RELATIVE EFFECTS OF FIVE DISPLAY DESIGN VARIABLES ON AIRCRAFT IDENTIFICATION RANGE IN DAYLIGHT

Charles J. Lloyd
L-3 Communications, Link Simulation & Training, Mesa, AZ

DeForest Joralmon, Ryan Amann, Chi-Feng Tai
L-3 Communications, Link Simulation & Training, Mesa, AZ

Logan Williams, Byron Pierce
711 HPW/RHA, Air Force Research Laboratory, Mesa, AZ

ABSTRACT

This paper summarizes the findings from two human factors evaluations conducted as part of the Immersive Display Evaluation and Assessment Study (IDEAS) program. For both evaluations experienced USAF F-16 pilots discriminated and positively identified distant fighter-sized aircraft. On each trial the ownship rapidly approached a pair of aircraft, one "friend" and one "foe," and the observers designated the foe as quickly and accurately as they could.

The first evaluation focused on variables expected to be primary determinants of motion-induced blurring (e.g., hold time and angular velocity) for modern display systems. The second evaluation filled out the data set required to validate a more complete model of the design variables expected to mediate task performance for very high resolution display systems. Across the two evaluations, task performance was measured as a function of 420 combinations of five practical display system design variables including: pixel hold time, angular velocity of the image, pixel pitch (resolution), display contrast, and display luminance.

Prior to conducting the evaluations a computational model was prepared and used to make quantitative predictions of the effects of these design variables. The correlation between the model predictions and the results of the first evaluation was high (e.g., $R^2 > 0.75$, $p < 0.001$, 109 df). After tuning three parameters in the model to the data the correlation increased significantly ($R^2 = 0.973$, $p < 0.001$, 106 df).

A significant benefit provided by the model is the quantification of the interactions among the design variables. Thus, the model is useful for examining the impact of design trades among the variables that affect task performance.

In a companion paper presented at this conference, we describe the design and initial validation of the model and provide examples of its use as a decision support tool by acquisitions professionals who prepare requirements and make source selection decisions, and by suppliers who wish to maximize the utility of their product offerings.

INTRODUCTION

This paper addresses the effects of five practical display design variables on the range at which pilots can identify aircraft: a visual task of great importance in the training of Air Force pilots. Few would argue that target identification range is not dependent on display resolution. The 5 m minimum dimension of a fighter sized aircraft viewed at a range of 3 km (2 nm) subtends an angle of 5.7 arcmin. At the typical resolution of training display systems of the past decade (e.g., pixel pitch = 2.5 arcmin) the minimum dimension of the aircraft would be 2.3 pixels, far less than the 13-ish pixels recommended by Johnson[8] for target identification tasks.

The use of Johnson's criteria assumes the threshold visual angle for target identification scales linearly with system resolution. Since Johnson's original paper, many similar studies have confirmed the utility of this simple method of analysis[5]. However, it has been pointed out that resolution requirements produced by the method are not precise as they depend on additional factors such as stimulus duration, background clutter, and observer capability[3].

Presented at the IMAGE 2011 Conference
Scottsdale, Arizona – June 2011

Two recent works have confirmed the linear scaling assumption for the case of relatively coarse pixel pitch where performance is limited primarily by display resolution. However, as pixel pitch is reduced performance becomes primarily limited by observer capability as illustrated in Figure 1. The upper curve in the figure shows the data from of Gaska et. al.[6] for a triangle orientation discrimination task. The lower curve shows the results of one of our preliminary evaluations for a Landolt C orientation discrimination task.

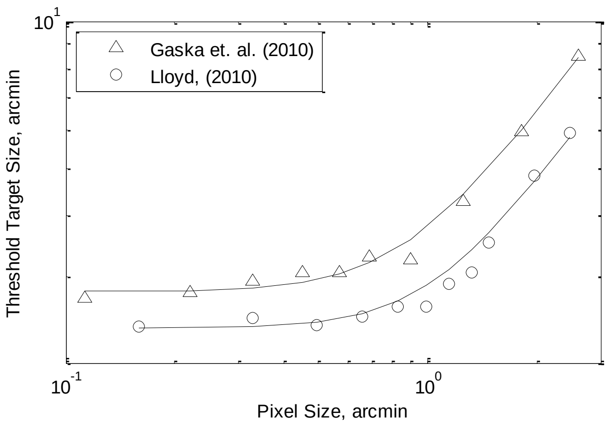

Figure 1. Threshold target size as a function of pixel pitch for triangle and Landolt C orientation discrimination tasks. Threshold target size is proportional to pixel pitch for pitches larger than about 1 arcmin. Threshold target size is constant for pitches below about 0.6 arcmin where it is mediated by observer acuity.

For pixel pitches larger than about 1 arcmin, these data show threshold target size is proportional to pixel pitch. In other words, the observer needs some minimum number of pixels across the target to accomplish the task. For pixel pitches below about 0.6 arcmin, pitch has no effect as performance is limited not by the display system but by the capability of the observer. For the case of static images, the expected effects of pixel pitch are by now well studied. In contrast, very little data are available which indicate how visual performance is affected by pixel pitch in the presence of image motion.

Motion Induced Blur

Motion induced blur has been recognized as a significant limitation of the “sample and hold” projectors (e.g., LCD, LCoS, and DLP) which are now being installed in many simulation trainers. Motion induced blurring occurs when an observer visually tracks a moving target that is drawn using pixels that remain on for a significant fraction of the frame time. Much research pertaining to the causes and remedies for motion induced blurring has been completed by researchers supporting the entertainment and advertising industries. Several recent papers provide overviews of the motion picture response time (MPRT) and related metrics and available methods for measuring the data required for computing them[2, 4, 14, 15, 16].

The International Committee for Display Metrology is expected to release their Display Measurement Standard[23] in the summer of 2011. This standard addresses the MPRT and related measures as well as several methods for their measurement. Concurrently with the development of these methods, the Air Force Research Laboratory (AFRL) in Mesa has conducted a series of evaluations that have focused on correlating a similar metric (hold time) with perceived blur and task performance[5].

Our preliminary evaluation of the standard indicates the measurement procedure should be no more complex than the AFRL-developed procedure. A strong correlation between the MPRT and hold time metric is anticipated as the MPRT is a measure of hold time convolved with the temporal step response of a display. In a future paper, we plan to address the relationship between MPRT and hold time more rigorously and expect to develop a conversion between the two methods of characterizing motion induced blurring so these literatures can be compared.

Pixel hold time

Hold time refers to the duration of time a pixel (and illumination system) is turned on at the commanded state during each frame period. A decade ago researchers at the AFRL developed a simple procedure for measuring hold time in which a fast photo sensor is used to measure a small portion of the screen. The luminance response of the display system is measured for a test pattern that alternates between full on and full off every other frame. The hold time is simply the width of the “on” time of the display device where width is defined using 50% peak luminance points on the measured curve. In the language used by the broader display community, the periodic temporal impulse response (TIR) of the display system is measured using a stationary pattern and stationary sensor. Hold time is computed as the half maximum width of the measured impulse response.

Correlation of hold time and perceived blur

A number of authors have demonstrated a strong correlation between MPRT and perceived blur[4]. Similarly, the data from a series of six evaluations at the AFRL demonstrate the strong relationship between hold time and perceived blur (Figure 2) as measured using a 2-line test pattern for which observers adjusted the width of the gap between the lines[5]. The AFRL evaluations indicate this relationship holds over a range of display technologies including CRT, LCoS, and DLP projectors.

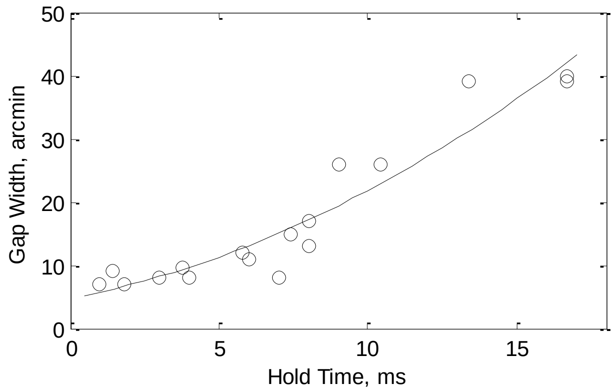

Figure 2 Gap Width measured using the 2-line perceptual blur test, as a function of Hold Time, showing a correlation of $R^2 = 0.91$ ($p < 0.001$, 16 df). Data are from six separate evaluations as summarized in Figure 8.3 of Gaska, et. al (2010) for a line speed of 40 deg/sec.

Correlation of hold time and task performance

While we expect hold time to correlate well with training task performance, relatively little work has been done to demonstrate this correlation. In a study conducted by Winterbottom et. al.,[18] aircraft roll detection threshold was measured as a function of hold time. The correlation obtained in this evaluation was moderate but statistically reliable ($R^2 = 0.4$, $p = 0.03$, 10 df). To date we have found no other papers describing evaluations in which task performance was measured as a function of hold time or MPRT. Thus, we do not yet have sufficient data to recommend the use of the hold time metric (or the MPRT) for the evaluation of simulation training display systems on the basis of task performance.

Model of Task Performance

In early 2010, work restarted on the development of a computational model of visual performance for display systems. This model is an extension of decades of image quality metric development work by Snyder, Barten, and their colleagues during the 80s and 90s[1, 11, 12, 13]. An overview of this model is provided in a companion paper at this conference[10] and more detailed descriptions of the validation studies summarized in this paper will be provided in technical reports in preparation[20, 21].

At the heart of the task performance model is the calculation of the limiting resolution of the display system. A primary input to this calculation is the modulation transfer function (MTF) of the display system which is typically computed from a measured line spread function (LSF) of the display (See Figure 5 for example). Other inputs include angular pixel pitch, hold time, target velocity, contrast, luminance, noise, and anti-aliasing. The parameters MTF, pixel pitch, hold time, target velocity, contrast, and anti-aliasing are used to compute the system MTF. The parameters luminance and noise are used to compute the contrast threshold function (CTF) of the observer. The crossover point of the system MTF and CTF is used to determine the limiting resolution of the display system which is used in the calculation of identification range.

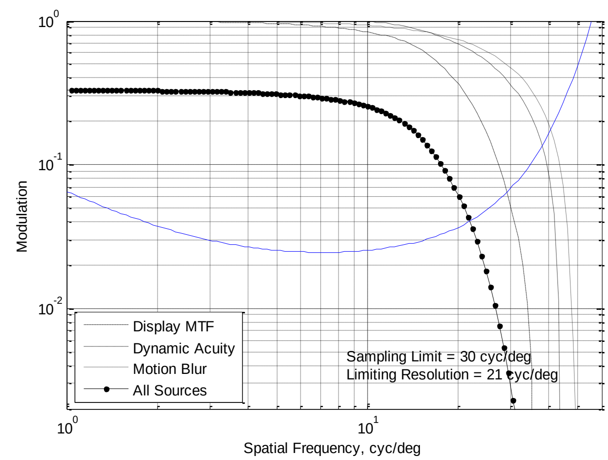

Figure 3. Illustration of the essential calculations performed within the task performance model. In this example the limiting resolution of the system is 21 cyc/deg for a pixel pitch of 1 arcmin, hold time of 8 ms, velocity of 5 deg/sec, luminance of 10 fL, and target CR of 2.2:1.

METHOD

Two separate evaluations were conducted in the AFRL laboratory in Mesa, AZ. The first evaluation focused on those parameters and interactions expected to be the primary determinants of motion induced blurring: Hold time, angular Velocity of the targets, and angular pixel Pitch. In the second evaluation eight combinations of these same three variables were again exercised but for each of 25 combinations of display Luminance and Contrast for a total 200 experiment conditions. Across the two evaluations, model predictions were computed and aircraft identification range was measured as a function of 420 combinations of the design variables. The predictions were compared with the model to assess the validity of the model. Once this comparison was made, the data were used to tune the model parameters to fit the data.

Participants

A total of eight male observers participated in the first evaluation. Seven of the eight observers were experienced USAF F-16 pilots. The observers ranged in age from 31 to 48 years with a mean age of 43 years. Prior to participation, the (far) visual acuity of each observer was measured using an Optec 2000 vision tester. All observers had a visual acuity of 20:17 or better. The single non-pilot observer was well practiced with the identification of the aircraft models and had an acuity of 20:12.

For the second evaluation, twelve male observers participated; eight of these were experienced pilots. Ten of the twelve participants completed 3 sessions each and two of the observers completed 2 sessions each. The mean age of the observers was 44 years. For the second evaluation visual acuity was tested using the Freiburg Visual Acuity test (FrACT) running on a laptop computer positioned 4 m from the observer. All observers had a visual acuity of 20:15 or better.

Evaluation Task

In both evaluations, a self-paced, two-alternative, forced choice procedure was used in which the participant selected the “foe” on each trial as quickly and accurately as practical. On each trial, the ownship started at a range of 3 to 6 km from a pair of aircraft, one friend and one foe. The starting position for each trial was set at 2.2 times the expected identification range and was randomized +/- 20 percent. Trial length was capped at a maximum of 15 seconds. On average the participants identified the foe after about 7-8 seconds and initiated the next trial immediately. Each observer participated in two experimental sessions on separate days.

On each trial, the aircraft traveled in straight and level flight at a ground speed of 250 to 300 knots, pointing either left or right as in Figure 4. The mean contrast of the aircraft against their background was 2.2:1.

At the typical identification range, the horizontal speed of the bogies would produce a negligible angular velocity from the point of view of the ownship.

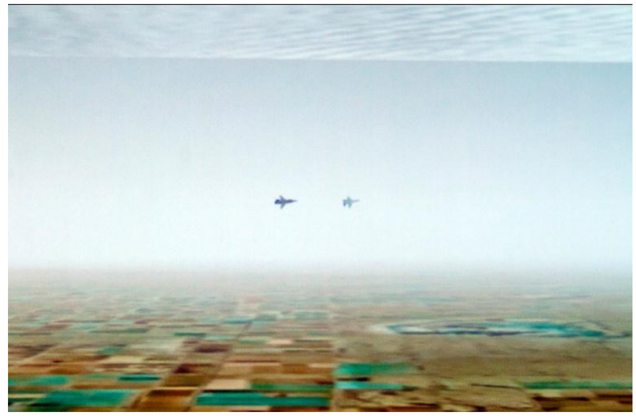

Figure 4. Photograph of a typical trial showing a pair of aircraft pointing to the right. The aircraft were always seen against the uniform portion of the sky.

The angular velocity of the targets/image was set to a constant and controlled level on each trial by changing the pitch and yaw of the ownship in a circular orbit. From the point of view of the observers, this gave the appearance of the ownship approaching the bogies in a spiraling motion. This orbital motion of the ownship allowed sustained high angular velocities for the duration of the trial while keeping the targets near the center of the screen. A second advantage of the spiraling motion was that it produced motion smearing in all orientations during the course of each trial. Angular velocity was controlled by the diameter of the orbit. The largest orbit used in the evaluations had a radius of 8 inches which kept the bogies within the central portion of the screen where our calibration of the hold time shutter was valid.

Prior to each experimental session, each observer studied larger images of the aircraft to become familiar with their appearance. Each session required about 50 minutes to complete. The experimental task was identical between the two evaluations except for the following procedural differences:

- The sizes of the sets of aircraft models was increased to 13 friends and 12 foes
- The orbital period of the yaw-pitch motion was increased from 1 to 1.4 sec.

Equipment

The same laboratory space and equipment was used in both evaluations except for the differences noted below.

Image generator

The IG computer ran the Windows XP operating system on a custom built computer, consisting of an Intel Core I7 - 920 processor 12 GB of ram. The graphics for the IG were driven by the Nvidia Quadroplex 2200 D2 model which provided the 4 channels required to drive the Sony SXRD projector at a resolution of 4096 X 2160. The IG software is MetaVR version 5.6. The Sim Host computer ran the MATLAB (The Math Works) software under the Windows XP operating system.

Projector and Screen

Both evaluations were conducted in the OBVA laboratory at the AFRL facility in Mesa AZ using an 8 Mpix Sony SRX-S110 LCoS projector. The image was projected on a flat screen measuring 2.28 x 1.27 m (90 x 50 in). The center of the screen was 1.88 m (74 in) above the floor. The projector was mounted overhead on a stand which position the lens 2.39 m (94 in) from the floor and 6.3 m (248 in) from the screen. The walls in the laboratory were painted black, thus, very little scattered light was present.

Motion blur reduction shutter

An LCD motion blur reduction shutter was purchased from VDC Display Systems in the fall of 2010. This device allows hold time to be controlled, from trial to trial, over a range of 1.5 to 14 ms.

Five levels of hold time, indicated in Column 1 of Table 1 were used in the first evaluation. The luminance of the projected image varied proportionally with the hold time setting of this shutter as expected. The luminance levels for the sky against which the aircraft were observed are indicated in Column 2 of Table 1.

Filters and Luminance

In the first evaluation display luminance was confounded with the hold time setting of the shutter. Since there would be no way to differentiate the effects of luminance and hold time, luminance was also independently manipulated. This was accomplished by doubling the number of conditions in the first evaluation and reducing luminance to half for the additional trials using a neutral density (ND) filter mirror. The resulting luminance levels for the filtered conditions are provided in Column 3 of Table 1.

Contrast and flood lighting

When viewed in the darkened lab, the mean contrast of the aircraft models against the sky background was 2.2:1 with all aircraft being darker than the background. In the first evaluation, the display system contrast (checkerboard) was fixed at 24:1. The dark laboratory and flat projection screen resulted in negligible scattered light from these sources. The primary source of scatter in the system was the hold time shutter.

Table 1. Mean luminance levels of the sky against which the bogies were observed for all combinations of the hold time and filter condition. These levels were approximately 85% of the peak white of the display system.

Hold Time, ms	Luminance with no Filter, cd/m^2	Luminance with Filter, cd/m^2
2.0	11	5.5
4.5	24	13
7.0	45	22
9.5	58	29
12.0	76	39

In the second evaluation, the contrast of the display system was manipulated with the use of computer controlled flood lights that uniformly illuminated the screen. Use of the flood lights allowed simulation of the unavoidable "washout" that occurs in most training display systems due to light scattered from the projection screen (and mirror) that illuminates other portions of the screen. Six levels of washout lighting were used as indicated along the top of the following three tables.

Comparing across these tables shows the combinations of filter and washout lighting produced display luminance levels ranging from 5.7 to 170 cd/m^2 and display contrast ratios ranging from 3.1 to 99. Most of the luminance variation in the second evaluation was not confounded with hold time as it was in the first evaluation.

Table 2. Measured peak display luminance and black level for each of the 6 washout conditions for a hold time setting of 6 ms. Luminance is in cd/m^2, display contrast is indicated in bold.

Filter Trans	0 cd/m^2	0.5 cd/m^2	1 cd/m^2	2 cd/m^2	4 cd/m^2	7.5 cd/m^2
0.61	42 1.72 **24**	43 2.26 **19.0**	43 2.9 **14.8**	44 3.8 **11.6**	47 5.8 **8.1**	49 9.2 **5.3**
0.37	25.7 1.05 **24**	26.2 1.57 **16.7**	26.8 2.22 **12.1**	27.7 3.2 **8.7**	29.5 5.2 **5.7**	33.0 8.7 **3.8**
0.14	10.5 0.45 **23**	10.9 0.97 **11.2**	11.4 1.63 **7.0**	12.4 2.58 **4.8**	14.2 4.6 **3.1**	
0.085	5.7 0.26 **22**	7.0 0.82 **8.5**	7.5 1.40 **5.4**	8.5 2.40 **3.5**		

Table 3. Measured peak display luminance and black level for each of the 6 washout conditions for a hold time setting of 12 ms. Luminance is in cd/m^2, display contrast is indicated in bold.

Filter Trans	0 cd/m^2	0.5 cd/m^2	1 cd/m^2	2 cd/m^2	4 cd/m^2	7.5 cd/m^2
0.61	86 3.5 **24**	86 4.1 **21**	87 4.6 **18.7**	88 5.6 **15.7**	90 7.6 **11.9**	93 11.0 **8.5**
0.37	52 2.14 **24**	52 2.67 **20**	53 3.3 **16.1**	54 4.2 **12.7**	56 6.2 **9.0**	59 9.7 **6.1**
0.14	21 0.89 **24**	21.5 1.42 **15.1**	22.1 2.04 **10.8**	23.0 3.0 **7.7**	25.0 5.0 **5.0**	28.5 8.5 **3.4**
0.085	11.0 0.48 **23**	12.2 1.03 **11.8**	12.7 1.64 **7.7**	14.0 2.65 **5.3**	16.0 4.7 **3.4**	

Table 4. Measured peak display luminance and black level for each of the 6 washout conditions with filters and hold time shutter removed. Luminance is in cd/m^2, display contrast is indicated in bold.

Filter Trans	0 cd/m^2	0.5 cd/m^2	1 cd/m^2	2 cd/m^2	4 cd/m^2	7.5 cd/m^2
1.0	163 1.65 **99**	163 2.18 **75**	164 2.80 **59**	165 3.8 **44**	167 5.7 **29**	170 9.2 **18.5**

Pixel Pitch and Resolution

In this evaluation, the practical variable viewing distance was used to control the angular pixel pitch of the display system. No other manipulations were made that independently affected the relationship between pixel pitch, viewing distance, and measured resolution (MTF), thus, the three variables Pixel Pitch, Viewing Distance, and Resolution were completely confounded and are used inter-changeably in this report.

For most training display systems, the linear Pixel Pitch (in mm) and viewing distance are clearly defined and relatively immutable attributes of the system. The angular Pixel Pitch (in arcmin) of the system is easily computed from these two quantities and is thus also clearly defined and not often misinterpreted.

However, the “effective” or “limiting” resolution of a display system is not nearly as easy to define or measure as is pixel pitch. This is primarily because this system attribute depends on a number of additional factors such as optical blur, pixel hold time, angular velocity, mis-convergence, luminance, contrast, anti-aliasing, and observer acuity.

For these evaluations, the pixel pitch (pixel-to-pixel spacing) measured at the center of the screen was 0.60 mm (100 pixels measured 60 mm). The vertical and horizontal pitches differed by no more than 2%.

For each trial, the observer was seated at one of five viewing distances indicated in Column 1 of Table 5. The angular pixel pitch of the display system corresponding to each viewing distance is provided in Column 2. To give the reader a sense for the number of pixels used to render images in this evaluation, Column 3 provides the number of pixels spanning the wingspan of the aircraft.

Table 5. Viewing distance, angular pixel pitch and number of pixels per wingspan for a fighter aircraft with an 11 m wingspan viewed at a range of 2.5 km (15 arcmin).

Observer – Screen Distance meters	Pixel Pitch arcmin	Number of Pixels per 15 arcmin wingspan
0.8	2.58	5.8
1.2	1.72	8.7
1.8	1.15	13
2.8	0.74	20
4.2	0.49	30

The line spread function (LSF) of the projected image was measured using a calibrated color camera (Canon G-9) positioned approximately 12 inches from the screen (see Figure 5). For this measurement, a pair of widely spaced single pixel wide white lines on a black background were projected on the screen and photographed. The space between the pair of lines was measured with a ruler and used to determine the sampling rate of the camera arrangement which measured 11.4 camera pixels per mm.

The MTF of the display system was used as an input to the model along with the settings of each of the five independent variables used in the evaluations. The limiting resolution of the display system and the expected threshold target size were computed separately for each of the 420 experimental conditions.

Aircraft Models

Table 6 lists the model names for the 11 friendly and 10 enemy aircraft used in the first evaluation. On each trial one aircraft was selected at random from each of these lists. The aircraft model numbers were recorded for each trial so that a table of the relative discriminability of model pairings could be constructed. An average of 31 trials were used to estimate the level of each aircraft pairing and these data were used to remove the variance due to the pairings.

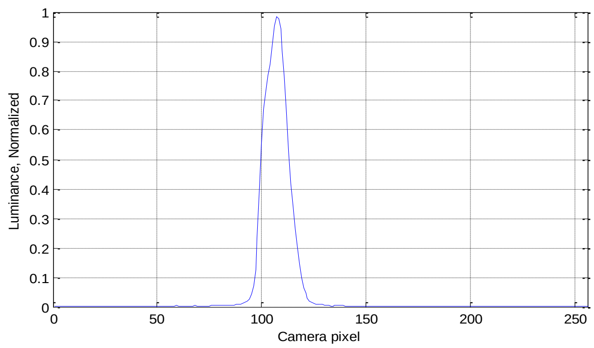

Figure 5. Line spread function for a single white line on a black background. Camera calibration was 0.088 mm / camera pixel, thus, about 25 camera pixels spanned the line spread function.

Table 6. Listing of the 11 friendly and 10 enemy aircraft models used in the evaluation. All models were supplied with the MetaVR image generator.

	Friendly Aircraft		Enemy Aircraft
120	F-16	131	Mig 29
121	F-16_Oman	132	Mig 29C
122	F-2	133	Su 30
123	F-18	134	Su 30mk
124	CF-18	135	Su 27
125	F-16_Low_res	136	Mig 25
126	F-35	137	Mirage_Iraq
127	F-15	138	Mig 23
128	AV8B	139	Mig 21
129	Tornado_F3	140	Mig 21_Iraq
130	Tornado_gr1		

Independent Variables

The primary goal for the experimental design was to cover the design trade space, making sure to gather enough data to fully characterize the expected interactions among the variables. The independent variables and levels used in the first evaluation were:

- Hold Time, 5 levels: 2, 4, 7, 9.5, and 12 ms
- Target Velocity, 5 levels: 1, 8, 15, 22, and 29 deg/sec
- Pixel Pitch: 2.57, 1.70, 1.14, 0.73, 0.49 arcmin
- Luminance (filter), 2 levels, 100 and 50%

A nearly full factorial design was used so that variance representing the interactions would be included in the data. 220 of the 250 possible combinations were measured; the remaining 30 combinations were not collected because the required orbit diameters were larger than the +/- 8 inch central portion of the screen where the blur reduction shutter calibration was considered valid.

The independent variables and ranges used in the second evaluation were:

- Pixel Hold Time, 2 levels: 6 and 12 ms
- Target Velocity, 2 levels: 3 and 12 deg/sec
- Pixel Pitch, 2 levels: 0.5 and 1.4 arcmin

Each of the four combinations of target Velocity and Pitch were used with the 44 conditions described in Tables 2 and 3 for a total of 176 conditions employing the hold time shutter. For the case of no hold time shutter, 24 more conditions (Table 4) were added consisting of 6 levels of Washout, 2 levels of Velocity, 2 levels of Pixel Pitch, for a total of 200 experimental conditions.

RESULTS

Evaluation One

At all stages of analyses, the distributions of data were clearly more symmetrical and Gaussian when a logarithmic transformation of threshold size was used as compared with either size or range. Thus, all of the statistical analyses described below were performed on the log10 transformation of the dependent variable to improve the accuracy of the statistical tests.

Prior to conducting the statistical analyses reported below, the effects of observer, practice, and aircraft model pairings, were removed from the data. Details of the data reduction procedures used for each evaluation are provided in the AFRL technical reports describing each evaluation[20, 21].

For the first evaluation, the data from 3511 trials were used in the analyses, for an average of 16 trials per experimental condition.

Fit of Initial Model

Prior to completing the data collection for this evaluation the model described in Lloyd et. al., (2011)[10] was used to compute the expected responses to the experimental variables. Prior to the optimization of model coefficients, the correlation between these predictions and the mean responses of the 8 observers was $R^2 = 0.78$ ($p < 0.001$, 109 df). In the analyses that follow, the model coefficients were optimized to maximize the correlation between the model and data.

Differences among observers

The eight observers in the first evaluation differed substantially in the maximum ranges at which they identified targets. The mean range for the three observers with the longest ranges was 1.6 times larger than the mean range of the three observers with the shortest ranges. The correlation between identification range and visual acuity was in the expected direction; however, it was not statistically reliable ($p > 0.05$, 7 df). Thus, it appears that factors in addition to acuity may be responsible for the differences among observers.

Effect of luminance

The independent variable luminance (filter) had a small but statistically reliable ($p < 0.01$, 109 df) effect on performance. When luminance was halved with the insertion of the ND filter, the mean size at which targets were identified increased by 2.5%. Halving the luminance of the display system using the filter essentially raised the surfaces plotted in Figures 6 and 7 by 0.011 log10 units without noticeably changing their shape.

Effects of Hold Time, Angular Velocity, and Pixel Pitch

Compared with the effect of Luminance, the variables Hold Time, Angular Velocity, and Pixel Pitch had much larger effects as predicted by the model. Since the interactions among these variables are strong and complex, the effects of all three variables are shown in the form of surface plots which illustrate the main effects and interactions in the same set of plots. For these plots the data have been averaged over the two levels of luminance which did not substantially change the shapes of the surfaces.

The results of the evaluation are illustrated in Figures 6 and 7 which show the effects of pixel pitch and hold time on the log10 of threshold identification size for two angular velocities. These plots illustrate the fit of the model (the surfaces) to the data. The circles on these figures indicate the mean threshold target size averaged across observer and filter condition (N = 16). The correlation between these data and the model was $R^2 = 0.973$ ($p < 0.001$, 106 df). The standard deviation of the residuals was 0.021. Converting from log10 of the residuals, the standard deviation was 4.9% of the target size (or range).

When the data were averaged across observer and not filter condition (N = 8), the correlation between data and model was $R^2 = 0.952$ ($p < 0.001$, 216 df). For this case the standard deviation of the residuals was 0.028. Converting from log10 of the residuals, the standard deviation was 6.6% of the target size (or range).

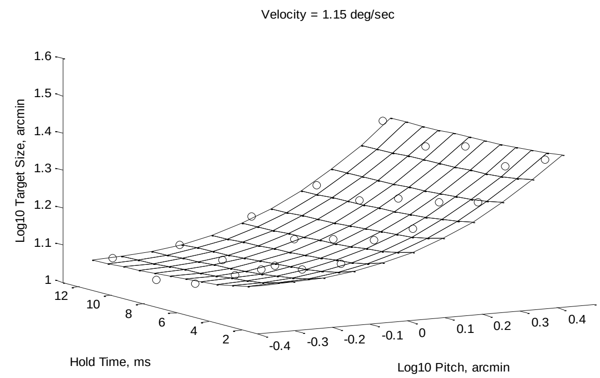

Figure 6. Threshold target size as a function of Pixel Pitch and Hold time for a velocity of 1 deg/sec. Note that hold time had little effect for slowly moving targets.

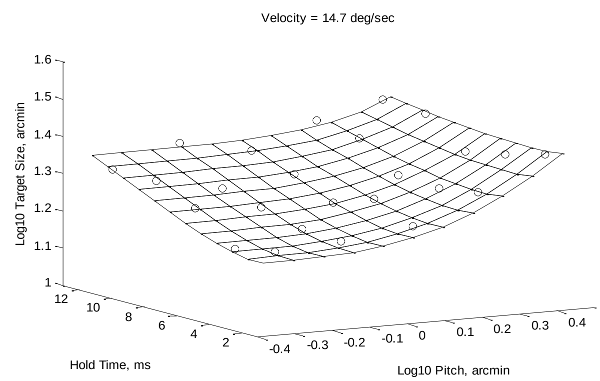

Figure 7. Threshold target size as a function of Pixel Pitch and Hold time for a velocity of 15 deg/sec. Note that at 15 deg/sec hold time has a large effect on threshold target size, especially for fine display pitch.

Evaluation Two

The data reduction and analyses for Evaluation 2 are identical to the first evaluation. The analyses below are based 6350 trials for an average of 31.7 trials per experimental condition. At the time this paper was prepared, the model parameters had not yet been optimized to fit the model to the data from the second evaluation. The parameters were left at the settings that maximized the correlation with the data from Evaluation 1. With these pre-determined settings the correlation between the model and the Evaluation 2 data is $R^2 = 0.911$ ($p < 0.001$, 199 df). The standard deviation of the residuals is 0.028. Converting from log10 of the residuals, the standard deviation is 6.5% of the target size (or range).

The results of the second evaluation are illustrated in Figures 8 and 9 which show the effects of the filter and washout conditions for selected levels of Velocity, Hold, and Pixel Pitch. These plots are designed to show the degree to which the model fits the data. Figures 10 to 12 provide more easily interpreted views of these effects.

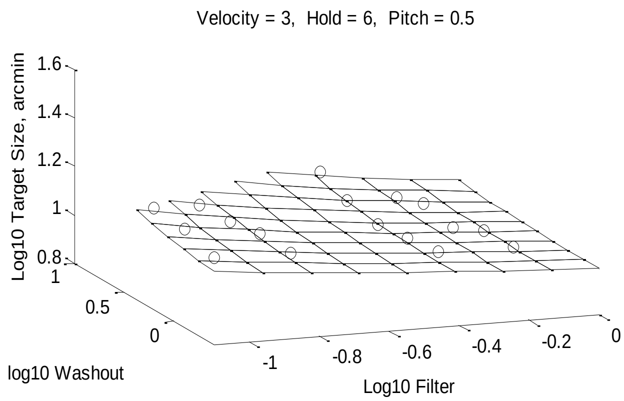

Figure 8. Threshold target size as a function of Filter transmittance and Washout Luminance for the case of very fine Pitch and low motion induced blurring.

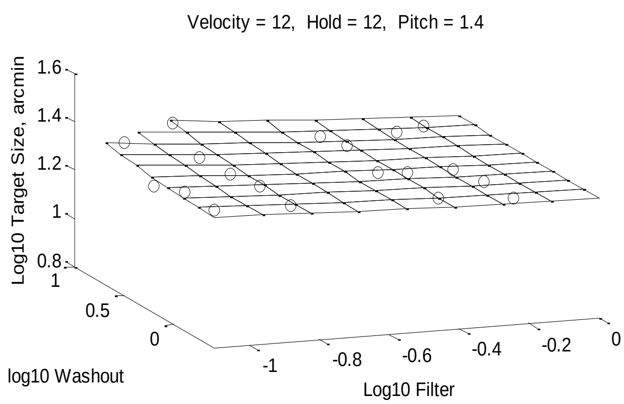

Figure 9. Threshold target size as a function of Filter transmittance and Washout Luminance for the case of coarser display pitch and moderate motion induced blurring.

DISCUSSION

In the previous sections, the data and model were shown together on surface plots that were scaled in a transform space that homogenizes the variance across the experimental conditions so that the fit of the model to the data could be assessed. In this section, we plot several views of the model in a transform space that is more immediately useful to a display system specifier or design engineer.

Figure 10 illustrates the effects of the three variables that had the largest impact on performance: Pitch, Hold time, and angular Velocity. These surfaces represent the mean performance of our 8 observers for a peak display luminance of 30 fL, a display contrast of 20, and a fighter-sized aircraft (11 m wingspan).

Figures 11 and 12 illustrate the effects of display luminance and contrast which had smaller effects on performance than did the first three variables. Figure 11 represents the case of dark targets against a bright background that is near the peak luminance of the display system. Figure 12 shows the effect of contrast is expected to be stronger when the target background is only 25% of the peak display luminance.

CONCLUSIONS & RECOMMENDATIONS

- These evaluations provide the simulation training community with far more data pertaining to hold time and task performance than were previously available.
- The expected effect of hold time on task performance has been confirmed with high statistical reliability across hundreds of combinations of parameter settings.
- The correlation between model predictions and the data were very high, confirming the validity of the model.
- The model accurately quantifies the interactions between the five practical design variables, thus, the model is well suited for supporting design trades among these variables.
- The data and model presented here indicate larger improvements in training task performance are available through decreased display pitch and hold time than are available through increased display luminance and contrast.

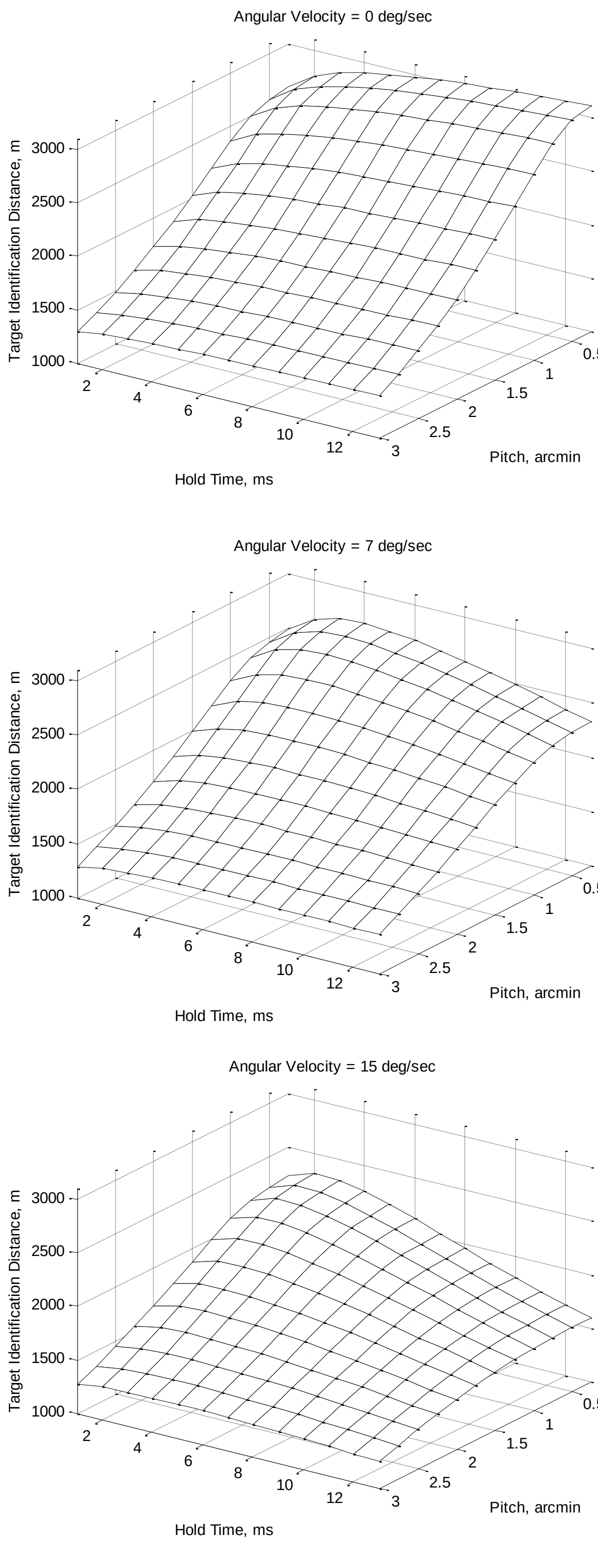

Figure 10. Threshold target identification range for fighter-sized aircraft as a function of Pitch and Hold time for three levels of target Velocity.

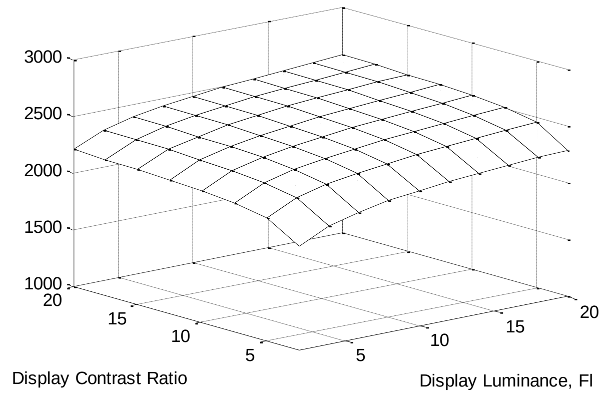

Figure 11. Aircraft identification range as a function of display Luminance and Contrast for a display Pitch = 1 arcmin, Hold time = 8 ms, Velocity = 5 deg/sec, and target background at 85% of the peak display luminance.

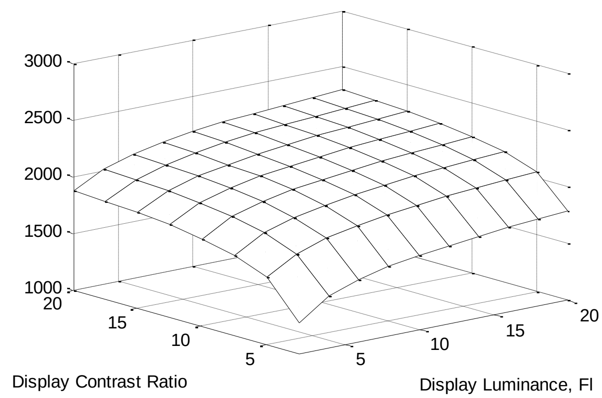

Figure 12. Aircraft identification range as a function of display Luminance and Contrast for a display Pitch = 1 arcmin, Hold time = 8 ms, Velocity = 5 deg/sec, and target background at 25% of the peak display luminance.

Beware of Flicker

To obtain near-eye limited resolution in the presence of even moderate image motion, hold times of only a few ms will be required (see the bottom panel of Figure 10). A substantial literature recommends the use of frame rates of 75 Hz or greater for short hold time displays (e.g., CRTs) to avoid the detrimental effects of flicker[24].

We know of no other means by which motion induced blurring can be reduced to inconsequential levels but to reduce hold time. Thus, it appears the simulation training industry will have to move to higher frame rates as we move towards eye-limited resolution.

ACKNOWLEDGEMENTS

We wish to acknowledge the following people for their significant contributions to the IDEAS program.

- Jim Evans, ASC/ENDR
- Tim Dwyer, ASC/WNSEA
- Jim Basinger, Superior Technical Services
- William Morgan, Boeing
- George Geri, L-3 Communications
- Joseph Riegler, Boeing
- James Gaska, L-3 Communications

Thanks also to VDC Display Systems for providing technical support for the projector and shutter used in these evaluations.

AUTHOR BIOGRAPHIES

Dr. Charles J. Lloyd has 25 years of experience in display systems and applied vision research at such organizations as the Displays and Controls Lab at Virginia Tech, the Advanced Displays Group at Honeywell, Lighting Research Center, Visual Performance Inc., BARCO Projection Systems, and FlightSafety Int. Charles is now the Lead Scientist for the IDEAS program at the Air Force Research Laboratory (L-3 Communications) where he manages the development and validation of display system metrics for simulation training. Charles has presented 65 papers in this arena.

REFERENCES

[1] Barton, P G. J. (2000) Contrast sensitivity of the human eye and its effects on image quality. SPIE Press.

[2] Becker, M. E. (2008) Evaluation of moving-line contrast degradation without motion. Soc. for Info. Display Digest of Tech. Papers.

[3] Donohue, J. (1991) Introductory review of target discrimination data. Report PL-TR-92—2129, Phillips Laboratory, Air Force Systems Command, Hanscom AFB.

[24] Farrell, J. E., Benson, B. L., and Haynie, C. R (1987) Predicting flicker threshold for video display terminals. SID Digest of Technical Papers.

[4] Feng, X., Pan, H., and Daly, S. (2008) Comparisons of motion-blur assessment strategies for newly emergent LCD and backlight drive technologies. Journal of the SID, 16. pp. 981-988.

[5] Gaska, J. P., Geri, G. A., Winterbottom, M. D., and Pierce, B. J. (2010) Evaluation of the spatial and temporal resolution of digital projectors for use in full-field flight simulation. In K. Nial Ed. *Vision and Displays for Military and Security Applications.* Springer, New York.

[6] Gaska, J. P., Winterbottom, M., Sweet, W., and Rader, J. (2010) Pixel size requirements for eye-limited flight simulation. IMAGE Society Annual Conference, Scottsdale, AZ.

[7] Holst, G. C. (2000) Electro-optical imaging system performance, Chapter 20, SPIE Optical Engineering Press, Bellingham, WA.

[8] Johnson, J. (1958) Analysis of image forming systems. Proc. Image Intensifier Symposium, pp. 249-273.

[9] Klompenhouwer, M. A. (2006) Comparison of LCD motion blur reduction methods using temporal impulse response and MPRT. SID Digest of Tech. Papers.

[10] Lloyd, C. J., Williams, L., and Pierce, B. (2011) A model of the relative effects of key task and display design parameters on training task performance. *Proceedings of the IMAGE 2011 Conference,* The IMAGE Society, Phoenix, AZ.

[11] Lloyd, C. J. and Beaton, R. J. (1990) Modeling suprathreshold visual responses for image quality evaluations of color displays**.** Proceedings of the SPIE's 43rd Annual Conference. The Society for Imaging Science and Technology, Springfield, VA

[12] Lloyd, C. J. (2002) Quantifying edge-blended display quality: Correlation with observer judgments. Proceedings of the IMAGE Society Annual Conference, Scottsdale, AZ

[13] Snyder, H. L. (1985) Image quality: Measures and visual performance. In L.Tannas Ed., *Flat-panel displays and CRTs,* vanNostrand Reinhold, New York.

[14] Someya, J. and Sugiura, H. (2007) Evaluation of liquid-crystal-display motion blur with moving-picture response time and human perception. Journal of the SID.

[15] VESA (2005) Flat panel display measurements standard, Version 2.0. FPDM Update, May 19, 2005. Video electronics standards association FPDM task group.

[16] Watson, A. B. (2009) Comparison of motion-blur measurement methods. SID Digest of Technical Papers.

[17] Winterbottom, M. D., Geri, G. A., Morgan, W., and Pierce, B. J. (2007) An integrated procedure for measuring the spatial and temporal resolution of visual displays. Proc. Of the I/ITSEC conference.

[18] Winterbottom, M. D., Geri, G. A., and Park, L. (2007) Task validation of display temporal-resolution measurements. Proc. Of the I/ITSEC conference.

UPCOMMING REPORTS

[19] Lloyd, C. J., et. al. (2011) Towards a decision support system for simulation training display requirements. To be presented at I/ITSEC 2011.

[20] Lloyd, C. J. et. al. (2011) Effects of hold time, angular velocity, pitch, and luminance on simulated aircraft identification range. AFRL technical report, to be distributed by Defense Technical Information Service (DTIC).

[21] Lloyd, C. J. et. al. (2011) Effects of luminance, contrast, hold time, angular velocity, and pitch, on simulated aircraft identification range. AFRL technical report, to be distributed by Defense Technical Information Service (DTIC).

[23] (2011) Display Measurement Standard, International Committee for Display Metrology. Society for Information Display. Release anticipated June 2011.