Evaluation of a head-worn display with ambient vision cues for unusual attitude recovery
Stephanie N. Nicholas¹, Jarvis (Trey) Arthur, Kathryn Ballard, Renee C. Lake, Kyle E. Ellis, Lawrence J. Prinzel, III
NASA Langley Research Center, Mail Stop 152, Hampton, VA, USA 23681

ABSTRACT
A Commercial Aviation Safety Team (CAST) study of 18 loss-of-control events determined that a lack of external visual references was a contributing factor in 17 of these events. CAST recommended that manufacturers should develop and implement virtual day-VMC display systems, such as synthetic vision (SV) or equivalent systems (CAST Safety Enhancement, SE-200). In support of this recommended action, CAST has requested studies to define minimum requirements for virtual day-visual meteorological conditions (VMC) displays to improve flight crew awareness of airplane attitude. NASA’s research in Virtual day-VMC displays, known as synthetic vision systems, are intended to support intuitive flight crew attitude awareness similar to a day-VMC-like environment, especially if they could be designed to create visual dominance.

A study was conducted to evaluate the utility of ambient vision (AV) cues paired with virtual Head-Up Display (HUD) symbology on a prototype head-worn display (HWD) during recovery from unusual attitudes in a simulated environment. The virtual-HUD component meets the requirement that the HWD may be used as an equivalent display to the HUD. The presence of AV cueing leverages the potential that a HWD has over the HUD for spatial disorientation prevention. The simulation study was conducted as a single-pilot operation, under realistic flight scenarios, with off-nominal events occurring that were capable of inducing unusual attitudes. Independent variables of the experiment included: 1) AV capability (on vs off) 2) AV display opaqueness (transparent vs opaque) and display location (HWD vs traditional head-down displays); AV cues were only present when the HWD was being worn by the subject pilot.

Keywords: Head-Worn Display, synthetic vision, unusual attitude recovery, attitude awareness, loss-of-control inflight, ambient vision

1. INTRODUCTION
Recent accident and incident data suggest that spatial disorientation and Loss-of-Control In-flight (LOC-I) are becoming an increasingly prevalent safety concern in domestic and international operations¹. Spatial disorientation is defined as the inability to correctly interpret aircraft attitude, altitude or airspeed in relation to the Earth or other reference point². A Commercial Aviation Safety Team (CAST) study of 18 loss-of-control events determined that a lack of external visual references was associated with flight crew loss of attitude awareness or energy state awareness in 17 events. CAST recommended that manufacturers should develop and implement virtual day-VMC display systems, such as synthetic vision or equivalent systems (CAST Safety Enhancement, SE-200). In support of this recommended action, CAST has requested studies to define minimum requirements for virtual day-visual meteorological conditions (VMC) displays to improve flight crew awareness of airplane attitude.

For the purpose of this safety enhancement, CAST has identified “virtual day-VMC displays” as systems that have the following elements:

- Are presented full time in the primary field-of-view
- Are presented to both flight crew members
- Include display of energy state cues, including flight path, acceleration, and speed deviation, in a manner similar to modern head-up displays³.

¹ Author information: Stephanie Nicholas, E-Mail: stephanie.n.nicholas@nasa.gov, Telephone: 1-757-864-8327
The National Aeronautics and Space Administration (NASA) is conducting research and development in support of these CAST SEs as part of the Airspace Operations and Safety Program (AOSP), under the Technologies for Airplane State Awareness (TASA) sub-project. Associated with this research toward the development and implementation of SV display systems, NASA is also investigating SV display concepts in support of new, intuitive methods for the prevention of spatial disorientation.

1.1 Head-up display vs head-worn display

Today’s commercial aviation industry is learning toward the incorporation of virtual-day VMC displays using the traditional primary flight display (i.e., a full color synthetic vision system) or via a Head-Up Display (HUD). A HUD is a display system which projects information to the operator onto an optical-quality piece of semi-transparent glass (i.e., the combiner) fixed to the vehicle directly between the operator and the windscreen in front of the operator. The information is projected onto the combiner, in a fixed field-of-view, from an image generation source and collimating optics.

The physical characteristics of the display device (brightness, luminance, color) and associated optics (e.g., collimation, focus) are critical determinants of the operators’ perceived vision image quality and its usability, acceptability, and performance. Head-Up Displays in commercial and business aircraft operations have many benefits, but two hold particular appeal:

- The HUD provides an eyes-out, conformal display of critical flight parameters/symbology. As such, HUD operations improve safety by reducing the visual accommodation and re-accommodation that the pilot would otherwise have between reading head-down instrumentation and transitioning to out-the-window visual references during operations such as approach, landing, taxiing, and takeoff. Based on a 2009 study conducted by the Flight Safety Foundation, researchers found that HUDs could have prevented or positively influenced 38% of accidents overall in modern glass cockpits. Of those accidents where the pilot was directly involved, such as takeoff or loss-of-control, the “safety advantage” of the HUD became much higher (69% and 57%, respectively).
- The explicit display of flight path and energy information on the HUD promotes precision touchdown capability and energy management – i.e., maximizing landing performance (minimizing aircraft wear-and-tear). Studies show that significantly lower touchdown sink rates are flown under manual than automatic landings’ and experimental evidence show that the HUD promotes better sink rate control and touchdown performance.

Because of these advantages, HUDs are becoming standard equipage. But because of the cost, volume, and weight of a HUD installation, alternative hardware options, such as HWDs, are being investigated. For commercial and business aircraft, the benefits of utilizing a HWD include:

- In many aircraft, a HUD installation is not possible or practical, principally from a volume standpoint. Current generation HUDs require a significant overhead volume for the projector. For this reason, many new aircraft and retro-fit applications of a HUD are impossible or impractical because the aircraft doesn’t have enough overhead space.
- The HWD presents a much easier installation and retro-fit path than a HUD. Since, by definition, the HWD must be lightweight and un-encumbering to the user, the installation does not require overhead hard-points from which to mount combiner glass or wiring harnesses for a HUD.
- The weight of current HUD systems is not insignificant. Current overhead and combiner technologies can easily weigh 25 pounds or more. This weight is far greater than the desired 4 pounds head-borne weight of the ideal HWD. The return-on-investment by saving weight is obviously different for each operator, but the possibility of saving 25 pounds for each HUD-equipped aircraft might create a reasonable return on investment.

1.2 Visual processing systems

There are two modes humans use to process visual information: 1) ambient and 2) focal. Based on a person’s visual observation, ambient mode normally answers the question of “where” the person is located in space as well as “where” a particular object is relative to the observer. Ambient visual processing typically occurs during the first few seconds of
viewing and involves short fixations and long saccades as people extract information from peripheral vision. Changes detected in ambient process provide optical flow (the pattern of motion between the observer and objects and surfaces in the observer’s visual field). Ambient mode processing is aligned with the vestibular and auditory senses to provide gaze stability, spatial awareness, and self-motion. The focal mode normally answers the question of “what” the perceived observed object, surface, or space is. The shift from ambient to focal processing creates longer visual fixations on an item of interest to allow for object/space recognition and distance/depth perception.

Research has established that the pilot’s ambient vision (AV) is, by far, the most important sensory system for establishing and maintaining spatial orientation during flight. Ambient vision is the primary determinant of visual dominance – that is, where the human’s visual perception dominates confounding or conflicting vestibular sensations. Historically, the design of an aircraft’s cockpit instrumentation has been tailored to address a crew’s need to fly under instrument conditions using foveal vision with instruments that show: 1) control (i.e., attitude); 2) performance (i.e., energy state); and 3) navigation (i.e., position relative to desired flight path) information. The potential for spatial disorientation is “trained out” of each instrument-rated pilot by reinforcement learning that the pilot must ‘trust their instruments’ – that is, to use their foveal view of abstract instrumentation to maintain aircraft control while suppressing or ignoring confounding, conflicting vestibular cues.

Research suggests the following general requirements for ambient vision and visual dominance from synthetic displays:

1) Wide FOV subtending at least 60 degrees and preferably greater than 100 degrees in extent, with ‘correct’ spherical projection of the horizon;
2) Optical scenes optically projected at a distance that exceeds the effective range of most binocular mechanisms (greater than 5m or preferably, larger);
3) Temporal resolution to match the realism and believability of the imagery compared to an out-the-window scene (less than 100 ms latency); without aliasing, flicker or other optical defects;
4) Spatial resolution is not critical in the periphery but must promote image realism (greater than 0.5 min of arc spatial resolution);
5) Sufficient content to both stimulate the ambient visual system and create the level of detail expected of a real-world daylight scene. The synthetic scene must believably represent the movement of Earth-fixed space.

For HUDs, the stimulation of ambient vision is not practical. An extended horizon line and compressed pitch ladder symbology are strategies to prevent and recover from an unusual attitude, respectively, but the horizontal extent of HUDs is limited to approximately 40 degrees at best. Previous research conducted by Sharkey showed possible improved pilot performance when using AV on HWDs. HWDs offer extended fields-of-regard, not practical in HUDs or head-down displays that offer the potential for ambient vision stimulation and visual dominance. Sharkey’s research, however, was very limited by dated technology.

NASA conducted a research study that utilized the lessons-learned from this previous research to evaluate display configurations representative of virtual day-VMC displays as they support attitude recognition, attitude alerting, and visual dominance. Specific display configurations evaluated included the presentation of virtual-HUD flight symbology on a prototype HWD as well as the presence of AV cueing.

2. METHOD

The specific objectives of the Augmented Flight Deck Countermeasures (AFDC) 5.0 experiment were to:

- Evaluate a prototype SV Head-Worn Display (HWD), as well as the presence of AV cueing, for attitude recognition, attitude alerting, and visual dominance. The HWD uses a virtual-HUD component intended to meet the requirement that the HWD may be used as an equivalent display to the HUD. The presence of AV cueing leverages the potential that the HWD has over the HUD for spatial disorientation prevention.
- Test and evaluate virtual day-VMC/SV displays under realistic flight operation scenarios capable of inducing reduced attention states in pilots and evaluate each display concepts’ efficacy to improve attitude awareness and prevent entry to unusual attitudes.
- Collect and evaluate pilot behavioral response data in nominal and off-nominal attention states across realistic flight operation scenarios.
The research evaluated four display concepts to meet these objectives: 1) Baseline head-down display (HDD); 2) Virtual-HUD; 3) Virtual-HUD with AV (transparent); and, 4) Virtual-HUD with AV (opaque). These display concepts were evaluated in a two-part combined experiment structure, with pilots flying a series of takeoff / landing tasks and plane-following task scenarios.

2.1 Evaluation pilots
A total of 24 commercial pilots from various US airlines were used as evaluation pilots (EPs). Each pilot held an active Airline Transport Pilot certificate and had accumulated at least 100 hours of HUD experience.

2.2 Simulation facility
The experiment was performed in the Development & Test Simulator (DTS) at NASA Langley Research Center (Fig. 1). The DTS is a medium-fidelity, fixed-based advanced all-glass transport with programmable sidestick control inceptors and panorama visual systems, providing approximately 210 degrees horizontal by 45 degrees vertical field-of-view (FOV). The aerodynamics and flight controls mimicked a Boeing 757 large commercial transport aircraft.

2.3 Head-Worn Display
A modified Microsoft HoloLens™ unit was used as the HWD in this experiment (Fig. 2). In order to provide ambient vision capability, two Scorpion® Display Modules² coupled with a Thales Visionix, Inc. InertiaCube™ inertial tracker were attached to the HoloLens™. Each Scorpion® Display Module was mounted at an angle of 79 degrees outboard from the nose bridge (as viewed from the top of the head) of the HWD. The specifications of a single Scorpion® Display Module and the Microsoft HoloLens™ are listed below in Table 1.

Table 1: HWD and Ambient Vision Display Specifications.

<table>
<thead>
<tr>
<th>Resolution</th>
<th>HWD - HoloLens</th>
<th>Single Ambient Vision Display Module</th>
</tr>
</thead>
<tbody>
<tr>
<td>1268 (H) x 720 (V)</td>
<td>800 (H) x 600 (V)</td>
<td></td>
</tr>
<tr>
<td>Field-of-View</td>
<td>30° (H) x 17.5° (V)</td>
<td>26° (H) x 20° (V)</td>
</tr>
</tbody>
</table>

² The use of trademarks or names of manufacturers in this report is for accurate reporting and does not constitute an official endorsement, either expressed or implied, of such products or manufacturers by the National Aeronautics and Space Administration.
The HWD is shown in Figure 2. In the lower picture, the ambient vision displays can also be seen with the opaque (left) and transparent (right) configurations. As the Scorpion® displays are see-through, the black background for the AV (opaque) concept provided more contrast and visibility. This concept was achieved by placing a piece of black-opaque film over the back of the Scorpion® lens, removing the outside world (simulator flight deck) from the pilot’s view; all AV imagery and symbology was still visible. The virtual-HUD with AV (transparent) concept utilized the Scorpion® displays with no opaque material over the lens, thus showing the outside world behind the AV imagery and symbology.

![Image of HWD](image)

**Figure 2: Modified Microsoft HoloLens™ HWD Hardware**

### 2.3.1 Head-Worn Display symbology

The HWD depicted a “Virtual-HUD” concept that was designed to mimic most features of a conventional commercial HUD. One symbology difference was the inclusion of a floating aircraft heading indicator initially positioned in the top left corner of the eye box (Fig. 3). The indicator was fixed when the pilot looked within the eye box of the virtual-HUD, and moved with the pilot all other times. Additional flight symbology included a runway outline and extended runway centerline during approach and landing.

When the aircraft maneuvered into an unusual attitude, flight symbology was switched to “compressed mode” (Fig. 4). This configuration created a non-conformal, compressed, pitch ladder accompanied by critical information (i.e., airspeed, altitude, and heading). The conditions in which compressed mode triggered were: pitch-up attitude greater than +25 degrees, or pitch-down attitude lower than -12.5 degrees, or a roll angle of greater than +45 degrees. Once the aircraft returned to straight-and-level flight, the flight symbology reverted to standard mode after maintaining the following conditions: pitch attitude between -5 degrees and +10 degrees for more than 2.5 seconds and a roll angle of less than or equal to 10 degrees for more than 5 seconds.
2.3.2 Ambient vision display

The AV displays were “always on” and nominally showed a synthetic vision view of the external topography (see Fig. 5), but an alerting feature was added.

When the aircraft entered an unusual attitude, the AV displays showed a yellow square grid scheme in the sky (i.e., above the horizon line). The conditions for activation and de-activation of the box-grid alerting scheme were same as the virtual-HUD compressed mode. The AV display information was earth-referenced, thus, the yellow-grid alerting scheme was presented on the “wing high” AV display. For instance, if the aircraft was positioned in a steep right turn, the right AV display would show the nominal SV terrain (right wing directed toward ground), while the left AV display would show the yellow square alerting scheme, and vice versa (Fig. 6). If the aircraft’s wings were level, but the nose was pitched too high, both the left and right AV displays would show the yellow grid above the horizon line and synthetic vision below the horizon line.
2.3.3 Head-down display

The head-down display suite is shown in Figure 8. A baseline “blue over brown” configuration was used during all experimental conditions. Altitude and enhanced call-outs were played over the flight deck speakers (for baseline takeoff and landing scenarios only). Enhanced callouts were generated to improve roll recovery in the event of unusual attitude. The system identified when the aircraft entered an unusual roll attitude, and provided auditory, as well as visual directive roll recovery guidance to improve pilot situation awareness and reduce the confusion commonly associated with autopilot roll saturation and large bank angles. These were used since they were part of the “baseline” condition used in other SE-200 research activities13.

2.4 Evaluation pilot training

Training consisted of briefings and in-simulator familiarization, as well as additional experiment-specific standard operating procedures.
2.5 Takeoff and landing scenarios

Four event scenarios (two takeoff events and two landing events) were developed that effectively led unsuspecting pilots to lose attitude awareness. The four scenario descriptions are as follows:

1. ILS RWY 36C approach with right runway rudder trim and right engine flame-out (Task L1): Initial condition positioned at 6500’ from RWY 36C, 1300’ altitude on track 358 degrees. Flaps 30 and gear down. 15kt crosswind from 270 degrees; Autopilot (A/P) engaged for autoland. At 20 seconds after the scenario start, simulated air traffic control (ATC) instructed the pilot to “go-around” and follow missed approach procedures. At 60 seconds, the subject pilot encountered full runaway right rudder trim coupled with a right engine flameout, inducing a large roll upset.

2. ILS RWY 36C approach with wind shear event (Task L2): Initial condition positioned at 6500’ from RWY 36C on track 358 degrees. Flaps 30 and gear down. 15kt crosswind from 270 degrees. A/P engaged for autoland. At 30 seconds after the start of the scenario, ATC will instruct pilot to “go-around” and follow missed approach procedures. At 70 seconds, the aircraft encounters a large wind shear event, inducing a pitch down upset.

3. ILS RWY 36L takeoff with wind shear event (Task T1): Initial condition positioned on surface in take-off position for RWY 36L. Aircraft is cleared for takeoff RWY 36L and instructed to climb and maintain 5000’ on heading 358 degrees in 1000’ Runway Visual Range (RVR) conditions. At 70 seconds after the start of the scenario, the aircraft encounters a large shear event, inducing a pitch down upset.

4. ILS RWY 36L takeoff with wind shear event (Task T2): Initial condition positioned on KMEM surface at take-off hold line. Aircraft is asked to hold short of RWY 36L as heavy aircraft completes takeoff. At 25 seconds after the scenario start, the aircraft is cleared for takeoff RWY 36L and instructed to climb and maintain 5000’ on heading 358 degrees in 1000’ RVR conditions. At 1000’ above ground level, the aircraft pilot encounters a large wake vortex, inducing roll upset.

2.6 Plane-following task scenario

The plane-following task (PFT) scenarios were designed to present a series of tasks and/or flight dynamics, inducing conditions similar to the incidents analyzed by the Joint Safety Analysis Team (JSAT) report, that lead to an increased probability of spatial disorientations through the use of pilot distraction, startle/surprise, and/or channelized attention without effective pilot intervention.

The PFT was developed from earlier Naval Medical Research Unit – Dayton (NAMRU-D) research. The piloting task was to follow a pre-recorded lead aircraft, shown on the out the window display, through a series of turns, climbs, and descents. Each scenario began with the aircraft straight and level, at 18,000 feet above mean sea level (MSL) with a separation of 2,000 feet from the lead aircraft. Weather was undercast (cloud deck) at 17,000 feet MSL, and provided a level and well-defined visual horizon. Visibility above the clouds was set to 4,500 feet. Pilots were instructed to fly in trail formation and to stay close enough to easily see the lead aircraft’s attitude.

Two lead aircraft flights (PFT-Left and PFT-Right) were recorded, each a mirror image of the other. During the flight, the lead aircraft executed four right (R) and four left (L) turns, each with a 90 degree heading change and at a 45 degree angle of bank, with the exception of a 60 degree bank executed on the final turn. In PFT-Left, the order of turns were L, L, R, R, R, L, R, and in PFT-Right the turns were reversed. For both flights, turns one, two, and four were level, turn three was ascending, and turns five through eight were descending. While in the clouds the lead aircraft remained visible but the horizon was completely obscured. Evaluation pilots were instructed that while flying in the clouds, the lead would disappear at an undisclosed time, whereupon they should level their wings and start a gentle climb. At approximately 5 minutes after the scenario started, the lead aircraft disappeared while established in the eighth turn, in the clouds and with no visible horizon.

2.7 Experiment design

The test matrix was a within-subjects design; each pilot flew all four take-off and landing scenarios and 4 PFT tasks (2 left and 2 right), with each display configuration using a randomized selection of the display type. For the “baseline” display
type, pilots did not wear the HWD and only flew using the traditional head-down display configuration. See Table 2 for a sample trial order.

Table 2: Example of scenario trial order (randomized).

<table>
<thead>
<tr>
<th>Pilot</th>
<th>Run Number</th>
<th>Display</th>
<th>Scenario</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Virtual-HUD</td>
<td>PFT-Left</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>Baseline</td>
<td>T1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>Virtual-HUD with AV - Opaque</td>
<td>PFT-Right</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>Baseline</td>
<td>PFT-Left</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>Virtual-HUD with AV - Transparent</td>
<td>PFT-Right</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>Virtual-HUD with AV - Opaque</td>
<td>T2</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>Virtual-HUD with AV - Opaque</td>
<td>PFT-Left</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>Virtual-HUD</td>
<td>PFT-Right</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>Virtual-HUD</td>
<td>L2</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>Virtual-HUD with AV - Transparent</td>
<td>L1</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>Baseline</td>
<td>PFT-Right</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>Virtual-HUD with AV - Transparent</td>
<td>PFT-Left</td>
</tr>
</tbody>
</table>

2.8 Questionnaires

Pilots were administered a post-run, subjective NASA Task Load Index (TLX)\(^{15}\) workload scale, to evaluate pilot workload during each scenario. The scale ranged from 0 – 100 for each of six categories, with 0 being ‘very low’ and 100 being ‘very high’. For the performance category, 0 meant a perfect performance while 100 meant a failure. Pilots were administered a post-run, qualitative situation awareness measured using the three question NASA Situation Awareness Rating Technique (SART)\(^{16}\) as well as a display efficacy rating. The scale for each of the SART questions as well as the efficacy rating was the same as the NASA TLX. Pilots were also administered a paired-comparison survey for the given displays and a Likert scale questionnaire for general HWD comfort, attitude awareness, and field-of-view.

3. RESULTS

Aircraft state data were recorded during each run. Pilot stick inputs are used to determine if the pilot made the correct input for recovery in all tasks. These metrics also quantify the duration of the recovery for each takeoff and landing scenario.

3.1 Takeoff and landing scenarios

3.1.1 Recovery performance

All of the pilots successfully recovered from the wind shear upset events (Task L2).

There were unsuccessful recoveries for the takeoff with wake encounters (Tasks T1 and T2) as well as the rudder and engine failure event on approach (Task L1). Figure 9 shows the failure to recover percentages by display.
3.1.2 Recovery time

The time to make first input in pitch or roll as well as the correctness of that input was analyzed. The time-to-first input was calculated as the time elapsed between when the upset occurred and when the pilot made their first input. A pilot’s first input was calculated by finding the first input in pitch for Tasks T1 and L1 or roll for Tasks T2 and L2 that went beyond ±10% of the initial stick position at the time of the upset. This was done to capture only intentional stick inputs. Figure 10 shows the time-to-first input data by scenario and display. Overall, the average time-to-first input was 3.94 seconds (SD = 2.38). There were eleven cases from the two wind shear scenarios (Tasks T1 and L1) where pilots relied on autopilot to recover so no input was detected. Those cases are not shown in Figure 10 and were not used to calculate the overall mean time-to-first input.
Because time-to-event data does not follow a normal distribution, the time-to-first input data was analyzed using a generalized linear mixed model with a log link. The display was a fixed effect in the model while the random effects were task and pilot because there was no interest in the differences between scenarios or pilots. All models discussed in the results section used the same configuration of fixed and random effects. Testing the model terms using analysis of variance revealed that the effect of display on time-to-first input was not statistically significant ($p = 0.47$).

The correctness of the first input was found by comparing the direction of the first stick input (pitch input for Tasks T1 and L1 or roll input for Tasks T2 and L2) with the proper direction required to recover. Pilots made the correct first input 68.5% of the time, the incorrect input 19.6% of the time, and no input 12% of the time. For the purpose of this analysis, the no-input cases were combined with correct input cases because the run was still recovered on autopilot and the pilot did not make the situation any worse by not making any input. With only two categories, correct or incorrect, the first inputs were analyzed using a generalized linear mixed model with a logistic distribution. The correctness of the pilot input after an upset was not affected by the display type ($p = 0.77$).

If the pilot did recover, the time to complete the recovery was recorded. The recovery criteria for takeoff and landing scenarios with pitch upsets was roll between -5 and 5 degrees and pitch between 5 and 15 degrees (since they are in the middle of a takeoff or go-around, the pitch should be positive to return to that stage of flight). Roll upsets had the recovery criteria of roll from -5 to 5 degrees and pitch from -5 to 10 degrees. If these criteria were not met, the time of recovery defaulted to the end of the run since the researcher called the end of the run when it was deemed either a recovery or a failure. Table 3 shows that the more difficult recoveries took longer to recover on average, as expected.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Display</th>
<th>Mean Time-to-Recovery</th>
<th>Std Dev</th>
<th>Runs Recovered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Takeoff - Wind Shear (Task T1)</td>
<td>Virtual-HUD Transparent AV</td>
<td>12.0</td>
<td>7.1</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Virtual-HUD Opaque AV</td>
<td>11.9</td>
<td>7.0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Virtual-HUD No AV</td>
<td>11.0</td>
<td>6.4</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Baseline</td>
<td>16.2</td>
<td>8.5</td>
<td>6</td>
</tr>
<tr>
<td>Approach - Wind Shear (Task L1)</td>
<td>Virtual-HUD Transparent AV</td>
<td>27.3</td>
<td>6.3</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Virtual-HUD Opaque AV</td>
<td>22.0</td>
<td>14.8</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Virtual-HUD No AV</td>
<td>19.2</td>
<td>11.8</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Baseline</td>
<td>27.9</td>
<td>15.2</td>
<td>6</td>
</tr>
<tr>
<td>Takeoff - Wake Encounter (Task T2)</td>
<td>Virtual-HUD Transparent AV</td>
<td>40.2</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Virtual-HUD Opaque AV</td>
<td>25.1</td>
<td>4.6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Virtual-HUD No AV</td>
<td>23.5</td>
<td>3.6</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Baseline</td>
<td>23.1</td>
<td>2.1</td>
<td>2</td>
</tr>
<tr>
<td>Approach – Rudder/Eng. Failure (Task L2)</td>
<td>Virtual-HUD Transparent AV</td>
<td>34.6</td>
<td>4.0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Virtual-HUD Opaque AV</td>
<td>21.0</td>
<td>1.3</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Virtual-HUD No AV</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Baseline</td>
<td>27.2</td>
<td>3.3</td>
<td>2</td>
</tr>
</tbody>
</table>

A generalized linear mixed model with a log link showed that the display type did have an effect on the time-to-recovery after an upset occurred ($p < 0.01$). In scenarios with wind shear inducing a pitch upset, pilots using the baseline displays took longer to recover (a 4.8 second average difference above other displays). The scenarios with a roll upset had too few recovered runs (about 35% were recovered) to make conclusions on the time-to-recovery based on display type.
3.2 Plane-following task results

The goal of the plane-following tasks was spatial disorientation. Almost all the pilots were guided into an unusual attitude of around 60° to the right or left (M = 59.5°, SD = 7.0°). Data from 15 runs were excluded because the pilot lost sight of the aircraft early.

Pilot input on the side stick after the lead aircraft disappeared was examined for direction and magnitude. Because the pilot was already using the stick when the lead aircraft disappeared, the first roll input beyond ±10% of the stick position at the time it disappeared was obtained to capture intentional inputs that were part of the recovery. There were 30 runs with an incorrect initial roll input (17.8%).

A generalized linear mixed model with a logistic distribution for the binary data (each run was either a success or failure in terms of making the correct initial input) was used to assess the significance of display type. Analysis of variance on the model shows there was no difference between the displays in terms of incorrect initial inputs (p = 0.70). Table 4 summarizes the recovery performance by display type.

Table 4: Summary of incorrect roll inputs during plane-following tasks

<table>
<thead>
<tr>
<th>Display</th>
<th>Task</th>
<th>Incorrect Roll Inputs</th>
<th>Total Runs</th>
<th>Percent Incorrect Roll Input</th>
<th>Mean Magnitude in Wrong Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>PFT-Left</td>
<td>3</td>
<td>22</td>
<td>13.6</td>
<td>0.28</td>
</tr>
<tr>
<td>Baseline</td>
<td>PFT-Right</td>
<td>4</td>
<td>23</td>
<td>17.4</td>
<td>0.40</td>
</tr>
<tr>
<td>Virtual-HUD No AV</td>
<td>PFT-Left</td>
<td>3</td>
<td>22</td>
<td>13.6</td>
<td>0.26</td>
</tr>
<tr>
<td>Virtual-HUD No AV</td>
<td>PFT-Right</td>
<td>7</td>
<td>20</td>
<td>35.0</td>
<td>0.41</td>
</tr>
<tr>
<td>Virtual-HUD Opaque AV</td>
<td>PFT-Left</td>
<td>5</td>
<td>21</td>
<td>23.8</td>
<td>0.38</td>
</tr>
<tr>
<td>Virtual-HUD Opaque AV</td>
<td>PFT-Right</td>
<td>2</td>
<td>21</td>
<td>9.5</td>
<td>0.41</td>
</tr>
<tr>
<td>Virtual-HUD Transparent AV</td>
<td>PFT-Left</td>
<td>2</td>
<td>20</td>
<td>10.0</td>
<td>0.29</td>
</tr>
<tr>
<td>Virtual-HUD Transparent AV</td>
<td>PFT-Right</td>
<td>4</td>
<td>20</td>
<td>20.0</td>
<td>0.23</td>
</tr>
</tbody>
</table>

3.3 Post-run questionnaires

The post-run questionnaires on workload, situation awareness, and efficacy were analyzed as well as post-experiment questionnaires on the comfort of the HWD, clarity of display, and general opinions.

The reported TLX workload ratings were averaged across the six subcategories (mental demand, physical demand, temporal demand, performance, effort, and frustration). The performance category was reversed, where 0 was a perfect performance and 100 was failure. The averaged TLX workload scores varied between tasks but not between display types. A linear mixed model was applied to the TLX workload data and after accounting for differences between tasks and pilots, no difference between display types was found (p = 0.88 for PFTs, p = 0.33 for takeoff and landing). The radar plots below (Figure 11) shows the average ratings by task for each workload subcategory.
Situation awareness scores were calculated from the SART questionnaire by the formula supply - demand + understanding. Pilots reported higher SART scores in the PFTs (M = 79.8, SD = 47.4) than the takeoff and landing scenarios (M = 25.1, SD = 53.7) but scores were similar among display types. Analysis of the mixed model did not reveal a difference in situation awareness between display types (p = 0.9 for PFTs and p = 0.51 for takeoff and landings).

Pilots rated the efficacy of each display on a scale from 0 to 100, 100 being most effective. All displays were rated slightly higher for the PFTs than the takeoff and landing scenarios (M = 74.2, SD = 19.6 for PFTs compared to M = 68.9, SD = 21.8 for takeoff and landings). However, analysis of variance on the mixed model of efficacy as a function of display did not show significant results for display type (p = 0.58 for PFTs and p = 0.09 for takeoff and landings).

After completing all runs, the pilots were asked to compare the four displays they had used in a series of paired contests assessing which display provided more situation awareness (SA). There were six contests to capture every combination. Figure 12 shows how many pilots preferred one display over another in terms of situation awareness for each contest. The first three rows show that around 80% of pilots felt the virtual-HUD, with or without AV, was superior to the baseline in providing SA. The next two rows show around half of the pilots found that having AV, either transparent or opaque, was superior to only having the virtual-HUD. The last row shows that 50% of pilots felt transparent AV was equal to opaque AV and the remaining pilots were split in their preference between the types of ambient vision.

To see if pilot preferences were statistically significant, the Bradley-Terry model was applied to the paired-comparison data. The Bradley-Terry model is a special generalized linear model that deals with situations where k items are compared to each other in paired contests. In this logistic model ties are not taken into account so the ‘equal’ category was ignored for analysis. Tests showed that the virtual-HUD without AV provided more situation awareness than baseline (p < 0.001)
and virtual-HUD with AV provided more situation awareness than the virtual-HUD without AV ($p < 0.001$). No difference was found between transparent or opaque ambient vision ($p = 0.78$).

Figure 12: Pilot responses to paired-comparison survey for the plane-following tasks, takeoff, and landing scenarios

Pilots also rated their experiences with the HWD in a Likert scale questionnaire. Figure 13 shows the distribution of responses for five questions. Sixty-six percent of pilots thought the HWD was moderately or extremely uncomfortable. Regarding display glare, 78% of pilots agreed or strongly agreed that they did not experience glare with the HWD.

Figure 8: Pilot responses for overall HWD experience
Even though nearly half of pilots experienced eye strain, 69% of pilots agreed or strongly agreed that they did not experience headaches from the HWD. Fifty-six percent of pilots agreed or strongly agreed that they considered the HWD to be equivalent to a HUD based on their experience.

**CONCLUDING REMARKS**

The goal of the research study was to evaluate the utility of AV cues paired with virtual-HUD symbology on a prototype HWD during recovery from unusual attitudes in a simulated environment. Use of a virtual-HUD component ensured the fulfillment of the requirement that the HWD may be used as an equivalent display to the HUD.

The quantitative analyses of the results did not generally yield statistically significant differences for the constructs investigated, with the exception of the scenarios related to wind shear inducing a pitch upset, wherein display type had a significant effect on the time-to-recovery after an upset occurred ($p < 0.01$). In those scenarios, pilots using the baseline displays took longer to recover (a 4.8 second average difference above other displays). The scenarios with a roll upset had too few recovered runs (about 35% were recovered) to make conclusions on the time-to-recovery based on display type.

With respect to situation awareness, a pairwise-comparison of display types revealed that the Virtual-HUD without AV provided more situation awareness than the Baseline HDD ($p < 0.001$), and Virtual-HUD with AV provided more situation awareness than the Virtual-HUD without AV ($p < 0.001$). No difference was found between transparent or opaque ambient vision lenses ($p = 0.78$). Collectively, this finding suggests that the presence of AV (regardless of transparent/opaque type) on the HWD is more effective at providing situation awareness than what was achieved by displays that did not feature AV in this study. This finding is on trend with results obtained by Arthur, et al.\textsuperscript{10}, whereby a HWD featuring AV was preferred by pilots over the traditional blue-over-brown baseline.

Based on qualitative analysis, many pilots felt the HWD system was uncomfortable (e.g., heavy, bulky, pressure on nose bridge) and attributed to, or had the potential for attributing to, eye strain, although most pilots did not report display glare or headaches. In terms of visual stability, four pilots reported experiencing intermittent double vision when using the HWD, typically during the plane-following tasks, and four pilots commented that they experienced some recurrent episodes of partial clipping or other obstruction of view regarding the HWD symbology. Nonetheless, fifty-six percent of pilots agreed or strongly agreed that they considered the HWD to be equivalent to a HUD. Additional comments revealed that while some pilots did not notice the transparent AV or thought it was distracting at first, it was noted by some pilots that during certain phases of flight and with additional training, it had the potential to be an effective alerting tool to increase situation awareness. More pilots stated that they preferred the opaque AV as it was brighter and easier to discern.

Additional research is warranted in this area of HWD technology. Obviously, pilot feedback reflected a general desire for a lighter weight design with improved weight balance/distribution and increased fore/aft adjustability (particularly for those wearing glasses). Additional feedback was suggestive that future testing of the HWD concept may also benefit from optimizing the fore/aft (azimuthal) placement of the AV lens location, as well as capturing an expanded range of view for the symbology. Other variables, such as investigating the effectiveness of alternate alerting schemes, as well as higher fidelity testing in a motion-based simulator environment would further identify and inform HWD requirements and serve to optimize future designs.

**REFERENCES**