Size-Constancy in the CAVE®

Robert V. Kenyon, Daniel Sandin, Randall C. Smith[1], Richard Pawlicki[1], and Thomas Defanti

Electronic Visualization Laboratory, Department of Computer Science, University of Illinois at Chicago
[1]General Motors Research and Development, Warren, MI

Abstract

The use of Virtual Environments (VE) for many research and commercial purposes relies on its ability to generate environments that faithfully reproduce the physical world. However, due to its limitations the VE can have a number of flaws that adversely affects its use and believability. One of the more important aspects of this problem is whether the size of an object in the VE is perceived as it would be in the physical world. One of the fundamental phenomena for correct size is size-constancy, i.e., an object is perceived to be the same size regardless of its distance from the observer. This is despite the fact that the retinal size of the object shrinks with increasing distance from the observer. We examined size-constancy in the CAVE and found that size-constancy is a strong and dominate perception in our subject population when the test object is accompanied by surrounding environmental objects. Furthermore, size-constancy changes to a visual angle performance (i.e., object size changed with distance from the subject) when these surrounding objects are removed from the scene. As previously described for the physical world, our results suggest that is necessary to provide surrounding objects to aid in the determination of an object's depth and to elicit size-constancy in VE. These results are discussed regarding their implications for viewing objects in projection-based VE and the environments that play a role in the perception of object size in the CAVE.

Will Appear in: *Presence,* 2007

Introduction

Projection-based virtual environments (VE)s, such as the CAVE, have made great inroads in the VE community. As a result, there is a large installed base of projection-based systems used for a variety of purposes in both scientific and commercial fields. A common use of VE relies on its ability to create scenes within the environment that faithfully replicate those in the physical world. However, due to limitations, the VE can have a number of flaws that adversely affects its use and the credibility of the environments that it offers. One of the more important aspects of this problem is whether the size of an object in the VE is perceived, as it would be in the physical world. One of the fundamental phenomena for correct size is size-constancy, i.e., an object is perceived to be the same size regardless of its distance from the observer.

Many studies of perceived size of objects in the physical world have been performed (see Sedgwick (1986) for review). Descartes (1637) 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 shrinks with increasing distance from the observer. Holaday (1933) showed that removal of various cues would change this behavior to one relying on the physical optics of the situation. He showed that as the number of cues to depth 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 (1941) confirmed these findings for objects 10-40ft from the observer. Harvey and Leibowitz (1967) showed similar results for objects 1-9ft from the observer. Furthermore, they and Leibowitz and Dato (1966) showed that removal of stereovision had little to no effect on performance and that performance was only affected by the removal of monocular depth cues.

Further studies of size-constancy have been performed in the physical world with consistently reproducible results. In addition, explorations of the characteristics that affect size-constancy have also been investigated quite exhaustively (see Cutting and Vishton, 1995 for review). However, in the virtual environment size-constancy has not been demonstrated. This may indicate that due to the characteristics and limitations of this synthetic environment some perceptions that appear readily the physical world may not materialize in the virtual world. A study by Eggleston et al. (1996) using a head mounted display (HMD), did not find size-constancy in their subjects who instead used a visual angle approach in sizing objects. That is, instead of the perceived size of the object remaining the same regardless of its distance, the object size perceived by the subject shrank with increasing distance of the object from the subject. Baitch and Smith (2000) showed similar results for a single object that was approximately 15 inches from the subject using a CAVE system that provided stereovision but no head tracking or surrounding items in the environment to support the object’s position in space. The absence of such a fundamental and essential percept in VE would be a significant detriment to the use of VE.

Unlike other electronic forms of visual display, VE can provide veridical size and distance cues to the user. Therefore, one would expect similar size-constancy changes to those reported in the physical world. However, Eggleston et al. (1996) and Baitch and Smith (2000) showed no size-constancy but visual angle performance. Nonetheless, we believe that these results are the consequence of either exceeding the visual limits of the VE or using a sparse environment that eliminated surrounding objects that provide cues to depth that others have shown to be so important when doing this task in the physical world.

Our objective in this research was to understand to what extent size-constancy could be experienced in this VE. Therefore, we placed objects at distances that subjects could easily appreciate in this VE and our experimental protocol paralleled those performed in the physical world that have been successful in demonstrating size-constancy. Thus, this experiment was not designed to delineate which cues to depth were prominent. Subjects were tested using a rich environment which was constructed so that a number of surrounding items in the environment accompanied the virtual object [i.e., a bottle], and a sparse environment where no surrounding items were provided and only the bottle was visible. Our results were similar to those described for subjects performing this task in the physical world. Size-constancy was more prevalent when the environment contained objects surrounding the test object in the scene to aid depth perception. Without these surrounding objects, most of the subjects adopted a visual angle (VA) performance when sizing the bottle.

Methods

These experiments were performed using a projection-base virtual environment know as the CAVE (CAVE Automatic Virtual Environment). The CAVE surrounded the viewer with three rear-projection screens arranged in a 10 ft. cube (Cruz et al. 1992). A fourth projector overhead pointed to a mirror, which reflected the images onto the floor. In order to create stereoscopic objects, two off-axis perspective images were consecutively displayed; one visible to the right eye, the next to the left eye. The visibility of images by each eye was controlled by the stereo glasses (Stereographics, Inc) which rapidly turned each lens on and off in synchrony with the corresponding images on the screen. The CAVE screens were driven by an SGI Onyx II with two Infinite-Reality graphics pipelines; each split into two channels to control the projected images. The image resolution was 1280x1024 with a refresh rate of 120 Hz, which resulted in an update rate of 60 stereo images per second. 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 six-degrees-of-freedom head-tracking system (Intersense IS900) provided real-time head position so that the correct stereoscopic perspective projections were calculated for each wall dependent on viewer position. The tracker, which had an end-to-end latency of 65ms, measured using a video technique (Ding et. al, 2000), was calibrated to an accuracy of ±0.5 inch for the tracking distances used in the experiment. A second sensor and buttons in a wand held by the viewer provided interaction with the virtual environment.

Construction of the Environments

The environments that we used in this study were constructed employing the most common elements of computer graphics applied to VE. These components were stereovision, perspective rendering, head tracked projection centers, texture mapping, diffuse and specular lighting effects using a Gouraud algorithm and hidden surface removal. In each environment, the object whose size could be changed was the 2-liter coke bottle. The virtual coke bottle was textured with an image from a physical 2-liter coke bottle label. In many of the size-constancy experiments performed previously in the physical world, researchers achieved cue reduction by restricting the subject's field-of-view using visual field stops which blocked the subject's view of a number of objects within the target's area of interest. In our experiments, we achieved this by removing objects from the environment. Therefore, in our virtual environment two different scenes were produced: rich and sparse. For the rich environment [ENV] (Figure 1a), additional environmental components accompanied the coke bottle, i.e. table and a floor extending to a horizon. We employed stereovision, tracked head position for computing perspective, Gouraud-shading, and occlusion of objects in the scene. In addition, texture mapped surfaces included tables used to rest the bottle on and a checkerboard floor for the table to stand on. A clear demarcation between the grey sky and ground plain was visible with the horizon line. As a result, a number of depth cues were available in the scene including stereopsis, perspective, occlusion, texture gradients, motion parallax due to head motion, relative size/motion/position, lighting effects due to diffuse and specular shading [no shadows], convergence eye movements, fixed accommodative stimulus, and eye height about the floor since the virtual floor was registered with the physical floor. For the sparse environment [No-ENV] (Figure 1b), we produced a scene that contained only this bottle against a grey background. The cues in this environment were confined to the bottle itself. However, cues such as relative size, motion, and position were eliminated along with eye height since there was no floor or horizon.

Rich Environment (ENV): This scene consisted of a gray green checkered floor with a wood textured table and a coke bottle on top of the table (Figure 1a). The height (30, 33, 36 inches above the floor) and appearance of the table was changed for different runs, and the distance of the coke bottle from a subject with each run. Subjects used the wand joystick to increase and decrease the size of the coke bottle and pressed a wand button to continue once they had finished sizing the virtual coke bottle. The head was tracked so the scene was drawn appropriate to the position of the subject's head. Motion parallax was available if the subject moved their head but we did not actively move the objects in the scene. The subjects could converge their eyes on objects at different [simulated] distances but the accommodative stimulus was fixed at the CAVE wall.

Sparse Environment (No-ENV): Subjects were presented with a gray background (Figure 1b). Only the virtual coke bottle was presented. It was suspended in mid air at different heights from the floor (corresponding to the table heights) and at the five different distances from the user as described in the previous section. The head was tracked identically to that described above. Motion parallax was available if subjects moved their head but our observations showed that subjects' head position was quiet and did not move during the experiments. The subjects could converge their eyes on objects at different [simulated] distances but the accommodative stimulus was fixed at the CAVE wall.

Experimental Protocol

The subjects' task was to adjust the size of the virtual object (2-liter Coke bottle) so that they perceived the virtual object's size as being identical to that of a physical coke bottle if placed at same distance from the subject (Figure 2). We purposefully chose to display a coke bottle since it is ubiquitous and its size was well known to subjects. However, to aid in our subject's sizing task, a physical 2-liter coke bottle was visible to the subjects for comparison to the virtual object. The 2-liter coke bottle was placed on a black plastic table at a height of 4 feet. The table was positioned at the front right hand side of the CAVE at an approximate distance of 4 feet from each subject. Both the physical and virtual coke bottle was 12 inches tall and 5.5 inches (maximum) wide (physical bottle: ~14° x ~6.5° H). The physical coke bottle was lit by a spotlight mounted on the top left hand wall of the CAVE at a height of 10 feet and at a distance of 8 feet from the front CAVE wall. The physical coke bottle was visible to the subjects by simply turning their head approximately 40° to the right.

The virtual coke bottle was displayed at distances of 2.0, 3.5, 5.0, 6.5, and 8.0 ft. from the subject. A virtual bottle at its correct size at these distances would yield a vertical visual angle at the eye of approximately, 26°, 16°, 11°, 9°, and 7°, respectively. These distances resulted in the virtual object being placed in front of, on, and behind the front screen of the CAVE. Each subject received an explanation of size-constancy to insure that they understood the task. In addition, familiarization runs which could have up to 50 trials were done initially where the subjects adjusted the size of the virtual bottle and took about 5-10 minutes to perform. Subjects were encouraged to take 5-minute breaks between runs as often as they needed to avoid fatigue. The total experiment time varied from 45 minutes to 60 minutes.

The two visual conditions (ENV and No-ENV) were randomly presented to each subject. In each visual condition, the subject was placed at three different viewing distances from the front screen: 6.5 ft. (FAR), 5 ft. (MID), and 3.5 ft. (NEAR) (Figure 3). This gave us six distinct conditions to test.

In each run (or trial), a subject was presented with the image of a virtual coke bottle at one of the five possible locations chosen randomly by the program. For the ENV condition runs, three different virtual tables were used that had the same length and distance from the subject but differed in tabletop texture, shape, and height (Figure 4). The computer randomly set the initial size of the virtual coke bottle from 0.2 to 2.0 times the normal size (12 inches) of the bottle then drew a bottle of that size at the specified distance. The user was then asked to change the size the virtual coke bottle to match the size of a physical coke bottle if placed at that distance. The subject repeated the coke bottle sizing operation 50 times (10 for each bottle location). In the No-ENV condition, only the bottle was displayed but the same bottle protocols were used as in the ENV condition.

Subjects

Twelve subjects were tested (EC1-EC12). The mixture of subjects included people who were experienced (40%) and inexperienced in VE. Experienced subjects had a minimum of 6 months of using the CAVE; for inexperienced subjects, this was their first time using the CAVE. All subjects were given a complete vision/oculomotor exam at the Illinois Eye and Ear infirmary. The following characteristics measured included: visual acuity, AC/A ratio, CA/C ratio, IPD, phoria, and stereo acuity. All subjects were corrected to 20/20 or better and had normal stereo vision.

Analysis

Subject performance was evaluated quantitatively using several measures based on the size of the virtual bottle set by the subject. One basic measure, called size-ratio, represented the relative size of the virtual bottle compared to the proper size of the physical bottle.

$$\text{Size–Ratio} = \frac{\text{Bottle size set by subject}}{\text{Correct bottle size}} \tag{1}$$

"Bottle size set by subject" corresponds to the size of the virtual bottle set by the subject performing the task and the "correct bottle size" was fixed at 12 inches (height of the physical 2-liter coke bottle). For example, the Size-Ratio values would be 1 at each bottle location if the subject sets the bottle size according to size-constancy.

Since in projection-based VE everything is drawn on the CAVE wall, we calculated the VA setting that would result if subjects perceived their distance to the bottle as being the distance they were from the CAVE wall regardless of the bottle's intended distance from the subject. Given this premise, we calculated the expected size-ratios that would result by calculating the expected bottle size set by the subject and then using it in the size-ratio formula above.

$$\text{Bottle size set by subject} = \tan\theta_{va} * \text{Distance to virtual bottle} \tag{2}$$

θ_{va} represented the visual angle that a bottle 12 inches high would make when the subject was at one of three distances from the CAVE wall.

$$\theta_{va} = \tan^{-1}\frac{\text{Correct bottle height on CAVE wall}}{\text{Distance to the CAVE wall}} \tag{3}$$

This produces three visual angles 15.9° (NEAR), 11.3° (MID), and 8.7° (FAR) that were used to obtain the theoretical size-ratio settings plotted in Figure 5.

The percentage relationship between the subject's size-ratio data regression slopes to that of the predicted VA slopes was calculated using the equation:

$$\text{Percent VA slope} = \left[\frac{\text{Subjects' regression slope}}{\text{VA slope}_{\text{theoretical}}}\right] * 100\% \tag{4}$$

The "$\text{VA slope}_{\text{theoretical}}$" was taken for the appropriate viewing distance from Figure 5. For example, if the regression slopes of the subject's data at MID distance from the CAVE wall were identical to that shown in Figure 5 (for MID), then the "Percent VA slope" would be 100% (the subject is showing no size-constancy). If the subject regression data showed strong size-constancy, the regression slope would be zero and the "Percent VA slope" would be 0%.

The absolute value of the error for each size judgment and the mean absolute value of the error was calculated to examine the differences between ideal performance and the size-ratio data collected from our population. Mean absolute value of the error averaged absolute value of the size-ratio judgment errors for a given bottle location and was computed using the following equation:

$$Mean\ absolute\ error = \frac{1}{n}\sum_{1}^{n}\left|SizeRatio(n) - 1\right| \qquad (5)$$

The need to use an absolute value derived from the fact that size-ratio errors can be above and below 1 (the ideal size-ratio value). Therefore, we could have a mean error of zero but an absolute value mean error that is much greater than zero. Since we are interested in the subject's performance (i.e., how much error there was in setting the bottle size) we adopted this approach to get this information from our data.

Subject data was analyzed separately and as a group. Statistical tools used were Microsoft Excel, SPSS for Windows (v11) and Data Desk (v6.1). Statistical methods used included single factor ANOVA, multiple factors ANOVA and Turkey post hoc t-tests.

Results

The ability of subjects to set the virtual bottle to the correct size (a size-ratio of 1) was best in the ENV condition where a number of objects accompanied and supported the position of the bottle. Not only was the performance consistent with that for size-constancy but also the task was easier to perform according to subject reports. Furthermore, we found that side to side head movements where small or absent in our subjects while performing this task in either environment indicating that motion parallax was not a strategy that subjects used to aid them in target sizing.

Population Performance

When we averaged the size-ratio settings for each bottle position across subjects and for each condition (ENV and No-ENV) we found that size-ratio settings were consistently closer to 1 in the ENV than the No-ENV condition. For the ENV condition (Figure 6a) subjects produced a mean size-ratio that hovered close to a size-ratio of 1 for different bottle positions. In contrast, the mean size-ratio for the No-ENV condition (Figure 6b) increased as the bottle positions receded from the subject. The size-ratio settings for this condition ranged between 0.4-1.6 for the bottle distance of 2-8ft from the subject.

We also examined the absolute value of the error for size judgments made in the ENV and No-ENV conditions in our population. The absolute value of the error in Figure 7 shows a clear difference between ENV and No-ENV performances. Examination of the absolute value of the error for all judgments showed that 68.5% of the errors were 0.2 (or 2.4 inches) and below with the ENV condition while only 47.4% of the errors fell within this range with the No-ENV condition. Furthermore, when we averaged theses errors for each bottle location from the CAVE wall, mean absolute value of the error, we found that the ENV condition (Figure 8a) showed errors below 0.2 for the bottle positions of 0 to -4.5 ft from the CAVE wall (i.e. inside the CAVE). A very different picture was portrayed in the No-ENV condition where values below 0.2 occurred only at the CAVE wall (Figure 8b). Consequently, the mean absolute value of the error was larger for the No-ENV condition than the ENV condition across all bottle positions.

Our subjects' performance was also quantified by examining the degree of similarity between regression slopes for their data and those computed for a theoretical VA performance (Figure 5). We found that the regression slopes obtained in the ENV condition more closely matched the slopes expected with size-constancy and conversely the slopes in the No-ENV viewing condition more closely matched those associated with VA performance. The percentages shown in Table 1 (ALL) for the No-ENV case show that the regression slopes under this condition has a closer similarity to VA slope than to a size-constancy slope. The No-ENV slopes were 3 or 4 times larger than those for the ENV condition. This illustrated once again that our population's performance in the ENV condition was very different from that in No-ENV condition.

Individual Subject Performance

When we examined individual subject data and compared their settings in the two visual conditions, we found several categories of performance that ranged from excellent size-constancy performances in both viewing conditions to poor size-constancy in both viewing conditions.

Different Performance in ENV and No-ENV: Eight of the twelve subjects tested showed size-ratio data that had shallow regression slopes (2%-25%) in the ENV condition. The size-ratio data for individual bottle distances were similar to that shown in the population average data (Figure 6). Typical performance for this group of subjects in the ENV condition is shown in Figure 9a where subject (EC-9) size-ratio settings are very close to the optimal value of 1. The resulting regression slopes for this subject are a very small percentage of the VA curves as seen in Table 1 (EC-9). In the No-ENV condition, nine of the twelve subjects displayed a gradual increase in size-ratio as the bottle receded from the subject. This was illustrated by the higher VA slopes percentage in the No-ENV condition whose values were 2-3 times those in the ENV condition. A typical example of this behavior is shown in Figure 9b for the same subject (EC-9). Comparing the two graphs, we found a dramatic change in the performance between the two conditions. The average regression slopes for the No-ENV (EC-9) condition in Table 1 show the VA slopes that are close to those predicted in Figure 5 (59-79%). When we calculated the regression slopes for all nine subjects' data, we found a significant difference ($p < 0.0001$) between the average regression slopes for the ENV (13%) vs. the No-ENV (43%) conditions.

Similar Performance in ENV and No-ENV: Despite our population's strong demonstration of maintaining a size-ratio close to 1 in the ENV condition, there were two subjects (EC-1, EC-8) who showed large changes in size-ratio with bottle position in the ENV condition. The resulting regression slopes were similar to those we measured in our population for the No-ENV viewing condition. In addition, for these subjects, both ENV and No-ENV conditions produced similarly large regression slopes. The average regression slopes for subject EC-8 (Table 1) were typical for this group of subjects. The regression slopes were large in both viewing conditions, with the No-ENV condition slope larger than the ENV condition. The data plotted in Figure 10a show a typical performance for subjects in this category. Specifically, subject EC-8 showed large deviations from the size-ratio value of 1 despite the presence of the ENV environment.

In contrast, three subjects (EC-2, EC-4, EC-5) show mean size-ratio data close to 1 in the No-ENV condition. Subject EC-4 (Table 1) showed average regression slopes that were small in both viewing environment conditions. As seen in Figure 10b, subject EC-4 demonstrates a performance indicative of size-constancy in the No-ENV condition even when the additional objects, to support the depth of the bottle, are absent.

Comparison of the visual examination data from our population showed no differences among these subjects' visual characteristics that might explain the differences we have described.

Effect of Subject Position in the CAVE on Performance

When the size-ratio regression slopes were large, the subject's viewing distance significantly affected size-ratio settings for bottles equi-distant from the subject. In the ENV condition, only two subjects, (EC-1, EC-9) who had large percent VA regression slopes (65-133%), showed a significant difference in their size-ratio settings with the subject's viewing distance ($p < 0.01$). In the No-ENV condition, nine of the 12 subjects had high regression slopes and a significant difference in the size-ratio values with respect to viewing distance. For instance in Figure 9b, we see a significant difference ($p < 0.01$) between the size-ratio settings for each bottle at the three different viewing distances. In our population, we found at least one bottle setting that was significantly different ($p < 0.01$) from that same bottle distance from the subject at another seating distance. The most common finding was that significant differences were seen in setting the far bottles (6.5 or 8ft from the subject) when compared to those closest.

In four subjects, when their performance in the No-ENV condition was plotted against the position of the virtual bottle in space relative to the CAVE wall rather than from its distance from the subject, we found that the size-ratio settings per bottle were almost uniform. In Figure 11 we have re-plotted a representative subject's data, EC-9, that shows this relationship. This subject showed a strong uniformity of size-ratio settings for bottles at a particular location in <u>space</u> rather than a location from the subject. This response may have some relationship to the quantization of images produced by the virtual environment at the CAVE wall.

<u>Discussion</u>

Effects of visual cues on performance

In our experiments, a vast majority of subject size judgments were close to a size-ratio of 1 when we provided objects in the scene to support the depth of the bottle (ENV condition). Most of the errors in the ENV condition are very small for bottles on or in front of the CAVE wall. As the bottles receded from the wall (i.e. outside the CAVE) the errors began to increase. In contrast, in the No-ENV condition the errors were smallest at the CAVE wall and increased in front of and behind the wall. Furthermore, if we associated a regression slope less than 25% with a size-constancy performance, we found 55% our population runs showed strong size-constancy with the ENV condition but only 25% with the No-ENV condition. That is, the slopes of the regression curves for the No-ENV condition increased to such an extent that their values were closer to those predicted by a visual angle performance rather than size-constancy performance. A similar effect was reported by Holway and Boring (1941) when subjects were tested in the physical world. When looking down the corridor with all the other cues to depth available, subjects' performance was remarkably similar to that predicted by size-constancy. When many of environmental objects and therefore supporting cues were removed by the use of a circular tube, the performance moved towards that predicted by visual angle. One of the interesting differences between our results and those of Holway and Boring is that their distances ranged between 10-40ft and ours is 2-8ft from the subject. We should also mention (the obvious) that our results were performed using a virtual object whereas Holway and Boring study was performed using physical objects. Also, given the size of the bottle in our experiments and the resolution of the CAVE wall in our VE system, it would be very difficult to replicate the distances used by Holway and Boring due to the small size of the image that would be produced on the CAVE wall when using those distances. In fact, Eggleston et al. (1996) reproduced Holway and Boring experiments using a HMD and found that their subjects showed VA performance for all distances from 10-40ft. It may be that the large distances that they used were beyond the fidelity range of their hardware. However, whether size-constancy is preserved in HMD systems is an empirical question that needs further attention.

When we compare our results to studies that have used similar distances in the physical world, we find close agreement in the data. Harvey and Leibowitz (1967) and Koh, and Charman (1999) examined the performance of subjects to size objects placed from 1-9ft from them. They also found size-constancy when environmental cues derived from associated objects were made available to the subjects. When these objects were occluded by a field-of-view stop, visual angle performance was observed. In addition, with and without supporting cues to depth, subjects in their study and ours would see objects closer to them as being too large relative to objects further away when visual angle performance was measured. To compare their results with ours we have transformed their data into our format and plotted each in Figure 12[1]. Comparing their results with ours shows basic agreement for both ENV (Figure 12a) and No-ENV (Figure 12b) conditions. In Figure 12a, the size judgments in the Harvey & Leibowitz study are slightly better than in the CAVE. That is, the size-ratios are closer to 1. In the Koh, and Charman (1999) study the performance is worse than in the CAVE except for the most distance point. Harvey and Leibowitz (1967) as well as Leibowitz and Dato (1966) showed that performance degraded when objects surrounding the target were removed using a field-of-view visual stop. Furthermore, they indicated that it was the information to depth provided by these accompanying objects that remained the most important element to maintaining size-constancy. This is also shown in our data; the removing objects that provided supporting cues to depth from the environment reduced the incidence of size-constancy performance of our population. These findings may explain the Baitch and Smith (2000) results where their sparse environment, a condition similar to our No-ENV condition, was an insufficient environment for subjects to exhibit size-constancy.

In the No-ENV condition, three subjects were able to use the cues confined to the bottle to perform the task consistent with size-constancy. Consequently, even when cues were spatially constrained to the bottle there was enough information remaining for some subjects to make correct size judgments. Conversely, two other subjects showed visual angle performance with and without the additional visual cues for depth. Although these subjects had no visual anomalies according to the eye exam, their performance did not substantially change with the different visual conditions. The mechanism(s) for this performance cannot be identified through these experiments.

Size-Constancy in Physical and Virtual Worlds

This research was not designed to reveal which cues were the most important for the establishment of size-constancy but to show whether this percept was possible in VE. Our findings turnout to be very similar to those established for the physical world. Consequently, our conclusions regarding the effectiveness of cues that were present in our environments verses those in the physical world experiments are similar. For example, stereovision was provided in each of our environments but did not seem to be sufficient for most of the subjects when surrounding objects were removed from the scene as in the No-ENV condition. This result is consistent with that shown by Leibowitz and Dato (1966) in the physical world stereovision alone is not a strong cue for size-constancy. As has been shown in the physical world, the global appearance of cues such as perspective, obstruction, relative size, texture are all important for this percept (Holaday, 1933). Future work will be examining the use of motion parallax in particular for the maintenance of size-constancy in the VE.

Effect of the CAVE wall on performance

Individual data from four subjects during No-ENV case (Figure 11) and the population's mean absolute judgment errors (Figure 8) showed a relationship not to the distance that the objects were from the subject but the position of the bottles from the CAVE wall. These figures showed that the object's size and errors were similar for a given distance the object was from the CAVE wall. Furthermore, as bottles were presented farther from the subject, the associated standard deviations of the data increased with the distance of the virtual object from the subject. There was a significant difference between the nearest and farthest bottle settings in the No-ENV conditions. In the physical world, we also find a similar pattern as shown in Figure 10 (Harvey and Leibowitz (1967) and Koh, and Charman (1999)). In VE, the added factor of quantization error may contribute to this effect. This led us to examine the effect of the quantization of the image at the wall as a limiting factor in both disparity and resolution of the objects. This effect has been discussed by Hodges and Davis (1993) and may have some bearing on our data and the objects that we used.

The CAVE uses digital projectors with a resolution of 1280x1024 pixels to create the virtual environment. This quantizing of virtual objects can lead to an error in a viewer's depth perception of virtual objects (Pfautz 1996; Hodges and Davis, 1993). Figure 13 shows a case where a point object is being displayed in the CAVE. In this very simplistic case, the size of the point object is less than a pixel. However in order to display the object the CAVE projection system has to light up two pixels (one for the left and one for the right eye) in order for a viewer to see the object in stereo. This allows the object **P** to be displayed at any depth between **A** and **C** before a pixel boundary is reached and the perceived depth of the object changes. Figure 14 shows the error interval for various viewer distances and bottle distances used in our experiment. In a worst-case scenario the depth error interval is 23.79 inches (object distance = 8 ft, view dist = NEAR). This quantizing of depth can adversely affect an observer's ability to judge the distance of objects in the CAVE. As a result, their judgments about object size are also affected. An interesting observation is that as viewers move away from a screen there is a reduction in the quantizing error.

Conclusions

Our findings have significance to designers of virtual environments. To achieve the perceptual consistency in the virtual world that exists in the physical world, one must provide a surrounding environment of objects that supports the properties of the object(s) of interest. We have found that in the CAVE the ability of subjects to use size-constancy is predicated on the inclusion of objects to visually support the position in space of the object-of-interest and aid in delineating the cues to depth. This is similar to results from the physical world despite the differences in the methodology. Our best subjects could perform size settings in the rich environment with a 1.2-inch error in height for a 12-inch bottle. Furthermore, we found a small number of subjects could still achieve size-constancy with only the bottle in the environment. Conversely, still fewer subjects did not show a size-constancy performance with accompanying objects in environment we presented. An open question that would be interesting to investigate is whether improving the rendering of the environmental scene using photo-realistic methods would improve performance. The effects of quantization on subjects performance still needs further exploration. Additional experiments, comparing the size judgments made by subjects for different CAVE resolutions (e.g. 800 x 600, etc.), could help us understand whether this effect plays a significant role in perceiving virtual objects to be larger. It would appear, then, that for an observer, the presence, or absence of size-constancy depends upon the complexity of the environment.

It is worth mentioning that in the physical world 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 ones that are more complex. A VE that has numerous cues to depth (monocular and stereo) 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. Understanding the relationships that exist between the physical and virtual environments will help us to better utilize this extraordinary technology by supplying the most important information to the user.

References

Baitch, L., Smith, R.C., Physiological Correlates of Spatial Perceptual Discordance in a Virtual Environment, in Proceedings Fourth International Projection Technology Workshop, Ames, Iowa, 2000.

Cutting, J. E., & Vishton, P.M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), Perception of space and motion (pp. 71-118). San Diego: Academic Press.

Cruz-Neira C, Sandin D, Defanti T, Kenyon R, and Hart J., The CAVE Audio-Visual Environment. ACM Trans. on Graphics, 35: 65-72, 1992.

Ding H, Fuhu L, Dave P, Greg D, Dan S. Video-Based Measurement of System Latency. International Immersive Projection Technology Workshop, Ames IA, June 19-20, 2000

Descartes, R. (1637) La Dioptrique, discours VI: de la vision, p. 140, The Philosophical Writings of Descartes, Vol. 1, Cottingham J, Stoothoff R, Murdoch D, Cambridge Univ. Press, Cambridge, England, 1985, p. 172.

Eggleston R.G., Janson W. and Aldrich K. (1996) Virtual reality system effects on size-distance judgments in a virtual environment. Proceedings of VRAIS, pp 139-146.

Harvey L.O and Leibowitz H.W. (1967) Effects of exposure duration, cue reduction, and temporary monocularity on size matching at short distances, J. Opt. Soc. Am., 57: 2, pp. 249-253.

Hodges L.H. and Davis E.T (1993). Geometric considerations for stereoscopic virtual environments. Presence 2: 34-43.

Holaday BE (1933) Die Grössenkonstanz der Sehdinge bei Variation der inneren und äusseren Wahrnehmungsbedingungen. Arch. ges. Psychol., 88: 419-486.

Holway, A. H. and Boring E. G (1941) Determinants of Apparent Visual Size with Distance Variant. American Journal of Psychology, 54, pp. 121-151.

Koh, L and Charman, W. N. (1999) Size constancy and angular size matching in size perception of near objects. Optometry and Vis. Sci. 76:4, 241-246.

Leibowitz HW and Dato RA. (1966) Visual size-constancy as a function of distance for temporarily and permanently monocular observers. Am. J. Psychol. 79: 279.

Pfautz, J. D. (1996) Distortion of Depth Perception in a Virtual Environment Application. Dept. of Elec. Eng. and Comp. Sci., Massachusetts Institute of Technology, pp 32-54.

Sedgwick, H. (1986). Space perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), Handbook of perception and human performance (Vol. 1, pp. 1-57). New York: Wiley.

Tables

Table 1. Percent VA of regression slopes for ENV and No-ENV performance.

Population Performance					
Distance	*Viewing Conditions*				Subject
	ENV		No-ENV		
	% VA (*Avg. slope*)	r^2	% VA (*Avg. slope*)	r^2	
FAR	10% (0.016)	0.95	47% (0.07)	0.97	ALL
MID	14% (0.029)	0.49	45% (0.09)	0.99	
NEAR	2% (0.004)	0.01	39% (0.11)	0.99	
FAR	3% (0.005)	0.24	79% (0.1215)	0.99	EC-9
MID	5% (0.0098)	0.71	71% (0.1421)	0.99	
NEAR	2% (0.005)	0.21	59% (0.1674)	0.99	
FAR	75% (0.115)	0.99	156% (0.24)	0.99	EC-8
MID	69% (0.1381)	0.99	170% (0.33)	0.99	
NEAR	50% (0.144)	0.97	93% (0.26)	0.98	
FAR	20% (0.03)	0.96	13% (0.02)	0.51	EC-4
MID	15% (0.03)	0.96	13% (0.025)	0.53	
NEAR	1% (0.004)	0.43	10% (0.028)	0.79	

Figures

Figure 1 (A) Visual scene for sizing bottles in the ENV condition. The position and initial size of the bottle is randomly presented to the subject. **(B)** Visual scene for sizing bottles in the No-ENV task. Both the table and the floor are removed leaving only the grey background and the bottle for the subjects to perform this task. The dark lines in the figure represent the edges of the CAVE walls.

Figure 2 Experimental setup for the ENV condition with the physical coke bottle to the right of the subject and the virtual bottle and table are projected in front of the subject. The subject holds a wand with a joystick to adjust the size of the virtual bottle.

Figure 3 A bird's eye view of the tables and the position of the subject in the three locations for these experiments. **(A)** The position where the subject was farthest from the front CAVE wall (8.5ft) (FAR), **(B)** the middle viewing position where the subject was 5ft. from the front CAVE wall (MID), **(C)** the nearest viewing position to the CAVE wall (NEAR) (3.5ft) for subjects performing these experiments.

Figure 4 The three different tables used in these experiments. Each table had different texture for its top and the height of the table was different in each case. Table **(A)** was 36 inches from the floor; Table **(B)** was 30 inches from the floor and Table **(C)** was 33 inches from the floor.

Figure 5 The expected Size-Ratio settings if the subject were to set the object size based on either size-constancy (dotted line) or the visual angle (VA) of the bottle at the front CAVE wall. Notice that the slope is zero and coincident with a size-ratio of 1 for all bottle and viewing distances for the size-constancy condition. For the VA performance, the slope increases as the bottle distance from the subject increases. The size-ratio is 1 when the object distance is at the CAVE wall. The symbols for each distance from the front wall are shown in the legend and apply to all subsequent figures as well.

Figure 6 (A) Population performance in the ENV condition. Size-ratio is plotted as a function of the position of the bottle from the subject. The size-ratio values maintain a value that hovers about 1 for the different bottle positions from the subject indicating approximately size-constancy performance. **(B)** Population performance in the No-ENV condition. The rise in the size-ratio with increasing bottle distance indicates that subjects are performing more like visual angle than size-constancy.

Figure 7 The absolute value of the error in size judgments for the population. The data obtained in the ENV condition is different with subjects performing better with fewer errors than in the No-ENV conditions. An error of 0.1 corresponds to a 1.2-inch error in height of the 12-inch bottle.

Figure 8 The distribution of mean absolute value of the errors in size judgments for all runs (or trials) as a function of the bottle's position from the CAVE wall in the **(A)** ENV condition and **(B)** the No-ENV condition.

Figure 9 (A) Typical subject (EC9) response to sizing the coke bottle in the ENV condition. A size-ratio close to 1 (size-constancy performance) is seen in these data for all the bottle positions. **(B)** Subject (EC9) performance with the No-ENV condition. The values of the average size-ratio change dramatically from one bottle to the next unlike this subject's ENV case performance. These data more closely represent visual angle performance rather than size-constancy.

Figure 10 These data show exceptions to the general behavior of our population. **(A)** Size-Ratio from subject (EC8) in the ENV condition that shows VA performance despite the inclusion of texture mapped table and floor in the scene. **(B)** Subject EC-4 results in the No-ENV condition show size-ratio settings that are grouped about the value of 1 despite the removal of many visual cues. This performance is consistent with that predicted for size-constancy.

Figure 11 Subject EC9 results from the No-ENV condition showing the consistency of the settings with respect to the distance bottles are from the front CAVE wall. "Inside" refers to in front of the front CAVE wall and "Outside" to the behind or beyond the CAVE wall.

Figure 12 **(A)** Transformed data from Harvey and Liebowitz (1967) [dashed line] to our size-ratio format [solid line]. A close relationship is seen in their data and ours from the CAVE. Comparison of these data to Koh and Charman (1999) [dotted line] shows some agreement but at close positions there is a large deviation from our data. **(B)** Comparison of Harvey and Liebowitz [dashed line] to our No-ENV [solid line] data show good agreement in data.

Figure 13 Spatial quantization due to discrete pixels results in a depth quantization. The depth of an object (in this case a point) in the CAVE is confined within the diamond shaped space ABCD.

Figure 14 Displays how the error interval increases for increasing object distances from the viewer along the midline of the subject for the nearest viewing distance. This quantization error may explain the deteriorating performance in that condition.

Figures

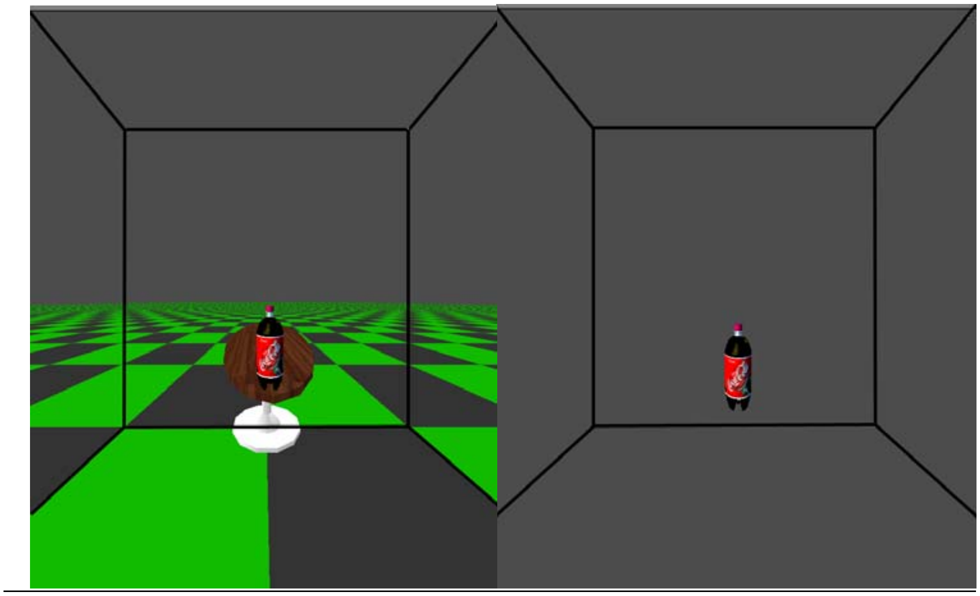

Figure 1 **(A)** **(B)**

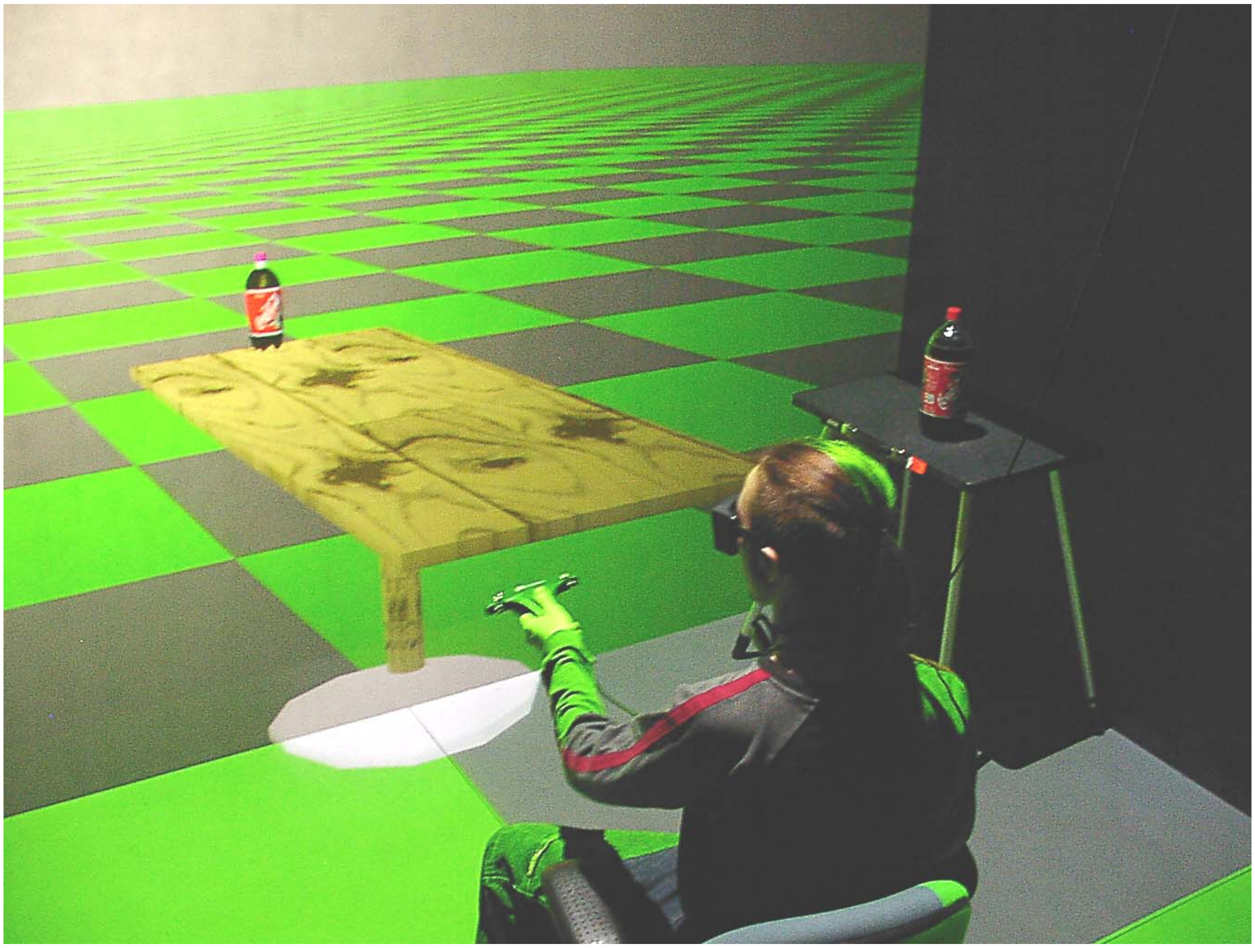

Figure 2

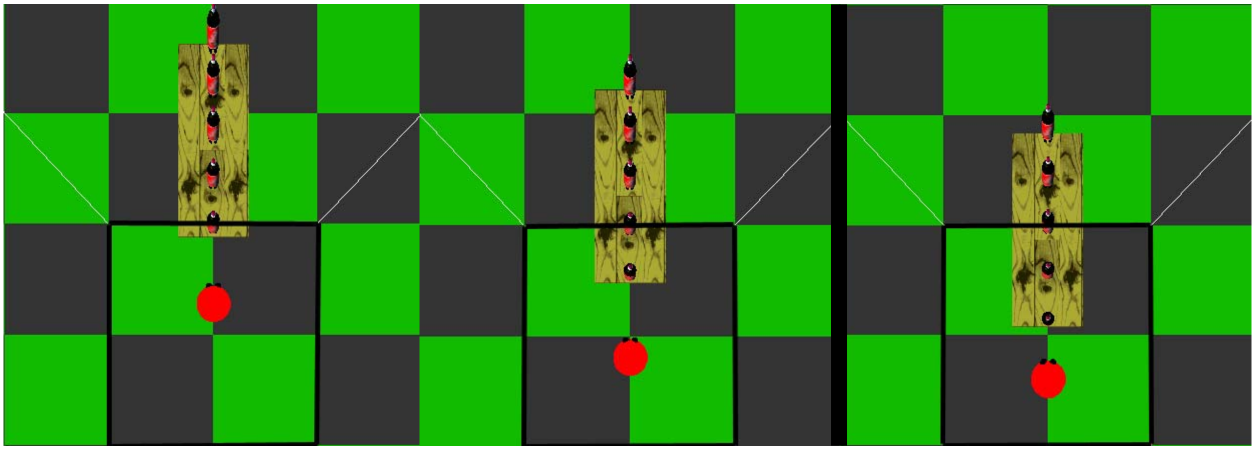

Figure 3 **(A)** **(B)** **(C)**

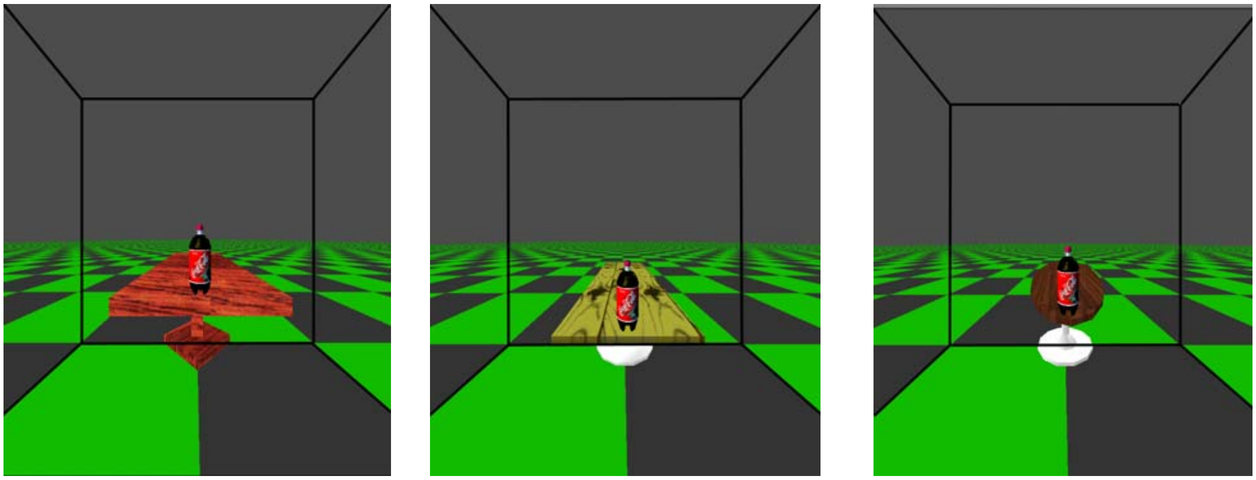

Figure 4 **(A)** **(B)** **(C)**

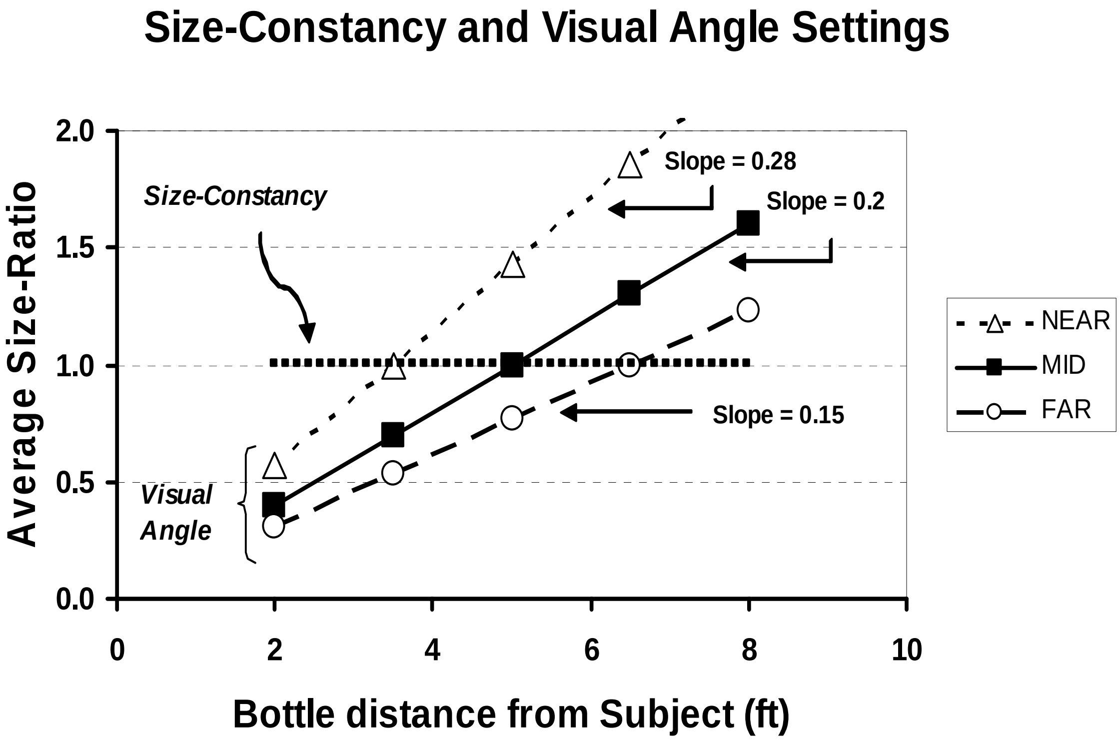

Figure 5

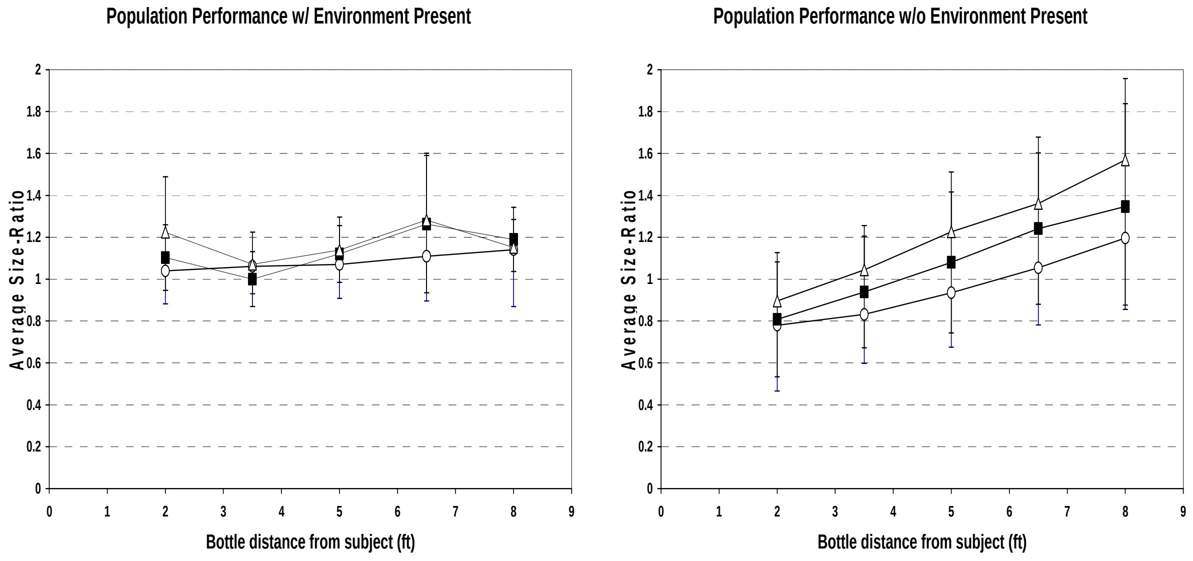

Figure 6 **(A)** **(B)**

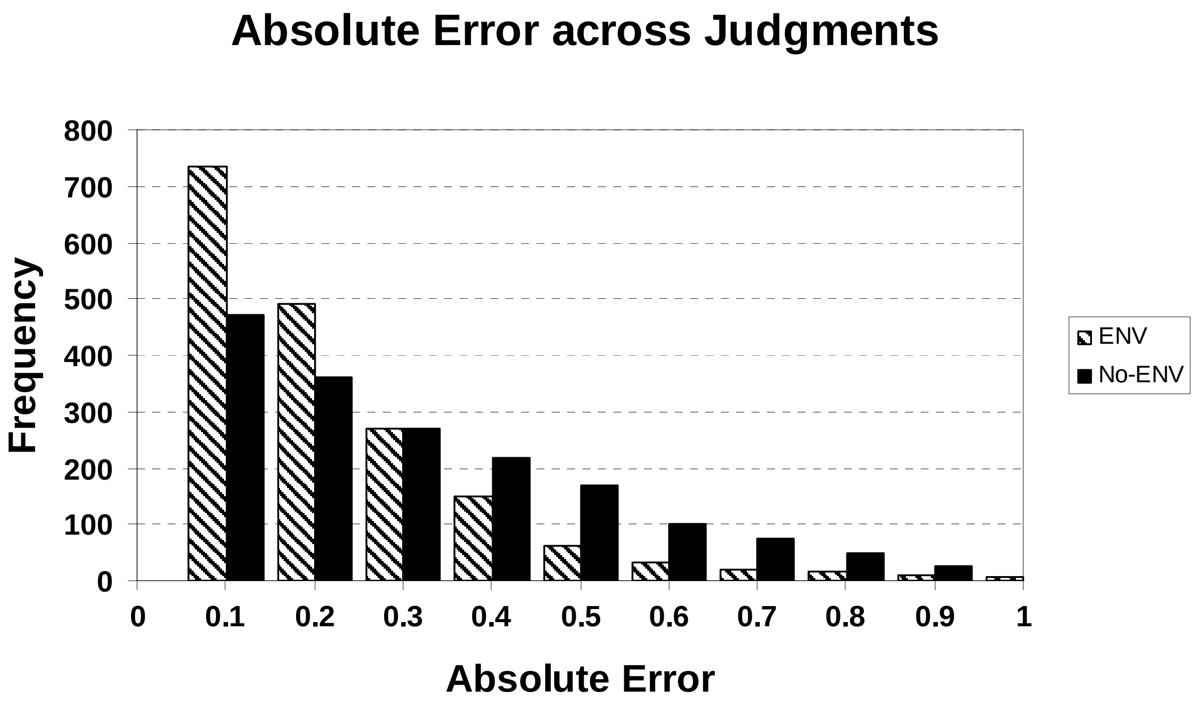

Figure 7

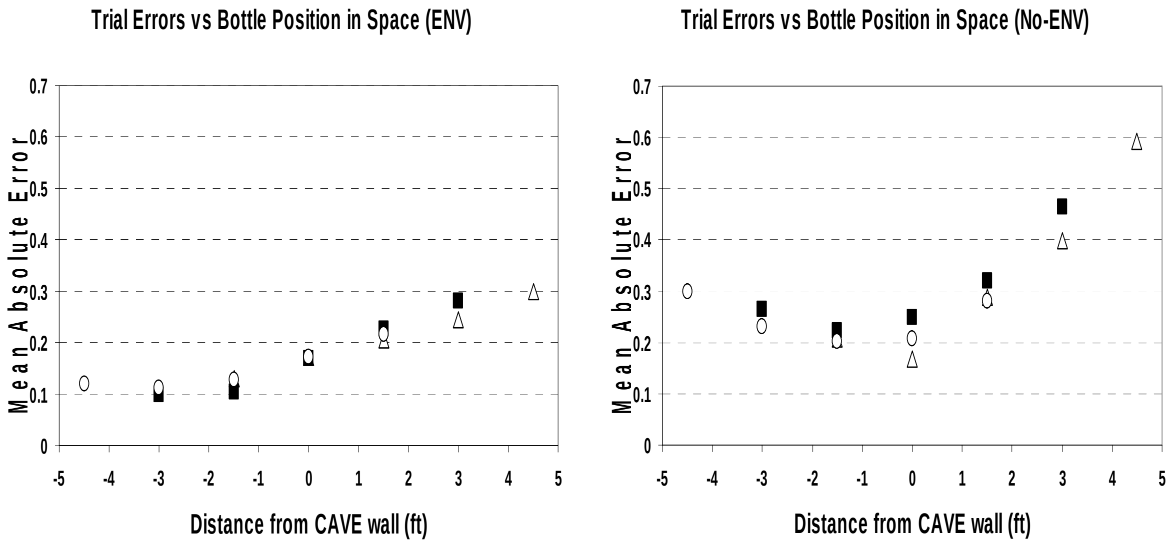

Figure 8 (A) (B)

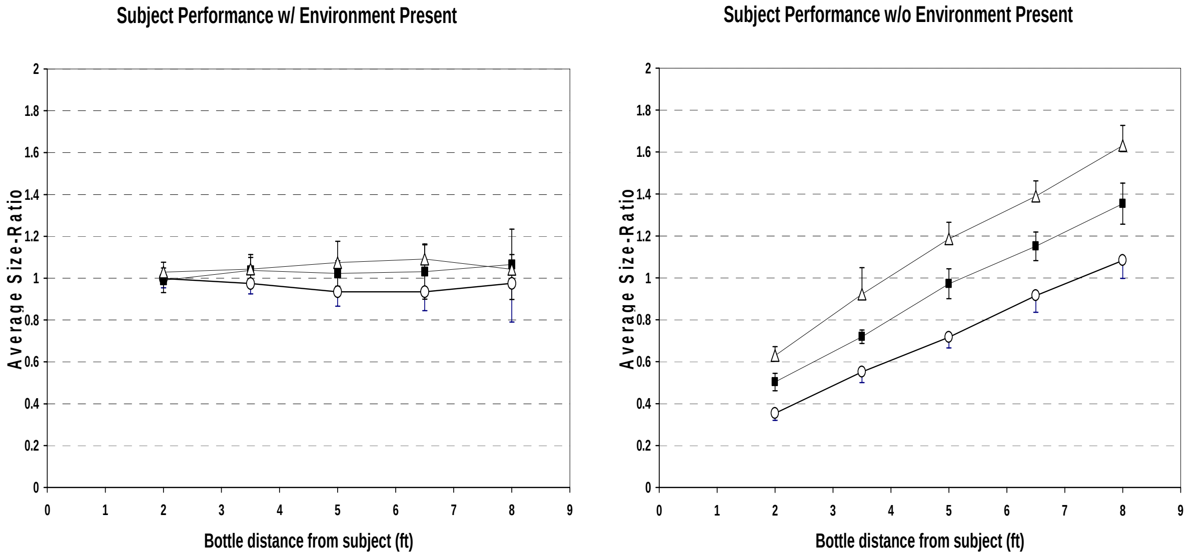

Figure 9 **(A)** **(B)**

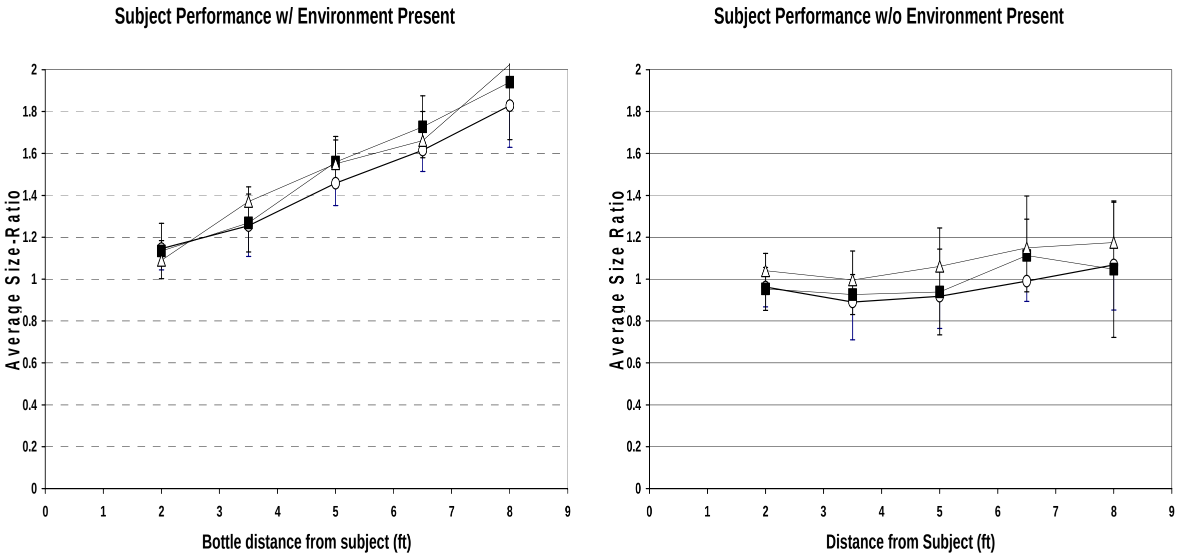

Figure 10 **(A)** **(B)**

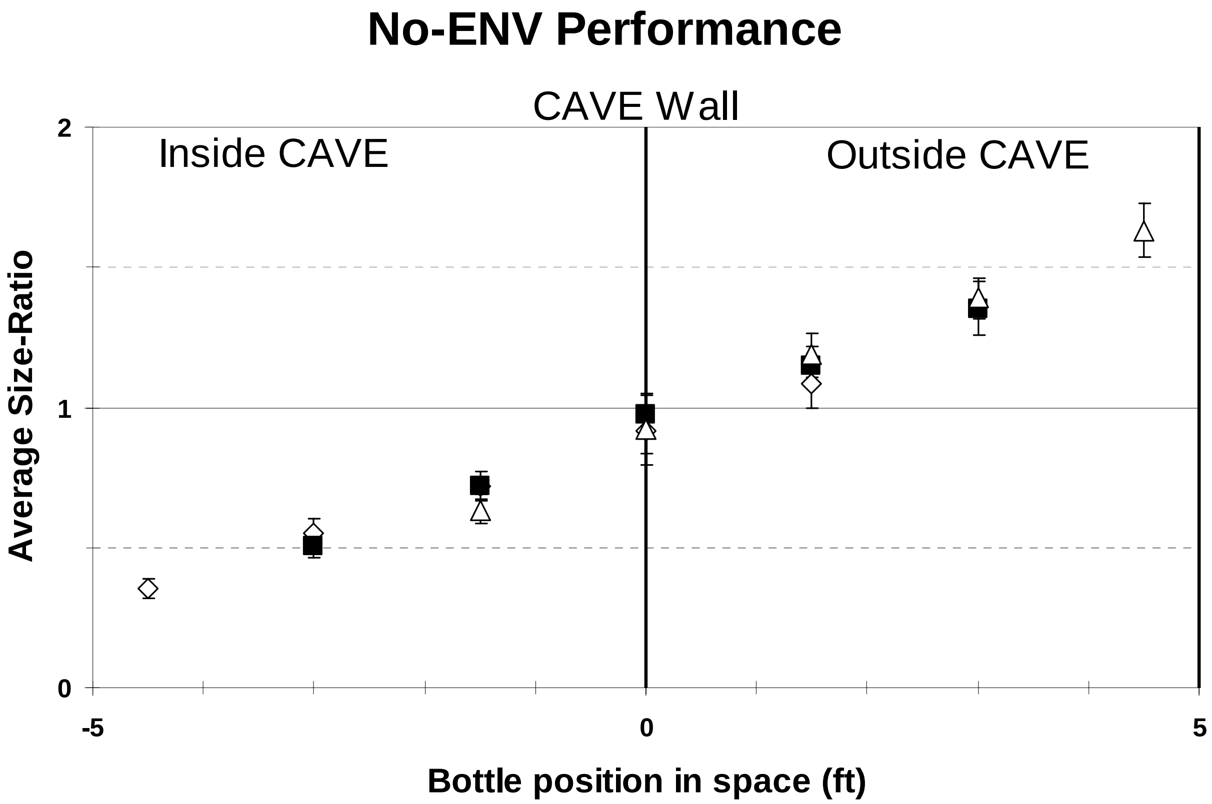

Figure 11

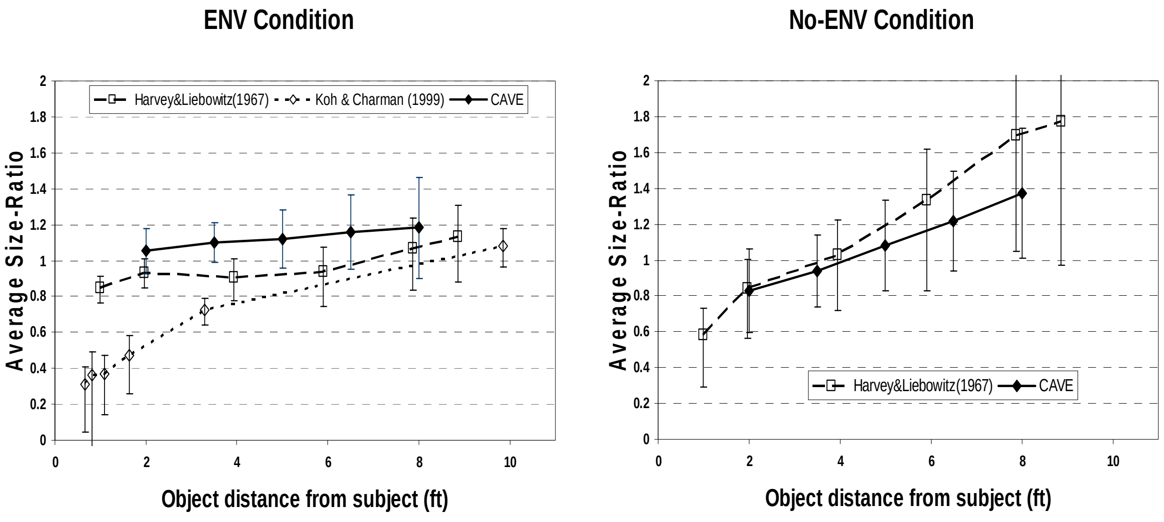

Figure 12 **(A)** **(B)**

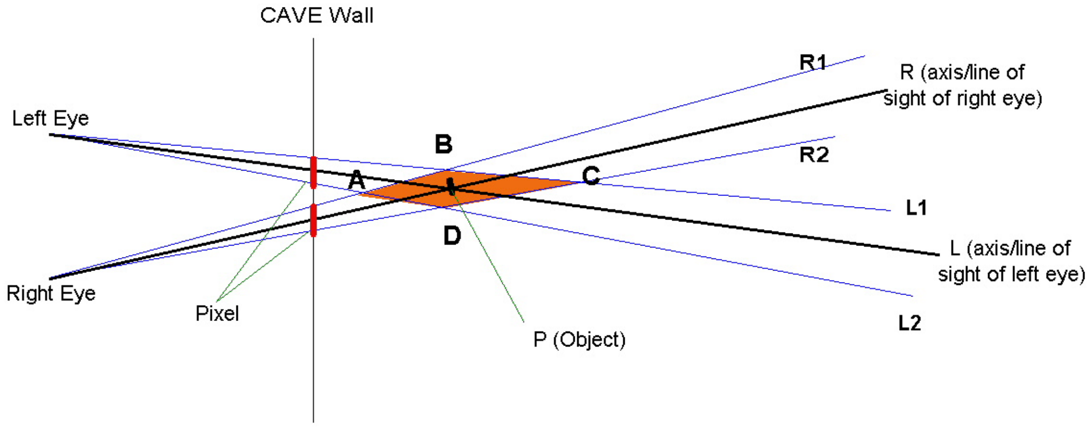

Figure 13

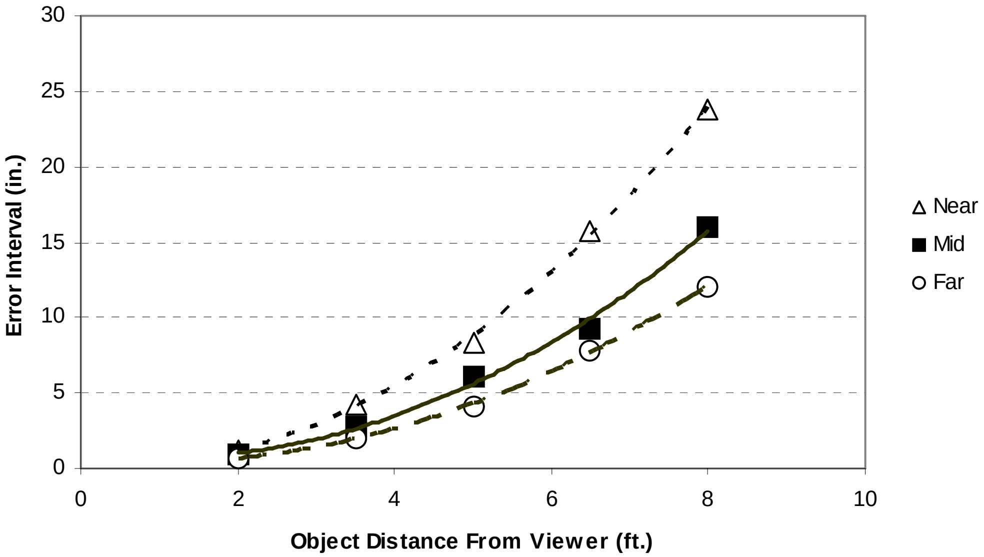

Figure 14

---

[1] The following formula was used to transform Harvey & Leibowitz (1967) and Koh & Charman (1999) data to our format: In their data objects of a fix size are presented at a various distances and subjects were asked to adjust a reference object placed at a constant distance. In our experiment we had a fixed size reference object at a constant distance and subjects were asked to adjust the size of objects placed at varying distances. We converted from their data to ours by multiplicative inversion. This is a hyperbolic conversion and causes asymmetry in their error bars.