

Simulation test of a head-worn display with ambient vision display for unusual attitude recovery

Jarvis (Trey) J. Arthur, III, Stephanie N. Nicholas, Kevin J. Shelton, Kathryn Ballard, Lawrence J. Prinzel, III, Kyle E. Ellis, Randall E. Bailey, and Steven P. Williams

NASA Langley Research Center, Hampton, VA

ABSTRACT

Head-Worn Displays (HWDs) are envisioned as a possible equivalent to a Head-Up Display (HUD) in commercial and general aviation. A simulation experiment was conducted to evaluate whether the HWD can provide an equivalent or better level of performance to a HUD in terms of unusual attitude recognition and recovery. A prototype HWD was tested with ambient vision capability which were varied (on/off) as an independent variable in the experiment testing for attitude awareness. The simulation experiment was conducted in two parts: 1) short unusual attitude recovery scenarios where the aircraft is placed in an unusual attitude and a single-pilot crew recovered the aircraft; and, 2) a two-pilot crew operating in a realistic flight environment with “off-nominal” events to induce unusual attitudes. The data showed few differences in unusual attitude recognition and recovery performance between the tested head-down, head-up, and head-worn display concepts. The presence and absence of ambient vision stimulation was inconclusive. The ergonomic influences of the head-worn display, necessary to implement the ambient vision experimentation, may have influenced the pilot ratings and acceptance of the concepts.

Keywords: unusual attitude, spatial disorientation, ambient vision, head-worn display, head-up display, synthetic vision, flight simulation

1. INTRODUCTION

NASA research, development, test, and evaluation (RDT&E) of flight deck interface technologies is being conducted to proactively identify, develop, and mature tools, methods, and technologies for improving aviation safety and operational efficiency in the Next Generation Air Transportation System (NextGen).

Recent accident and incident data suggests that Spatial Disorientation (SD) and Loss-of-Energy State Awareness (LESA) for transport category aircraft are becoming an increasingly prevalent safety concern in all domestic and international operations.¹ SD is defined as an erroneous perception of aircraft attitude that can lead directly to a Loss of Control (LOC) event, resulting in an accident or incident. LESA is typically characterized by a failure to monitor or understand energy state indications (e.g., airspeed, altitude, vertical speed, commanded thrust) and a resultant failure to accurately forecast the ability to maintain safe flight. The leading consequence of LESA is aircraft stall. A Commercial Aviation Safety Team (CAST) study of 18 loss-of-control events determined that a lack of external visual references (i.e., darkness, instrument meteorological conditions, or both) was associated with flight crew loss of attitude awareness or energy state awareness in 17 events. CAST recommended that, to provide visual cues necessary to prevent loss-of-control resulting from flight crew spatial disorientation and loss of energy state awareness, manufacturers should develop and implement virtual day - Visual Meteorological Conditions (VMC) display systems, such as synthetic vision or equivalent systems. In support of this implementation, CAST has requested studies to support definition of minimum requirements for virtual day - VMC displays to accomplish the intended function of improving flight crew awareness of airplane attitude. In support, NASA is studying the effectiveness of virtual day-VMC displays in improving flight crew awareness of airplane attitude as a function of various display system characteristics, including but not limited to: a) field-of-view; b) presentation/removal of virtual day-VMC displays in unusual attitudes; c) image minification d) optical flow cues, including the display elements over water or featureless terrain and the use of color and texture; and, e) potential unintended consequences (e.g., attentional issues, Crew Resource Management (CRM) impacts). NASA

Further author information: Jarvis (Trey) Arthur, E-mail: Trey.Arthur@nasa.gov, Telephone: 1-757-864-6609

is primarily studying the design requirements for implementation of Synthetic Vision (SV) for unusual attitude recognition and recovery on the head-down, Primary Flight Display (PFD) but attention is also directed toward Head-Up Displays (HUDs) due to the emergence and growing prevalence. Further, Head-Worn Displays (HWDs) are envisioned as a possible equivalent to a HUD in commercial and general aviation and appropriate research attention is being given.

A simulation experiment was conducted to evaluate whether the HWD can provide an equivalent or better level of performance to a HUD in terms of unusual attitude recognition and recovery. The HWD was outfitted with ambient vision displays which were varied (on/off) as an independent variable in the experiment to examine the effects of a large-horizontal field-of-view HWD for attitude awareness. The simulation experiment was conducted in two parts: 1) short unusual attitude recovery scenarios where the aircraft is placed in an unusual attitude and a single-pilot crew recovered the aircraft; and 2) a two-pilot crew operating in a realistic flight environment with “off-nominal” events to induce unusual attitudes.

1.1 Unusual attitude recognition and recovery

Experienced commercial flight crews are trained and are familiar with unusual attitude recognition and recovery using a typical “blue-over-brown” PFD. This recognition and recovery task is a learned behavior, primarily using a PFD as an abstract representation of the aircraft attitude in relation to the earth/horizon, on a display that subtends only within the pilot’s foveal vision. The addition of SV on the PFD, per the CAST initiative, is intended to remove the confound of this abstraction, eliminating any inside-out/outside-in attitude indicator design predicament.² However, there are some unique conditions with a SV-enhanced PFD that are being researched in terms of attitude awareness. For example, over-water flights with aircraft equipped with SV-enhanced PFD could create a “blue-over-blue” display that may cause attitude confusion.

NASA is also investigating an SV presentation that spans the entire horizontal field-of-view of a glass flight deck display that could potentially enhance attitude awareness by stimulating the pilot’s visual dominance through the mechanism of Ambient Vision (AV).³ This presentation⁴ – termed the Background Attitude Indicator (BAI) – is an extension of the Malcolm Horizon in that it is a combination of the Malcolm Horizon with SV^{5,6} as well as previous BAI work with fighter aircraft designs.⁷

1.2 Ambient vision - HUD and HWD

For HUDs, the stimulation of AV is not practical. An extended horizon line and compressed pitch ladder symbology are strategies to prevent and recover from an unusual attitude, respectively.⁸ The use of SV for unusual attitude recognition and recovery on the HUD is an active research issue.

NASA has been researching the use of HWDs⁹ on commercial flight decks as a HUD equivalent display.¹⁰ The HWD research focused on surface operations and low visibility approach and departure operations.¹¹ The overall goal of this research is to evaluate a HWD system as a replacement for a standard flight HUD. If this equivalence can be shown, then the unique capabilities of the HWD - that is, unlimited field-of-regard head-up operations for piloted surface operations¹¹ - can be capitalized.

Sharkey¹² showed the potential of AV for attitude awareness using a Helmet-Mounted Display (HMD). The present research extends the BAI concept and the concept of AV stimulation - placing peripheral displays in the pilot’s view to improve spacial awareness - using the HWD and SV.

2. EXPERIMENT

The main objective for this experiment is to determine if a HWD system can provide equivalent performance compared to a HUD for unusual attitude recognition and recovery. The data from this work should help quantify the characteristics that define an equivalent display as well as potentially enhanced an aircrew’s attitude awareness to prevent LOC. Secondary objectives include the influence of color symbology on the HWD.

2.1 Simulation facility

This experiment was conducted in the NASA Langley Research Center (LaRC) Research Flight Deck (RFD) simulator albeit *without* motion cueing (Fig. 1). The RFD was configured to mimic the instrument panel of current state-of-the-art commercial transport aircraft, with four 10.5” vertical by 13.25” horizontal, 1280x1024 pixel resolution, color displays tiled across the instrument panel. Also, the RFD included a mode control panel, Flight Management System (FMS), control display units, and hydraulic-actuated side-stick control inceptors. A collimated Out-The-Window (OTW) scene provided approximately 200° horizontal by 40° vertical Field-Of-View (FOV) at 26 pixels per degree.



Figure 1. The Research Flight Deck simulator at NASA Langley Research Center.

The test facility had navigation and communication facilities simulation that replicated realistic voice communication during approach, taxi-out, and takeoff scenarios. The communications included pre-recorded voice communications to stage most of the scenarios. A researcher also provided simulated Air Traffic Control (ATC) directives and information to augment the pre-recorded ATC audio, as well as intervention when required (i.e., when a crew queried ATC with questions regarding a clearance).

2.1.1 Head-Worn Display

The HWD used in this experiment is shown in Fig. 2. The HWD system was a tiled array of four Scorpion Display ModulesTM coupled with a single head tracker known as the HOBITTM sensor. The front displays immediately in front of the eyes were configured for bi-ocular viewing (same image both eyes) with inter-ocular pupil adjustment. The AV displays (i.e. the two outer displays) were centered approximately 50° off of boresight. The display modules were mounted to a headband and were counter balanced with weights mounted on the back of the headband. The overall weight of the head-worn portion of the HWD was 2.0 kg and pre-test fitting showed that pilots were able to wear the HWD for 1.0 - 1.5 hours with little discomfort. The characteristics of a single HWD display module and the HUD are given in Table 1.

Table 1. Display Specifications

	HWD Single Display Module	HUD
Resolution	800 (H) x 600 (V)	1400 (H) x 1050 (V)
Field-of-View	26° H x 20° V	46° H x 34.5° V

2.1.2 Head-Up Display

The HUD used in this experiment was a Rockwell Collins HGS-6700.

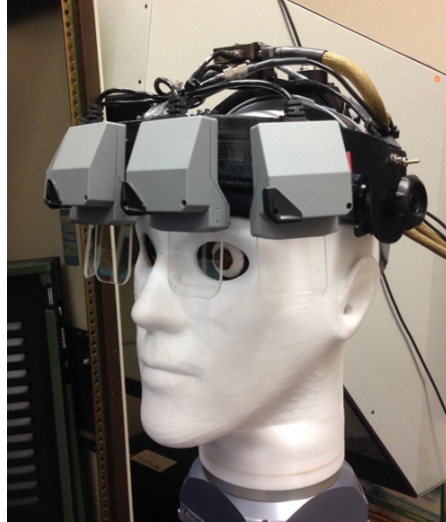
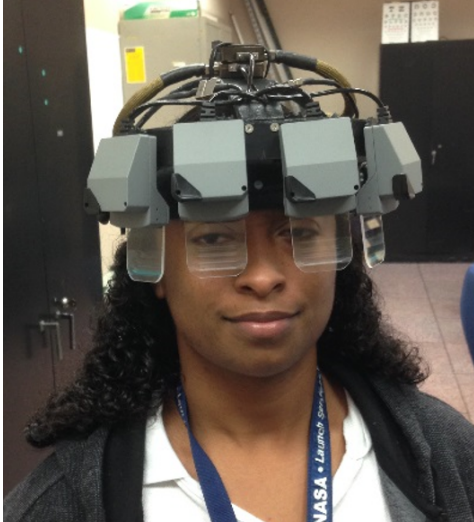


Figure 2. The HWD system used in the experiment. The HWD system consisted of 4 tiled Scorpion optics and a single coupled head tracker.

2.1.3 HWD/HUD Symbology

The HWD was designed to replicate the primary HGS-6700 HUD symbology functionality (Fig. 3). The HWD depicted a “Virtual-HUD” concept where imagery and typical HUD symbology were rendered aircraft-referenced – displaying HWD information as if the pilot were looking at the area where a HUD combiner glass would be mounted. In addition to typical flight symbology, a SV terrain image was rendered on the HWD and the HUD which was conformal to the outside world. The HWD did not completely replicate all HGS-6700 symbology (e.g., airspeed and altitude were depicted using dials versus HUD tape formats) due to field-of-view and other limitations; however, the primary functions were replicated as confirmed by pre-test pilot checkout and acceptance testing.

The flight symbology set during unusual attitude conditions was typical symbology for a commercial transport HUD. When the aircraft was in an unusual attitude, the pitch ladder symbology changed to non-conformal pitch ladder presentation (compressed) on the HWD and the HUD (Fig. 4). For the HWD, the unusual attitude symbology also changed to display-referenced (i.e., the waterline symbology was centered on the display and the head tracker had no effect on the display). The displays would remain in the unusual attitude mode until aircraft recovered to a nominal state (i.e., straight-and-level flight). Once recovered, the symbologies on the HWD and HUD would return to a conformal state and the HWD would return to a “Virtual-HUD” state.¹³

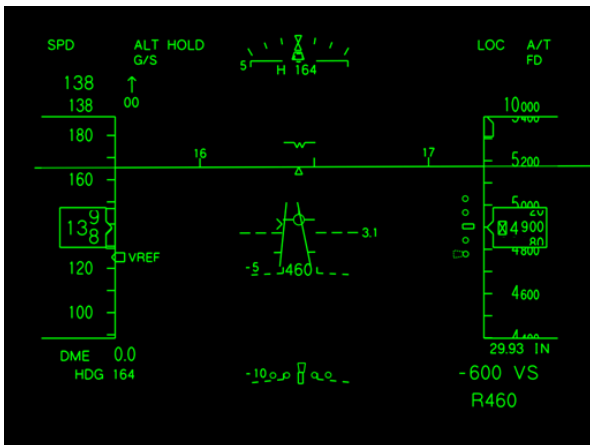


Figure 3. The HUD flight symbology (left) and the HWD flight symbology (right).

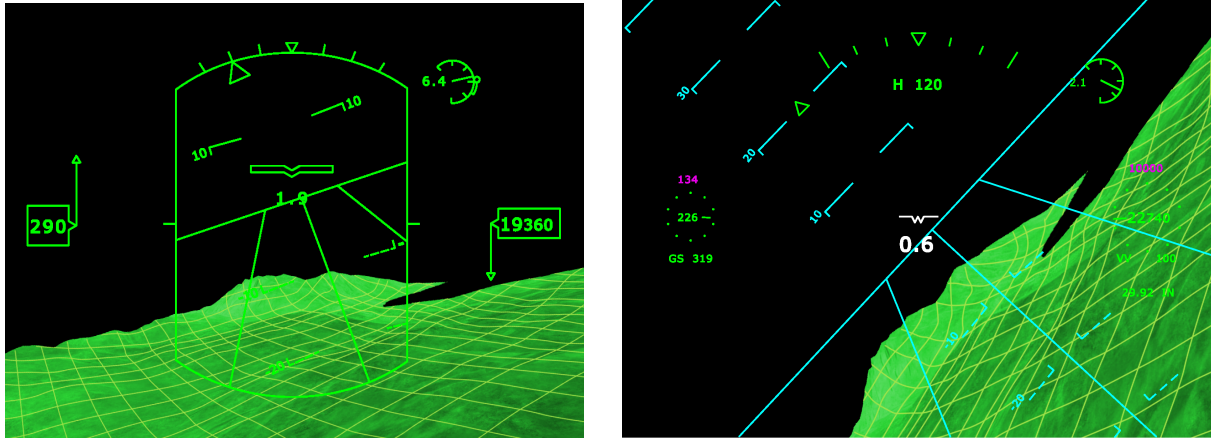


Figure 4. The upset recovery symbology and synthetic terrain for the HUD (left) and the HWD (right).

For departures, a typical takeoff symbology set was used, identical for both the HUD and HWD. The takeoff symbology set is very similar to the flight symbology with the addition of the ground localizer line to aid in centerline tracking during takeoff roll.

2.1.4 Head-Down Displays

The variation of the head-down displays used in the experiment are shown in Figs. 5-7. Fig. 5 shows the Baseline head-down concept without SV. Fig. 6 shows the SV-BAI concept. Fig. 7 shows the head-down display SV-Head-Down Display (HDD) concept (no BAI).

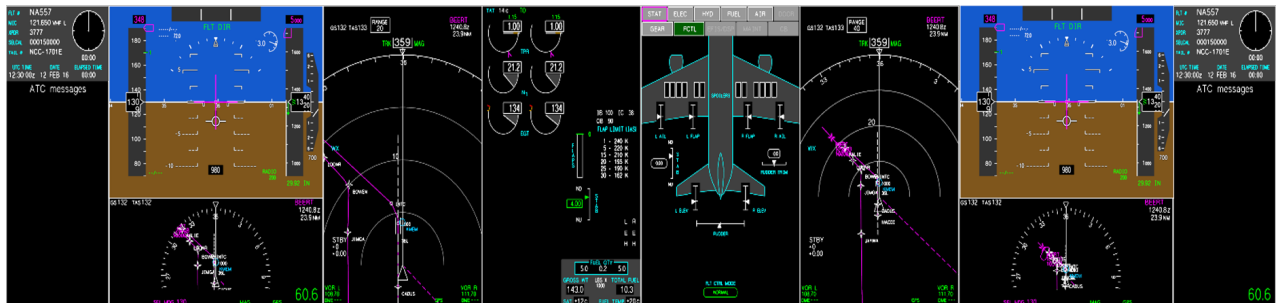


Figure 5. The HDD Baseline displays (HDD-No SV).

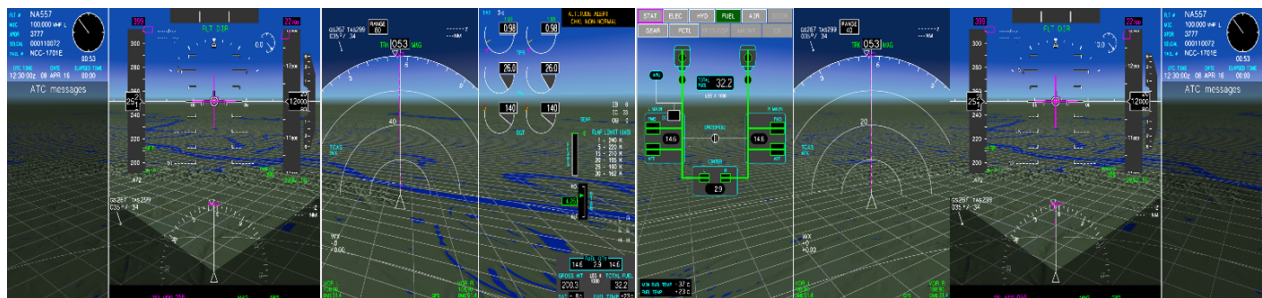


Figure 6. The HDD SV-BAI displays.

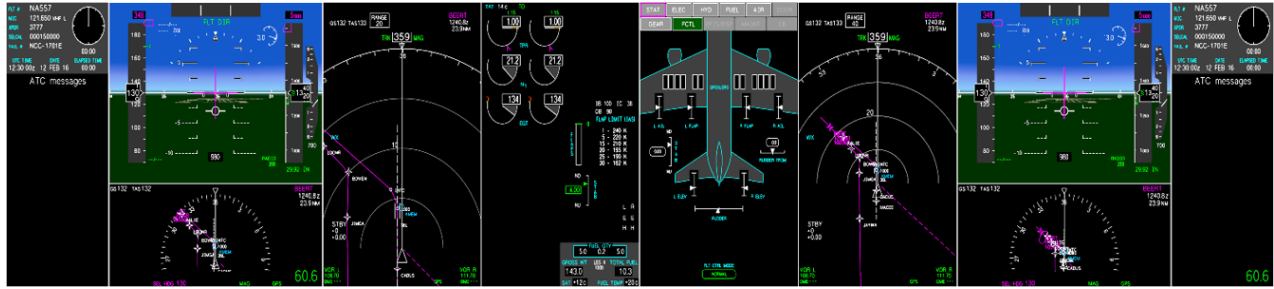


Figure 7. The HDD-SV display concept.

2.2 Evaluation pilots

Twelve commercial flight crews from various US airlines participated in the experiment. The Evaluation Pilots (EPs) were paired based upon their current employer to minimize inter-crew differences in Standard Operating Procedures and CRM procedures. The Captain was the Pilot Flying (PF) and sat in the left seat. The First Officer was the Pilot Monitoring (PM) for the duration of the experiment; thus, crew members did not switch roles for the duration of the experiment. Each pilot held an Airline Transport Pilot rating. Captains had an average of 33 years experience, all with HUD flight time which averaged 1,326 hours, though 5 Captains had less than 1000 hours of HUD flight experience. First Officers had an average of 30 years experience.

2.3 Flight crew training

The EPs were given a 45-minute classroom briefing to explain the display concepts and the evaluation tasks for the experiment. Evaluation crews then received familiarization training in the RFD before data collection began. At the end of each familiarization training block, pilots were asked if they wanted to repeat any of the training. At the end of the day, a post-test interview was conducted to solicit the crew's comments on the experiment. The total duty time for an evaluation crew was approximately 8 hours.

2.4 Methodology

The experiment was conducted in two parts:

1. Unusual Attitude Recovery (UAR) data collection runs: 24 short (about one minute duration), single-pilot data collection runs where the aircraft was placed in an unusual attitude and the Captain was asked to recover.
2. Special Purpose Operations Training (SPOT) data collection runs: eight operationally realistic scenarios of which 4 scenarios were designed to induce unusual attitudes.

Upset recovery data collection runs were conducted using three display conditions: 1) HDD; 2) HUD; and the 3) HWD. The data runs were blocked by display device (HDD, HUD, or HWD) within an operation block (UAR/SPOT). The experiment was blocked by display to minimize the need for EPs donning and removing the HWD between runs. During data trials using the HWD, the HUD was stowed. The EPs (Captains only) wore the HWD for approximately 45 minutes.

2.5 UAR Evaluation task

All expected procedures and appropriate protocols were briefed prior to the test for each crew, and training was provided to familiarize crews with operational procedures prior to data collection.

For the UAR data collection runs, there were four initial condition states of the aircraft (see Table 2). At the start of the UAR scenario, all aircraft displays were blanked as the aircraft automatically maneuvered into an unusual attitude. Since the displays were blanked, EPs were not asked to close their eyes. During the blank times, EPs were allowed have their hands on the controls. When the desired attitude was reached, the displays

Table 2. Aircraft initial condition for UAR scenarios.

Scenario	Pitch	Roll
A	Up 30°	Left 60°
B	Up 30°	Right 60°
C	Down 30°	Left 130°
D	Down 30°	Right 45°

were unblacked and an audible tone was played to alert the pilot to recover the aircraft. The displays and flight conditions were dynamic. Pilots were instructed to recognize and recover from the unusual attitude and verbally call out “recovered” at which time the run would end.

The UAR experiment matrix is shown in Table 3. The absence/presence of SV terrain on the HDDs was also varied. The HUD display condition used an SV presentation that was identical to the monochromatic HWD. The SV terrain used in this experiment consisted of a typical industry-type terrain with grid lines, range rings and major bodies of water. For the HWD, color and the absence/presence of AV were varied.

Table 3. UAR experiment matrix showing all data points across all pilots.

Scenario	HDD		HUD	HWD			Total
	no SV	SV-BAI		color/no AV	color/AV	green/no AV	
A	12	12	12	12	12	12	72
B	12	12	12	12	12	12	72
C	12	12	12	12	12	12	72
D	12	12	12	12	12	12	72
Total	48	48	48	48	48	48	288

2.6 SPOT Evaluation task

The SPOT scenarios were designed by NASA and subject matter experts using a sequence of off-nominal events that created challenging flight and workload conditions that may ultimately lead to an unusual attitude without timely pilot intervention. For the SPOT data collection runs, there were four unusual attitude scenarios (see Table 4). The unusual attitude conditions from which the EPs recovered were not as precisely controlled as the UAR evaluation task, but they were performed within an operational context, making the evaluations more realistic, challenging, and multi-faceted. Details are provided in Ellis et al.¹⁴

Eight (8) SPOT scenarios in total were flown; four nearly identical “distracting” scenarios were created which mirrored the four SPOTs with off-nominal events but with the off-nominal events removed. These additional scenarios were not challenging and did not lead to unusual attitude conditions. SV terrain was present on all Display Concepts and there was no HDD Baseline condition for the SPOTs. The SV-BAI concept was not used for any of the SPOT tasks. Scenarios lasted on average ten minutes.

Post-run questionnaires were given to both EPs after each scenario, and consisted of a 3-point Situation Awareness Rating Technique (SART)¹⁵ form and a NASA Task Load Index (TLX) workload rating.¹⁶ These questionnaires were given immediately after the end of each data trial. At the end of the day, a post-test questionnaire was administered to the evaluation Captain.

3. RESULTS

3.1 Quantitative results

Quantitative data (i.e., correct recovery inputs, time-to-first correct inputs, etc.)¹⁷ as well as qualitative measures (i.e., questionnaires, workload and situation awareness metrics, pilot opinion) were recorded and used in a detailed

Table 4. SPOT experiment matrix showing total data points.

SPOT	Off-Nominal Event	HDD	HUD	HWD no AV	HWD with AV	Total
1	radar altimeter fail	3	3	3	3	12
2	fuel leak	3	3	3	3	12
3	degraded autopilot	3	3	3	3	12
4	wake encounter	3	3	3	3	12
Total		12	12	12	12	48

data analysis to answer the purpose of the research.

3.1.1 UAR performance

A scoring system to used to assess the pilot's ability to recover from an unusual attitude. This UAR score was derived from three control elements: pilot's initial pitch input, pilot's initial roll input and the pilot's initial throttle input. If the pilot made the correct initial input to recover from the unusual attitude, a score of 1 was given. An incorrect input resulted in a score of -1. The total UAR score was calculated by adding the three control elements resulting in a score range from -3 to 3. In addition to the UAR score, the time-to-first correct input was measured.

The following comparisons can be made for the quantitative variables of interest:

1. Effect of BAI: The difference between Baseline (no SV) and SV with BAI performance data in the Full Color HDD case.
2. Effect of HWD Color: The difference between Monochrome and Full Color performance data in the SV HWD no AV case.
3. Effect of AV (HWD): The difference between the HWD with AV and the HWD with no AV performance data in the SV Full Color HWD cases.
4. HWD Equivalence to HUD: The difference between HUD and the monochrome HWD performance data. The differences are tested using t-tests and a threshold p-value of 0.05/4 to account for multiple testing (collapsing the data across the 4 UAR scenarios).

With the exception of time to pitch input for Scenario D (see Table 2), there were no significant differences for any of the four comparisons above. For the one statistically significant result for time-to-first correct pitch input for the Scenario D, the HUD time was found to have a mean value of 0.77 seconds compared to a mean value of 1.45 seconds for the HWD with a difference of 0.68 seconds.

Quantitative measures of nose-up unusual attitude recovery testing has been found to be problematic because of ambiguity in defining the correct inputs to successfully recover large transport category aircraft from these conditions.¹⁸ Nose-low recoveries contain none of these ambiguities and clearly show the display effect differences, if present. In the following, the UAR recovery scores, time-to-first correct pitch input, and time-to-first correct roll inputs are shown to highlight the results:

- In Figure 8, the data indicates that adding SV with a BAI on the HDD produces almost no differences in performance. Nearly all UARs were correct (scores of 3) but with four outliers and five outliers for the BAI and no-SV conditions, respectively, where one incorrect input was made by the EPs. The times-to-first pitch and roll correct inputs were nearly identical. The roll inputs were less than 1.0 second on average with very little deviation.
- In Figure 9, the effect of color on the HWD is shown for UAR performance and again, nearly identical performance data are shown albeit with more scatter about the mean data for the color condition.

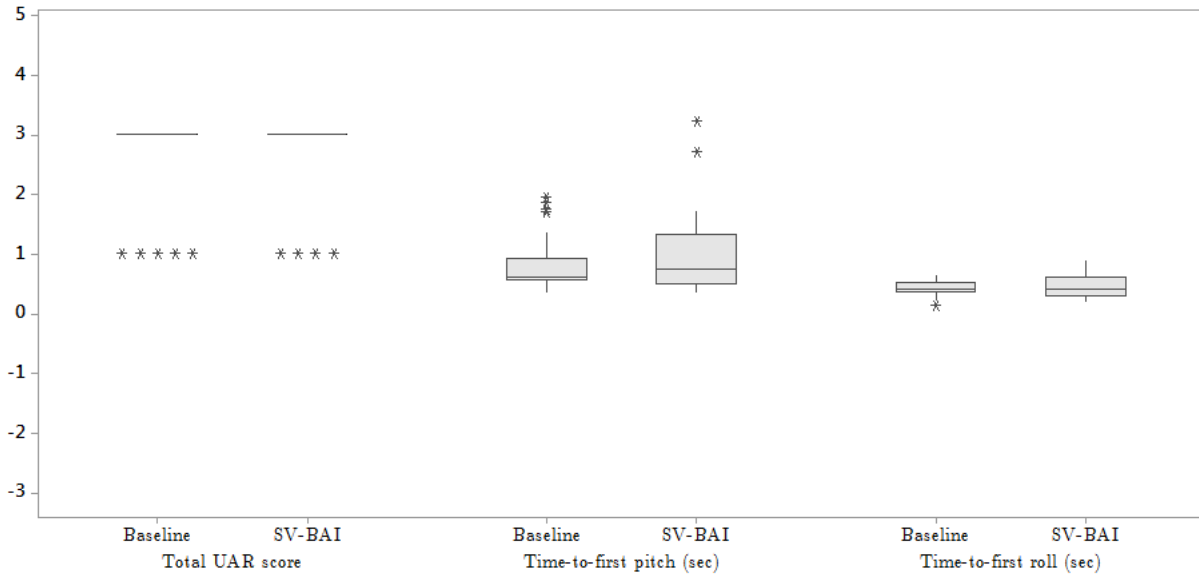


Figure 8. Boxplot data for the Baseline (no SV) versus SV-BAI HDD.

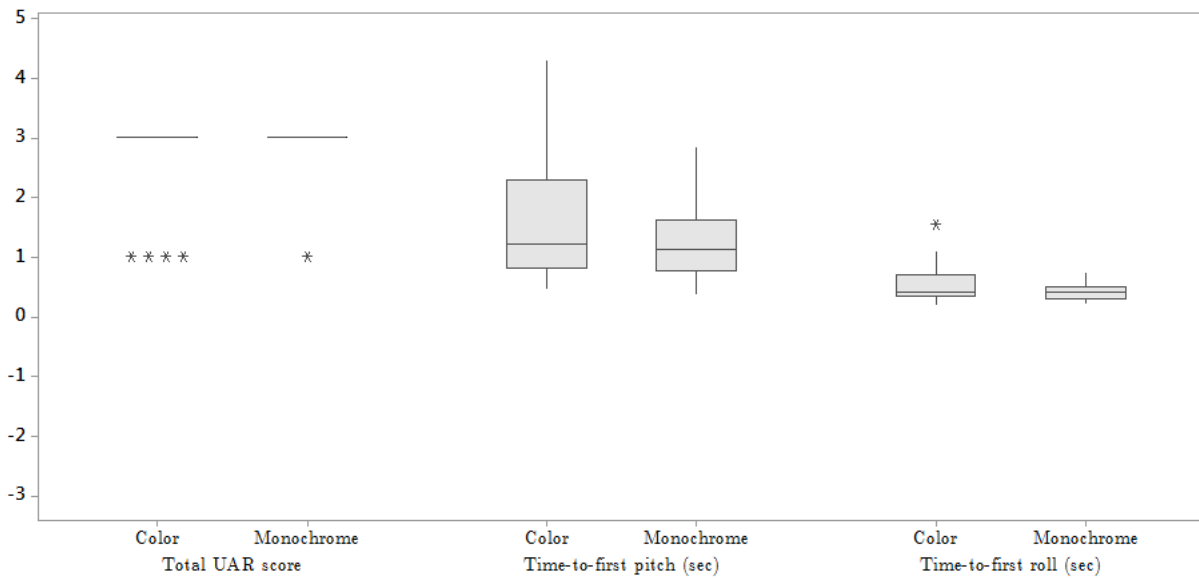


Figure 9. Boxplot data for the HWD Color versus Monochrome Display Concept.

- In Figure 10, the trend data suggests that the AV caused quicker time-to-first correct roll and pitch inputs, but the UAR correctness score was slightly worse than without the AV stimulation on the HWD.
- In Figure 11, the trend data shows nearly identical performance between the HUD and HWD but the HUD data shows evidence of more incorrect inputs when flying the HUD than the HWD.

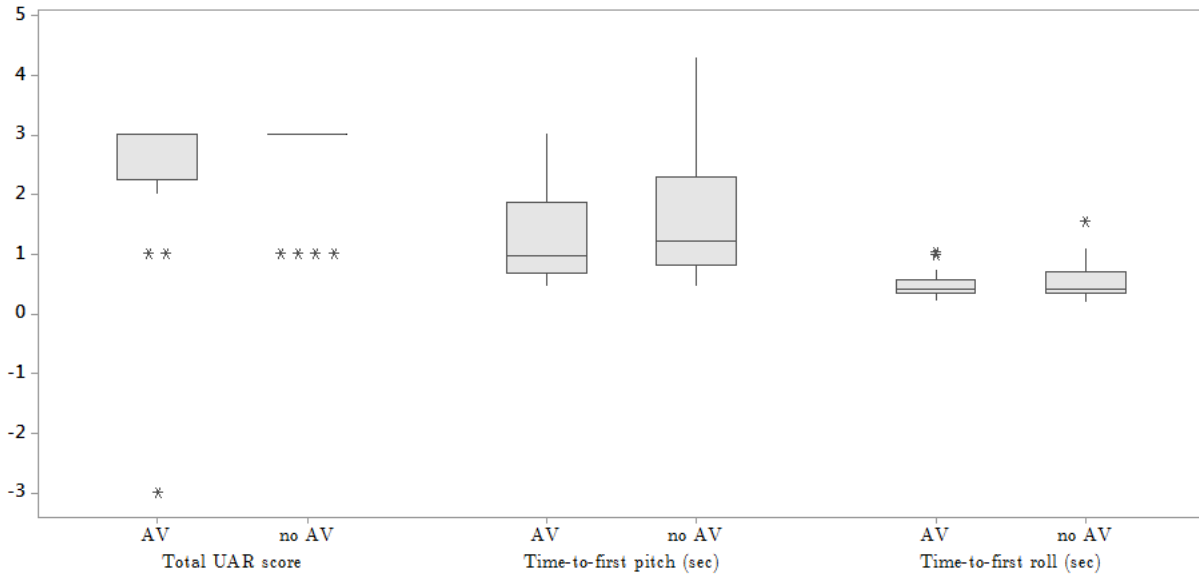


Figure 10. Boxplot data for the HWD with AV versus HWD without AV Display Concept.

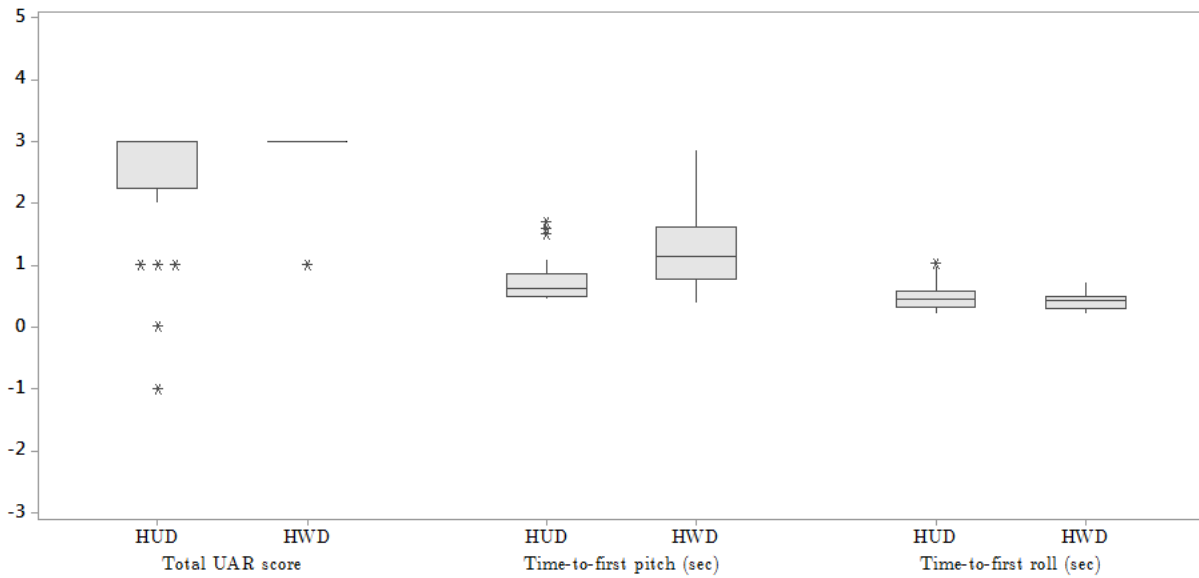


Figure 11. Boxplot data for the HWD Monochrome (no AV) versus the HUD Display Concept.

3.1.2 SPOT performance

Analysis of Variance (ANOVA) showed no statistically significant results for any of the four constructs (UAR score, time-to-first correct pitch input, time-to-first correct roll output, and number of roll reversals) for the four display conditions (HDD, HUD, HWD with AV and HWD with no AV) within each SPOT event.

3.2 Qualitative results

3.2.1 UAR Situation Awareness

A 3-point SART was administered after each run. SART provided an assessment of the Situation Awareness (SA) based on the pilot's subjective opinion of three dominant components: demand on the pilot's resources, supply of resources, and understanding of the situation. Pilots rated their perception of the impact of these components using scales from 0 to 100. A total SART score was derived using the formula: $SA = Understanding - (Demand - Supply)$. The range of scores from the application of the formula is from -100 for extremely low SA to 200 for extremely high SA.

Hypothesis testing was done using t-tests for the four UAR scenarios to determine if the mean score difference among all pilots. Because the hypothesis tests were done simultaneously, the threshold for significance was adjusted using the Bonferroni correction, which divided the usual threshold of 0.05 by the number of tests. If the p-value of a test was less than $0.05/4 = 0.0125$, then the result was deemed significant, or unlikely to be due to chance, and the null hypothesis can be rejected.

None of the t-tests were significant for situation awareness for any of the comparisons (Baseline to SV-BAI, Monochrome to Full Color, AV to no AV, and monochrome HWD to HUD) across the three display types (HDD, HWD, HUD).

3.2.2 UAR Workload

Workload was assessed via the NASA TLX subjective workload scale.¹⁶ The NASA TLX consisted of six scales associated with mental, physical, temporal demand, performance, effort, and frustration level. The raw TLX scoring was used based on Hart and Staveland.¹⁶

For workload, with the exception of the color comparison for Scenario C, all other comparisons were not statistically significant. The t-test for UAR scenario C was significant showing the Full Color condition to have a significantly higher workload than the Monochrome condition (see Fig. 12). The average pilot workload for that scenario was 13.1 points higher with the Full Color display than with Monochrome. The other scenarios did not show evidence of a difference in workload based on color type in the HWD display.

3.2.3 SPOT Situation Awareness and Workload

For the SPOT scenarios, an ANOVA test with the Bonferroni correction was conducted to determine the effects of the four display locations (HDD, HUD, HWD no AV, and HWD with AV) on SA and workload.

There were no statistical differences between any of the four display concepts for the SPOTs in terms of SA or workload.

3.2.4 Crew coordination workload

At the end of each SPOT data trial, both crew members were asked to rate the other crew member's workload on a scale from 0 to 100 (0 meaning they thought the other pilot had nothing to do and 100 meaning the other pilot was completely task saturated). Pilots were instructed not to discuss or share their rating of the other pilot.

No significant effect was found for Display, $F(3, 63) = 0.473, p > 0.05$ or SPOT scenario, $F(7, 63) = 0.994, p > 0.05$.

3.2.5 Paired comparisons

Post-test, paired comparison questionnaires^{19,20} were administered following the UAR data trials to assess the constructs of 1) synthetic vision head-down, 2) color symbology on the HWD and 3) absence/presence of AV. After the SPOT data trials, a second paired comparison was administered with the construct of the absence/presence of the AV for the HWD condition. For each of these UAR and SPOT constructs, EPs compared the three display conditions: 1) HDD, 2) HUD and 3) HWD.

For the UAR scenarios, there was a significant main effect for display for paired comparison ratings of UAR performance, $F(5, 55) = 9.148, p < 0.01$. Pilots provided paired ranking across display concepts on efficacy of

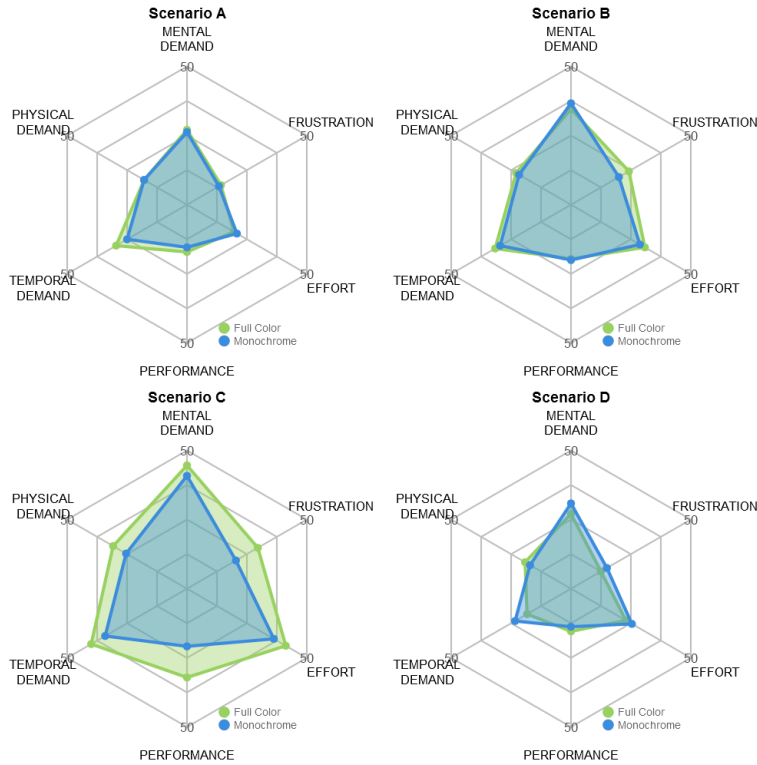


Figure 12. Radar plot of the NASA TLX ratings for UAR comparison between the Monochrome and Full Color HWD display concept.

display for unusual attitude recovery. Post-hoc analysis revealed that pilots ranked the HWD Color with AV and head-down SV with the BAI concepts significantly higher than all display concepts (but not significantly different from each other). The HWD Color without AV was also found to be ranked higher than the HWD Monochrome display concept, but not significantly from the Baseline or the HUD. The Baseline, HUD, and HWD Monochrome were not significantly different from each other.

For the SPOT scenarios, there was a significant main effect found for geomeans on the paired comparison for display concept, $F(3, 33) = 11.408, p < 0.0001$. A Bonferroni post-hoc test revealed that pilots rated the head-down SV concept (BAI off) to be significantly better for unusual attitude recovery than the head-up display (with unusual attitude recovery symbology) ($p = 0.012$). The HUD was a monochrome SV display with UAR symbology and the HDD was an industry representative SV PFD. No other effects were found suggesting that the head-worn display concepts, with and without augment vision, were rated comparable to the SV PFD for the SPOT scenarios.

3.2.6 Post-test questionnaire

At the end of all simulation trials, pilots were given a questionnaire to rate their agreement using a 5-point Likert scale. The questionnaire contained the following 12 questions:

1. Please rate the clarity of the HWD symbology.
2. Please rate overall comfort of the HWD.
3. Please rate the ease of use regarding the brightness control.
4. Please rate the usefulness of the color symbology (versus monochrome symbology).

5. I did not experience display glare while using the HWD.
6. The HWD's vertical field-of-view was sufficient to provide conformal display of flight guidance information throughout the intended operational envelope.
7. The HWD's horizontal field-of-view was sufficient to provide conformal display of flight guidance information throughout the intended operational envelope.
8. The synthetic vision was helpful in increasing situation awareness during tasks.
9. I did not experience eye strain while using the HWD system.
10. The HWD system did not cause me to experience headaches.
11. The HWD peripheral display provided attitude awareness.
12. Based on my overall experience with HUDs, I would consider this HWD equivalent to a HUD.

EPs responses are shown in the boxplots in Figs. 13 - 24.

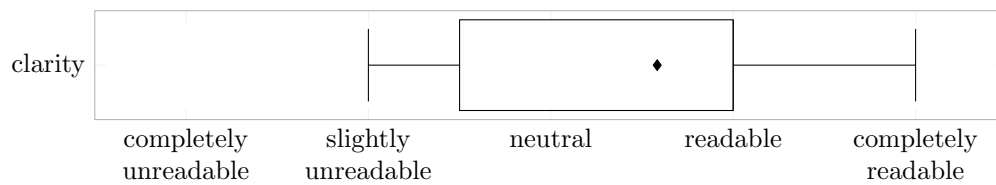


Figure 13. EPs rating of the clarity of the HWD symbology. $N = 12$

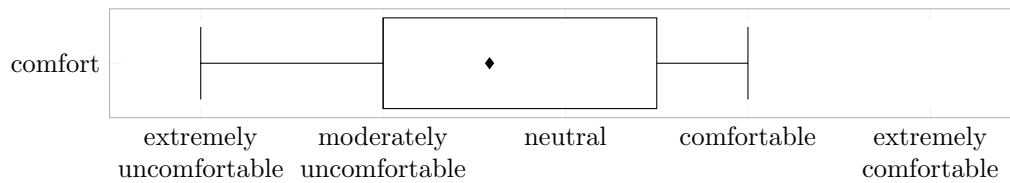


Figure 14. EPs rating of the comfort of the HWD. $N = 12$

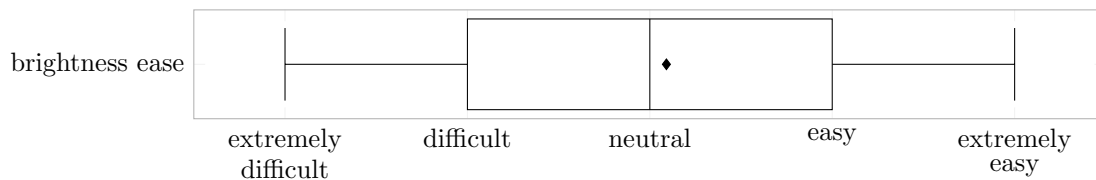


Figure 15. EPs rating of the ease of using the HWD brightness control. $N = 12$

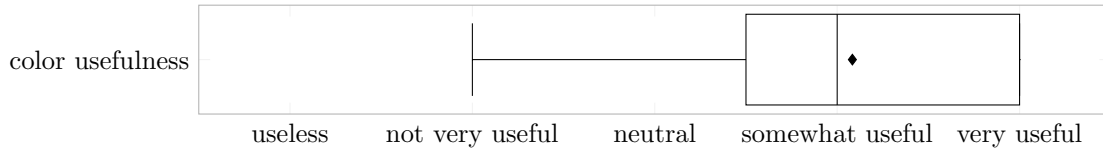


Figure 16. EPs rating of the usefulness of color on the HWD. $N = 12$

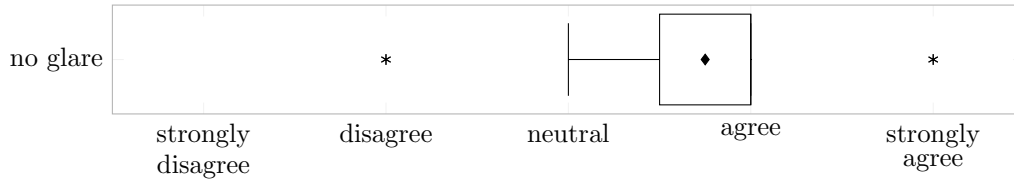


Figure 17. EPs rating of not experiencing glare on the HWD. $N = 12$

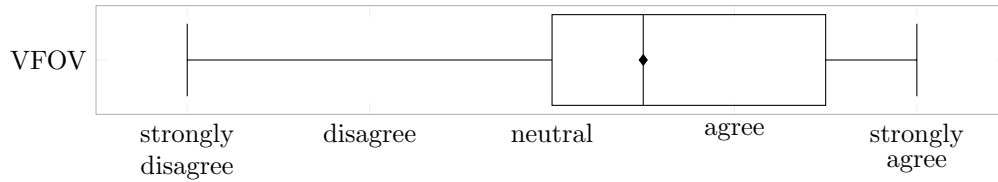


Figure 18. EPs rating of the sufficiency of the vertical field-of-view of the HWD. $N = 12$

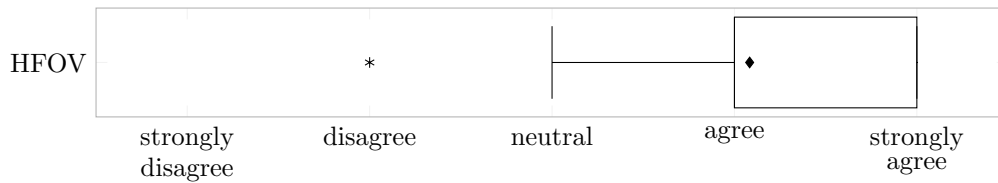


Figure 19. EPs rating of the sufficiency of the horizontal field-of-view of the HWD. $N = 12$



Figure 20. EPs rating of the helpfulness of synthetic vision. $N = 12$

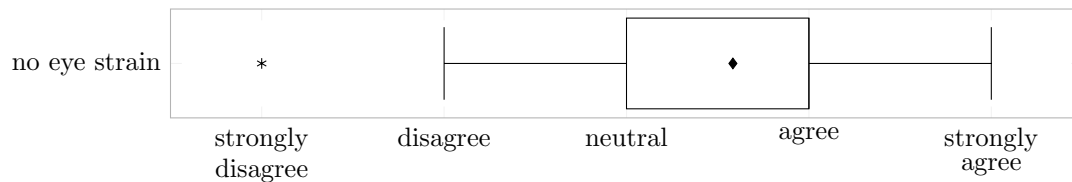


Figure 21. EPs rating of not experiencing eye strain. $N = 12$

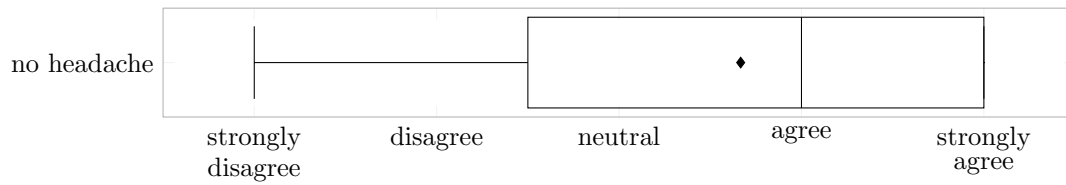


Figure 22. EPs rating of not experiencing a headache. $N = 12$

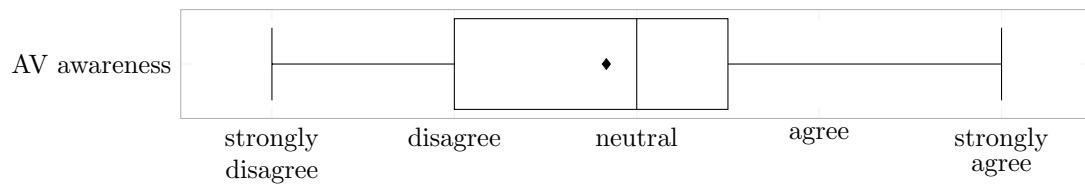


Figure 23. EPs rating that the AV displays provided attitude awareness. $N = 12$

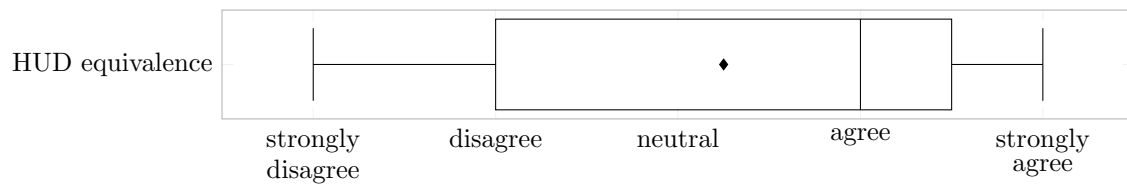


Figure 24. EPs rating that the HWD is equivalent to the HUD. $N = 12$

4. DISCUSSION OF RESULTS

This study was conducted to examine the effects of ambient vision on unusual attitude recovery. In addition, secondary objectives of the experiment were to compare the effects of color on the HWD, the equivalence of the HUD to the HWD and the effects of SV terrain on upset recovery. The quantitative data showed no statistically significant effects for any of the constructs except for time-to-pitch between the HWD and the HUD for Scenario D (initial condition of down 30 deg, rolled right 60 deg). After reviewing the video and audio data from the Scenario D HUD/HWD data runs, there were no operationally significant differences between the HUD and the HWD. Further, pilots did not comment on any differences between the HUD and HWD for that particular scenario.

There were no statistically significant results between any of the display condition in terms of SA for either the UAR or the SPOT data trials. In other words, display type (HDD, HUD, HWD), color and head-down SV terrain did not affect the pilots' SA ratings. For workload, there was 1 statistically significant result for the UAR Scenario C (initial condition of down 30 deg, rolled left 130 deg) where pilots rated the color HWD concept to have a higher workload compared to the monochrome HUD. Scenario C was also rated as being the highest workload compared to the other 3 scenarios. In the post-test questionnaire, the pilots' mean rating indicated they thought color on the HWD was "somewhat" helpful. This suggests that pilots like the decluttering aspects of color; but, the color scheme used for the HWD recovery symbology may not be optimized for the task. Further study would be needed to draw conclusions on the advantages and disadvantages on using color for high workload unusual attitude recovery. The post-test HWD questionnaire responses to Questions 4 (color usefulness) and 8 (SV usefulness) elicited "agree" to "strongly agree" responses.

The pairwise comparison ratings show that pilots preferred the head-down SV display with the BAI and the color HWD with AV. Some pilots commented that the HWD AV and the head-down SV with BAI display helped in commanding roll rate by having a large horizon in their view. Other pilots commented that they did not notice the AV displays on the HWD; thus, AV was not a benefit to them. The mean for the HWD AV display providing attitude awareness (question 11) was just below "neutral" in post-test rating with a wide range of comments regarding the AV display (see Fig. 23). Future studies could examine the content, location and size of the AV displays. This awareness is also a factor of the vertical and horizontal field-of-view of the HWD.

The post-test questionnaire allowed pilots to give their opinion after seeing all concepts. The mean pilot rating for the HWD was between neutral and moderately uncomfortable (Question 2). The extra weight of the AV displays was a trade-off in order to collect data with a wide FOV. Question 7 (horizontal field-of-view) elicited "agree" to "strongly agree" responses. All pilots agreed with the statement that SV on the HWD was helpful in upset recovery with the mean trending towards "strongly agree." On the question of HWD equivalence, pilots' ratings were nearly equally distributed between agree and disagree. While some pilots felt the HWD allowed them to do the upset recovery, the discomfort and latency effects caused them to rate the HWD lower than the HUD.

Overall, pilots were able to recover from an unusual attitude for all of the display concepts. There were no performance, SA, or workload advantages for any particular display. Pilots in this experiment were experienced and well trained on recovery using the a typical "blue over brown" attitude indicator on the PFD. The data suggest that the HWD was equivalent to today's displays for upset recovery; however, the HWD used in this test would require ergonomic improvements for a typical commercial operation.

5. CONCLUSIONS

The goal of the research presented in this paper was to investigate operational equivalence between a head-tracked HWD system and a HUD in terms of unusual attitude and upset recovery. The results showed that there were few statistical differences in the crew's performance across the display concepts tested. Future work includes determining the ideal size, placement, and content of the AV displays. Future flight testing of the HWD system should include determining latency requirements and the effects of turbulence under various flight deck lighting conditions. Though a Virtual HUD concept for the HWD seems intuitive for fastest certification and acceptance, the latency and ergonomic requirements will need to be understood before full certification of a HWD system could be realized. NASA is continuing to research the advantages of an unlimited field-of-regard HWD. Previous

NASA studies have shown the potential advantage of a HWD for surface operations in particular, but the full gamut of commercial flight operations has not been fully explored.

REFERENCES

1. Commercial Aviation Safety Team, "Airplane state awareness joint safety analysis team final report," tech. rep., CAST, Washington, DC, 2014.
2. F. Previc and W. Ercoline, "The "outside-in" attitude display concept revisited.," *The International Journal of Aviation Psychology* **9**, Issue 4, pp. 337–401, October 1999.
3. F. Previc and W. Ercoline, *Spatial Disorientation in Aviation (Progress in Astronautics and Aeronautics)*, American Institute for Aeronautics and Astronautics, 2004. ISBN 978-1563476549.
4. R. Bailey, K. Ellis, and C. Stephens, "Test and evaluation metrics of crew decision-making and aircraft attitude and energy state awareness," in *2013 AIAA InfoTech@Aerospace Conference*, AIAA, (Boston, MA), 2013.
5. *Peripheral Vision Horizon Display (PVHD)*, **CP-2306**, (Edwards, CA), NASA, 1984.
6. R. Comstock, L. Jones, and A. Pope, "The effectiveness of various attitude indicator display sizes and extended horizon lines on attitude maintenance in a part-task simulation," in *Human Factors and Ergonomics Society 47th Annual Meeting*, Human Factors and Ergonomics Society, (Denver, CO), October 2003.
7. K. Liggett, Reising, J.M., and D. Hartsock, "Development and evaluation of a background attitude indicator," *The International Journal of Aviation Psychology* **9**, Issue 1, pp. 49 – 71, 1999.
8. S. M. Cone and J. A. Hassoun, "Attitude awareness enhancements for the F-16 head-up display," Tech. Rep. ASC-TR-92-5017, Air Force Materiel Command, Wright-Patterson AFB, Ohio, October 1992.
9. J. Arthur III, R. Bailey, S. Williams, L. Prinzel III, K. Shelton, , D. Jones, and V. Houston, "A review of head-worn display research at NASA Langley Research Center," in *Head- and Helmet-Mounted Displays XX*, D. D. Desjardins, K. R. Sarma, P. L. Marasco, and P. R. Havig, eds., **9470**, SPIE, (Bellingham, WA), April 2015.
10. R. Bailey, K. Shelton, and J. Arthur III, "Head-worn displays for NextGen," in *Enhanced and Synthetic Vision 2011*, **8041**, SPIE, (Bellingham, WA), April 2011.
11. J. J. Arthur III, L. Prinzel III, R. Bailey, K. Shelton, S. Williams, L. Kramer, and R. Norman, "Head-worn display concepts for surface operations for commercial aircraft," Tech. Rep. NASA/TP-2008-215321, NASA Langley Research Center, Hampton, VA, June 2008.
12. T. J. Sharkey, R. T. Hennessy, and T. W. Marlow, "Display of aircraft state information for ambient vision processing using helmet mounted displays," Tech. Rep. 00-D-20, U.S. Army Aviation and Missile Command, Wright-Patterson AFB, OH, September 2000.
13. T. Frey and H. Page, "Virtual HUD using an HMD," in *Helmet- and Head-Mounted Displays VI*, R. J. Lewandowski, L. A. Haworth, H. J. Girolamo, and C. E. Rash, eds., **4361**, SPIE, (Bellingham, WA), April 2001.
14. K. Ellis, L. Prinzel III, J. Arthur III, and S. Nicholas, "Design considerations for attitude state awareness and prevention of entry into unusual attitudes," in *19th International Symposium on Aviation Psychology*, (Dayton, OH), May 2017. (in press).
15. R. M. Taylor, "Situational awareness rating technique (SART): The development of a tool for aircrew systems design," in *Situational Awareness in Aerospace Operations, AGARD Conference Proceedings*, pp. 3-1 – 3-37, (Aerospace Medical Panel Symposium, Copenhagen), October 1990.
16. S. G. Hart and L. E. Staveland, "Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research," in *Human mental workload*, P. A. Hancock and N. Meshkati, eds., pp. 139–183, Elsevier, (Amsterdam, The Netherlands), 1988.
17. L. Le Vie, "Survey of quantitative research metrics to assess pilot performance in upset recovery," Tech. Rep. NASA/TM-2016-219191, NASA Langley Research Center, Hampton, VA, April 2016.
18. V. Gawron, R. E. Bailey, and A. Randall, "Differences in in-flight unusual attitude recovery performance and workload," in *HFES Aerospace Systems Technical Group*, 2009.

19. M. A. Vidulich, "The use of judgment matrices in subjective workload assessment - the Subjective WORKload Dominance (SWORD) technique," in *Human Factors and Ergonomics Society, 33rd annual meeting*, pp. 1406–1410, (Denver, CO), October 1989.
20. M. A. Vidulich and E. R. Hughes, "Testing a subjective metric of situation awareness," in *Human Factors Society 35th Annual Meeting*, pp. 1307–1311, Human Factors Society, (Santa Monica, CA), 1991.