Annals of Biomedical Engineering, Vol. 36, No. 2, February 2008 (© *2007*) pp. 342–348
DOI: 10.1007/s10439-007-9414-7

Accommodation and Size-Constancy of Virtual Objects

Robert V. Kenyon, Moses Phenany, Daniel Sandin, and Thomas Defanti

Electronic Visualization Lab, Department of Computer Science (m/c/ 152), University of Illinois at Chicago, SEO Building, rm 1120, Chicago, IL 60607, USA

(Received 13 November 2006; accepted 27 November 2007; published online 8 December 2007)

Abstract—Accommodation has been suspected as a contributor to size illusions in virtual environments (VE) due to the lack of appropriate accommodative stimuli in a VE for the objects displayed. Previous experiments examining size-constancy in VE have shown that monocular cues to depth that accompany the object are a major contributor to correct size perception. When these accompanying cues are removed perceived size varied with the object's distance from the subject, i.e., visual angle. If accommodation were the dominant mechanism contributing to a visual angle response [due to its action to keep physical objects clear] in this condition, an open-loop accommodation viewing condition might restore size-constancy to this condition. Pinhole apertures were used to open-loop accommodation and examine if size-constancy might be restored when few accompanying monocular cues to depth were present. Visual angle performance when viewing a low cue environment was found with and without the use of the pinhole apertures. Thus, these results signify that accommodation does not play a dominate role in the loss of size-constancy in sparse visual environments often used in VE. These results suggest that size-constancy is driven by the inclusion of the remaining monocular cues to depth in VE as it is in the physical world.

Keywords—Accommodation, Pinhole, Size-constancy, Virtual environment.

INTRODUCTION

Users of the CAVE® (The CAVE is a registered trademark of the Board of Trustees of the University of Illinois; CAVE Automatic Virtual Environment) and other virtual environment (VE) systems have reported that they perceive virtual objects to be incorrectly sized.[1,6] This effect may be attributed to a variety of factors including hardware errors, software errors, and perception errors. Recently, Kenyon *et al.*[12] have shown size-constancy in a VE (CAVE) and that monocular cues to depth play a major role in size-constancy performance. These results mirrored the results from many studies on the perceived size of objects in the physical world that have been performed. Descartes[5] first described the phenomenon known as ''size-constancy'' where an object is perceived as being the same size regardless of its distance from the observer even though the retinal size of the object gets smaller with increasing distance from the observer. Holaday[9] showed that removal of various cues would change this behavior to one relying on the physical optics of the situation. She showed that as the number of monocular cues to depth [e.g., shadows, motion parallax, *etc*.] is reduced, performance suffers and subjects adopt a size judgment that is based on the visual size of the object on the retina also know as visual angle (VA) size judgments. Holway and Boring[10] confirmed these findings for objects from 10 to 40 ft from the observer. Harvey and Leibowitz[8] showed similar results at distances of 1–9 ft from the observer. Furthermore, they and Leibowitz and Dato[13] showed that removal of 3D cues to depth (i.e. Stereovision) had little to no effect on performance and that performance was only affected by the removal of monocular depth cues.

Consistent with the experiments performed in the physical world, Kenyon *et al.*[12] also showed that a subject's loss of size-constancy in VE occurred most frequently when scenes did not contain numerous monocular cues to depth. Nevertheless, the amount of accommodative demand for an object at the same distance in the virtual and physical worlds is not necessarily the same.[11,15,17] Consequently, the CAVE and other forms of virtual environments can cause the user to endure conflicts between accommodation and vergence for objects in the scene or experience perceptual errors due to the lack of an appropriate relationship between accommodation and the distances of the virtual objects.[16,18,20] For example, in projection-based virtual systems like the CAVE, accommodation is stimulated by the distance the user is from CAVE wall, regardless of where the virtual objects appear in 3D

Address correspondence to Robert V. Kenyon, Electronic Visualization Lab, Department of Computer Science (m/c/ 152), University of Illinois at Chicago, SEO Building, rm 1120, Chicago, IL 60607, USA. Electronic mail: kenyon@uic.edu

space.[11] Hence, the accommodative information does not necessarily correspond to the other monocular and 3D distance cues that the virtual object may possess when it is drawn by the computer. Indeed, this discrepancy between accommodative demand in virtual and physical worlds has been an important topic in the design and use of VE and augmented reality (AR) systems.[19] Therefore, because of this lack of synergy between the accommodation and other visual information, and accommodation's role in a person's estimate of distance to an object[14] (an important component in object size judgments), it is possible that this mismatch could be an important factor in the lack of size-constancy when many monocular cues are absent from the virtual scene (stereovision remains).

This study examines the role of perception errors that might be induced due to inappropriate accommodative stimuli when scenes do not contain numerous monocular cues to depth. By opening the feedback loop of accommodation using a pinhole viewing system we can release accommodation from its constraint to maintain a clear image of the objects on the VE screen.[1] If this fixed accommodation is mainly responsible for the visual angle responses in sizing objects then by creating an open-loop accommodative condition we should be able to allow accommodation to vary as it is driven by vergence and see if that release can restore size-constancy performance.

METHODS

The CAVE

The CAVE is a projection-based VE system,[4] where four screens are arranged in a 10 ft cube composed of three rear-projection screens for walls and the fourth projector (overhead) points to a mirror, which reflects the images onto the floor. A viewer is supplied stereovision through stereo shutter glasses (Stereographics, Inc). Proper perspective images are drawn for each eye using head position from a six-degrees-of-freedom head-tracking device (Intersense 900) calibrated to an accuracy of ±0.5 in. Subject's interpupillary distance (IPD) was measured (R.H. Burton Digital P.D. Meter) and used by the CAVE program to generate the stereo images for that subject. A joystick and buttons in a hand held "wand" provided the needed interaction with the VE to change the size of the virtual object.

An SGI Onyx with two Infinite Reality graphics pipelines, each split into two channels controlled the projected images. The image resolution was 1024 × 1024 pixels (Marque 5000 projector) with a refresh rate of 120 Hz, i.e., an effective stereo refresh rate of 60 Hz.

Pinhole Aperture

Pinhole apertures worn less than 2 mm from the cornea, open-looped accommodation and reduced the effect of the CAVE wall's accommodative stimulus during the size judgments. Each pinhole aperture provided a single 1 mm diameter opening for each eye that resulted in an approximate 35° field of view (FOV). The depth of focus produced by this aperture would open-loop accommodation beyond the range of the CAVE wall (0.5D, 0.65D, 0.93D) and the bottles distances from the subject (0.4 to 1.64D).[3,21] Users could adjust the distance between the pinholes (horizontal direction) for their specific IPD. They could also adjust the pinholes in the vertical direction. Users were asked to adjust the pinholes so that the image for the two views had a 100% binocular overlap.

Neutral Density & Reduced Field of View Filter

To provide the same viewing conditions as with the pinhole aperture but without open-loop accommodation affects, an aperture and a neutral density filter were fixed to the outer surface of the shutter glasses to equal the light intensity and FOV produced by the pinhole aperture. The circular aperture (6.35 mm dia.) placed over the shutter glasses approximated the 35° FOV provided by the pinhole aperture. The neutral density filters (3f stops) placed over this circular opening were used to approximate the reduced light intensity encountered when using the pinhole aperture. The circular opening could be adjusted along the horizontal and vertical axes. Users were again asked to adjust the apertures as described above.

The Physical World

A 2-L coke bottle was placed on a black plastic table at a height of 4 ft. The table was positioned at the front right hand side of the CAVE at an approximate distance of 4 ft from the subject. The height of the coke

---

[1] To be more specific, when a subject looks at the VE scene projected on to the screen in front of them, the blur signal that drives accommodation is determined by the distance the subject is from the screen. Other forces may try to change this level of accommodation such as the influence of vergence but the need to maintain a clear image would cause accommodation to resist such changes and remain close to the dioptric value for the wall. In the case of VE, accommodation can be considered fixed at the projection screen and resists any vergence-accommodation influences due to the necessary convergence of the eyes to prevent diplopia. However, if we open-loop accommodation so that regardless of the level of accommodation the retinal image remains clear then changes in accommodation due to vergence may occur without a concomitant change in the blur of the physical target.

FIGURE 1. **The ENV visual scene contained a number of monocular cues to depth. The patterned floor and textured table provided information for the subject to make size judgments. The floor and table were absent in the NoENV visual scene so that only the bottle and a grey background was present (Image not to scale).**

bottle was 12 in. tall and 5.5 in. (maximum) wide. The coke bottle was lit by a spotlight mounted on the left hand wall of the CAVE at a height of 10 ft and at a distance of 8 ft from the front CAVE wall.

The Virtual World

The virtual world simulated 2 different scenes. (a) The ENV scene consisted of a gray green-checkered floor with a wood textured table and a coke bottle on top of the table (Fig. 1). The coke bottle was textured with an image of the real coke bottle. The height (30, 33, and 36 in. above the floor) and appearance of the table were changed for different sets of measurements, as was the distance of the virtual coke bottle from a subject (2, 3.5, 5.0, 6.5, and 8.0 ft). If these were physical objects the accommodative demand would range between 1.64D and 0.41D respectively. Subjects used the wand's joystick to increase and decrease the size of the coke bottle and a wand button to continue once they had finished sizing the virtual coke bottle. The size of the virtual coke bottle changed as it would in the physical world i.e., from its position on the tabletop. The bottle did not penetrate the table but rested on the tabletop as it changed size. (b) In the NoENV scene subjects were presented with only a gray background and few monocular cues to depth (only those associated with the bottle itself) but stereovision was presented as in the ENV scene. The Coke bottle appeared suspended in mid air at the same distances and heights (from the floor) from the user as in ENV condition. In the NoENV scene, the bottle changed size in the same manner as in the ENV scene even though the table was not drawn.

Procedure

Three viewing conditions were tested for each subject: (1) (REG) Subject's vision was unobstructed and only the shutter glasses were worn (FOV: 100°H × 50°V). (2) (PIN) Subject wore *pinhole* apertures and wore the shutter glasses over the pinhole apertures (FOV: 35°H × 35°V). (3) (ND) Subject wore the apertures and *neutral density* filter over the shutter glasses (FOV: 35°H × 35°V).

For each viewing condition described above, subjects were tested using 2 scene environments, ENV and NoENV, under which they had to size the virtual coke bottle. Furthermore, for each of the conditions, subjects sat on a chair facing the front wall of the CAVE and were placed at 3 different viewing distances from the front screen: 6.5 ft (FAR) 0.5D, 5 ft (MID) 0.65D, and 3.5 ft (NEAR) 0.93D. Consequently, at each viewing distance one bottle was drawn as if it were at the CAVE wall. These conditions were randomly sequenced for each subject. The physical coke bottle was visible to subjects by simply turning their head. The initial size of the virtual coke bottle was randomly varied from 0.2 to 2.0 of its correct size and the subject was required to size the virtual coke bottle to match the size of the real coke bottle placed at that distance from the subject. Ten coke bottle sizing operations were performed for each virtual bottle location.

The first run in each experiment was a trial run using the ENV scene. This allowed the subjects to familiarize themselves with sizing the virtual coke bottle. Data was collected but not used in the analysis. Subjects were encouraged to take 5 min breaks as often as they needed to avoid eye fatigue. The total experiment time varied from 45 to 75 min.

Subjects

The four subjects tested were volunteers aged between 22 and 33 years. All subjects were healthy with normal oculomotor function, visual and stereo acuities (measured at UIC Eye Clinic). Each of the four subjects was familiar with the operation and environment of the CAVE. Each subject was unaware of the hypothesis being tested. The subjects' task was to adjust the size of the virtual object (2-L Coke bottle) so that they perceived the virtual object's size as being identical to that of a physical coke bottle placed at the same distance from the subject.

Analysis

The subject's setting of the virtual object size was stored by the computer and then compared to the true size of the object. To evaluate subject performance, we developed a measure called size-ratio that represents the size of the perceived bottle set by subjects divided by the correct bottle size.[12] Consequently, when the subject is sizing the bottle according to size-constancy the size-ratio values will be 1 at each bottle location. Otherwise, a virtual bottle of the correct size that appeared too large would be reduced by the subject resulting in size-ratios less than 1. Conversely, that same coke bottle perceived too small would be enlarged and results in size-ratios greater than 1.

Data was analyzed using the statistical tools from Microsoft Excel and SPSS for Windows. A repeated-measures analysis of variance (ANOVA) was used with bottle distance as the within-subject factor, and scene conditions [ENV and NoENV] and viewing conditions [REG, PIN, ND] as the 2 between-subject factors. The analysis to test for similar regression slopes was performed using a paired *t*-test. The mean size-ratio (for 10 sizing operations) for each bottle location served to create regression slopes using the least squares method. Since our population was small, a power analysis of the paired *t*-test was computed using SAS 9.1 (SAS, Inc). A power estimate of 0.8 or above is a realistic value to assume that the decision to reject the null hypothesis is correct.[2]

RESULTS

Comments from subjects indicated that determining the correct size of the bottle was the easiest to perform while viewing the ENV scene under the REG condition while the NoENV scene under all viewing conditions was substantially more difficult. Both the PIN and ND viewing conditions were the most difficult to perform; subjects took longer to set the bottle size.

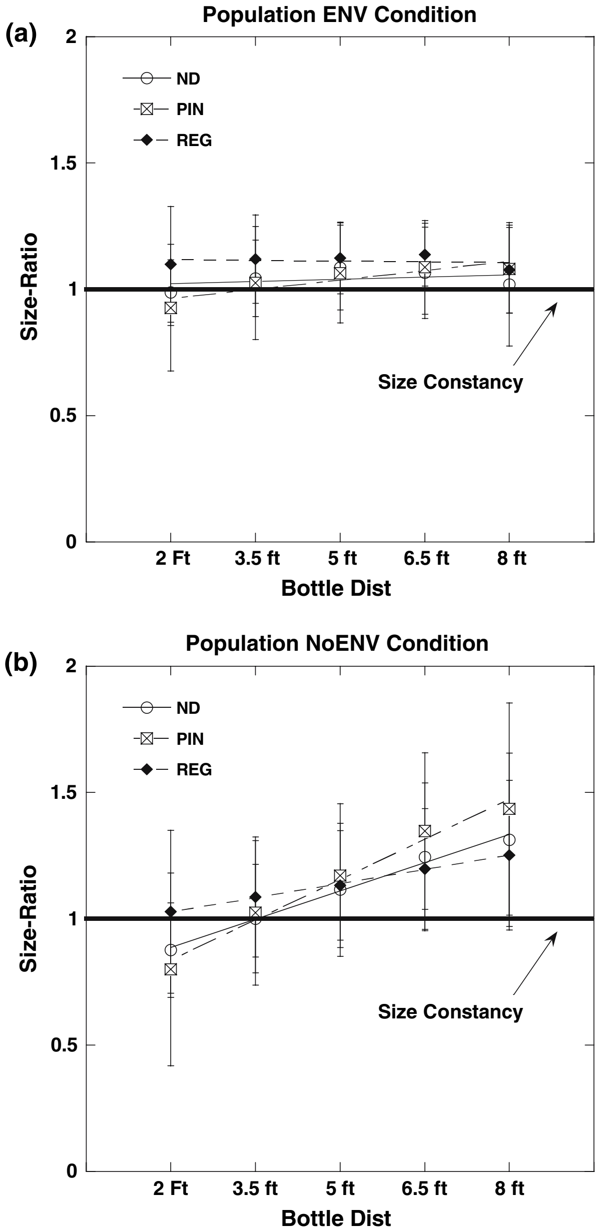

FIGURE 2. Population's performance for the ENV and NoENV scenes under REG, ND, and PIN viewing conditions [shown in legend]. The size-ratio average and standard deviation are plotted as a function of the distance of the bottle from the subject. The theoretical size-ratio slope for size-constancy (solid line) is plotted in all the graphs. Population's performance using the REG, PIN, ND viewing conditions and viewing the ENV (a) and NoENV (b) scenes. Size-ratio values with the ENV scene tended to maintain a value that was slightly above 1 for the different viewing conditions but were not significantly different with bottle distance from the subject. The NoENV scene caused a deviation from size-constancy as shown by the slope of the data. Notice that viewing with a pinhole (PIN) did not restore size-constancy performance. Size-ratio values for each viewing condition scene shows a similar responses, i.e., no size-constancy.

Our population's average size-ratio settings for each bottle at its given distance are shown along with their standard deviations in Fig. 2. Figure 2a shows the population's performance when the ENV scene was viewed. These data show that in all viewing conditions the data remain almost horizontal. A mixed ANOVA was conducted to assess whether there were bottle distance and view condition differences in the sizing of the bottle. With the scene fixed at ENV our results showed a significance main effect of view condition ($F[2, 165] = 4.12$; $p < 0.018$) but no effect of bottle distance with no interactive effects. This indicates that the sizing was consistent from bottle to bottle (i.e., size-constancy) even if there was an offset component due to viewing condition. The data in Fig. 2a shows this consistency of response from bottle to bottle within each viewing condition. The data in Table 1 shows that

TABLE 1. The percent similarity between our population's size-ratio regression slopes (in parentheses) and that predicted by VA regression slopes (in parentheses next to FAR, MID, NEAR) at each viewing distance.

Population performance: regression slopes					
		Scene		Paired *t*-test power	
Viewing condition	Distance (ideal slope)	ENV %VA (Avg. slope)	No-ENV %VA (Avg. slope)	ENV	No-ENV
REG	FAR (0.153)	–3% (–0.004)	18% (0.027)*	–	0.45
	MID (0.2)	2% (0.003)	13% (0.026)*	–	0.20
	NEAR (0.28)	–2% (–0.004)	20% (0.057)*	–	0.40
PIN	FAR	16% (0.024)*	53% (0.080)*	0.82	0.94
	MID	12% (0.024)*	53% (0.106)*	0.56	0.98
	NEAR	9% (0.024)*	46% (0.131)*	0.80	0.80
ND	FAR	3% (0.005)	34% (0.052)*	–	0.35
	MID	1% (0.002)	42% (0.083)*	–	0.80
	NEAR	3% (0.009)	31% (0.088)*	–	0.99

Statistical power for each test where the null hypothesis was rejected is given in the last column.
*Significantly different than REG-ENV slope.

the regression slopes for REG and ND conditions are between –3 and 3% of the VA slope for the respective distance from the screen. While the PIN condition slope was higher at 9–16%. Examination of the sizing performance while viewing the NoENV scene shows a dramatically different response in Fig. 2b. Under all viewing conditions the slopes of the lines have increased compared to the ENV condition. The ANOVA with scene fixed at NoENV showed a significance main effect for bottle distance only ($F[4, 165] = 17.23; p < 0.0001$) with no interactive effects. This indicates that sizing of the bottle changed with distance from the subject and as shown in Fig. 2b bottle size increased with distance from the subject, a trait of VA performance.[12] Comparing the effect of the two scenes on subject performance shows that the regression slopes for the NoENV scene, Table 1, are significantly higher than those at REG-ENV in all conditions. The paired samples *t*-test indicated that PIN had on average a significantly higher slope than REG-ENV condition $t(11) = -8.671$, $p < 0.001$, $d = 2.5$. The NoENV slope values are 2.8–43 times higher than those measured when viewing the ENV scene. In fact the highest slope is found in the PIN condition using the NoENV scene. This is contrary to the expectation that the PIN-NoENV condition would produce a slope closer to that of the REG-ENV condition and a slope lower than that produced in the REG and ND condition for the NoENV scene. The power analysis for PIN conditions showed that all but the MID-ENV case had a power above 0.8.

DISCUSSION

Our results suggest that opening the loop on accommodation and allowing accommodation to be driven to any state by the vergence system [to fuse the bottles] or other systems does not restore our subjects' size judgments back to size-constancy. Comparing the REG, PIN, and ND conditions when viewing the NoENV scene shows higher regression slopes than that expected under size-constancy performance. Furthermore, these slopes were more like those expected for VA performance[10] and were significantly different than those obtained viewing the REG-ENV scene. In addition, the PIN-NoENV slope was greater than the REG-NoENV slope. If accommodation driven by the need to focus on the projection screen was playing a dominant role in producing the size errors found in the REG-NoENV scene condition then we would expect the subject's performance to return to or approach a size-constancy slope rather than VA.[2] However, this was not the case in our subjects. The data clearly shows that the PIN condition did not restore the sizing behavior to that expected for size-constancy as seen in the REG-ENV condition.

The current study was not intended to specifically identify if there was a difference in the cues used for

---

[2] To explain further, since accommodation is open-loop when viewing the image through the pinhole aperture, accommodative demand [i.e., accommodative blur signal] over a large range of distances should be reduced due to the large depth of focus. In addition, vergence driven accommodation could be manifested with little resistance from the accommodative feedback loop used to reduce blur. Therefore vergence demands needed to fuse targets could drive accommodation essentially unencumbered. Even though it is unlikely that this resulting accommodative state would be identical to a physical world condition, the release from a fixed accommodative state using a pinhole might change size and distance judgments about virtual objects. That is, releasing accommodation from being fixed to the distance of the physical wall and allowing it to vary with vergence could produce conditions that are more in concert with other visual information and show an improvement in size setting performance.

physical world verses the virtual world. Consequently, we did not examine this task in the physical world. However, Frisby *et al.*[7] have shown that cue integration and weighting can be different if we view a physical object verses one displayed using a stereogram (a visual condition similar to that in VE). They found that cues such as ridge gradients and accommodation appear to behave differently in each environment. They state that there are important differences between the stereogram viewing and that in the physical world and that one must be careful in equating cue integration in the two conditions. It is important to note that our work was done comparing VE conditions and not VE with the physical world conditions as did Wann *et al.*[20] However, Frisby's point is very important and one that is often lost on some that work with VE. Virtual environments simulate the physical world as much as the technology will permit but it does not equate to the physical world in many respects. In many cases this simulation is a good approximation of the physical world but each simulation needs to be evaluated independently. As Frisby points out some conditions may be appropriate only in the physical world.

Finally, it is worth mentioning that in the physical world monocular cues to depth are natural and usually abundant. In fact, it takes effort to arrange a situation that would diminish these cues to the subject. In VE, displaying less complex scenes is easier than showing more complex ones. Creating a VE that has numerous cues to depth (monocular and stereovision) takes time to program and computer-time to generate. Thus, it is more expensive to generate a complex world compared to a sparse world in terms of cost, programming time, and display time. Since it may be many years before VE will incorporate accommodative stimuli (if ever) understanding whether perception of the true size of an object is mainly meditated by accommodation can have an important impact on the use and development of the technology. From these experiments we infer that cues other than accommodation are paramount in generating size-constancy in a VE. Future experiments designed to explore the relationships that exist between the physical and virtual environments will help us better utilize this extraordinary technology by helping us understand what important cues need to be presented to the user.

ACKNOWLEDGMENTS

The authors thank our subjects for their diligence and tireless interest in these experiments. We also thank Rajesh Balasubramanian, and Tim Portlock for their assistance in conducting these experiments. The authors also thank the anonymous reviewers for their insightful and constructive comments. This research was supported by General Motors Research contract #2-5-37457.

REFERENCES

[1] Baitch, L., and R. C. Smith. Physiological correlates of spatial perceptual discordance in a virtual environment. In: Proceedings Fourth International Projection Technology Workshop, Ames, Iowa, 2000.

[2] Bausell, R. B., and Y.-F. Li. Power Analysis for Experimental Research. A Practical Guide for the Biological, Medical and Social Sciences. Cambridge UK: Cambridge University Press, 2002.

[3] Charman, W. N., and J. A. Jennings. Properties of annular artificial pupils. *Vision Res.* 19:947–949, 1979.

[4] Cruz-Neira, C., D. Sandin, T. Defanti, R. Kenyon, and J. Hart. The CAVE audio-visual environment. *A.C.M. Trans. Graphics* 35:65–72, 1992.

[5] Descartes, R. La Dioptrique, discours VI: de la vision, Vol. 1. Cambridge, England: Cambridge Univ. Press, p. 1637, 1985.

[6] Eggleston, R. G., W. P. Janson, and K. A. Aldrich. Virtual reality system effects on size-distance judgments in a virtual environment. In: Proceedings of VRAIS, 1996, pp. 139–146.

[7] Frisby, J. P., D. Buckley, and J. M. Horsman. Integration of stereo, texture, and outline cues during pinhole viewing of real ridge-shaped objects and stereograms of ridges. *Perception* 24:181–198, 1995.

[8] Harvey, L. O., Jr., and H. W. Leibowitz. Effects of exposure duration, cue reduction, and temporary monocularity on size matching at short distances. *J. Opt. Soc. Am.* 57:249–253, 1967.

[9] Holaday, B. E. Die Grössenkonstanz der Sehdinge bei Variation der inneren und äusseren Wahrnehmungsbedingungen. *Arch. Gesch. Psychol.* 88:419–486, 1933.

[10] Holway, A. H., and E. G. Boring. Determinants of Apparent Visual Size with Distance Variant. *Am. J. Psychol.* 54:121–151, 1941.

[11] Kenyon, R. V., T. A. DeFanti, and D. J. Sandin. Visual requirements for virtual environment generation. *J. Soc. Inform. Display* 3:211–214, 1995.

[12] Kenyon, R. V., D. Sandin, R. Smith, R. Pawlicki, and T. Defanti. Size-constancy in the CAVE. *Presence: Teleop. Virt. Environ.* 16:172–187, 2007.

[13] Leibowitz, H. W., and R. A. Dato. Visual size-constancy as a function of distance for temporarily and permanently monocular observers. *Am. J. Psychol.* 79:279–284, 1966.

[14] Leibowitz, H., and D. Moore. The role of changes in accommodation and convergence in the perception of size. *J. Opt. Soc. Am.* 56:1120–1123, 1966.

[15] Neveu, C., T. Blackmon, and L. Stark. Evaluation of the Effects of a Head-mounted Display on Ocular Accommodation. *Presence: Teleop. Virt. Environ.* 7:278–289, 1998.

[16] Peli, E. Real vision and virtual reality. *Opt. Photon. News* July:28–34, 1995.

[17] Robinett, W., and J. P. Rolland. A computational model for the stereoscopic optics of a head-mounted display. *Presence: Teleop. Virt. Environ.* 1:45–62, 1992.

[18] Rolland, J. P., W. Gibson, and D. Ariely. Towards quantifying depth and size perception in virtual environments. *Presence: Teleop. Virt. Environ.* 4:24–29, 1995.

[19] Rolland, J. P., M. Krueger, and A. Goon. Multi-focal planes in head-mounted displays. *Appl. Opt.* 39:3209–3215, 2000.

[20] Wann, J., S. Rushton, and M. Mon-Williams. Natural problems in the perception of virtual environments. *Vision Res.* 35:2731–2736, 1995.

[21] Ward, P. A., and W. N. Charman. On the use of small artificial pupils to open-loop the accommodation system. *Ophthal. Physiol. Opt.* 7:191–193, 1987.