Effects of Spatial Resolution and Antialiasing on Stereoacuity and Comfort

Charles J. Lloyd[1]
Visual Performance LLC, Ellisville, Missouri, 63021

The flight simulation training industry has considered the use of stereoscopic displays for decades, yet their use is still limited to a small fraction of training applications. Primary reasons for their slow adoption include high cost and complexity, discomfort, and the belief that stereopsis is not useful at distances greater than about 10 m. The cost and complexity of stereoscopic displays is rapidly declining due to heavy investments by other industries, minimizing this barrier to their use for simulation training. The barriers of discomfort and utility are addressed in this paper. An evaluation of standard and stereoscopic display systems designed for air refueling boom operator training was conducted with operators estimating the distance between receiver aircraft and the boom nozzle. Relative distance estimates were significantly more precise for the stereoscopic conditions, even at the 20 to 25 m working distance of the boom operators. This finding is at odds with the findings of researchers who concluded stereopsis is of little use at this distance, but who used low resolution displays in their evaluations. Our evaluation of spatial resolution and antialiasing showed that stereoacuity thresholds as low as 5.5 arcsec are attainable. For antialiasing filter widths less than about 1.2 pixels, both stereoacuity and comfort are heavily dependent on pixel pitch over the range of 0.5 to 3.0 arcmin. For filter widths greater than about 1.6 pixels, pixel pitch had a much smaller effect on stereoacuity and comfort. These results suggest stereoscopic displays have been undervalued by previous researchers due to the low resolution and insufficient antialiasing used in their evaluations. The results imply appropriately designed stereoscopic displays are more useful than previously thought for air refueling and rotary wing tasks involving working distances of 20+ m.

Nomenclature

Kf	=	coefficient that controls height of the floor region of the model
Kh	=	coefficient that controls height of the hill region of the model
Kb	=	coefficient that controls height of the blur region of the model
Kp	=	coefficient that controls position of the cliff region of the model in the pitch dimension
Kw	=	coefficient that controls position of the cliff region of the model in the width dimension
Ks	=	coefficient that controls slope of the cliff region of the model
Wd	=	critical width of the antialiasing function, based on disparity threshold
Wc	=	critical width of the antialiasing function, based on comfort ratings
Wdc	=	critical width of the antialiasing function, based on both threshold and comfort ratings

I. Introduction

A. The Promise of Stereoscopic Displays

In a 1990 paper titled "Stereopsis takes off in flight simulation," Tidwell described the state of the art in training display systems and concluded "Stereopsis is a practical reality today in flight simulation" and stereoscopic displays would provide more effective training[16]. In making his case, he cited studies done at NASA Langley[12] that led to the conclusion "The applications of stereo 3-D to pictorial flight displays within the program have repeatedly demonstrated increases in pilot situational awareness and task performance improvements." Tidwell also cites

[1] President, Visual Performance LLC, Charles.Lloyd@VisualPerformance.us, Senior Member AIAA

evaluations conducted at CAE[6] in which the authors conclude "Stereo imagery had a stronger effect on performance in formation flight than we expected. In tasks involving close approach to other objects (i.e., formation, air to air refueling, landing, and perhaps close air combat) it may be of significant value".

It would seem the case for the use of stereoscopic displays for simulation training had been made by 1990. But, some 22 years later, we find stereoscopic displays in only a small fraction of simulation trainers in current use. This paper examines an important design challenge that may be a primary reason stereoscopic displays have not gotten off the ground for simulation training applications. In this paper we describe an evaluation of the effects of antialiasing and spatial resolution and show that these variables are strong determinants of stereoscopic effectiveness and viewing comfort.

B. Conventional Wisdom

In the process of preparing display system requirements for the next generation KC-10 Boom Operator Trainer (BOT) upgrade, we examined the requirements and specifications for the recently awarded visual system upgrade for the KC-135 Boom Operator Weapons Systems Trainer (BOWST). Discussions with suppliers and experienced users revealed that stereoscopic display had not been considered for the program because it was believed the benefit of the stereopsis cue at the 20+ meter working distance of the boom operator would not justify the increased cost and complexity of the display system. Further discussion revealed that this assertion was based on display design experiences and rules of thumb commonly used in the broader virtual environments industry.

C. Expectations from the Vision Science Literature

Surprised by the assertion that the stereopsis cue would provide little benefit for the air refueling task, we conducted a literature review to establish a defensible estimate of the precision with which people can discriminate depth from binocular disparity at the working distance of the boom operator. The results of this review produced a number of papers[1, 2, 5, 9, 11 & 15] that indicate observers with normal stereoscopic vision can discriminate between 3 and 10 arcsec of disparity under photopic levels of illumination, for objects with a spatial extent of at least a half degree, and for objects that are separated by no more than about one degree. Using the 8 arcsec threshold described by Howard as representative of the normal observer, we expect boom operators to be able to discriminate distance differences of 0.24 m (9 inches) at 20 m using binocular disparity. Thus, we expected the stereopsis cue to be quite useful to boom operators in the refueling task.

To validate our expectations, we conducted an evaluation of the effects of stereopsis, collimation, and head tracking on the precision with which an experienced boom operator could estimate the distance between the nozzle on the refueling boom and the receptacle on approaching receiver aircraft[7]. As expected, the results of this evaluation showed a large and highly reliable improvement in distance estimation performance with the use of the stereopsis cue. Eight of the eight boom operators who participated in the testing indicated they preferred the stereoscopic display over the non stereoscopic. The standard deviation of the distance estimates for the stereoscopic (and collimated) display conditions was 0.61 ft which corresponds with a disparity threshold of 6 arcsec at 20 m. This precision is consistent with the disparity thresholds of 3 to 10 arcsec reported in the vision science literature.

D. Experience with Practical Stereoscopic Displays

While the research results from the vision science literature should make us confident boom operators can use the stereopsis cue to their advantage in the aircraft, the results of our evaluation with the ground based simulation displays are clearly at odds with the findings of many others. For example, two separate papers published in 1989-90 report a stereoscopic threshold of about 2.3 arcmin (140 arcsec) for evaluations using electronic displays[16 & 18]. This estimate of the disparity threshold has been accepted as reasonable by other researchers[13] since that time. Note that this threshold estimate is 17 times higher than the threshold estimate we derived from the vision science literature. It would appear these authors have assumed that the *practical* stereoscopic threshold attainable with electronic displays far exceeds the threshold that can be measured in the laboratory with the use of real objects. We suspect the assertion stereopsis provides little benefit at the boom operator working distance is based more on experiences with electronic displays than it is on visual capability in the real world.

The idea that practical electronic displays are incapable of producing stereopsis cues comparable with the real world has been implicit in the work of many researchers. For example, in a commonly cited paper summarizing the relative utility of potential depth cues for television displays, Nagata[10] states "binocular parallax is most effective at a distance of less than 1 m." and provides a figure indicating the relative sensitivity provided by the binocular parallax (disparity) cue to drop below the motion parallax, texture, and brightness cues at a distance of 10 m. In a similar paper Cutting & Viston[3] show the relative utility of the stereopsis cue to falling below the motion parallax, relative size, and visual field height cues at a distance of 10 m.

In a report describing the effects of stereoscopic display on task performance, Singer[14] cites a 1993 unpublished Army report by Rinalducci that indicates "the stereoscopic disparity cue is functional only for a rather limited range, at a maximum of ten to twenty meters from the observer." This position is consistent with that of Gregory[4] who concluded "For near objects the different views of the two eyes are used to compute depth, but the baseline between the eyes is too small for distances beyond 50 ft or so, when we are effectively one eyed." The results of Singer's own evaluation led him to conclude the stereoscopic cue provides benefit only at distances of less than about 10 ft and "should not be used unless required" because it can be so uncomfortable.

The overall conclusions we must draw from the literature describing practical display systems is that we cannot expect to achieve real world levels of stereoscopic disparity using practical electronic displays. The results of this review provide strong support for the position of the BOWST acquisition team and suppliers who did not consider the use of stereoscopic display for this training device. And these results indicate we should not expect the use of stereoscopic displays to provide a performance benefit for boom operators using the next generation BOT. However, the results of our previous evaluation[7] indicated the use of the stereopsis cue produced a large and highly statistically reliable improvement in ability of experienced boom operators to estimate the distance between the nozzle and receptacle. It would appear the results of our previous evaluation were unusual. Thus, we designed the current evaluation to replicate this finding and to address the reasons we were able to successfully demonstrate the utility of stereopsis cue at this distance where others had not.

E. Expectations and Objectives

Prior to conducting our previous evaluation we realized a display system that produced eye limited stereopsis cues would have to be capable of drawing image features such as points, lines, and polygon edges with great geometric precision. A typical simulation training display system of today might have a pixel pitch of 130 arcsec (2.2 arcmin). To produce disparity cues as fine as 3 to 10 arcsec, the display system would have to reliably position image features with an error no greater than a few percent of the pixel spacing. We therefore surmised that to be successful, we would need to use a fine pixel pitch and a very good antialiasing function. For that evaluation we used a pixel pitch of 1.5 arcmin and the "adaptive super sampling" method available on the Vital X image generator by FlightSafety. In the previous paper we held the pixel pitch and level of antialiasing constant as these design variables were not the subject of that evaluation. In the present paper we exercise these variables over a significant range so that we can quantify their effects on stereoscopic display performance.

The evaluation described in this paper was designed to accomplish four objectives:

- Replicate the finding that stereoscopic disparity thresholds comparable with those reported in vision science literature (i.e., 3 to 10 arcsec) can be attained using practical electronic display systems.
- Validate the expectation the disparity threshold increases with increasing pixel pitch.
- Validate the expectation the disparity threshold increases with decreasing antialiasing filter width.
- Collect sufficient data and construct a mathematical model that describes the trade space for the pixel pitch and antialiasing filter width design variables.

II. Method

F. Equipment

This evaluation was conducted using the stereoscopic display shown in Fig. 1. With this system, observers looked through a binocular optical system containing two mirrors and a lens for each eye (see Fig. 2). Accommodation distance of the display system was set using +0.50 diopter ophthalmic lenses (sphere) with anti-reflection coatings. The display distance was set to 1.82 m (-0.55 diopter), thus, the net result of these two settings produced a system focus distance of 20.0 m (-0.05 diopter).

Two Dell (Model 2007FPb) LCD monitors with a native resolution of 1600 x 1200 pixels were used to present the images. Prior to the evaluation the luminance response of the monitors was measured using a Minolta LS-100 luminance meter. These data were used to create an electro-optical response correction that linearized the luminance response of the displays.

Figure 1. Photograph of the stereoscopic display system used in the evaluation. Observers viewed a pair of LCD monitors through an optical system that set the accommodation and vergence distances to 20 m.

At the selected viewing distance the native angular pixel pitch of the monitors was 0.48 arcmin. The binocular FOV of the display system was 12.8 x 9.6 deg. Apertures were placed within the system such that the observer could see only the active display surfaces surrounded by black. Since two completely independent displays and optical systems were used, the leakage (crosstalk) between the right and left eye channels was zero.

The use of a long display-observer distance and low-power lenses produced an optical system with very large exit pupils and eye relief. With this system there was no evidence of geometric distortion, chromatic aberration, or luminance roll off as a function of eye movements across the exit apertures. Both the accommodation and vergence distances were independent of the inter-pupillary distance (IPD) of the observer. The optical system was suitable for observers with IPDs ranging from 60 and 75 mm.

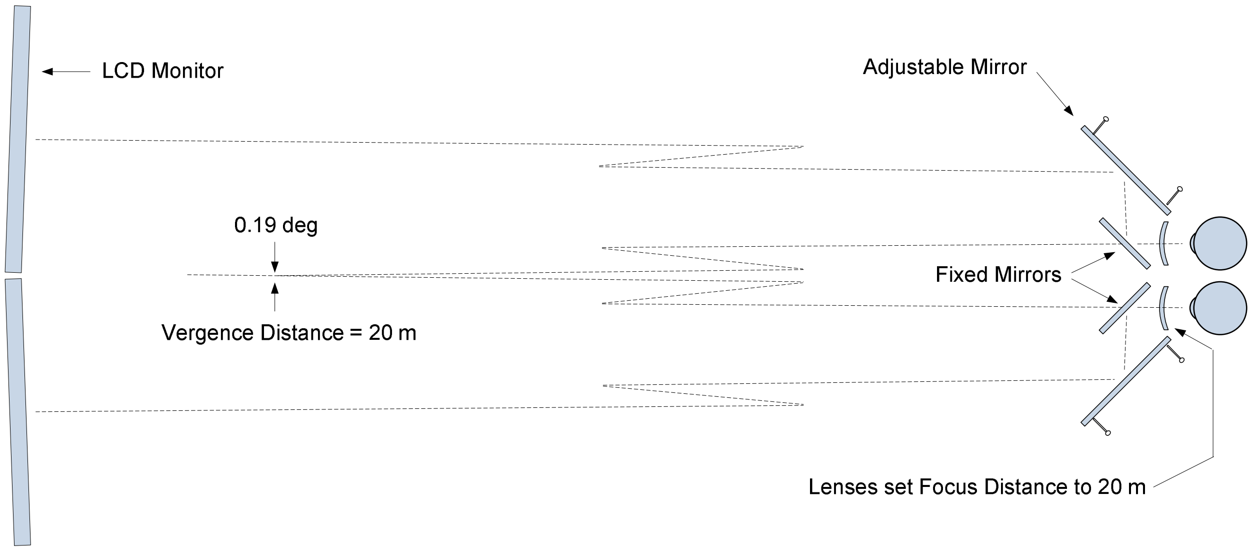

Figure 2. Diagram of the optical layout of the display system. The outboard pair of mirrors are mounted on 3 point mounts used to set vergence distance to 20 m. The +0.5 diopter lenses extend 1.82 m display distance to a focus distance of 20 m.

The binocular vergence angle of the system was calibrated using fine adjustment screws on one of the mirrors in the optical path of each eye. Prior to calibrating the display system, the laser vergence gauge shown in Fig. 3 was set to indicate 20 m. by shining it on a wall at that distance and carefully adjusting the angle between the red and green lasers to be coincident. This gauge was then mounted on a tripod and directed through the optical system at the LCD monitors while an alignment test pattern was displayed. The binocular display system mirrors were adjusted so the laser spots lined up with the alignment marks on the monitors. The accuracy of this alignment process is estimated to be approximately 0.5 arcmin.

The relative roll of the right and left eye images was adjusted to zero using a flat, partially reflecting mirror that was temporarily placed in front of the display optical system. A bright white horizontal line positioned above the display system could be seen in reflection from the mirror. This line was superimposed on the test patterns seen through the binocular optical system and was used to null the relative roll of the right and left images to within a tolerance of approximately 0.2 deg.

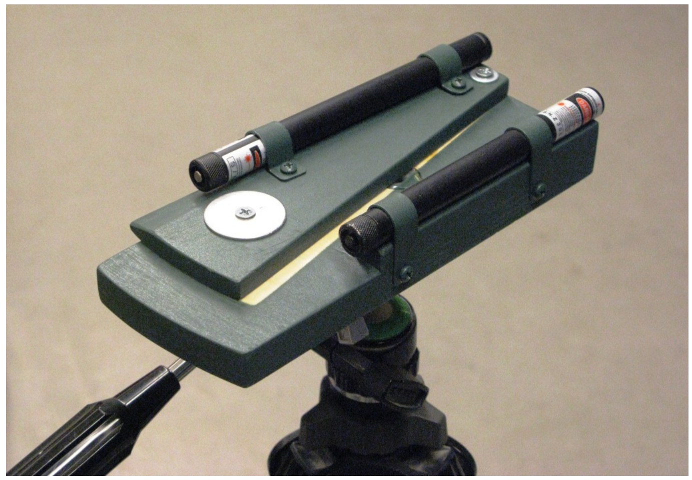

Figure 3. Photograph of the laser vergence gauge used to calibrate and check the binocular vergence distance of the display system.

G. Observers

Four paid observers, one female, participated in this evaluation. The observers were screened using a variant of the Titmus Circle test that is one of the accepted tests for screening Air Force pilot and boom operator candidates. On each trial of the test, observers were presented the pattern shown in Fig. 4 and indicated which of the four circles (top, bottom, left, or right) differed in depth as well as the direction of the difference (near or far). To participate in this evaluation observers had to respond correctly to at least 16 out of 20 trials with the disparity set to 25 arcsec. Twelve observers were screened for this evaluation. The four who passed the test produced 20 correct responses. The eight observers who did not pass the test were re-tested with the disparity set to 50 arcsec. None of these eight observers passed the retest.

H. Stimuli

On each trial the observer's task was to view a small rectangle positioned in a dark gap between two edges (see Fig. 5) and to indicate if the rectangle was in front of or behind the edges. The task is much like the classic "Howard-Dolman" or three-bar tests of stereopsis that have been used for almost 100 years[5]. In the version of the test used here, edges rather than bars were used for the comparison as this eliminated the possibility of using relative size (width) as a cue to distance. To further mask any monocular cues that might be used, the position of the rectangle was randomly moved from side to side on each trial. The height of the rectangle remained constant across all trials.

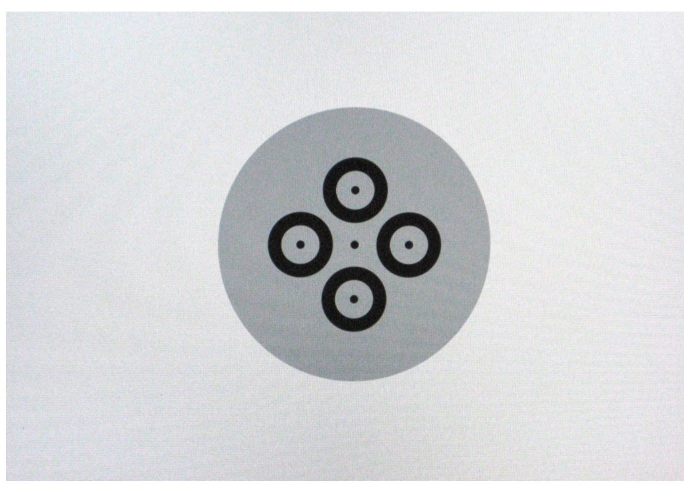

Figure 4. Image of the modified Titmus circle test used for screening observers. On each trial of the test, observers indicated which of the four inner circles (top, bottom, left, or right) was in front or behind the rest of the pattern.

The angular size of the rectangle was set to the approximate size of the receptacle on a receiver aircraft when viewed at the contact distance. The mean width of the rectangle was 0.25 deg while the height was 0.50 deg. The mean width of the gap between the edges was 1.25 deg, thus, the mean distance of the rectangle from an edge was 0.5 deg. The slope of the edges and rectangle was randomly varied from trial to trial over a range of +/- 1.5 deg relative to vertical. On trials for which antialiasing was insufficient and the pixel pitch was coarse, the observer could see definite spatial sampling artifacts (i.e., "stair stepping") on the edges and rectangle as in Fig. 6. On each trial, the positions of the images were randomly jittered relative to the pixel mosaic so that a different pattern of sampling artifacts would be produced on each trial. For those conditions where sampling artifacts were present, they were spatially uncorrelated between the right and left eye images.

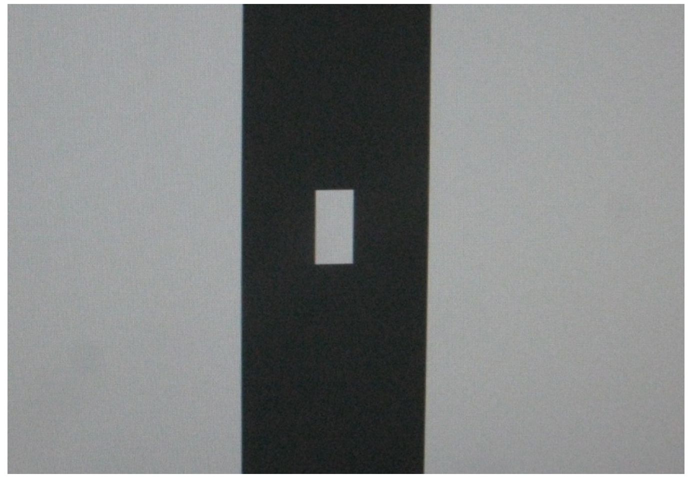

Figure 5. Photograph of the rectangle and edge pattern used in the evaluation. The rectangle subtended an angle of 0.25 x 0.5 deg for all experimental conditions. For this photograph the pixel pitch was set to 0.5 arcmin and the AA filter width was set to 1.5 pixels.

The luminance of the white edges and rectangle was 18 fL (60 cd/m2) and the dark gap was 2.2 fL for an edge contrast ratio of 8:1 (modulation = 0.78). On each trial the stimulus was presented for 1.4 sec, long enough to allow 2 or 3 fixations on the pattern. The results of pretesting indicated thresholds began rising as the stimulus duration was reduced below about 1 sec. The antialiasing filter used to generate the images had a Gaussian shape and was applied to both the rectangle and the two vertical edges. The width parameter used in this paper refers to the half-maximum width of the Gaussian that was applied to the edges prior to spatial sampling.

I. Disparity Thresholds

Binocular disparity thresholds were measured using a two-alternative, forced choice (2AFC) psychophysical staircase procedure that produced an average rate of correct responding of 0.75. Fixed-length sequences of 50 trials were used with the first five trials excluded from the threshold calculation. For the practice and first third of the trials, the starting level of disparity was set to 100 arcsec. For the remaining trials the starting level was set to 4 times the expected threshold based on a model fit to the initial data collected from that observer. Trial sequences were paced by the observer which allowed them to take a break at any time. The measurement of one threshold required approximately 2.5 minutes to complete and approximately 20 thresholds could be measured in one hour. Sessions were limited to one hour per day to limit fatigue. Each observer participated in 8-9 sessions that were spread over a period of 3 weeks.

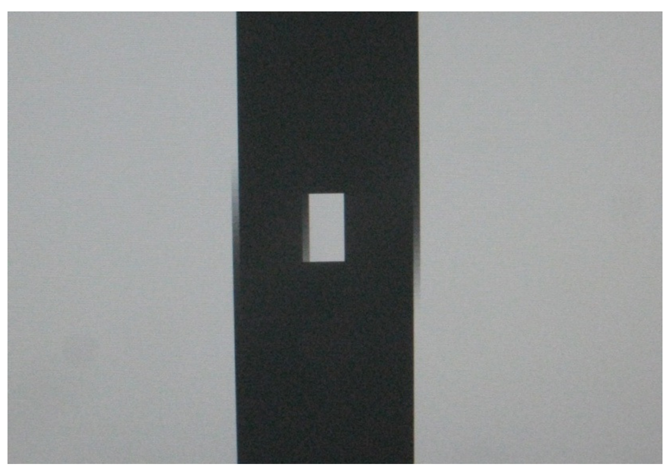

Figure 6. Photograph of the rectangle and edge pattern showing spatial sampling artifacts. For this photograph the pixel pitch was set to 3 arcmin and the AA filter width was set to 0.7 pixels.

J. Ratings of Comfort

Immediately after completing each threshold measurement (sequence of 50 trials), observers provided a rating of the visual comfort of that condition using a rating scale ranging from 1 to 5. Two verbal anchors were used for the scale: a 1 on the scale indicated the condition was "comfortable" and a 5 indicated the condition was "very uncomfortable." Observers were instructed to assume they would have to use a display system of that design for several hours per day.

K. Experimental Design

Threshold measurements were made for 42 combinations of pixel pitch and antialiasing filter kernel width. Six levels of pixel pitch were used and these were created using "pixel replication" such that the resulting pitches were integer multiples (1 to 6) of the 0.48 arcmin native resolution of the display system. Between six and eight levels of antialiasing kernel width were used at each level of pixel pitch. The specific levels of width selected were based on the results of pre-testing and were positioned to most efficiently reveal the shape of the surface.

For each observer the threshold for each experimental condition was measured at least 3 times. For those conditions that produced the highest variance, one or two additional threshold measurements were made to stabilize the estimates for those conditions. All of the measured thresholds were included in the analysis. On average, 147 threshold measurements were made for each observer for an average of 3.5 estimates per condition.

III. Results

L. Disparity Thresholds

Prior to model fitting, the data of each observer were reduced by calculating the median threshold for each of the 42 experimental conditions. The median data were then averaged across the observers, thus, the threshold estimates at each condition were based on an average of 14 measurements.

In a series of preliminary analyses, a stepwise multiple linear regression approach was taken using a wide range of 2-D polynomial models with second and third order terms in the main effects and interactions, and nonlinear transformations of both the dependent and independent variables. While some of these models produced a higher correlation with the measured data than did our final model, they were rejected as their behavior was considered to be implausible when extrapolated just beyond the range of the data. Our final model was fitted by minimizing the sum of the squared residual errors using the multi-dimensional nonlinear optimizer function (fminsearch) available in the MATLAB software.

Examination of the threshold data revealed distinct regions within the overall surface created by the data. For those conditions where a wide antialiasing filter was used, there was a distinct "floor" in the threshold data that gradually increased up a "hill" as the pixel pitch increased (see left side of Fig. 7). As the width of the antialiasing filter was reduced, the threshold first remained level and then climbed a steep "cliff" as it was reduce further. Given the distinctness of these regions, a multi-component model was fit to the data and the contributions of each independent component were combined to form the complete model. The nonlinear parameter optimizer was allowed to simultaneously vary all six parameters of the model. The first three model parameters were *Kf, Kh*, and *Kb* which controlled the height of the floor, magnitude of the hill, and magnitude of the blur (angular width of the antialiasing filter, in arcmin). Three separate parameters are used to set the position and slope of the cliff region in the model: *Kp, Kw,* and *Ks* which controlled the position in the pitch dimension, position in the width dimension, and slope of the cliff.

The magnitude of the floor of the model is simply the first parameter which is scaled in units of arcsec. The second and third parameters are used to scale the effects of pixel pitch and the product of pitch and width:

$$H = Kh * \text{pitch} \tag{1}$$

$$B = Kb * \text{pitch} * \text{width} \tag{2}$$

The last three parameters are used to calculate the "cliff" region of the model:

$$C = Ks * [Kw * \log10(Kp + \text{pitch}) - \text{width}] \quad (3)$$

Anywhere the quantity C goes negative it is set to zero:

$$C(C < 0) = 0 \quad (4)$$

The four model components are then combined in quadrature to calculate the disparity threshold surface:

$$\text{Disparity} = \text{sqrt}(Kf\text{^}2 + H\text{^}2 + B\text{^}2 + C\text{^}2) \quad (5)$$

The fitting process produced the following levels for the six free parameters in the model:

Kf = 5.544
Kh = 2.038
Kb = 1.112
Kp = 0.763
Kw = 3.830
Ks = 22.15

Using these levels, the correlation between the model and the mean threshold data was R^2 = 0.880 ($p < 0.0001$, 35 df). The standard deviation of the difference between the measured data and the model (RMSE) was 18% of the threshold level. Fig. 7 shows a surface plot of the resulting model while Fig. 8 shows the same model in the form of a contour plot.

The blue dots shown in Fig. 8 indicate the critical width of the antialiasing filter as a function of pixel pitch. The critical width indicates where the threshold was one RMSE (18%) above the minimum threshold for each pitch. As antialiasing filter width is decreased below this critical level the threshold increases rapidly. The formula for this critical level is given in Eq. 6.

$$Wd = 0.801 + 2.164 * \log10(\text{pitch}) \quad (6)$$

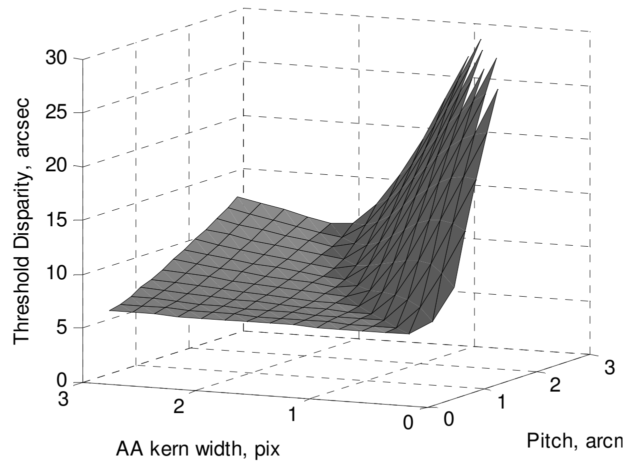

Figure 7. Model of threshold disparity as a function of pixel pitch and antialiasing filter (half-max) width. The correlation between the model and the mean data of four observers was R^2 = 0.880 ($p < 0.0001$, 35 df).

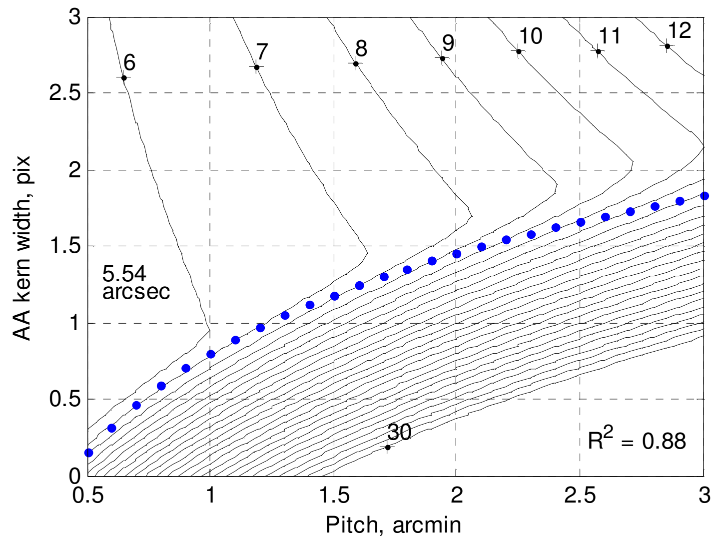

Figure 8. Contour plot of the disparity threshold model presented in Fig. 7. The blue dots in the figure indicate the critical width of the antialiasing filter as a function of pixel pitch (Eq. 1).

M. Viewing Comfort Ratings

Prior to fitting the model of viewing comfort, the rating scale data were reduced by averaging across replicates and observers. Thus, the means of between 14 and 15 estimates of comfort were used for each of the 42 experimental conditions. These mean comfort data were then used to fit a model of the same form used to describe the threshold disparity data. The results of the multidimensional fitting process produced the following levels for the six free parameters in the model:

Kf = 1.352
Kh = 0.846
Kb = 0.341
Kp = 1.474
Kw = 3.315
Ks = 2.786

Using these levels, the correlation between the model and the mean comfort rating data was R^2 = 0.964 ($p < 0.0001$, 35 df). The standard deviation of the difference between the measured data and the model (RMSE) was 0.149 rating scale points. Fig. 9 shows a surface plot of the resulting model while Fig. 10 shows the same model in the form of a contour plot.

The blue dots shown in Fig. 10 indicate the critical width of the antialiasing filter as a function of pixel pitch. The critical width indicates where the mean comfort rating was one RMSE (0.149 points) above the minimum rating for each pitch. Discomfort increases rapidly as the antialiasing filter width is decreased below this critical level. The formula for the critical level is given in Eq. 7.

Wc = 1.043 + 1.106 * log10(pitch) (7)

N. Observer Comments

During the evaluation observer comments were noted that revealed two distinct reasons for the discomfort produced by the sub optimal combinations of pixel pitch and antialiasing filter width. In the upper right corner of the surface (large pitch and large antialiasing filter width), consistient complaints of excessive blurring of the rectangle were noted (see Fig. 11). Along the bottom of the surface (small antialiasing filter width) observers complained consistiently of significant discomfort, however, they had a hard time describing the nature of the discomfort. Careful examination of these conditions revealed obvious spatial sampling artifacts were present and these artifacts did not match across the right and left images. It appears that the discomfort caused by these conditions was due to the binocular rivalry induced by the salient and mis-matching spatial sampling artifacts produced when insufficient antialiasing was employed.

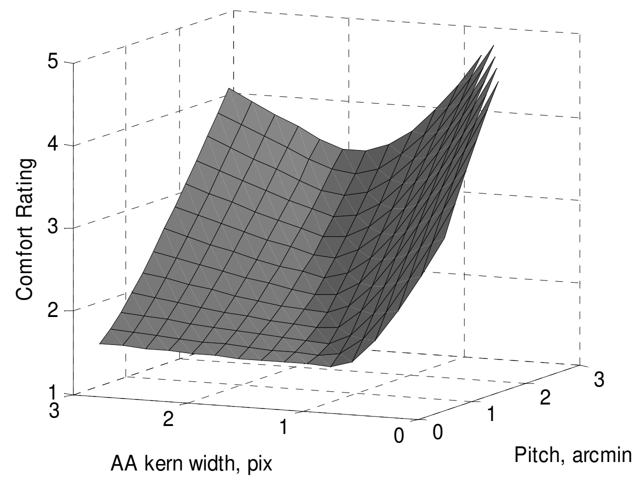

Figure 9. Model of viewing comfort as a function of pixel pitch and antialiasing filter (half-max) width. The correlation between the model and the mean data of four observers was R^2 = 0.964 ($p < 0.0001$, 35 df).

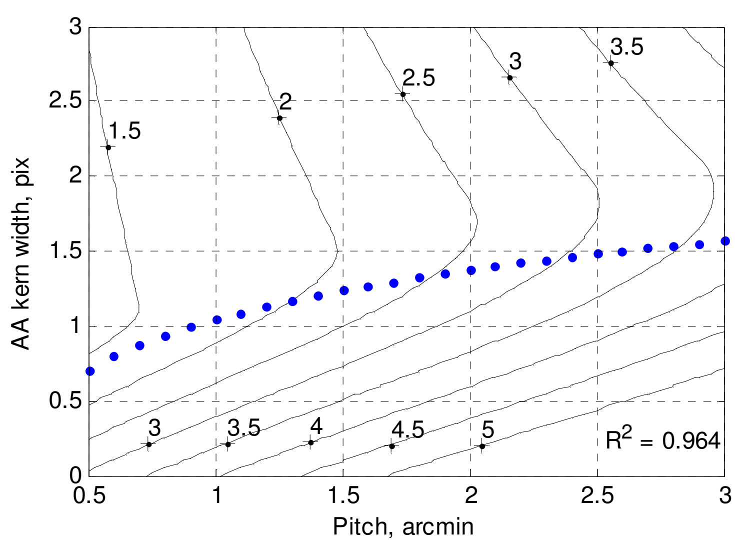

Figure 10. Contour plot of the viewing comfort model presented in Fig. 9. The blue dotted line indicates an increase of 1 RMSE (0.149 rating scale points). This line indicates the critical width of the antialiasing filter as a function of pixel pitch (Eq. 2).

IV. Discussion & Conclusions

The mean binocular disparity threshold for the four observers who participated in this evaluation was 5.54 arcsec in the floor region of the model. These data clearly confirm our expectation that thresholds in the range of 3 to 10 arcsec can be obtained using electronic displays when sufficient resolution and antialiasing are employed. Recall that the papers cited in the introduction[17 & 18] reported a disparity threshold of 140 arcsec (2.3 arcmin) for the display systems they used. This level is 25 times higher than the thresholds we obtained.

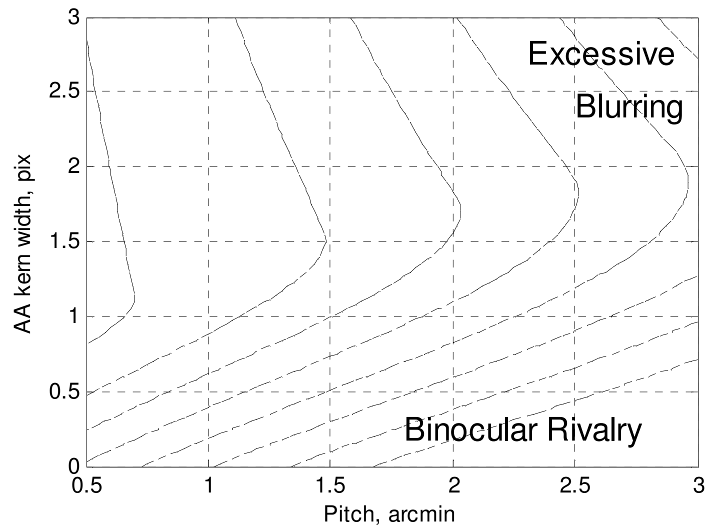

Figure 11. Contour plot of viewing comfort ratings indicating two distinct zones with apparently different causes of discomfort.

Perhaps the most important finding from this evaluation is that it illustrates the large difference these two design variables have on performance. Examination of Fig. 7 reveals the slope of the hill due to increasing display pitch is much shallower than the slope of the performance cliff caused by insufficient antialiasing. When we consider that the cost of reducing the pixel pitch is much greater than the cost of increasing the amount of antialiasing applied by an image generator, we can see the performance/cost ratio of improved antialiasing is far greater than that of improved resolution. It is surprising that the influence of antialiasing on stereoscopic display effectiveness has received so little attention by scientists and applications engineers over the past two decades.

Comparison of the comfort model in Fig. 9 with the disparity threshold model in Fig. 7 reveals these design variables have more equal effects on viewing comfort. With the comfort model the slope of the hill region is higher and the slope of the cliff region is lower than they are in the disparity threshold model, leading to a less well defined crease between the regions.

Considering both the disparity threshold and comfort models, we can see that real world levels of stereoscopic performance and reasonably comfortable display systems can be achieved using pixel pitches as coarse as 1.5 to 2 arcmin, but only if sufficient antialiasing is applied. Assuming a particular pixel pitch, the minimum filter width required to avoid a significant increase in the disparity threshold can be computed from Eq. 6. Similarly, the minimum filter width required to avoid a significant increase in ratings of comfort can be computed from Eq. 7. For many applications the design goal is to maximize performance as well as comfort, thus, a combined model of the critical antialiasing filter width was created by fitting a curve to the maximum of either the disparity threshold or the comfort rating curves described above. This combined curve is provided in Eq. 8 and is plotted as the dashed line in Fig. 12.

$$Wdc = 0.025 + \text{pitch}^{0.54} \quad (8)$$

O. Comparison with light point quality data

In 2007 a series of three human factors evaluations[8] was conducted at FlightSafety designed to quantify the effects of pixel pitch and antialiasing on lightpoint quality for fixed matrix displays. In those evaluations many combinations of the levels of pixel pitch and antialiasing filter (lightpoint spread function) width were used and response surface models were fitted to describe the results. The data presented in Figs. 4, 5, and 6 of that report provide us with estimates of the threshold for the detection of three common artifacts associated with spatially sampled display systems. The blue triangles plotted in Fig. 12 indicate the threshold for variations in the size and shape of lightpoints as they are slowly moved across the pixel grid. The blue circles represent the threshold for variations in the brightness (twinkling) of slowly moving lightpoints. The squares represent the threshold for variations in the rigidity (straightness) of lines of lightpoints such as a runway centerline or edge lights. Comparison of the curves in Fig. 12 shows that the critical antialiasing filter width determined in the present evaluation (dashed line) is quite consistent with the results of the lightpoint quality evaluations.

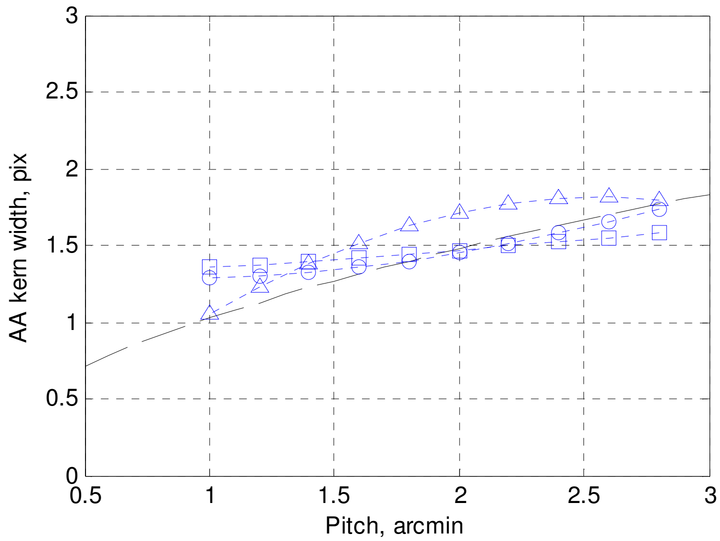

Figure 12. Critical width of the antialiasing filter as a function of pixel pitch (dashed line) for the simultaneous optimization of disparity threshold and comfort ratings (Equation 8). The blue symbols/curves indicate the threshold for common sampling artifacts associated with lightpoints rendered on fixed matrix displays[8].

Appendix A: Instructions to Observers

In this evaluation we will be evaluating the effects of several display system design variables on stereoscopic depth perception and viewing comfort. During the evaluation you will view the display through this optical system that sets the effective viewing distance at 20 m which matches the working distance of air refueling boom operators.

- Have observer take a seat and adjust the seat height

During the trials please keep your head an inch or two from the head rest. On each trial you will see two vertical white “edges” and a small rectangle between the edges. Your task is to indicate if the rectangle is nearer or farther than the two edges. After each observation you push the lever away from you to indicate the rectangle was farther than the two edges. You would pull the lever toward you to indicate the rectangle was nearer than the two edges. To begin the next trial press the “Enter” key with your thumb.

- Have the observer practice for a few trials

From trial to trial you will notice the tilt of the vertical edges changes, and the box moves around within the two edges. Please ignore these changes and pay attention only to the distance of the rectangle relative to the two edges. On average the rectangle should be “near” or “far” about half the time.

- Have the observer complete the series of trials

After completing each series of trials please provide me with a rating of visual comfort of the experimental condition. For this rating use a scale of 1 to 5 where a 1 indicates the condition was Comfortable and a 5 indicates the condition was Very Uncomfortable. When making these ratings, please assume you would have to use a display system of this design for many hours per day.

You may take a break at any time during the evaluation. When you wish to rest simply do not press the Enter key to initiate the next trial. You may find it helpful to occasionally look away from the display system at some distant object for a few seconds as this rests your eyes.

A series of 50 trials is used for each experimental condition. The software that controls the trials makes adjustments to the amount of depth so that you get about 75% of the responses correct. In other words, the system will adjust the difficulty of the task so your error rate is about 25%. Please make your best guess on each trial, you are expected to make errors about 25% of the time.

Participation in this evaluation is voluntary. You may stop your participation at any time for any reason.

You will be paid $25/hr for your participation in this evaluation. This rate applies for the preliminary instruction and practice, time spent in data collection, and for short breaks between trials. We will limit each experimental session to about 1 hour as this task can be tiring. You can participate in more than one session per day as long as you take a rest of at least 1 hour between sessions. You will not be paid for the breaks between sessions.

Participation in this evaluation is voluntary. You may stop your participation at any time for any reason.

References

1. Allison, R. S., Gillam, B. J., & Vecellio, E. (2009). Binocular Depth Discrimination and Estimation Beyond Interaction Space. *Journal of Vision, 9*(1).
2. Amigo, G. (1963). Variation of Stereoscopic Acuity with Observation Distance. *JOSA, 53*(5), 630-635.
3. 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.) *Handbook of Perception Aand Cognition, Vol 5; Perception of Space and Motion.* (pp. 69-117). San Diego, CA: Academic Press.
4. Gregory, R. L. (1965). Seeing in Depth. *Nature 207*(4992), 16-19.
5. Howard, H. J. (1919). A Test for the Judgment of Distance. *Transactions of the American Ophthalmological Society, 17*, 195-235.
6. Kruk, R., & Runnings, D. (1989). Low Level Flight Performance, and Air Combat Maneuvering Performance in a Simulator with a Fiber Optic Helmet Mounted Display System. Presented at the AIAA Flight Simulation Technologies Conference and Exhibit. Boston, MA.
7. Lloyd, C. J., Nigus, S. G. (2012) Effects of Stereopsis, Collimation, and Head Tracking on Air Refueling Boom Operator Performance. Presented at the IMAGE Conference. Scottsdale, AZ.
8. Lloyd, C. J. (2007). Rendering High Quality Lightpoints on Fixed Matrix Displays. Paper presented at the IMAGE Conference, Scottsdale, AZ.
9. McKee, S. P. (1983). The Spatial Requirements for Fine Stereoacuity. Vision Research 23(2), p. 191-198.
10. Nagata, S. (1991). How to Reinforce Perception of Depth in Single Two-Dimensional Pictures. In S. R. Ellis (ed), Pictorial Communication in Virtual and Real Environments. (pp) 527-545. New York, NY: Taylor & Francis.
11. Palmer, S. E. (1999). Vision Science: Photons to Phenomenology. Cambridge, MA: MIT Press.
12. Parrish, R. V., Busquets, A. M., Williams, S.P. (1990). Recent Research Results in Stereo 3-D Pictorial Displays at Langley Research Center, Proceedings of the IEEE/AIAA/NASA Digital Avionics Systems Conference. (pp. 529-539). New York: Institute of Electrical and Electronics Engineers, Inc.
13. Pfautz, J. D. (2001). *Sampling Artifacts in Perspective and Stereo Displays*. Massachusetts Institute of Technology, Cambridge, MA.
14. Singer, M. J., J. Ehrlich, et al. (1995). Task Performance in Virtual Environments: Stereoscopic Versus Monoscopic *Displays and Head-Coupling*. Alexandria, VA, U.S. Army Research Inst. for the Behavioral and Social Sciences.
15. Teichner, W. H., Kobrick, J. L., Dusek, E.R. (1956). Effects of Target Separation and Distance on Commonplace Binocular Depth Discrimination. *Journal of the Optical Society of America 46*(2), 122-125.
16. Tidwell, R. P. (1990a). Stereopsis as a Visual Cue in Flight Simulation. *Journal of Aircraft 27*(8), 731-732.
17. Tidwell, R. P. (1990b) Stereopsis Takes Off in Flight Simulation. *Proceedings of the IEEE Southeastcon Conference,* 578-582.
18. Yeh, Y. Y. and L. D. Silverstein (1989). Using Electronic Stereoscopic Color Displays: Limits of Fusion and Depth Discrimination. *Proceedings of the SPIE.*