High Dynamic Range in Visual Simulation for Training Applications

Harry Streid
The Boeing Company
St Louis, MO
harry.streid@Boeing.com

Shawn Stafford
Full Sail University
Orlando, FA
shawns@fullsail.com

ABSTRACT

The human eye is capable of adapting to displayed information with precision over a range of 14 orders of magnitude in luminance, from bright day to starlight. The best visual displays used in simulation and training today offer a range less than 4 orders of magnitude, and their peak luminance is a tiny fraction of that of the real world scenes we are trying to simulate. Now that affordable "eye-limited resolution" visual systems are becoming a reality, the training community needs to be aware of how visual performance in the simulator may still be limited by factors other than resolution, and how luminance and precision limitations can be removed by emerging technologies.

Commonly in the simulator today, visual details intended to represent part of a daytime visual scene are actually rendered at luminance levels only encountered at night in the real world. As is well known to vision scientists, the eye responds very differently at different luminance levels and training fidelity may suffer when display luminance is inappropriate for the training scenario. Fortunately, Hollywood and the consumer electronics industry are leading the way, as they did in HDTV and video game technology, with their embrace of the next big thing in visual displays, High Dynamic Range (HDR).

High Dynamic Range (HDR) is a confluence of technologies that offers visual displays with higher luminance and precision, as well as a better fit to the human visual system. But what will be the value to the warfighter and what tradeoffs will be encountered in bringing HDR to training?

In this paper we will discuss the scientific rationale for HDR, examine its maturity level relative to Warfighter training needs, analyze the potential it offers for higher visual fidelity in the simulator, and present the results of tactical decision making trials comparing HDR to conventional visual display systems.

ABOUT THE AUTHORS

Harry Streid is a visual systems engineer and Technical Fellow for Boeing Training Systems and Government Services (TS&GS). He develops flight simulation visual systems and designs large screen displays for these applications. He has been awarded numerous display related U.S. and foreign patents.

Shawn Stafford has a Ph.D. in Applied Experimental and Human Factors Psychology from the University of Central Florida. He runs the User Research Lab at Full Sail University and teaches several classes in user research and perception. He has focused his career on sensation and perception, user research and physiology.

High Dynamic Range in Visual Simulation for Training Applications

Harry Streid
St Louis, MO
Harry.Streid@Boeing.com

Shawn Stafford
Full Sail University
Orlando, FA
shawns@fullsail.com

INTRODUCTION

There is no question that capability of the display devices that we use every day is extraordinary. Driven largely by the highly competitive $100 billion world-wide consumer market for mobile device displays and large screen home entertainment, the pace of innovation in electronic displays has never been greater. Basic scientific discoveries such as OLEDS and Quantum Dots which languished in the lab for decades are suddenly enabling products found in our living rooms and in our pockets/purses. While the Warfighter has benefited enormously from the introduction of the digital projectors used in most simulators today, most of the benefits have been in better reliability and life cycle cost while basic performance in terms of dynamic range and dynamic acuity has only recently begun to approach the levels that were commonplace with the CRTs that were replaced by digital projectors. The recent advances will allow us to meet current requirements for visual training systems cheaper and more reliably, but can they also expand the envelope of training tasks that can be performed in the simulator? Can they change what we see and how we see in ways that provide greater fidelity in terms of increased magnitude and precision of luminance, color and contrast we are trying to the real world scenes we are trying to simulate and how can this provide better training? In this paper we will try to answer these questions.

History of High Dynamic Range

It has always been recognized that looking at pictures was different from looking at the real thing. Artists have always used their various media to create interpretations of real and imagined scenes for artistic effect. The limitations of photographic media in capturing and displaying the range of luminance available in the real world were well understood and were be dealt with by multiple exposure photography and compositing printing methods as early as 1800's. More recently a similar method, tone mapping was developed to allowed display of high dynamic digital content on low dynamic range electronic displays. Such methods create a synthetic interpretation of the real world by remapping certain colors into different colors which can then be displayed on a low dynamic range display. The simulation and training community developed one of the earliest true high dynamic range displays over 30 years ago by combining the full black capabilities of CRTs with very high intensity calligraphic light points to produce very realistic night scenes. The desire to mimic the performance of CRTs with digital projectors inspired the development of the dual modulation light valve projector, patented by Blackham, (USPO, 1999), which was one of the first digital electronic displays that could truly be considered as having high dynamic range. None of these flight simulation displays produced daytime luminance levels approaching real world values, however. The first electronic displays developed for the purpose of displaying realistic images with contrast and luminance approaching real world levels were those described by (Seetzen, 2004). Those who have had the opportunity to view HDR images realize that this is a different seeing experience qualitatively. The reasons for this qualitative difference has yet to be fully understood. Since these early demonstrations advances in electronic displays including quantum dots, OLEDs and high luminance LEDs have converged with an HDR infrastructure for content creation, standard interfaces and media distribution required for the imminent introduction by the consumer electronics industry of a myriad of high dynamic range products.

Quantifying High Dynamic Range

There is no consensus in the display industry as to what constitutes high dynamic range, other than that it needs significantly higher luminance than we have been accustomed to as well as better contrast and color depth (up to 12 bit precision). The once ubiquitous CRT and now the LCD monitors and televisions which replaced CRTs have always had a peak luminance of about 100 candelas per square meter (cd/m^2 or nits). For many years the FAA has set minimum requirements for display luminance of simulators used to train commercial airline pilots to be barely more than 20 cd/m$^{2.}$ Even today, flight simulator displays for military training rarely exceed 30 cd/m$^{2.}$ The displays as described in Seetzen had luminance on the order of several thousand nits. The UHD alliance, a consortium of consumer electronics manufacturers, content creators and distributors and trade groups, recently published standards for their consumer product rating "HDR Premium" for televisions to be e*ither more than 1000* cd/m^2 *peak luminance and less than 0.05* cd/m^2 *black level, or more than 540* cd/m^2 *peak luminance and less than 0.0005* cd/m^2 *black level.* (UHD Alliance press release, 2016).

The UHD Alliance also recommends HDR Premium devices to use SMPTE Standard ST2084 EOTF for color encoding. Based upon Weber's Law and derived using Barten's contrast sensitivity function (Barten, P. 1992), ST2084 provides the most efficient use of digital data and allows an entire dynamic range from .001 to 10,000 cd/m^2 to be encoded into digital pixels values of 12 bits per color with quantization steps below the JND threshold of human vision (SMPTE, 2014).

Comparing Standard and High Dynamic Range Displays With the Capabilities of Human Visual System

As shown in figure 1, the range of luminance in the real world environment that the eye is able to adapt to is about 14 orders of magnitude (Kunkel and Reinhard, 2010). The simultaneous dynamic range, which is the range over which the eye can detect details while in a state of adaptation, is much less and although there is less agreement about how to measure it, is thought to be no more than about 4 orders of magnitude, (Kunkel and Reinhard 2010). In figure 1 we also see that the range of luminance of electronic display devices typically used to simulate the real world is far less than the range of possible adaptations of the eye and even less than any single adaptation state would be.

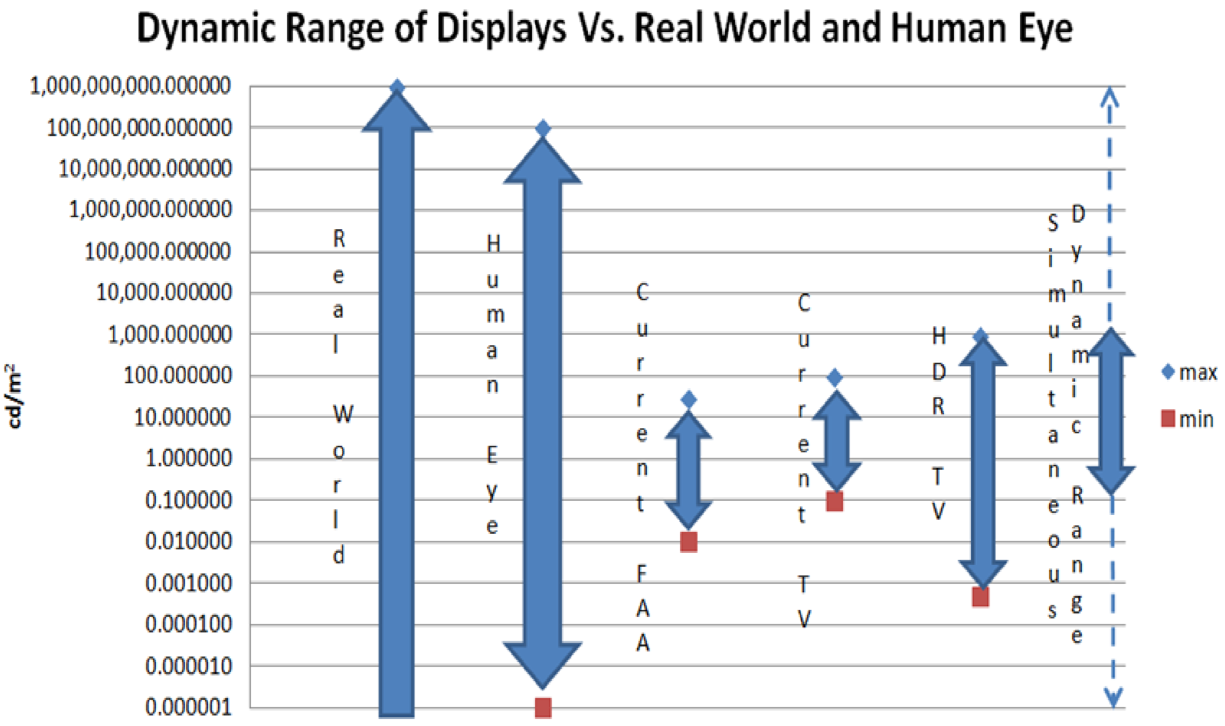

Figure 1 Dynamic Range of Standard and High Dynamic Range Vs Real World And Human Eye Adaptation Range

APPLICATION OF HIGH DYNAMIC RANGE TO WARFIGHTER TRAINING REQUIREMENTS

The discussion of Figure 1 above makes it clear that that visual systems in use today do not come close to matching the dynamic range of luminance in the real world. However, the human visual system responds to increasing intensity of stimulation with reduced precision in a logarithmic fashion (Weber's Law) to maintain human performance at the same level over a very wide range. If high dynamic range is to be useful in training systems it must be shown first of all to expand the range of psychophysical conditions of illumination and contrast that can be simulated in the trainer beyond current levels. If this is the case then the Human Visual System (HVS) will be taxed in new ways that alter human performance corresponding to the ways in which it would be altered under comparable conditions in the real world. The second qualification for it being desirable to add high dynamic range to the trainer is that the expanded range of illumination conditions can be used to enhance operational efficiency in specific tasks that are important to warfighter safety or mission success. In this paper we will attempt to demonstrate that the first criteria is achieved by predicting how the human visual system will respond to high dynamic range using established models for HVS performance and by performing man-in-the loop experiments to test human response to visual decision making tasks. With the data illustrated in figure 1 as reference, we can begin to anticipate what luminance and dynamic range might be required for a High Dynamic Range display system for Warfighter training.

Predicting Human Performance as a Function of Contract Sensitivity

Not shown in the figure 1 is the fact that the contrast sensitivity of the eye changes significantly with luminance. Figure 2 plots the contrast sensitivity of the eye at minimum and maximum luminance for range of luminance of a current FAA level D display system and for a high dynamic range display that would meet requirements of the UHD Alliance "HDR Premium rating. Figure 2 shows that the range of contrast sensitivities that can be simulated by the FAA display will be limited to the shaded area between the max and min luminance for this display.

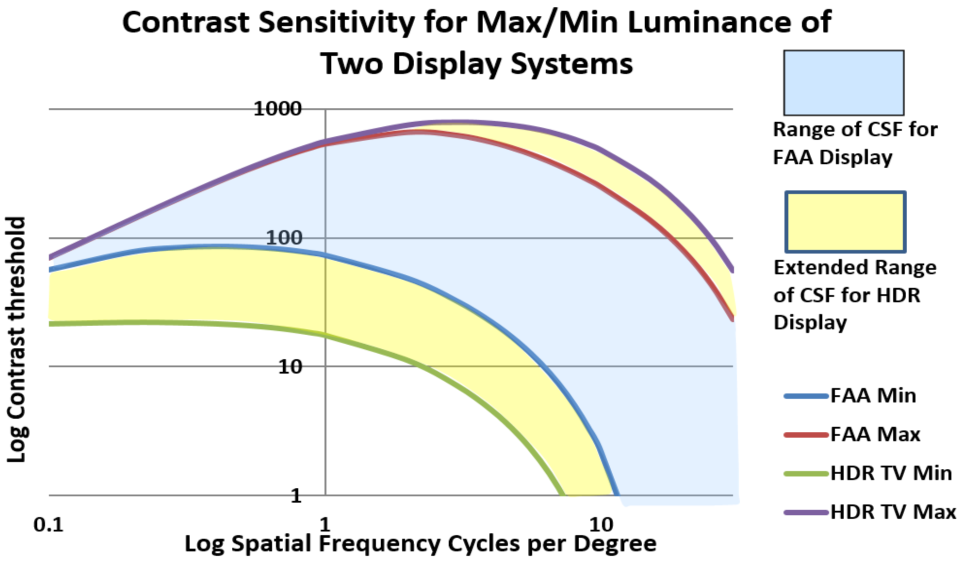

Figure 2 Contrast Sensitivity for Max/Min Luminance of Two Display Systems

In a training environment, negative training may result if we try to use such a display to simulate luminance levels beyond what it is capable of, since the performance of the trainee's human visual system will either be too good or too poor in the simulator relative to the real world conditions we are trying to simulate.

The contrast sensitivity data used in figure 2 was derived using Barten's model, which describes the ability of the eye to detect the contrast modulation of sinusoidal gratings in a state of adaptation. It can only approximate the response in an observer viewing a more complex scene and in scenarios where adaptation time is limited.

Predicting Human Performance as a Function of Stimulus Intensity

Another factor which will be greatly influenced by the dynamic range of the display will be response time. Decision making that is dependent upon retinal processing is highly sensitive to luminance since response time is determined by Pieron's Law, figure 3, which says that response time is inversely proportional to intensity of the stimuli.

In order to explore the potential of high dynamic range to expand the envelope of Warfighter training, we elected to try to isolate luminance as an independent variable. This is because the luminance of current visual training display systems is far below that which is considered in the referenced studies and industry standards to be high dynamic range. Achieving significantly greater luminance, if it is required, is therefore likely to greatly impact the design of new simulators and retrofit of existing ones if they are to benefit from high dynamic range.

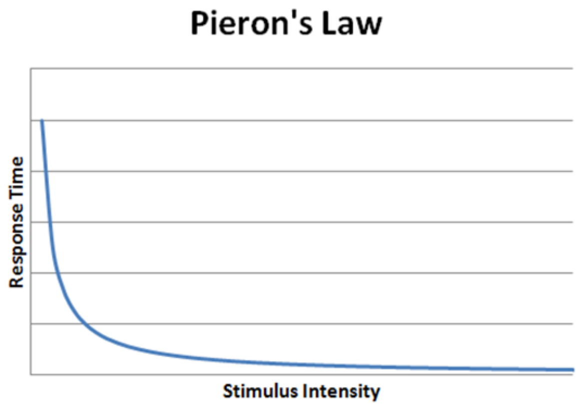

Figure 3 Pieron's Law for Reaction Times

The experiments isolated the luminance variable by using a display which is capable of very high luminance and having a fixed dynamic range of 10,000:1, which has been shown to be at the limit at which adaptation is possible.

The luminance of this display was then controlled by the addition of neutral density filters which altered the luminance optically by varying degrees without affecting the resolution, color or dynamic range of the display. In order to demonstrate the dependency of response time upon luminance, the stimuli presentation time was also varied at each luminance level. The specific luminance values were selected to cover a range of values which included that which is required for certification as "UHD Premium" by the UHD Alliance (1000 cd/m^2), down to a luminance more typical of standard dynamic range consumer electronic displays.

EXPERIMENT

Methodology

In order to explore our research questions 3 Levels of luminance were selected for this research study. A between subject's design was chosen to eliminate any learning effects. A series of pre-planned t-tests were aligned with each set of stimuli and luminance groups before participants were ran. Although t-tests were conducted our team would suggest the use of effect size (power of significance) is the more appropriate measure. The chosen method of analysis for power in addition to creating easy to understand visuals was the used of Cohen's *d*. Cohen's *d* was selected because it easy to understand, measures the mean against a comparison mean, is sensitive to standard deviations, is a widely used measure, and adequately sheds light on the strength of an effect.

Participants

An N of 90 was chosen for this study to allow for appropriate calculation of effect sizes. Each light condition was experienced by 30 participants (little n = 30). Each participant experienced only 1 light condition. Participants were selected and allowed to participate if they had no self-reported visual problems and had 20/20 vision or 20/20 corrected vision as determined by a Snellen Eye Chart test. Participants were faculty and staff at Full Sail University and were compensated for their time with a gift card or extra credit.

Apparatus

Super Lab 5.0 was used to present all stimuli. Super Lab 5.0 was loaded on to a high end gaming computer with a multiple output video card providing full RGB resolution and capable of rapidly and accurately presenting the stimuli. Super Lab 5.0 allowed us to effectively and accurately apply classic signal detection theory for this image presentation experiment. Super Lab 5.0 allows keyboard keys to be assigned to targets and non-targets. Participant input can be computed to hits and misses for targets (threats in our experiment) and correct rejections and false alarms for non-targets (non-threats in our experiment).

The projector was a JVC native 4K SH-7 projector tuned to maximum luminance along with brand new projector bulbs installed just before data collection to output 4000 lumens. This particular projector had a native sequential contrast ratio of 12,000:1. A high gain projector screen with 2.4 gain and 80 degree horizontal half-angle made by MicroliteScreen was chosen to minimize loss of contrast associated with low gain (lambertian) screens. A standard office chair was situated approximately 4 feet from the center of the screen. The participant sat in the office chair and the chair was adjusted so the head of participant was sufficiently lined up with the lens of the projector but far enough way not to feel any heat from the projector. Luminance was controlled by using high quality filters that reduce the amount of light reaching the screen. We choose ¼ inch glass ND filters of optical density .3 for use in this experiment. 0 filters were used for light condition 1, 2 filters were used for light condition 2, and 3 filters were used for light condition 3. The resulting filter combinations produced luminance ratios for each element in the test images of 1, ¼ and 1/8 respectively. The resulting peak luminance of the displayed images for each of three filter combinations was 2400, 600 and 300 cd/m^2 respectively. These luminance values were selected to bracket the UHD Alliance's requirement for consumer televisions meeting the specification for the HDR Premium rating, as described in a paragraph above.

Stimuli

Three sets of two images each, (one a threat image, and one a non-threat image), were to be presented at these 3 different luminance levels. The use of optical filters to control luminance would ensure that, under controlled ambient illumination conditions, contrast and signal-to-noise ratios were the same at each luminance. Subject matter experts with multiple combat tours were selected to make and consult on the 3 sets of images that vary only in a small component of each set for threat/no-threat signal detection purposes. This small variation would allow each image in a set to be a threat or non-threat while holding all other aspects of the image set constant. We asked the subject matter expert to select real images that were taken in low light situations (dusk, dawn, or dimly lit streets) as this is where signal detection is quite difficult for humans. Image set 1 was a dark beach scene with a man sitting with a shovel (non-threat) and the same man sitting with an AK-47 assault rifle (threat), (figure 4). Image set 2 was dark street scene with a man holding an AK-47 assault rifle (threat) and the same man holding a selfie stick (non-threat), (figure 5). Image set 3 was a dark alley scene with a man in cold weather clothing and mask kneeling down and holding a AK-47 assault rifle (threat) or petting a dog (non-threat) (figure 6). (Figures 4, 5 and 6 have had artificially enhanced luminance for reproduction in this paper). Each image was a 4K image and was tested in Super Lab 5.0 to ensure the image could quickly be presented to the participant. Figure 6.5 (this exemplar and reduced resolution image contains the entire scene of image set 3, including 1 of the manipulations shown in figure 6) attempts to show the complexity and color in the images used for the experiment but is much smaller and less detailed than the actual image. Likewise, the images used in image set 1 and 2 were much more complex than can be shown in this paper. The average presentation time was several milliseconds and was well within any signal detection experiment tolerance. Zoomed in examples of each variation for an image set are shown below. Note that these images have been enhanced for this publication to make them easier to see in print form. . The actual images are too large and too dark to be accurately presented in this paper but can be requested from the authors.

Figure 4: Enhanced example of variation in image set 1 (shovel vs. rifle)

Figure 5: Enhanced example of variation in image set 2 (dog vs. rifle)

Figure 6: Enhanced example of variation in image set 3 (selfie stick vs. rifle),

Figure 6.5: Smaller version of the actual image with enhanced variations shown in figure 6 (image set 3).

Design

Participants were randomly assigned to one of the three luminance levels before arriving. After signing an informed consent as required by our IRB the participant was thanked for being a part of this research. The participant was told the total time of this research would be approximately 10 minutes. The participant was then told he/she will be shown 3 images sets of dark scenes on a JVC projector at 1 luminance level. After sitting in the chair the participant was trained using the experimental stimuli and projector to know the variation of each image and whether that image in each set was a threat or non- threat. The participant could ask questions at any time during training and was allowed to study each image as long as they wanted to. The participant was told to respond on the keyboard with a 1 if the image is a threat, 0 if it is a non-threat image or space bar if they didn't know. The participant was told they would have 3 seconds to make a choice after the image disappeared and was replaced by a blank neutral gray screen. After 3 seconds the next image would appear on the screen and they would again have an additional 3 seconds after the image disappeared to make the next selection. This continued until each image was presented 18 times (6 times for 500 ms, 6 times for 1250 ms, and 6 times for 2000ms) in a random fashion. After 108 images were presented the participant was able to ask questions about the experiment and were compensated for their time before leaving.

RESULTS

Only a selection of the results have been presented for this paper due to restrictions on length. Cohen's d effects < .2 are considered small, .2 to .5 are considered moderate/medium in strength, and effect sizes >.5 are considered strong. Hits for a threat suggest the participant was correct in labeling the threat a threat, while correct rejections for non-threats suggest the participant was correct in selecting the non-threat a non-threat. A miss for a threat is when the participant labels a threat a non-threat, and a false-alarm for a non-threat is when the participant selects a non-threat as a threat. In summary, hits and correct rejections suggest good performance, and misses and false alarms suggest poor performance. Not-sures and no-responses also suggest poor performance. Figures 7, 8,9,10, 11 & 12 use the following color convention: green means good performance by the participants, all other colors are suggesting some level of poor performance.

Light condition 1 – No ND Filters

The brightest light condition threat images had the following combined Hit %, Miss %, Not Sure % and No Response % (see figure 7). The brightest light condition non-threat images had the following Correct Rejection %, False Alarm %, Not Sure %, and No Response % (see figure 8).

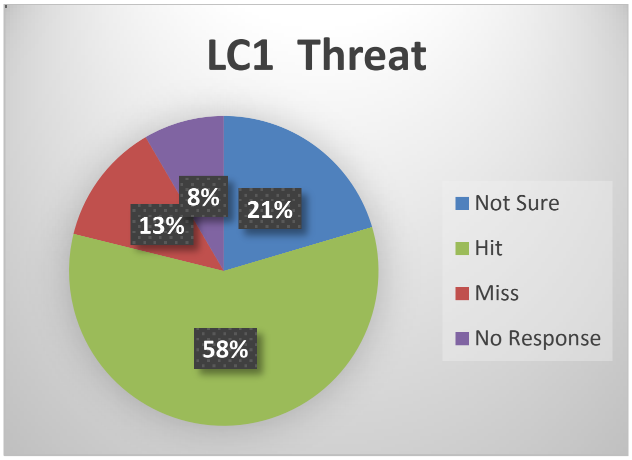

Figure 7: Light Condition 1 Threat Images (all 3 sets averaged together).

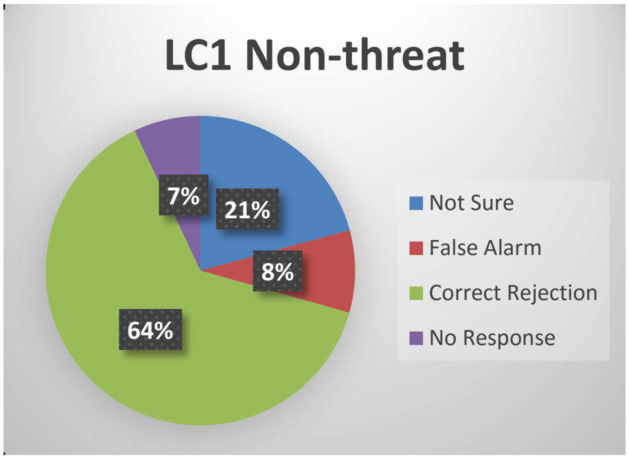

Figure 8: Light Condition 1 Non-Threat Images (all 3 sets averaged together).

Light condition 2 – 2 ND Filters

Light condition 2 threat images had the following combined Hit %, Miss %, Not Sure % and No Response % (see figure 9). As shown below in figure 9 when compared to figure 7, going from Light condition 1 (no filters) to Light condition 2 (2 filters), a 17% decrease in Hit Rate was found and a 12% increase in Miss Rate was found. The pre-planned independent-samples t-test was significant for Light condition 1 Hit Rate (58%) and Light condition 2 Hit Rate (41%): t (48) = .00, p < .005. Cohen's *d* showed a moderate effect size of .401 which suggest a medium strength effect.

Light condition 2 non-threat images had the following Correct Rejection %, False Alarm %, Not Sure %, and No Response % (see figure 10). As shown below in figure 10 when compared to figure 8, going from Light condition 1 (no filters) to Light condition 2 (2 filters), a 12% decrease in Correct Rejection was found and a 6% increase in False Alarms was found. The pre-planned independent-samples t-test was significant for Light condition 1 Correct Rejection Rate (64%) and Light condition 2 Correct Rejection Rate 52%): t (48) = .00, p < .005. Cohen's *d* showed a moderate effect size of .288 which suggest a medium strength effect.

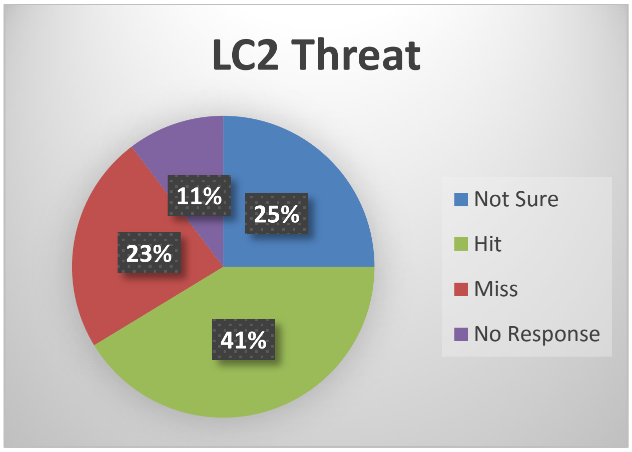

Figure 9: Light Condition 2 Threat Images (all 3 sets averaged together).

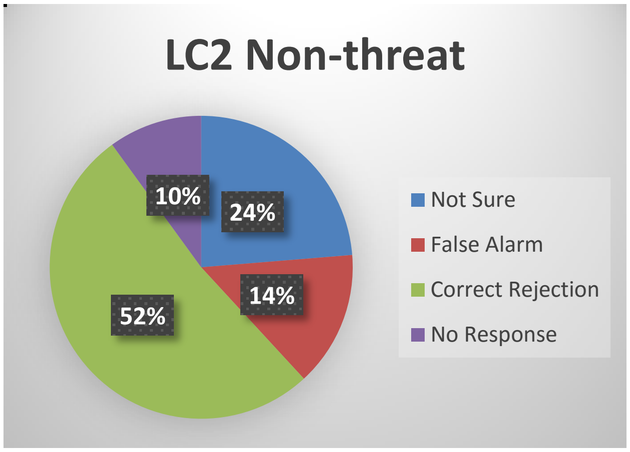

Figure 10: Light condition 2 Non-Threat images (all 3 sets averaged together).

Light condition 3 – 3 ND Filters

Light condition 3 threat images had the following combined Hit %, Miss %, Not Sure % and No Response % (see figure 11). As shown below in figure 11 when compared to figure 7, going from Light condition 1 (no filters) to Light condition 3 (3 filters), a 28% decrease in Hit Rate was found and a 13% increase in Miss Rate was found. The pre-planned independent-samples t-test was significant for Light condition 1 Hit Rate (58%) and Light condition 3 Hit Rate (30%): t (48) = .00, p < .005. Cohen's *d* showed an effect size of .696 which suggest a strong effect.
Light condition 3 non-threat images had the following Correct Rejection %, False Alarm %, Not Sure %, and No Response % (see figure 12). As shown below in figure 12 when compared to figure 8, going from Light condition 1 (no filters) to Light condition 3 (3 filters), a 17% decrease in Correct Rejection was found. The pre-planned independent-samples t-test was significant for Light condition 1 Correct Rejection Rate (64%) and Light condition 3 Correct Rejection Rate 47%): t (48) = .00, p < .005. Cohen's *d* showed a moderate effect size of .405 which suggest a medium strength effect.

Light condition 2 when compared to light condition 3 for both hits and correct rejections show as significant independent samples t-test and medium Cohen's d for LC1 Hit to LC2 Hit (Cohen's *d* = .280) and small effect for correct rejections (Cohen's *d* = .117).

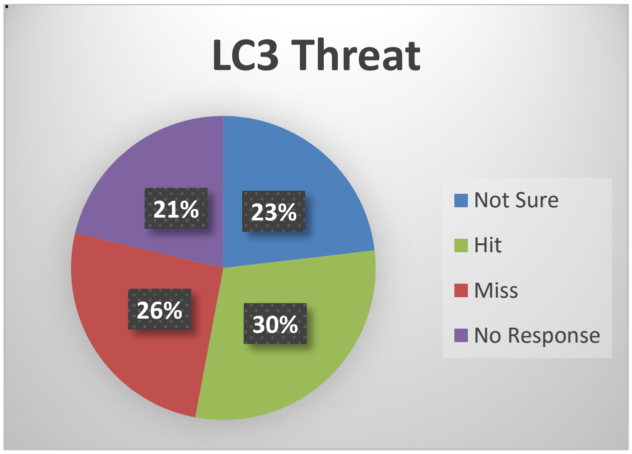

Figure 11: Light Condition 3 Threat Images (all 3 sets averaged together).

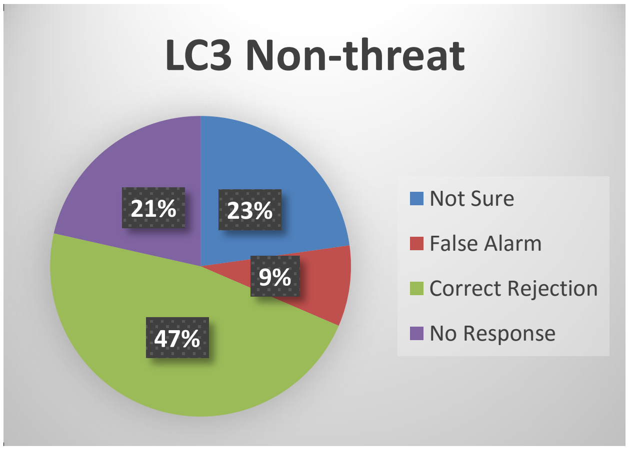

Figure 12: Light Condition 4 Non-Threat Images (all 3 sets averaged together).

DISCUSSION

As hypothesized from the analysis of CSF Vs. frequency comparing standard and high dynamic range display systems, (fig 1), our tests have shown that the human visual system is very sensitive to luminance for fixed contrast and fixed dynamic range images in low light situations. Low light signal detection situations using digital imagery and projection are common across a wide variety of training domains which include law enforcement, military and homeland security. Furthermore this relationship holds true for realistic images which are complex and include very bright areas even in close proximity to the low light target areas.

We also noticed the lack of an increase in false alarms regardless of light conditions. This may reflect a bias in participants to not over react by labeling non-threats as threats and instead opt for increases in not-sure and no response choices. Not responding or not categorizing non-threats may in fact seem like the lesser of two-evils when given the alternative of saying non-threats are threats. While this idea has some merit we would suggest that inaction in law enforcement, homeland security and the military can be extremely hazardous and lead to poor decisions making as the complexity of real world events unfold. Not knowing is never a good situation and accurate knowledge of threat and non-threat stimuli is the only way to have a complete picture of operational situation.

Our team acknowledges that a small part of the overall visual complexity involving human performance and HDR is explored with our experimentation and that more experiments are warranted. However, it is apparent that the displays used for most simulation and training applications are very limited in peak luminance and dynamic range relative to the real world scenes that we are trying to simulate with them. Since we have demonstrated the high degree of sensitivity of the eye to small changes in brightness when observers are performing tactically significant tasks, we see the need for more research into increased luminance accuracy in the displays that would be used to train such tasks. Emerging consumer electronic displays and the infrastructure and industry standards that support them appear to have the potential to satisfy this need and should be studied further for potential application to Warfighter training. We hope that we have shed some light on the performance gains that can be generated by carefully selecting and applying such displays for various training and surveillance applications in this dangerous 21st century.

FUTURE RESEARCH

To fully understand the complexity of human visual performance in signal detection experiment settings we suggest a variety of manipulations should be explored including but not limited to varying color depth, distance to the screen, environmental/ergonomic viewing conditions, dynamic video (vs. static imagery) and settings. Because our military and law enforcement personnel rarely complete one task at a time the loading and assignment of multiple tasking should also be studied (having more than one task to complete). Further analysis of this data is also warranted. Our team plans to explore our data set for effects in presentation times, reaction times and light conditions.

REFERENCES

Blackham, et al, (1999). Image Display Apparatus with Modulators for Modulating Picture Elements in an Image. United *States Patent Office (USPO),* United States Patent 5,978,142.

Seetzen, H. et al. (2004). High Dynamic Range Display Systems. *ACM Transactions on Graphics* (TOG), *Volume 23 Issue 3*, pages 760-768.

Barten, P. (1992). Physical Model for the Contrast Sensitivity of the Human Eye. *Proceedings of SPIE, Vol. 1666*, pages 57 -72.

Standard Reference. (2014). High Dynamic Range Electro Optical Transfer Function of Mastering Reference Displays, *Society of Motion Picture and Television Engineers, SMPTE Standard, ST 2084:2014*

Kunkel, T. & Reinhard, E. (2010). A Reassessment of the Simultaneous Dynamic Range of the Human Visual System, *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pages 17-24.

Press Release. (2016). UHD Alliance Defines Premium Home Entertainment Experience, *UHD Alliance,* Retrieved Aug 19, 2016, from http://www.uhdalliance.org/uhd-alliance-press-releasejanuary-4-2016/#more-1227