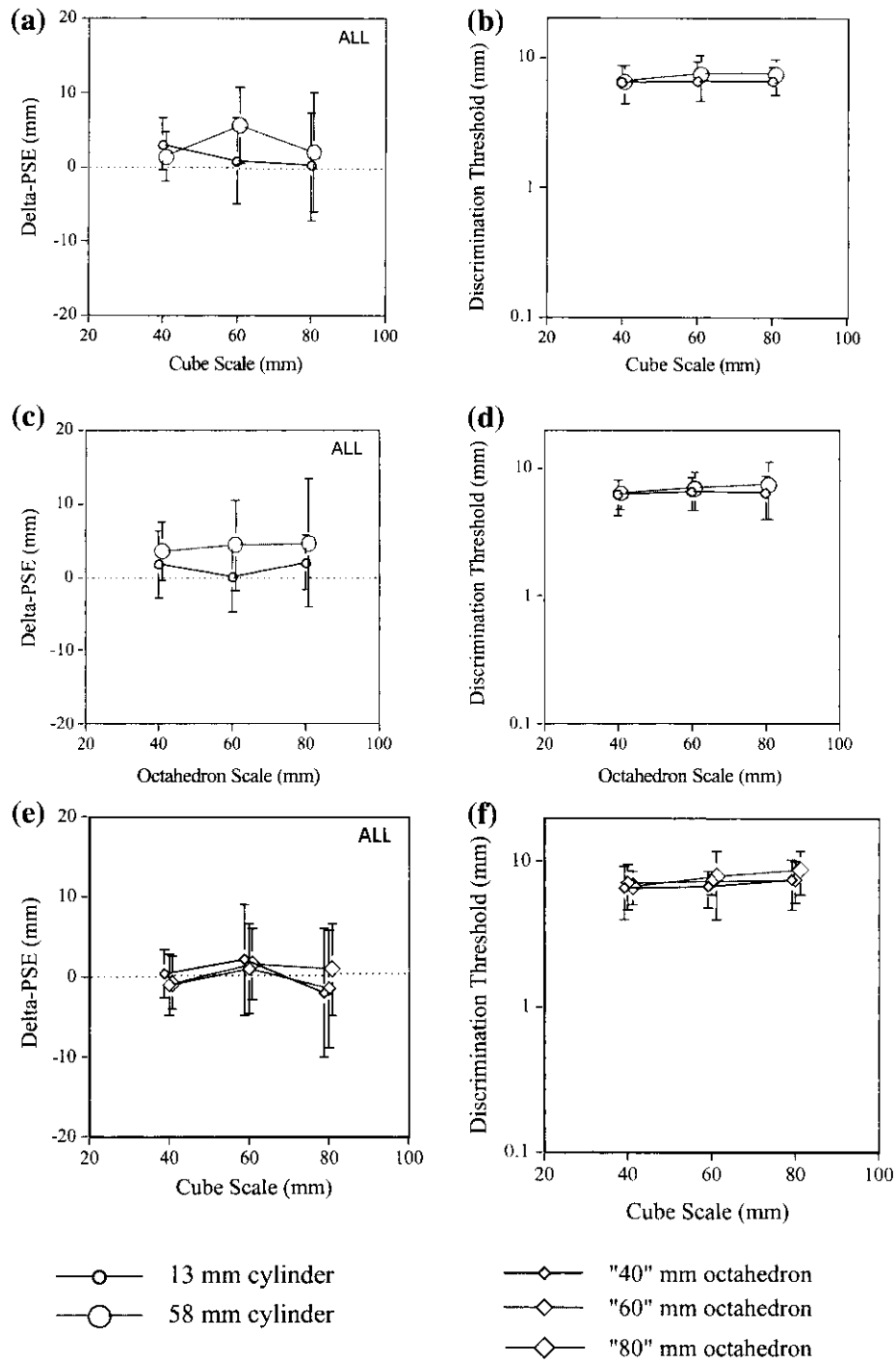
ing the precision of rendered depth in HMDs, values of up to  $12\ \mu\text{m}$  would be sufficient when depixelization is applied.

Results further show that the accuracy of measured rendered depth is fairly independent of the stimuli form, except for the larger cylinder. Adding structure to the cylinder, which also necessitated increasing its size, did not yield higher performance. To the contrary, results show that the large, faceted cylinder generally yields higher bias in measured rendered depth than the smooth, small cylinder. This finding is shown by an overall upper shift of the data points in figures 3(a) and 3(c). The shift indicates that, although the facets on the cylinder may subjectively give more 3D percept to the cylinder, as indicated by the human subjects, human subjects do in fact judge the 3D location of the smooth, small cylinder more accurately than that of the larger, faceted cylinder. Employing an octahedron instead of a cube did not improve performance, contrary to what we expected as a consequence of adding structure, and neither did increasing the size of the cube or the octahedron affect end performance significantly.

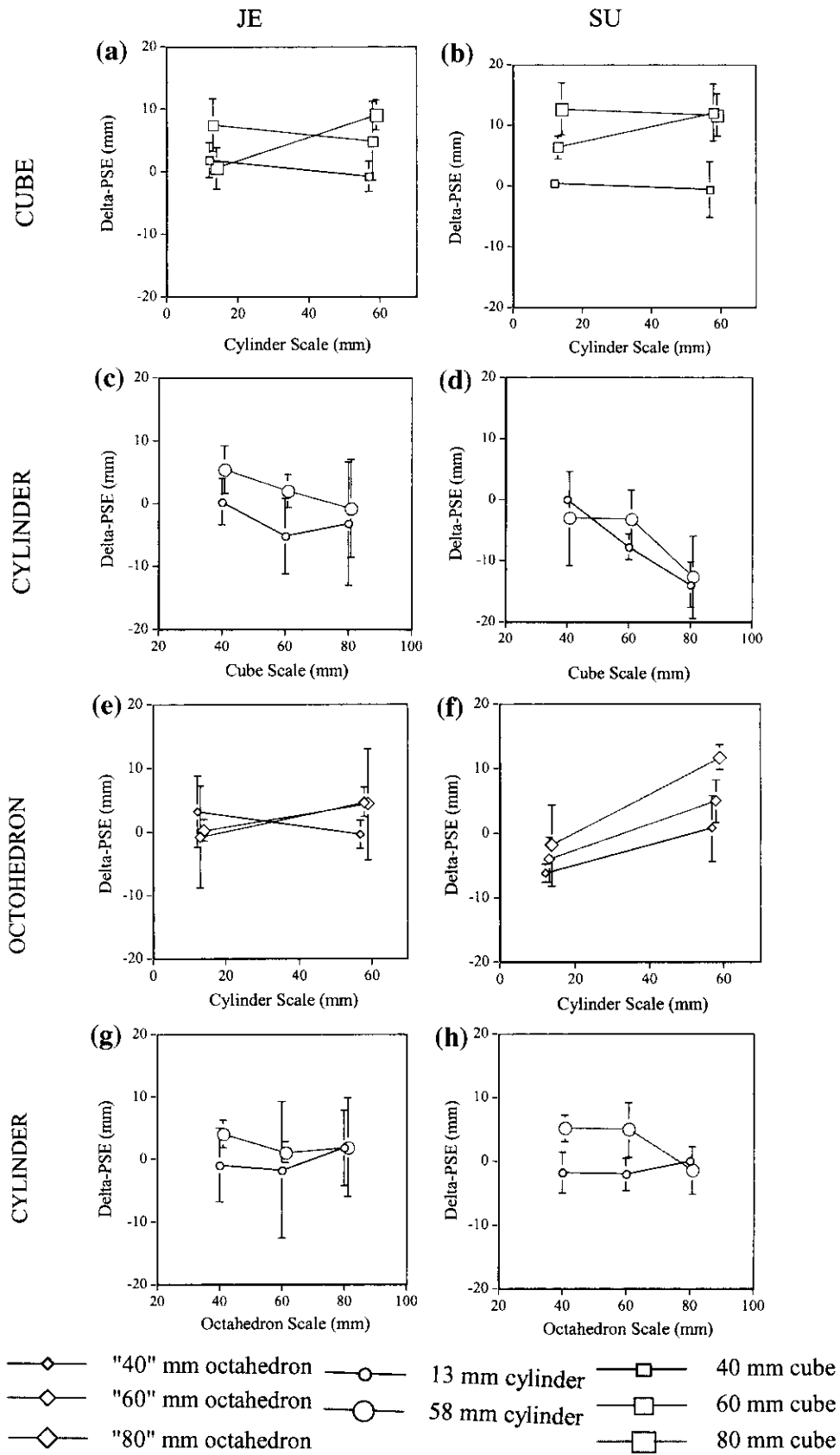
In spite of minimizing conflicts of accommodation and convergence, variability in the data was still found to be higher than expected, especially within-subjects' variability. Typical data for two individual human subjects are reported in figure 4 for a few conditions. Overall human subjects' variability were roughly equivalent for the cylinder-cube and the cylinder-octahedron conditions. Variability was slightly higher for the cube-octahedron case compared to cases using a cylinder. Because the task to be performed 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 through time (not plotted here) indeed shows that depth errors for all pairs of stimuli varied uncorrelated within individuals. Variability over time for equivalent trials was random.

Surdick et al. (1994) as well as Cutting and Vishton (1995) have shown the importance of relative size of objects as a strong cue to depth perception within a viewing distance of 1 m. We infer that the observed variability in the data must result from varying the size of one of the two displayed objects by  $\pm 15\%$ . The varia-

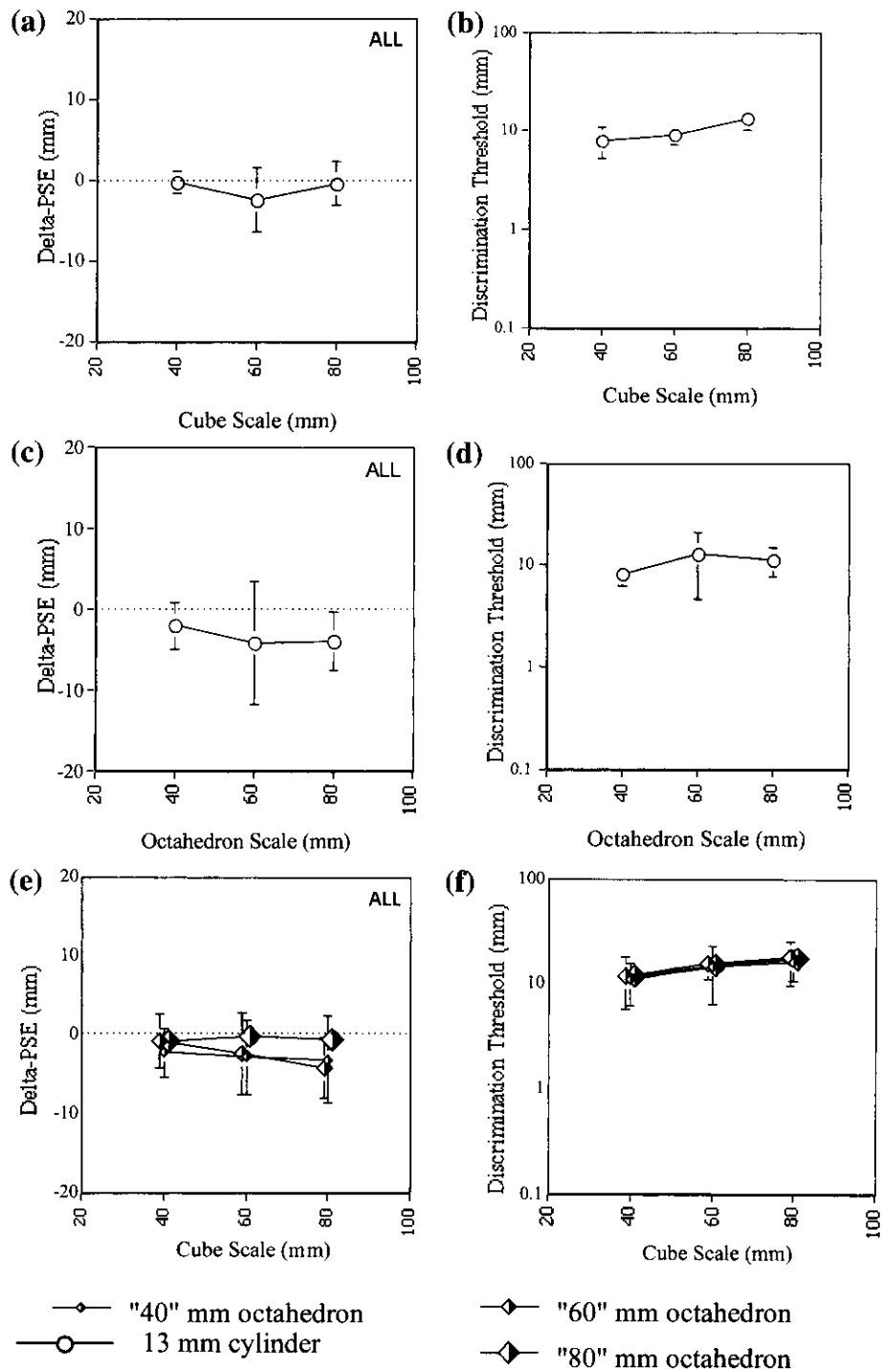




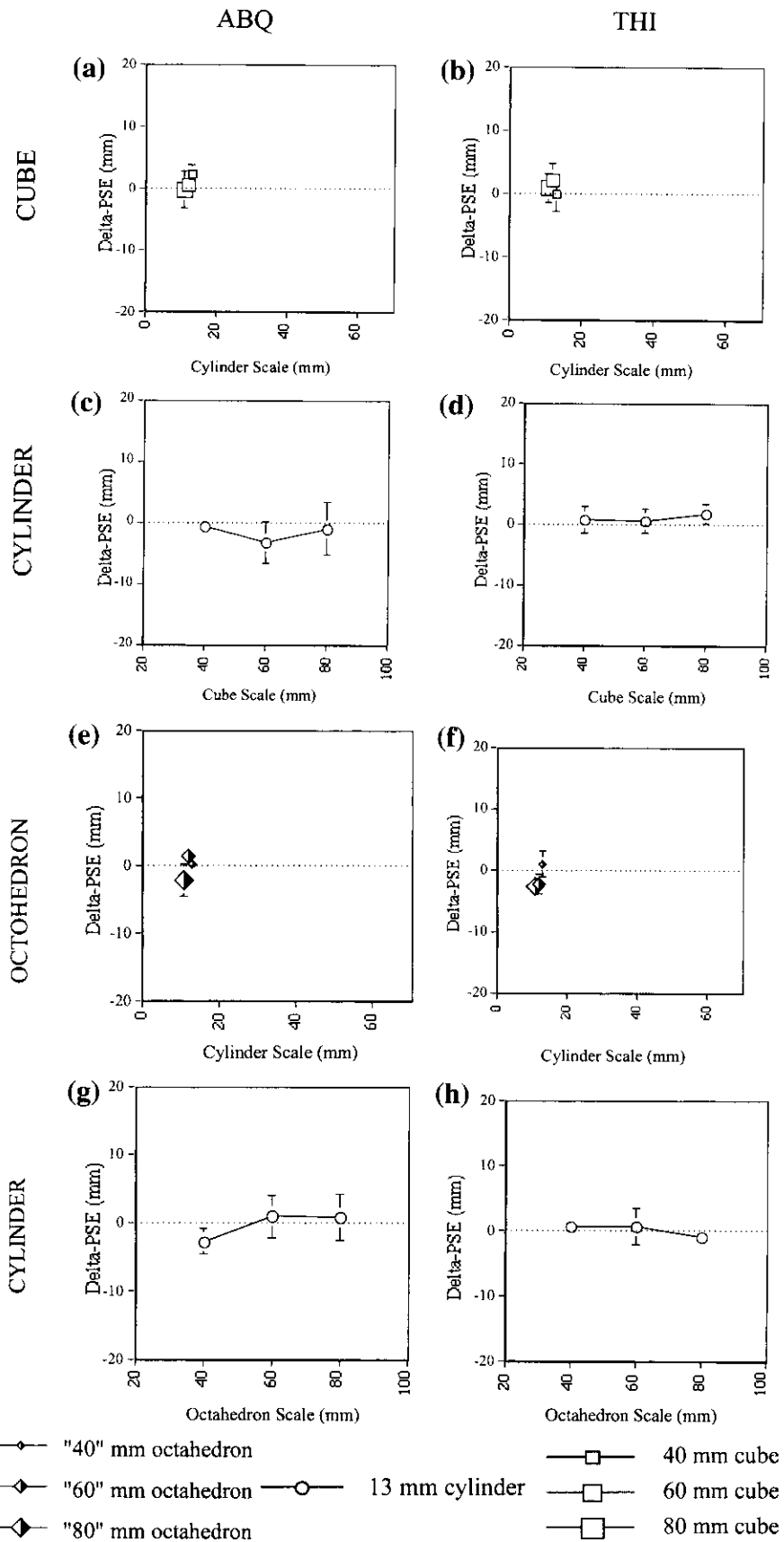
**Figure 3.** [Method of constant stimuli] Left: Average accuracy of measured rendered depth across mirror pairs of virtual stimuli (such as a cube with a cylinder and a cylinder with a cube) when judging the depth of one with respect to the other. Right: Precision of measured rendered depth for the same conditions.



**Figure 4.** [Method of constant stimuli] Accuracy of measured rendered depth for two human subjects (JE and SU) across four pairs of stimuli. Within-subject variability is shown.



**Figure 5.** [Method of adjustments] Left: Average accuracy of measured rendered depth across mirror pairs of virtual stimuli (such as 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 measured rendered depth for the same conditions.



**Figure 6.** [Method of adjustments] Accuracy of measured rendered depth for two human subjects (ABQ and THI) across four pairs of stimuli. Within-subject variability is shown.

tion in size applied on the objects was necessary as previously stated to encourage a strategy for depth judgment. Therefore, such a finding suggests that other methods should be considered for assessing depth perception in HMDs.

## 7.2 Results Using the Method of Adjustments

In figures 5 and 6, results show similar plots for the judgments of perceived depth using the method of adjustments. Figures 5(a) and 5(b) compare a cube and a cylinder, figures 5(c) and (d) compare an octahedron and a cylinder, and figures 5(e) and (f) compare a cube and an octahedron. Figure 6 shows typical data for two individual human subjects under four of the conditions tested.

We note that variability decreased significantly over all conditions, and, as a consequence, a stable measure of accuracy could be acquired in less time. Results also show invariance of measured accuracy with different forms and sizes of the stimuli. However, precision of rendered depth varied slightly for different pairs of stimuli between 8 mm and 12 mm, in favor of the smaller stimuli. We conclude that a small cube or octahedron, in combination with a thin cylinder, yields equivalent best stimuli among those considered to conduct such assessments. Possibly adding texture to the stimuli (a condition not tested in this paper) may yet bring forth finer structures without the need to increase the stimuli sizes. Finally, we observed that the method of adjustments yielded the least bias and variability in the data among the two methods investigated, and was more engaging for the participants than the method of constant stimuli. The method of adjustments is indeed equivalent to performing a localization task, which brings the methodology closer to usability of the technology.

## 8 Conclusion

We have reported results from psychophysical experiments on the quantification of rendered depth in

HMDs using either the method of constant stimuli modified with a random size presentation or the method of adjustments. Modifying the method of constant stimuli was necessary to eliminate the tendency of participants to judge depth solely on variation of size around an estimated mean imposed by the method. Various shapes and sizes were also considered as part of the investigation. A main finding of this study is that the modified method of constant stimuli leads to reduced bias compared to the unmodified method of constant stimuli but higher variability in the data compared to the method of adjustments. Furthermore, the method of adjustments lead to greater involvement on the part of the participants and brought the methodology closer to usability. Accuracy of rendered depth followed theoretical predictions (the 2 mm benchmark). Precision of rendered depth was measured to be in the order of 8 mm, which sets a benchmark for the evaluation of comparable systems. Finally, we established a factor-of-two gain in depth resolution accompanying the depixelization of the displays, an improvement that is significant when tradeoffs of field of view and resolution are often called for in HMD technology. Interesting future work could include the quantification of rendered depth in other stereo displays based on theoretical predictions for these displays and measures of accuracy and precision that could then be compared across stereo techniques.

## Acknowledgments

We thank Microsharp Technology for donating the depixelization screens for the LCD displays. We further thank Norbou Buchler and Alexander Quinn for their assistance with human subjects and data analysis, and our participants for their cooperation and commitment to this project. In a previous paper, we benefited from Christina Burbeck's suggestions on methodology that served as a basis in the method of constant stimuli results presented, and, regarding this specific paper, we further benefited from insightful discussions on methodology from Robert S. Kennedy of RSK Assessments, Inc. This research was supported by the Office of Naval Research under grants N00014-94-1-0503 and N00014-02-1-0261, a First Award 1-R29-LM06322-01A1 from the National Institute of

Health, and NSF/EIA-99-86051 and NSF/ITR Award 0082016.

## References

- Biocca, F., & Rolland J. P. (1998). Virtual eyes can rearrange your body: Adaptation to virtual eye location in see-thru head-mounted displays. *Presence: Teleoperators and Virtual Environments*, 7(3), 262–277.
- 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 and S. Rogers (Eds.), *The perception of space and motion*, (pp. 69–117). San Diego: Academic Press.
- Ellis, S. R., & Bucher, U. J. (1994). Distance perception of stereoscopically presented virtual objects superimposed by a head mounted see through display. *Proceedings, 38th Annual Meeting of the Human Factors and Ergonomics Society*, 1300–1305.
- Ellis, Stephen R., Bucher, Urs J., & Menges, Brian. (1995). The relationship of binocular convergence and errors in judged distance to virtual objects. *Proceedings of the International Federation of Automatic Control '95*, 297–301.
- Ellis, S. R., & Menges, B. M. (1997). Judged distance to virtual objects in the near visual field. *Presence: Teleoperators and Virtual Environments*, 6(4), 452–460.
- Engen, T. (1972). Psychophysics: I. Discrimination and detection. In J. W. Klins & L. A. Riss (Eds.), *Sensation and perception: Woodworth and Schlosberg's Experimental Psychology*, 3<sup>rd</sup> ed. New York: Holt, Reinhert & Winston.
- Gescheider, G. A. (1976). *Psychophysics: Method and theory*. New York: Wiley.
- Guilford, J. P. (1954). *Psychometric methods*. 2<sup>nd</sup> ed. New York: McGraw-Hill.
- Howard, H. J. (1919). A test for the judgment of distance. *American Journal of ophthalmology*, 2, 656–675.
- Hua, H., Girardot, A., Gao, C., & Rolland, J. P. (2000). Engineering of head-mounted projective displays. *Applied Optics*, 39(22), 3814–3824.
- Lit, A., & Finn, J. P. (1976). Variability of depth-discrimination thresholds as a function of observation distance. *J. Opt. Soc. of Am. A.*, 66(7), 740–742.
- Marran, L., & Schor, C. (1997). Multiaccommodative stimuli in VR systems: Problems and solutions. *Human Factors*, 37, 382–388.
- Palmer, E., & Cronn, F. W. (1973). Touchdown performance with computer graphics night visual attachment. *Proceedings of the AIAA Visual and Motion Simulation Conference*, 1–6.
- Randle, R. J., Roscoe, S. N., & Pettitt, J. (1980). *Effects of accommodation and magnification on aimpoint estimation in a simulated landing task* (Tech Paper NASA-TP-1635). Washington DC: National Aeronautics and Space Administration.
- Reading, R. W. (1983). *Binocular vision: Foundations and applications*. Woburn, MA: Butterworth-Heinemann.
- Robinett, W., & Rolland, J. P. (1992). A computational model for the stereoscopic optics of a head-mounted display. *Presence: Teleoperators and Virtual Environments*, 1(1), 45–62.
- Rolland, J. P., Ariely, D. & Gibson, W. (1995). Towards quantifying depth and size perception in virtual environments. *Presence: Teleoperators and Virtual Environments*, 4(1), 24–49.
- Rolland, J. P., & Fuchs, H. (2000). Optical versus video see-through head-mounted displays in medical visualization. *Presence: Teleoperators and Virtual Environments*, 9(3), 287–309.
- Rolland, J. P., & Hopkins, T. (1993). *A method of computational correction for optical distortion in head-mounted displays* (Tech. Rep. TR-93-045). Dept. of Computer Science, University of North Carolina at Chapel Hill.
- Rolland, J. P., Krueger, M. & Goon, A. (2000). Multi-focal planes in head-mounted displays. *Applied Optics*, 39(19), 3209–3215.
- Roscoe, S. N., Olzak, L. A., & Randle, R. J. (1976). 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*, A5.1–A5.9.
- State, A., Livingston, M. A., Garrett, W. F., Hirota, G., Whitton, M. C., Pisano, E. D., & Fuchs, H. (1996). Technologies for augmented reality systems: Realizing ultrasound-guided needle biopsies. *Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH*, 439–446.
- Surdick, R. T., Davis, E. T., King, R. A., Corso, G. M., Shapiro, A., Hodges, L., & Elliot, K. (1994). Relevant cues for the visual perception of depth: Is where you see it where it is? *Proceedings of the Human Factors and Ergonomics Society 38th Annual Meeting*, 2, 1305–1309.
- Sutherland, I. (1968). A head-mounted three-dimensional

- display. *Fall Joint Computer Conference, AFIPS Conference Proceedings*, 33, 757-764.
- Utsumi, A., Milgram, P., Takemura, H., & Kishino, F. (1995). Effects of fuzziness in perception of stereoscopically presented virtual object location. *Proc. of the SPIE: Telemanipulator and Telepresence Technologies*, 2351, 337-344.
- Wann, J., Rushton, S., & Mon-Williams, M. (1995). Natural problems in the perception of virtual environments. *Vision Research*, 35, 2731-2736.
- Wheatstone, C. (1838). Contributions to the physiology of vision. Part the first. On some remarkable and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, 128, 371-394.

