Adapting Flight Training Device Visual System Testing Methods to Extended Reality Near-Eye Displays

Benito Graniela
NAWCTSD
Orlando, FL
benito.graniela@navy.mil

Robert Calvillo
NAWCAD
Patuxent River, MD
robert.calvillo@navy.mil

ABSTRACT

The popularity and advancing capabilities of commercial head mounted displays (HMD) has opened the opportunity for the innovative application of virtual reality (VR), augmented reality (AR), and mixed reality (MR), aka extended reality (XR) devices to support Naval aviation training applications. In recent years, the US military has been looking at introducing new training methods to speed up pilot training by using a combination of XR, artificial intelligence, and modeling and simulation technologies within a new airman-training curriculum. In theory, the near-eye displays (NED) in XR headsets are just another type of display device available for flight simulators. However, even though HMDs share similar components with traditional flight simulator displays, there are significant differences in scale and performance of these components. As a result, traditional visual display system (VDS) test equipment cannot be used to measure the performance of a fully integrated XR display system. This paper provides background on conventional Navy VDS requirements and testing methods and presents a set of preliminary HMD parameters and methodology for the evaluation of this technology within a flight simulator. This paper also presents information on the initial implementation of VR methods as well as data collected (e.g., field of view (FOV), static resolution, and latency) for several VR headsets. Finally, the paper concludes with future research plans for the development of requirements and methodology for the evaluation of a MR HMD within a flight simulator.

ABOUT THE AUTHORS

Benito Graniela is currently a subject matter expert in visual systems at the Naval Aviation Warfare Center Training Systems Division (NAWCTSD). For 10+ years, Dr. Graniela has supported the acquisition of flight training devices, generated multiple SBIR topics and served as technical point of contact, and primary investigator of several Naval Innovated Science and Engineering (NISE) topics. Dr. Graniela has 15+ years of experience in Modeling and Simulation (M&S) with emphasis in virtual and constructive distributed synthetic environments, 20+ years of experience in the design and development of software applications, and 10+ years in the design and testing of military electro-optical systems. Dr. Graniela holds a BSE in Electrical Engineering from the University of Puerto Rico at Mayaguez, a MSE in Computer Engineering from the UCF and a PhD in M&S from UCF.

Robert Calvillo received a Bachelor of Science degree in Mechanical Engineering in 2008 and a Master of Engineering degree in Aerospace Engineering in 2013; both from Old Dominion University. After finishing undergrad, Robert started working for Naval Facilities Engineering Command while working on his Master’s. After grad school, Robert went to work for Naval Air Systems Command where he currently works in the Flight Vehicle Modeling and Simulation (FVMS) branch. Within FVMS, Robert divides his time between M&S of flight control systems and integrating mixed reality technologies with Navy flight simulators.

Adapting Flight Training Device Visual System Testing Methods to Extended Reality Near-Eye Displays

Benito Graniela
NAWCTSD
Orlando, FL
benito.graniela.civ@us.navy.mil

Robert Calvillo
NAWCAD
Patuxent River, MD
robert.calvillo@navy.mil

INTRODUCTION

The popularity and advancing capabilities of commercial head mounted displays (HMD) has opened the opportunity for the innovative application of virtual reality (VR), augmented reality (AR), and mixed reality (MR), aka extended reality (XR) devices to Naval aviation training applications. HMDs have been around for a long time (e.g., 48 years) but recent advances in display panel technology and graphics hardware have improved performance and created interest for many applications previously only possible in laboratories. Although advancements in HMD technology have increased field of view (FOV) and static resolution, reduced latency, and improved comfort, other areas such as vergence-accommodation conflict, simulator sickness, and cost remain major challenges for the adoption of HMD technology (Biggs et al., 2018).

One of the US Naval aviation training applications for XR headsets is flight training device (FTD) displays. It is well known that the traditional use of flight simulators in lieu of the aircraft results in safer flight training, cost reductions in operations, and benefits in fuel conservation. FTD requirements and evaluation procedures are implemented to ensure the device complies with the requirements, which validates its use within a training program. As such, the simulator needs to duplicate a specific aircraft cockpit, and closely represent the actual aircraft through various ground and flight regimens. From a visual systems point of view (e.g., visual database, IG, display), a military FTD must pass a set of validation tests, which demonstrate that the device compiles with a set of functional and performance requirements. The FTD XR requirements and components necessary to comply with FTD application requirements is under active investigation by the US Navy.

In theory, the near-eye displays (NED) in XR headsets are just another type of display device available for training devices. However, even though NED share similar components with traditional high-end FTD displays, there are significant differences in scale and performance of these components. Traditional high-end FTD displays include real image and collimated displays (with a screen or mirror radius in the order of 10 feet) that span a FOV larger than 180 degrees in the horizontal and 60 degrees in the vertical dimension. Differences in technology and design between high-end FTD displays and XR headsets prevents the use of traditional FTD visual system (VS) metrics, procedures, and equipment.

For the purposes of this paper, XR will be used to describe the whole range of VR, AR, and MR technologies (Milgram, Takemura, Utsumi, & Kishino, 1994). VR technology uses an HMD to completely replace a user's view of the real world with visuals of a 3-dimensional computer generated environment. AR refers to overlaying computer generated objects on top of real world visuals (Milgram & Kishino, 1994). MR is defined as an environment "in which real world and virtual world objects are presented together within a single display, that is, anywhere between the extrema of the RV continuum" (Milgram et al., 1994) where RV is the reality-virtuality continuum which includes many different classes of displays. This means a user can interact with both the real world and virtual objects at the same time. This interaction in both AR and MR HMDs can be provided using optical or video pass through displays to show the real world to the user. Optical pass through HMDs, such as the Hololens, overlay virtual objects on top of real world views through the use of semi-transparent displays. Video pass through HMDs, such as the Varjo XR-3, combine a VR HMD with external cameras to display video of the real world inside the HMD. Image processing software is then used to combine the virtual objects with the video feed from the two cameras.

The authors consider that VR and MR are the two main XR technologies that have the most potential for use in FTD within the Navy. The following section provides background on the FTD types and the important parameters and techniques that characterize performance (i.e., Previous Work) in traditional display systems. Then the paper moves on to describe some of the work that has been done to adapt these FTD visual system metrics and procedures to XR headsets (i.e., Adapted Metrics and Procedures). After that, the paper describes other efforts at Naval Air Warfare Center Training Systems Division (NAWCTSD) and Naval Air Warfare Center Aircraft Division (NAWCAD) which are looking at using XR technology (i.e., Collaboration). The paper concludes with a section on future work.

PREVIOUS WORK

An FTD needs to replicate to some specified degree the aircraft cockpit, the real world, and the interaction between the two. The display equipment, regardless of complexity or cost, is generally incapable of replicating the real world. The imagery is always less than faithful by way of FOV, brightness, resolution, or some other image quality characteristics. However, if the quantity of cues and overall fidelity of the scene is sufficient, then the human visual system (HVS) will fill in the additional information necessary for providing the pilot the feeling of being in the real world. In this section, we will first provide background on high-fidelity FTD visual system types and important parameters, requirements, and metrics.

Traditional High-Fidelity FTD Display Systems

FTDs have employed a diverse variety of visual displays to support their training tasks. Advancements in technology, evolving requirements, and training objectives have resulted in different display solutions for the FTD. These display solutions can be categorized into four main types: dome display, cross-cockpit collimated display, rear projection mosaic display, and HMD (Joseph, 2002). The characteristics to consider in the display's ability to support training tasks include but are not limited to: FOV, resolution, luminance (i.e., brightness), contrast, collimation, and distortions. FTD display types with multiple viewpoints (i.e., pilot and co-pilot) can be classified based on scene perspective errors and FOV limits (Joseph, 2002). Scene perspective errors are encountered in real image displays because the displayed images are calculated for one viewer's perspective (e.g., pilot eyepoint) but are viewed from another location in the cockpit (e.g., copilot eyepoint). Field of view determines the visibility of visual cues out the window (OTW), and may limit the ability of the crewmembers to perform training tasks. The comparison of the display types to XR HMDs will be limited to these features.

Real Image Dome Displays

This type of display is characterized by a large dome (e.g., 10+ ft. radius) located around the aircraft cockpit. A real image is projected on the inside surface of the dome using multiple projectors. These images are blended together to provide a continuous image. The key advantage of the dome display is a large continuous field of view. The main disadvantage of the dome is that the image is not collimated. This means that the image can only be presented in proper perspective from a single location in the cockpit. For aircraft with pilot and co-pilot configurations, if the image is computed for the pilot's perspective, then it will appear distorted from the copilot's perspective.

Cross-Cockpit Collimated Displays

Cross-cockpit (i.e., collimated) displays share most of the characteristics of real image dome displays with the exception that instead of a real image a virtual image is displayed. These collimated light displays provide a virtual image which provides a true perspective for both the pilot and copilot simultaneously (Joseph, 2002). Limitations due to geometry of the mirror and screen have traditionally limited the vertical FOV coverage (e.g., ~60 degrees).

Real Image Rear Projection Mosaic Displays

This type of display consists of an array of rear projection panels, around the cockpit. This type of display provides a smaller footprint than dome and collimated image displays. This display is capable of projecting a very high resolution image to the viewer using inexpensive projector technology. Due to the small size of the arrangement, the display is only suited for a single viewpoint.

Head Mounted Displays

HMDs consist of a near eye display (NED) with VR or MR capabilities. MR HMDs can use optical pass through optics or video pass through cameras to provide a collimated stereoscopic image of the virtual world combined with the real world to the viewer. HMDs use a head tracking device to match the visual scene position and orientation to the viewing direction and therefore provide the proper perspective for all crewmembers (e.g., pilot and co-pilot). Head tracking techniques includes external devices or sensors that track the headset and accessories to determine position and orientation (outside-in) and more recently sensors on the headset (inside-out) look out to determine how positon and orientation changes in relation to the environment. HMDs also provide an unrestricted FOV with head movement or field of regard (FOR) to the crewmembers. In MR applications a real world model of the cockpit may be used to provide high fidelity interactions with buttons and controls and provide virtual out-the-window (OTW) views through virtual cockpit windows. In this configuration MR FTD require very little space beyond the cockpit yet they provide unlimited FOR around the viewer.

FTD DISPLAY PARAMETERS, METRICS AND PROCEDURES

This section describes at a high level some of the most important parameters, metrics and measurement procedures for traditional high-resolution FTD displays. The scope of this paper is limited to the most important parameters which are field of view (FOV), static resolution, dynamic resolution, latency, luminance, contrast, color, and geometric distortions (Joseph, 2002).

Field Of View

Field of view (FOV) is the extent of the observable world as seen in any given moment. FOV requirements vary greatly with the application and are based on training requirements, aircraft, technology limitations and cost. Figure 1 shows a histogram of Navy and USMC FTD fielded by the NAWCTSD Visual and Sensor Simulation branch as of October 2019. The data includes a total number of 81 fixed wing aircraft (FWA) and 57 rotary wing aircraft (RWA). The data shows that FTD FOV vary from 150 to 360 horizontal FOV (HFOV) and from 50 to 135 vertical FOV (VFOV). Another data point on FTD FOV comes from FAA requirements. FAA FOV requirements range from 45° x 30° (level AB) to 176° x 36° (level CD) for fix wing aircraft and 75° x 30° (level B) to 176° x 56° (level D) for helicopters.

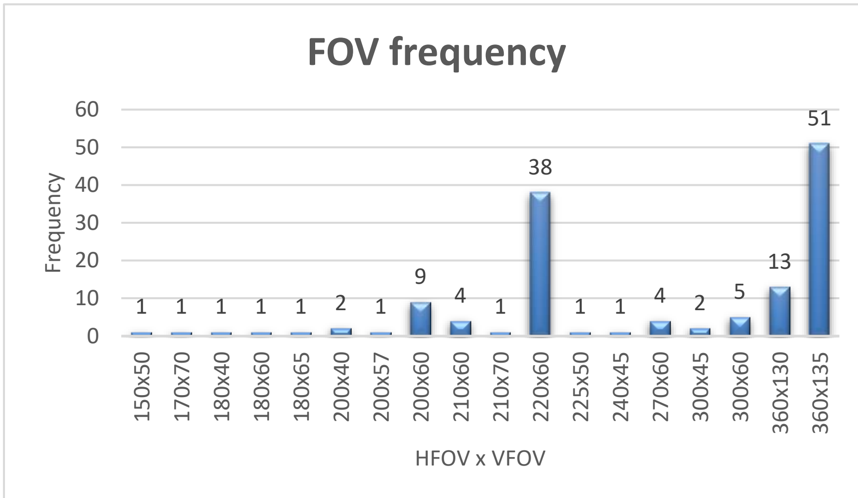


Figure 1- NAWCTSD fielded FOV extent histogram

The FOV metric is an angular measurement (i.e., degrees) which is traditionally measured using a theodolite or other device capable of measuring angles. Procedures for measuring FOV normally include measuring the display sustained angle along the horizontal axis. Vertical FOV is measured up and down (i.e., +/- elevation) with respect to the horizontal zero degree line. For obvious reasons, the use of a device such as a theodolite will not work for measuring FOV on HMDs.

Spatial Resolution

Spatial resolution is a measure of the smallest object (in arcminutes/optical line pair) that can be resolved by the visual system. Spatial resolution requirements vary from platform to platform and are based on training requirements, technology limitations, and cost. Figure 2 shows that the Navy and USMC FTD static resolution requirements fall in the 3.5 to 9 arcminute per optical line pair (arcmin/OLP) range. It is interesting to note the discontinuity in the graph at both extremes. The curve from AH-1W to C-2 is a smooth curve that includes 62% of all the FTD where resolution is between 4.6 and 6 arcmin/OLP. FAA level requirements for surface resolution are around 4, 8, and 10 arcmin/OLP (NSP 2018).

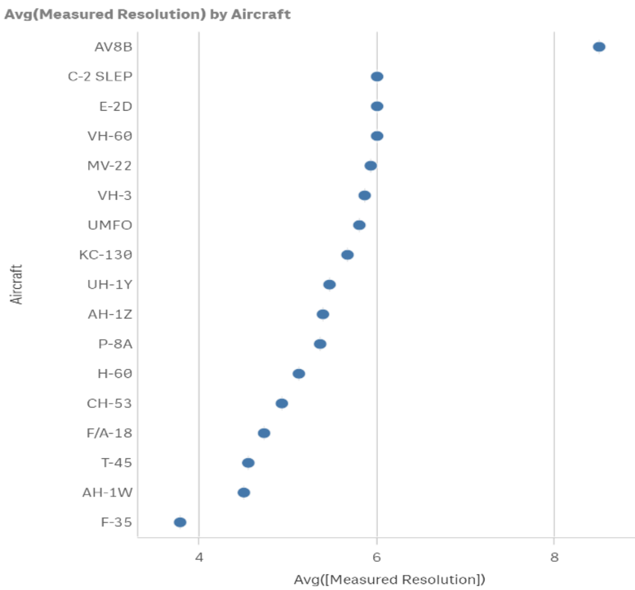


Figure 2 - NAWCTSD fielded visual system static spatial resolution as of October 2019

Spatial resolution is normally measured using a line pair test pattern (i.e., optical line pairs or bars) where the pattern distance to the viewer is increased until the line pair modulation is barely discernable. The test pattern is generated by the image generator using similar rendering settings as those used during normal FTD simulator operation. The procedure normally includes measuring static resolution on each one of the visual system display channels. Similar techniques can be used in a virtual setting to measure static resolution in HMDs.

Luminance

Luminance is the measurement of choice for measuring the luminous intensity per unit area of light in FTD. A valuable characteristic of luminance is that its measured surface value is optically invariant and independent of the measurement distance. The measurement of luminance is typically reported in units of the foot Lambert (fL) and corresponds to emission of one lumen of optical flux (power) per square foot from a perfectly diffuse radiator with a reflectivity of $p = 1$. Navy FTD requirements for minimum surface luminance range from 7 to 12 fL while FAA requirements vary from no less than 4.1fL (Level 7 FTD) to 5.8 fL (full flight simulator - FFS). Display luminance metrics may include average, maximum, minimum, dynamic range, and uniformity. Luminance should be measured on a white raster using a 1° spot photometer or luminance meter. Luminance is traditionally measured with a spot photometer but other computer vision methods (e.g., calibrated camera) may be used.

Unfortunately, the physical size and optical characteristics of most HMDs prevent the use of traditional light measuring devices such as spot photometer and most spectrometer systems. In particular, the collection optics typically used with spectrometers have a relatively large entrance pupil (e.g., 20 mm to 40 mm) and cannot focus at the HMD display distances. If the exit pupil of the source under test (i.e., HMD) is smaller than (under fills) the entrance pupil of the light measurement device (e.g., photometer and spectrometer), then there will be significant errors in the absolute accuracy of the measurements (Austin, 2018). Specialized equipment is necessary to allow for measuring luminance in HMDs. Furthermore the alignment of the exit pupil of the source under test and the entrance pupil of the light measurement device is critical. For obvious reasons, alignment of a measuring devices of less than 5 mm requires specialized equipment (Austin 2018).

Contrast

Contrast is a measure of the dynamic range of luminance provided by a visual display system. Images with low contrast appear washed out, have reduced color intensities, and have reduced visual resolution. There are many different types of contrast measurements including dynamic contrast, sequential contrast, and checkerboard contrast. For Navy visual systems, contrast is normally measured using the American National Standards Institute (ANSI) contrast test pattern image which consists of black and white squares. The luminance meter is placed at the display system viewing volume and is pointed at the test pattern. Luminance is measured at each white and black square. Then the corresponding average luminance for all white and all black square is computed. The ANSI contrast is the ratio of the average white and black squares. Contrast requirements vary significantly for different display types (e.g., real image and collimated) and FOV. Navy ANSI contrast ratios specifications range from 8:1 to 40:1 on real image displays, and from 5:1 to 12:1 on collimated displays. Whereas, the FAA surface contrast ratio requirement is no less than 5:1. Measurement device limitations similar to luminance also apply to contrast measurements as well. Furthermore, the measurement device not only needs to fit within the HMD eyebox (Austin 2018) but it must also have the ability to point to different locations on the display's instantaneous FOV to collect light within the black and white squares of the test pattern.

Geometric Distortion

Geometric distortion refers to the ability of the display to accurately portray shapes and positions of objects in an image. The FTD display system must be geometrically aligned to compensate for distortions that arise from the inherent optical system design, optical system component variances, and visual system integration. Local non linearity of the displayed images can affect the pilot's ability to accurately judge distance, speed, orientation and height. Navy requirements include two types; absolute geometry error which is error in azimuth and elevation from the computed eyepoint and the rate-of-change error for adjacent areas (Long 2010). Navy requirements for absolute geometry vary from 0.5° to 1.5° for real image displays and from 1° to 3.0° on collimated displays. FAA requirements range from for a 5° even angular spacing is +/- 1° from either pilot eye (full flight simulator - FFS) to" no distracting discontinuities" for FTD Level 7. Absolute geometry is measured using a test pattern filling the entire visual scene with lines or points which define a grid of 5° squares. Absolute geometry is generally measured by using a theodolite. Angular measurements (i.e., azimuth and elevation) are taken at various locations on the displayed grid test pattern. Azimuth and elevation of the grid points are measured and compared to the modeled grid-point locations as defined by the test pattern. Any angular deviation from the test pattern's correct location represents a geometric error. Traditional measurement equipment (e.g., theodolite) is typically too large to fit within the available space of a HMD.

Latency

Latency is the delay between input stimuli and the perception of the change in the visual display system. Significant latency can produce perceptual confusion, nausea, and disorientation (i.e., simulation sickness) when viewing a simulated OTW display (So & Griffin, 1992). The Navy requirements for a FTD running at 60 Hz update rate system is normally 100 milliseconds (msec), while the FAA requirements range from 100 to 120 msec (NSP, 2018). Latency is normally measured in the FTD by instrumenting the system and measuring the delay (NSP 2016) in response from the flight deck control input to the reaction of the instruments, motion systems, and visual systems. The effects of latency in XR headsets are far more severe than latency in FTD. So & Griffin (1992) found that lags on the order of 80-100 msec between head movements and corresponding shift of displayed imagery produced sensory conflict which leads to simulation sickness. Padmos and Milders (1992) suggest that for immersive displays, the latency of update should be less than 40 msec whereas Keller and Colucci (1998), and Winterbottom (2005) recommended that for see-through AR HMDs the lag should be no greater than 16 msec. Jerald (2009) concludes that people react very differently to latency where some people hardly notice a 100 msec delay while other people are able to perceive 3 to 4 msec of latency. At this point in time, the commercial industry is providing leading edge improvements in XR headset technology to reduce latency. Therefore the most recent advances, guidelines and reports on latency are coming from developers of XR equipment rather than scientists (Raaen, 2015). John Carmack (developer of XR equipment) states that a latency of 50 msec feels responsive, but recommends latencies under 20 msec (Raaen, 2015).

ADAPTED METRICS AND PROCEDURES

NAWCTSD realized that measuring display performance would require a set of parameters which could be measured on the integrated FTD XR system (i.e., field tests) as well as a set of parameters that would have to be measured in a laboratory setting (i.e., lab tests). The set of field tests would measure performance on the integrated system and does not required sophisticated or costly laboratory equipment. The field tests would include measurements of important parameters such as FOV, static resolution, and latency, while other important image quality parameters will need to be validated by subjective means. These set of field tests include subjective assessments (i.e., based on human perception) such as luminance and color uniformity, geometric distortions, color convergence, and other distracting artifacts. At this point the other set of parameters (i.e., lab tests: luminance, contrast, and color) require specialized measuring equipment.

The rest of this section presents the work done on the adaptation of FOV, spatial resolution, and latency metrics, and measurements during the field tests. Each section contains information on the developed test patterns, metrics, and procedures as well as some preliminary results.

FOV and Resolution Methodology

This section provides information on the measured FOV and static resolution for three commercially available VR headsets. The data was collected between February and March of 2020. Static resolution was measured at the center of the FOV and the extents of this static resolution were determined along the horizontal and vertical axis. Static resolution was also measured at 75% of the horizontal and vertical FOV limits to obtain an idea of how the resolution degraded across the FOV. An initial attempt to measure static resolution at the extreme corners of the FOV was performed, but degradations due to optics, eye gaze, and other factors required target sizes that were deemed too large. Instead, the collected static resolution data was used to develop a static resolution model, which describes the resolution across the entire FOV.

Field of View Methodology

A wireframe sphere test pattern with lines every 5 degrees (i.e., geometric test pattern) was displayed and then the headset was boresighted to the pattern (i.e., centered at the headset FOV). The observer was asked to provide the maximum viewable FOV azimuth angles (left and right) as well as the maximum elevation angles (top and bottom).

Max Static Resolution Methodology

A set of horizontal and vertical 3 line pair test patterns (i.e., static resolution test pattern also known as 3-bar test pattern) was displayed and boresighted. All static resolution measurements were made using the barely resolvable criteria. The tester increased the distance from the observer eye-point to the test pattern until the observer indicated that the test pattern was just barely resolvable as three line pairs.

Max Static Resolution Extents Methodology

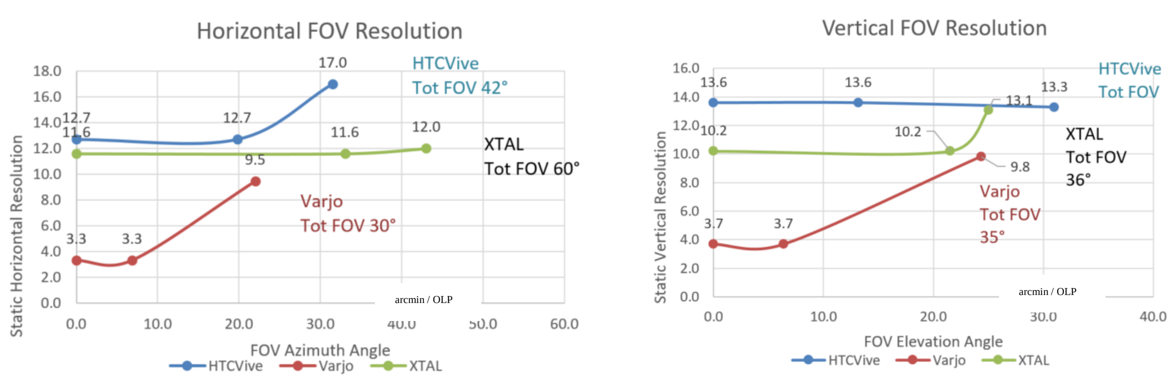
The main purpose of this test was to measure the extents for which the maximum static resolution was valid. First, the static resolution test pattern was boresighted. The previously measured white bar static resolution distance was set. The observer was asked to rotate the head left while keeping an eye on the test pattern and to stop when the three line pairs were no longer resolvable. When the observer reached the max degradation angle the geometric test pattern was turned on and the angle was recorded. The procedure was repeated for right, up, and down head rotations about the origin.

Static resolution at 75% of FOV Methodology

This data point provides information on degradations in static resolution as a function of line-of-sight (LOS). First the resolution test pattern was displayed and boresighted. Then the test pattern was positioned at an angle corresponding to the 75% FOV and static resolution was measured by moving the pattern until it was barely resolvable. This process was repeated left, right, top, and bottom of the boresighted location.

FOV and Static Resolution Results

Figure 3 a) shows the average horizontal static resolution performance across the horizontal FOV for each of the headsets. The Varjo VR-1 headset has the highest static resolution (i.e., lowest number) of all the headsets. However



a) b)

Figure 3 - Average static resolution across the horizontal and vertical FOV

the higher static resolution was only valid over a small horizontal angle (max 7.7 degrees) about the center of the FOV. The XTAL 5K headset provides a relative uniform static resolution (min 11.6 arcmin/OLP) across the FOV. The Vive headset provided the worst resolution across the FOV with the best resolution measured at 12.7 arcmin/OLP over an average of 19.8 degrees. However resolution degrades to 17 arcmin/OLP at 31.5 degrees.

Figure 3 b) shows the average vertical static resolution performance across the vertical FOV for each of the headsets. Again the Varjo VR-1 headset has the highest static resolution of all headsets. However the Varjo VR-1 static resolution at the center of the FOV quickly degraded at angles greater 6.4 degrees. The Vive headset provides the worst resolution across the FOV at 13.6 arcmin/OLP. The XTAL 5K provides a medium static resolution at 10.2 arcmin/OLP with a quick degradation at 25 degrees to 13.1 arcmin/OLP.

The data was processed in MATLAB using a locally weighted non-parametric regression model to generate surface color patterns of static resolutions across the FOV. Figure 4 shows the model predictions for static resolution across the FOV for each tested HMD.

Figure 4 shows that the model plot for the HTC Vive Pro (top) has the worst static resolution and that this resolution is somewhat smoothly distributed across the FOV from the low 13s (green) to the low 20s in arcmin/OLP. The Varjo XR-1 (middle) model plot shows that the static resolution is very good (3.4 arcmin/OLP) at the center 20 degrees and then degrades very rapidly to the high 20s in arcmin/OLP. This makes sense as the Varjo XR-1 uses a 35° azimuth and 20° elevation (35° x 20°) high-resolution insert. The XTAL 5K model plot displays a relative uniform static resolution across the FOV with a range of 10.9 to 13.3 arcmin/OLP.

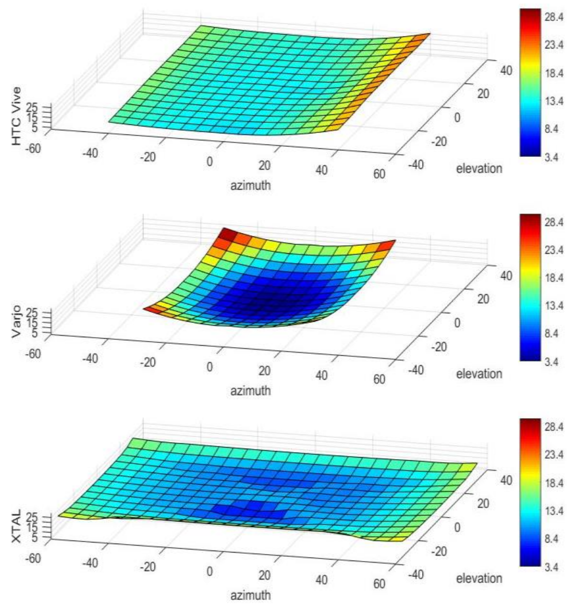
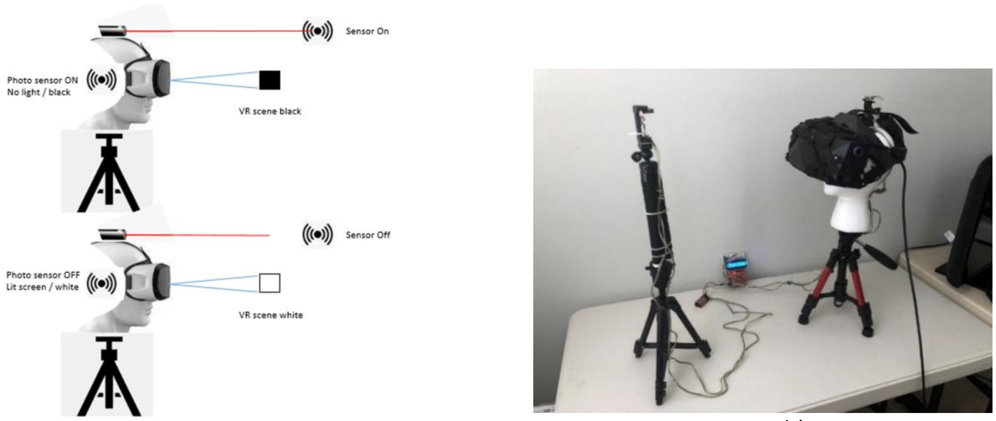


Figure 4 - Static resolution across FOV model

FOV and Static Resolution Discussion
The data shows the HTC Vive Pro has the worst static resolution of all the headsets with its best resolution at 13.1 arcmin/OLP. This resolution was valid for an angle of 29.5° in azimuth and 26.3° in elevation. The maximum FOV provided by the HTC Vive Pro was 84.5° in azimuth and 81.3° in elevation. The Varjo VR-1 provides the best static resolution of all the headsets with the best resolution at 3.5 arcmin/OLP. Unfortunately, this resolution is only valid for an angle out to 13.7° in azimuth and 12.6° in elevation. The maximum FOV provided by the Varjo was the smallest of all the headsets at 60° in azimuth and 70.3° in elevation. The XTAL 5K headset shows that it provides a static resolution of 10.9 arcmin/OLP across a FOV angle of 61° in azimuth and 42° in elevation. The XTAL 5K provided the largest FOV of all the headsets at 120° in azimuth and 67.5° in elevation.

It is important to note that the above data for FOV and static resolution shows differences between what was measured and the commercially available product data. The testing techniques used for measuring FOV and static spatial resolution are consistent with current test methodologies and not just focused on optimum system performance.

VR Latency Methodology
The following section describes the approach taken to measure VR headset latency. Latency is the time it takes between a user triggered action, such as turning the head, and the results are visible on the NED. This test method is based on the work of Raaen and Kjellmo (2015) which closely matches the way latency is measured on traditional visual systems. Figure 5 b) shows the latency measurement apparatus.



a) b)

Figure 5 - Latency measurement device concept and prototype

The physical setup consists of the VR headset mounted on a Styrofoam (Figure 5) head which in turn is mounted on a camera tripod. A light sensor in the Styrofoam head at the eye location monitors changes in the NED screen state (white/black image). A laser transmitter attached to the Styrofoam head and a light sensor on another tripod provides a reference alignment of the VR headset. The laser, laser alignment sensor, and light sensors are connected to an Arduino Uno microcontroller. A Unity 3D application transitions the NED image from black to white (or white to black) when the VR headset is rotated. The measurement procedure starts by mounting the VR headset on the system and aligning the laser with the laser alignment sensor. The Unity 3D application (version 2019.3.6.f1) displays the test pattern on the NED display which stimulates the light sensor in the Styrofoam eye. The microcontroller state at this point is ready to measure the transition of the laser alignment and the test pattern displayed on the NED. Rotation of the head causes the laser light sensor to transition from high to low in order to signal the triggered action time (T1). At the same time the Unity 3D application receives the head rotation signal and transitions the test pattern from black to white. The test pattern signal is sent to the NED which transitions the display from back to white. The light sensor at the Styrofoam eye location detects the NED transition from black to white and the microcontroller records the display time (T2). The microcontroller determines the time between the laser alignment change and the NED display change (i.e., Latency = T2 – T1 in milliseconds) and outputs it on the microcontroller LCD. The computers included a desktop and laptop. The desktop was an Alienware with i7 9590 CPU, NVIDIA GeForce RTX 2080 graphics card and SSD running Windows 10. The laptop was an Alienware model with i7 9750 2.6 GHz CPU and NVIDIA GeForce RTX 2070 card and running Windows 10.

Latency Results

The average measured latencies ranged from 20 msec to 70 msec for all headsets. Latencies on Oculus Rift and HTC Vive Pro were close to the expected 20 msec. Perhaps due to significant runtime optimizations by the companies. Latency was measured on the Varjo VR-1 headset using OpenVR (average 70 msec and standard deviation 15 msec) as well as the native Varjo SDK (average 62 msec and std of 12 msec) implementation using an Alienware desktop. The most efficient and smallest amount of latency is expected through native SDK and runtime. OpenVR has an SDK and API that supports many different commercially available headsets. Communication with OpenVR requires an extra layer of communication which may add additional latency. As expected there was a 13% increase in latency using OpenVR. Differences due to rendering platform were tested by measuring latency using an Alienware lap top and desktop. The Oculus Rift latencies increased by 13% when using a laptop vs. a desktop. It is important to note that the latency data for the Varjo (Native SKD : 62 msec) and XTAL (Native SKD : 49 msec) headsets shows differences between what was measured and the commercially available product data. Once again it is important to note that there are significant differences between what was measured and the product data. The testing techniques used for measuring latency are consistent with current test methodologies and are representative of the integrated system performance.

COLLABORATIONS

NAWCTSD has many different science and technology efforts seeking applications of XR technology. Of interest to this paper are several Naval Innovative Science & Engineering (NISE) projects. The NAVAIR NISE project 219WFD-SG-19-022: Strategic Development of Near-Eye Display Performance Metrics for Naval Aviation Training Applications whose objective is to develop techniques, tools, and procedures for measuring performance of near eye display systems, such as virtual reality, augmented reality, and mixed reality headsets. Another NISE project using XR technology is 219TT-19-002: Integration of Virtual Reality / Augmented Reality with a cockpit for Piloted Simulation whose objective is to investigate the use of VR/AR to provide fields of regard beyond those produced by FTD's at NAVAIR Manned Flight Simulator (MFS). Other technology development efforts include the Small Business Innovative Research (SBIR) N192-087: Headset Equivalent of Advanced Display Systems (HEADS) which is looking at developing a novel VR, AR, and/or MR headset that performs equivalent to or better than current flight simulator display systems. Also of interest are several Cooperative Research and Development Agreement (CRADA) such as the Flight Safety International (FSI), NCRADA-NAWCTSD-19-067 Standardized Near-Eye Display System Performance Metrics as well as the NVIDIA, NCRADA-NAWCTSD-20-070 Collaboration on Near Eye Displays Roadmap.

CONCLUSION

In theory, the NED in XR headsets are just another type of display device available for flight simulators. However, although they share similar components there are significant differences in scale and performance. This paper provides a description of traditional high-fidelity display systems and HMD as well as the procedures for measuring performance and typical NAWCTSD performance levels. It is expected that if HMD are to replace high-fidelity display systems this level of performance and equivalent metrics will be needed. Of course, there are other significant issues associated with VR and MR that need to be addressed including vergence-accommodation conflict and simulator sickness. As discussed, traditional VDS test equipment cannot be used to measure the performance of a fully integrated XR display system. It is expected that methods for measuring performance on HMD will likely end up as a combination of field and laboratory methods and metrics. The author adapted and explored methods for measuring FOV, resolution, and latency in VR headsets in the field on a fully integrated system. The results are significant as they show that resolution is not uniform across the FOV, and that the advertised and measured performance varied greatly for FOV and static resolution. The reasons for the deviations in FOV may be attributed to the user's face characteristics, fit, and optic design. Nevertheless, the measured results are representative of the performance that will be delivered in a fully integrated system by the end user. A method for measuring latency was also explored and results presented. Once again, the measured performance varied but is expected that performance would be characteristic of the integrated HMD system. This paper only looked at VR metrics but MR is expected to open up a lot of other applications for HMD in flight simulators which combine cockpit and OTW views. Further research and development is needed in standardizing XR metrics and procedures so that results can be consistently compare across the industry and characteristics of the actual system performance.

FUTURE WORK

Several efforts are planned for future research on areas such as XR laboratory testing, use of motion platforms, and use of cockpit hardware for MR applications. Future research includes development of XR headset lab testing equipment for the evaluation of luminance, color, and contrast. The equipment includes an imaging colorimeter with special NED lens design simulates the size, position, and field of view of the human eye and the design of a low cost XR mount to get access the FOV from the eyebox. Private industry has begun exploring the use of MR technologies with motion platforms, but the Navy has limited experience using and verifying the effectiveness of combining these technologies. The co-authors' future efforts will combine MR technology with a 6 DOF motion platform to assess the effectiveness.

Chief of Naval Air Training (CNATRA) has asked NAWCTSD to investigate the feasibility of developing a T-45 Mixed Reality (MR) flight trainer. Within the next couple of years the primary author will concentrate on MR testing and how this engineering performance translates to training effectiveness evaluations. Some of the pressing concerns include XR headset FOV for situational awareness and training, combination of NED and camera spatial resolution necessary to support cockpit interaction and training, comfort (e.g., weight and inertia of the headset for comfort and fit), and simulation sickness. This effort will not only generate procedures and metrics to properly verify the performance of MR headsets, but also provide recommendations for visual requirements for future XR FTD.

DISCLAIMER

The views expressed herein are those of the authors and do not necessarily represent the views of DoD or its Components.

REFERENCES

Austin, R. L., Denning, B. S., Drews, B. C., Fedoriouk, V. B., & Calpito, R. C. (2018). Qualified viewing space determination of near-eye and head-up displays. Journal of the Society for Information Display, 26(9), (pp. 567-575).

Biggs A. T., MSC, Geyer D. J., Schroeder V. M., Robinson E. and Bradley J. L. (2018), Adapting Virtual Reality and Augmented Reality Systems for Naval Aviation Training, NAVAL MEDICAL RESEARCH UNIT DAYTON report NAMRU-D-19-13.

Davis, S., Nesbitt, K., & Nalivaiko, E. (Dec. 2014). A systematic review of cybersickness. In Proceedings of the 2014 conference on interactive entertainment (pp. 1-9).

Keller, K. & Colucci, D. (1998). Perception in HMDs: What is it in head mounted displays (HMDs) that really make them all so terrible. In Proceedings of SPIE Vol. 3362: Helmet- and Head-Mounted Displays III, (pp. 46-53).

Jerald, J. J. (2009). Scene-motion-and latency-perception thresholds for head-mounted displays (Doctoral dissertation, The University of North Carolina at Chapel Hill).

Jerald, J., Whitton, M., & Brooks Jr, F. P. (2012). Scene-motion thresholds during head yaw for immersive virtual environments. ACM Transactions on Applied Perception (TAP), 9(1), (pp. 1-23).

Joseph D, Burch T & Connolly R (2002), Comparison of Display System Options for Helicopter Aircrew Tactical Training Systems, I/ITSEC, Orlando, FL.

McCoy-Fisher, C., Mishler, A., Bush, D., Severe-Valsaint, G., Natali, M., Riner, R. (Sep. 2019), Student Naval Aviation Extended Reality Device Capability Evaluation, NAWCTSD report NAWCTSD-TR-2019-001.

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (Dec. 1995), Augmented reality: A class of displays on the reality-virtuality continuum. In Telemanipulator and telepresence technologies (Vol. 2351, pp. 282-292). International Society for Optics and Photonics.

National Simulator Program (NSP) (2016), 14 CFR part 60, Federal Aviation Administration (FAA).

Long, J. L., Lloyd, C. J., Beane, D. A., (2010). PRACTICAL GEOMETRY ALIGNMENT CHALLENGES IN FLIGHT SIMULATION DISPLAY, In Proceedings of the 2010 IMAGE Society conference.

Padmos, P. & Milders, M. V. (1992). Quality criteria for simulator images: A literature review. Human Factors, 34, (6), 727-748.

Raaen, K., & Kjellmo, I. (Sep. 2015), Measuring latency in virtual reality systems. In International Conference on Entertainment Computing (pp. 457-462). Springer, Cham.

So, R. H., & Griffin, M. J. (1992). Compensating lags in head-coupled displays using head position prediction and image deflection. Journal of Aircraft, 29(6), (pp. 1064-1068).

Winterbottom, M. D., Patterson, R., & Pierce, B. J. (2005). Helmet-Mounted Displays for use in Air Force Training and Simulation.