PREFACE

The Have Rhino test team wishes to thank Tom “Mach” Schnell, Major Patrick “Dice” Highland, Mathew “Beaker” Cover, Carl “Bender” Richey, Zeke “Chip” Gunnink, Jon “Hitch” Stoltz, Brad “Roto” Parker, Nate “Bluto” Smith, Colton “Colt” Thompson, and Katharine “Woody” Woodruff for their outstanding contributions. The hard work of the University of Iowa Operator Performance Laboratory (OPL) team throughout every step of the test program from planning, implementation, execution, and data processing was essential. OPL represents a clear state-of-the art in airborne HSI testing and engineering effectiveness. Without their support this project would have simply been the hopes and dreams of six TPS students.

The test team would also like to thank Noel Cox and Riley Shanley from the Air Force Research Laboratory (AFRL) for providing assistance with prototyping and early simulation of the HSI methods. And of course, the team thanks Eric Geiselman and Paul Havig, also of AFRL, for entrusting this project to us and providing support throughout planning and execution.

This page was intentionally left blank.

EXECUTIVE SUMMARY

This technical memorandum presents the results for Have Rhino, a United States Air Force (USAF) Test Pilot School (TPS) test management project (TMP). Have Rhino compared three different human-system interfaces (HSIs): eye-tracking, rhino-pointing (a form of head or nose tracking), and traditional HSI. Testing was requested by the 711th Human Performance Wing (HPW) of the Air Force Research Laboratory (AFRL) at Wright-Patterson Air Force Base (AFB), Ohio. The lead developmental test organization (LDTO) was the Air Force Test Center at Edwards AFB, California. The executing test organization was Class 19B of the USAF Test Pilot School, 412th Test Wing, Edwards AFB. Testing was conducted at Iowa City Municipal Airport in a modified L-29 Delfín, tail number N429GC, owned and operated by the Operator Performance Laboratory (OPL) of the University of Iowa. Testing consisted of 14.5 ground test hours and 14 sorties totaling 21 flight test hours under job order number MT19B800.

The size and interactivity of multifunctional displays (MFDs) grew in the decade prior to 2020, while aircrew continued to drive cursors on displays using hands-on throttle-and-stick (HOTAS) actuators. Touchscreen augmented HOTAS, but required aircrew to remove their hands from flight controls. Consequently, aircrew expressed interest in higher performance, lower workload HSIs for large displays. Eye-tracking and rhino-pointing, which respectively used gaze location or head movement to drive a cursor on an MFD were proposed, though the technological readiness of such technologies remained unproven. The overall test objective for Have Rhino was to compare eye-tracking (ET), rhino-pointing (RP) and traditional tactile input as components of a large area display HSI. This overall test objective was met.

The three HSI methods comprised the system under test. An evaluator station in the rear cockpit of the L-29 hosted the system. This station had a large, programmable, touchscreen display, a helmet with head-tracker and eye-tracker, physiological monitoring, and controls for flight. It also transmitted test and flight data in near real-time to a control room. A team of seven operators with a variety of backgrounds flew two sorties each, executing tasks from the evaluator station. Tasks included designation of a single track symbol, expansion and designation in a dense symbol environment, data entry, and a simulated air-to-air intercept. Aircraft maneuvers varied load factor and workload during task accomplishment.

Overall, all three HSI methods demonstrated clear military utility. All three also required improvement.

ET was fast to act initially, but inaccurate and unpredictable. The inaccuracies required use of the cursor-slew switch to correct initial error, which slowed overall task completion. ET calibration also drifted. When properly calibrated, ET enabled rapid, unconscious cursor relocation, which enhanced task accomplishment and decreased operator workload. If the calibration drifted due to head movement or load factor, ET performance ranged from unacceptable to completely unusable for interaction with the screen.

RP nearly equaled ET in terms of initial speed. However, operators spent unnecessary time refining head orientation to place the cursor precisely, even when this wasn't necessary for optimal task accomplishment. RP also imposed physiological burdens. Cursor refinement in the center of the screen did not cause strain, but moving a cursor to the screen extremities was unpleasant and strained the neck musculature, particularly under G. Though workload varied significantly amongst operators and maneuvers, the display of the rhino boresight proved mentally taxing and degraded performance.

Traditional HSI (TH), including both HOTAS and touchscreen, was familiar to all operators and enabled task accomplishment. However, when large slew distances were required, TH was slow and would benefit from augmentation. The accuracy of cursor-slew designations was the best of the three methods. However, the accuracy of touchscreen was no better than RP or ET in conjunction with a cursor slew. Under challenging flight conditions and workloads, TH remained methodical and reliable, if slow and distracting.

Each HSI method excelled in particular areas and fell short in others. Augmenting existing HSI with improved versions of the novel methods holds the most promise for operator performance optimization.

This page was intentionally left blank.

TABLE OF CONTENTS

1 INTRODUCTION 7
1.1 GENERAL 7
1.2 BACKGROUND 7
1.2.1 New HSI Concepts 7
1.3 TEST-ITEM DESCRIPTION 9
1.3.1 Eye-Tracking Implementation 9
1.3.2 Rhino-Pointing Implementation 10
1.3.3 Traditional Human System Interface (TH) Implementation 12
1.3.4 Operationally Flight Representative Helmet 14
1.3.5 Research Vehicle 15
1.3.6 Aircrew Flight Equipment (AFE) 20
1.3.7 Workload Measurement Systems 21
1.3.8 Test Range 23
1.3.9 Test Support Facilities and Equipment 23
1.3.10 Instrumentation Requirements 24
1.3.11 HSI Operators 25
1.4 TEST OBJECTIVES 26
1.4.1 Specific Test Objectives 26
2 TEST AND EVALUATION 31
2.1 TEST CONDITIONS AND MANEUVERS 27
2.1.1 Overall Test Methods 27
2.1.2 Overall Statistical Methods 29
2.2 COMPARISON OF EYE-TRACKING, RHINO-POINTING AND TRADITIONAL HSI 31
2.2.1 Single Track Designation Characterization 31
2.2.2 Dense Symbol Environment Designation Characterization 40
2.2.3 Data Entry Characterization 52
2.2.4 Operationally Representative Task Comparison 61
3 RECOMMENDATIONS 67
4 REFERENCES 69

APPENDIX A - DETAILED TEST ITEM DESCRIPTION A-1
APPENDIX B – DETAILED TASK DESCRIPTION B-1
APPENDIX C – FLIGHT LOG C-1
APPENDIX D – STATISTICAL ANALYSIS D-1
APPENDIX E – ADDITIONAL PLOTS E-1
APPENDIX F – HSI USABILITY SURVEY RESULTS F-1
APPENDIX G – LIST OF ACRONYMS G-1
APPENDIX H – ABBREVIATED DATA ANALYSIS PLAN H-1
APPENDIX I – DIGITAL DATA APPENDIX I-1
APPENDIX J – FUTURE WORK J-1
APPENDIX K – LESSONS LEARNED K-1

This page was intentionally left blank.

1 INTRODUCTION

1.1 GENERAL

This technical memorandum presents the results for Have Rhino, a United States Air Force (USAF) Test Pilot School (TPS) test management project (TMP). Have Rhino compared three different human-system interfaces (HSIs): eye-tracking, rhino-pointing (a form of head or nose tracking), and traditional HSI. Testing was requested by the 711th Human Performance Wing (HPW) of the Air Force Research Laboratory (AFRL) at Wright-Patterson Air Force Base (AFB), Ohio. The lead developmental test organization (LDTO) was the Air Force Test Center at Edwards AFB, California. The executing test organization was Class 19B of the USAF Test Pilot School, 412th Test Wing, Edwards AFB. Testing was conducted at Iowa City Municipal Airport in a modified L-29 Delfín, tail number N429GC, owned and operated by the Operator Performance Laboratory (OPL) of the University of Iowa. Testing was comprised of 14.5 ground test hours and 14 sorties totaling 21 flight test hours under the job order number MT19B800.

1.2 BACKGROUND

As of 2020, many aircraft cockpits, including fifth-generation fighters, had large multifunctional displays (MFDs) that spanned nearly the full width of the forward console. These MFDs replaced disconnected individual gauges, providing fused head-down information to the aircrew.

Legacy aircraft commonly used hands-on throttle and stick (HOTAS) to interact with MFDs and operate cockpit functions during phases of flight that required constant flight control inputs. HOTAS referred to the buttons and switches located at the fingertip positions of the stick and throttle flight controls. HOTAS allowed the aircrew to perform higher workload tasks by preserving cognitive resources that may not have otherwise been available. HOTAS also increased aircrew situational awareness by minimizing distracting motions, and reduced reaction time.

To move a cursor fixed to an MFD, traditional HOTAS used a small joystick, typically actuated by a single finger, referred to as a “cursor slew switch.” This cursor slew switch enabled the operator to move a cursor on an MFD much like a mouse moved an icon on a computer screen. A cursor slew switch enabled very precise inputs; however, such precision required slew rate limitations, making fast movements across large displays time consuming. This increased time interacting with the cursor increased workload commensurately.

In some aircraft, touchscreens replaced tactile buttons as a means to operate cockpit functions. In such aircraft, virtual buttons appeared on a large touchscreen MFD; the aircrew made selections by touching those buttons with a fingertip. However, to make selections on touchscreen MFDs, aircrew removed their hands from the primary controls, which violated the principle behind HOTAS and increased workload. Additionally, dynamic flight conditions reduced touchscreen precision without any means to brace or steady the outstretched arm.

Traditional HSI (TH), for the purposes of Have Rhino, consisted of a combination of HOTAS and touchscreen as it was used in modern aircraft, at a high technological readiness level (TRL).

1.2.1 New HSI Concepts

Aircrew expressed interest in other HSI systems that did not require hand movement. Innovative concepts developed by AFRL and the OPL involved augmenting cursor functionality with eye-tracking and head-tracking technologies to improve the HSI and decrease aircrew workload in modern aircraft cockpits.

1.2.1.1 EYE-TRACKING (ET)

Eye-tracking (ET) was the process of measuring the point of gaze (where one is looking) by measuring the angle of pupil deflection relative to the eye socket. Eye-trackers measured rotations of the eye in one of several ways, but principally they fell into one of three categories: (1) measurement of the movement of an object attached to the eye; (2) measurement of electric potentials using electrodes placed around the eyes; or (3) optical tracking without direct contact to the eye. Have Rhino used this third implementation, as described in section 1.3.1. The focus-boresight vector as depicted in Figure 1 was calculated from the eye rotation angle.

The operator could make inputs on an MFD display by manipulating gaze location, which moved the point in which the focus-boresight intercepted the screen (as depicted by the orange circle). However, the technology required to deploy reliable and accurate eye tracking in the operational environment remained at a low technological readiness level (TRL) between four and five at the time of this testing according to AFRL experts (reference 6).

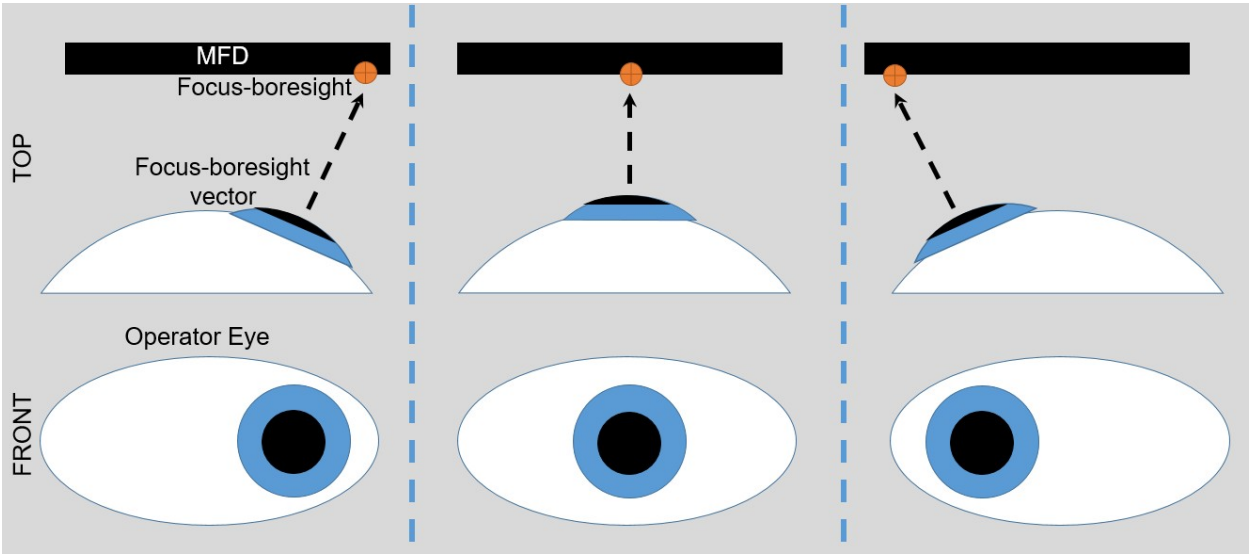


Figure 1. Focus-boresight Vector

1.2.1.2 RHINO-POINTING (RP)

A new concept developed by AFRL proposed cursor movement caused by the motion of the head itself. Such an HSI method calculated an intercept point on an MFD by tracking a vector passing through the center point of the aircrew flight helmet. Since this point roughly coincided with the location of the operator's nose, this HSI method was variously termed nose- or "rhino-pointing" (RP) as shown in Figure 2.

RP required precise tracking of the operator's head orientation. Head orientation data was gathered by either cameras that optically tracked reference points on the operator's head, or electromagnetic sensors that tracked magnetic field shifts induced by helmet-mounted magnets, the latter being the implementation used in this testing and is further described in section 1.3.2. The rhino-boresight vector as depicted in Figure 3 could be displayed directly on the MFD as a single point rhino-boresight (as depicted by the orange circle) and was moved by the aircrew member across the MFD through head and neck movements. The operator could manipulate a display-fixed cursor and interact with display entities by deliberately moving this rhino-boresight. In principle, RP enabled stable, fast, hands-free operation of all functions available on the MFD without limiting traditional HOTAS methods. Head tracking was already being used in various relevant environments which led AFRL experts to establish a TRL of six for RP at the time of testing (reference 6).

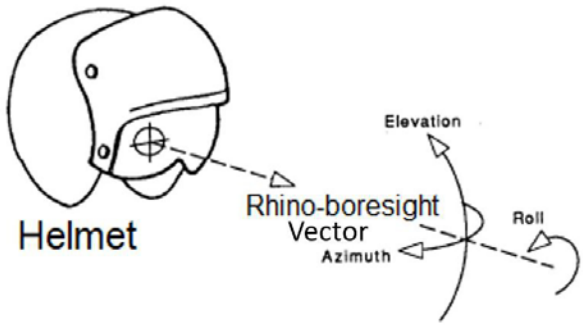


Figure 2. Rhino-pointing

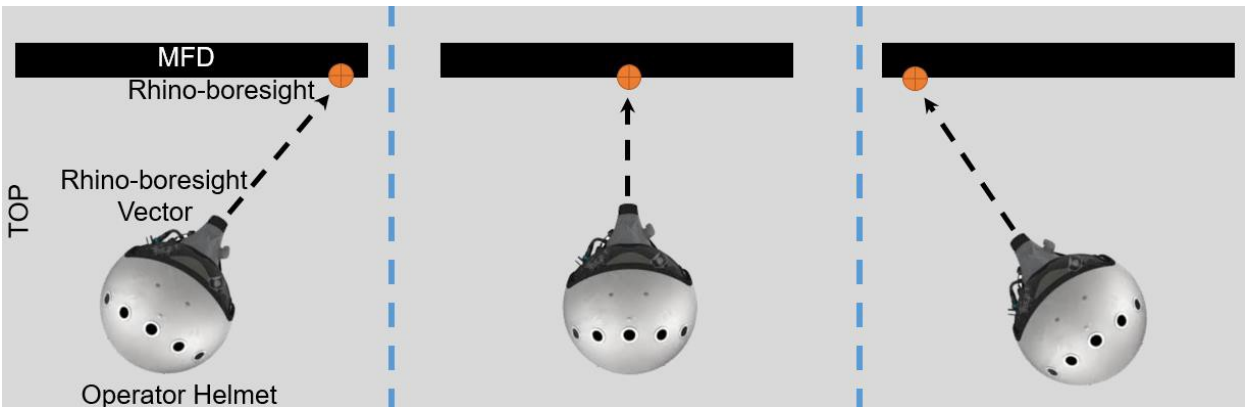


Figure 3. Rhino-boresight Vector

1.3 TEST-ITEM DESCRIPTION

The system under test consisted of the three HSI methods (eye-tracking, rhino-pointing, and traditional HSI) implemented on a test aircraft equipped with a large area display. The particulars of the implementations of these HSI methods were not directly under test, however as they incidentally affected the examination of the system under test, the following sections detail those implementations.

1.3.1 Eye-Tracking Implementation

The Dikablis Professional ET system used two eye cameras located in front of the wearer’s eyes and a scene camera mounted facing outward on the front of the mask. The eye cameras were aligned with the wearer's eyes and mounted on the clear visor of the helmet. The eye camera used black and white video to track the pupil of the wearer as shown in Figure 4 and calculate the focus-boresight. Only the right eye camera was used during testing to increase system reliability, due to a defect in the left eye camera discovered a few days prior to test. Additionally, the Dikablis ET system used only the apparent orientation of the pupil to track eye direction, and did not use glint spot correction from IR illuminance. Such glint spot correction would have, in theory, corrected for motion of the eye tracker relative to the eye due, for example, to helmet slippage. In the Have Rhino project, helmet slippage did affect test results. See, for example, section 2.2.3.2.1. Also see APPENDIX A – DETAILED TEST ITEM DESCRIPTION for elaboration on the technological readiness of the ET implementation.

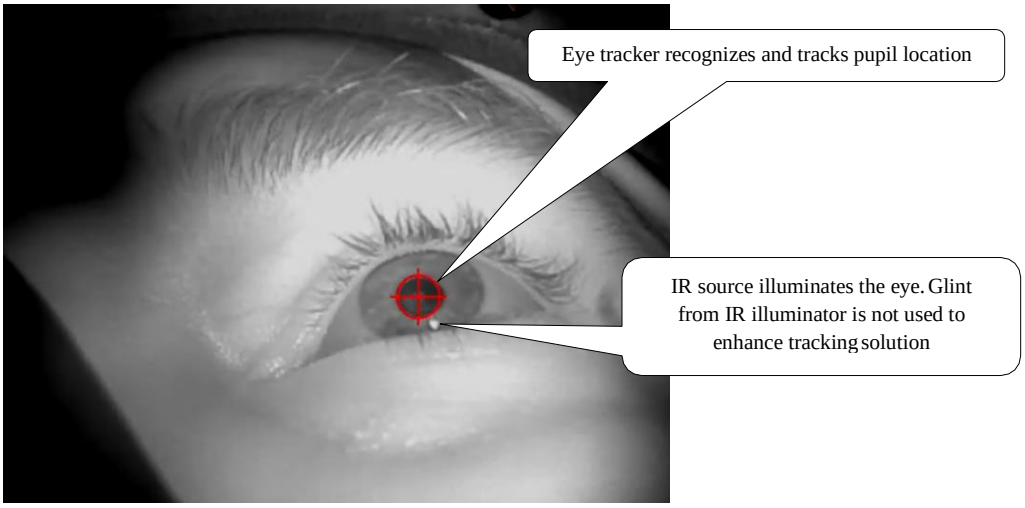


Figure 4. Eye Tracker Camera View

An initial ET calibration occurred before every flight to link the focus-boresight to the pupil angle. During the calibration, the eye tracker located the pupil in a reference frame fixed to the eye tracker. Then, the head tracker was used to determine how the eye tracker reference frame transformed into a reference frame fixed to the cockpit.

This calibration typically took 2 to 3 minutes. To perform the calibration, the operator focused on each corner of the MFD, while a technician verified alignment of the reference frames. After the calibration, the focus-boresight intercept point became visible, allowing the operator to get a sense of the ET system accuracy. Additional in-flight calibrations were performed as necessary. An in-flight calibration was conducted if the operator noticed a persistent bias in the cursor snap command or if an ET system monitor noticed the focus-boresight grossly out of alignment. The operator snapped the cursor around the MFD several times prior to conducting any ET test points to verify alignment. Testing continued if the operator subjectively determined the ET accuracy to be acceptable.

The OPL had successfully executed airborne tests using eye-tracking to evaluate operator performance during previous studies. The main difference to earlier studies was that new software developed by OPL mapped the helmet-based eye gaze coordinates to a cockpit-fixed global coordinate system, which allowed projection of a virtual focus-boresight onto the MFD. Otherwise, the Have Rhino implementation was largely identical to the setups used in those studies (Reference 1). ET use is further described in sections 1.3.1 and 1.3.5.1.4. Additional details on the ET test hardware and its calibration process can be found in APPENDIX A – DETAILED TEST ITEM DESCRIPTION.

1.3.2 Rhino-Pointing Implementation

The face vector was defined as a forward vector, normal to the plane of the face with an origin at the bridge of the nose. The rhino-boresight vector was offset from the face vector by a depressed elevation angle to optimize for head-down viewing of the MFD as shown in Figure 5. This angular offset varied based on operator preference and was fixed after the initial RP calibration. The angular offset was between 7 and 10 degrees for most operators. Motions of the head (and, thus, the rhino-boresight vector) were obtained from a commercial off-the-shelf electromagnetic Polhemus Scout head tracker installed in the test aircraft. This head tracker was accurately calibrated within the L-29 cockpit by Polhemus engineers. The particular approach used in this test, illustrated in Figure 6, involved a small externally mounted AC electromagnetic transmitter with three orthogonal electromagnetic coils and a small helmet mounted magnetic receiver, also with three orthogonal search coils. The head tracker system used measured variations in the electromagnetic

field to find and relay the instantaneous six degree of freedom position and orientation of the helmet to calculate the rhino-boresight vector.

Rhino-boresight calibration was performed during preflight and as required during the sortie if a bias was detected. During calibration, the operator held his head at a preferred orientation that allowed some head-up forward viewing while looking down his nose to view a small calibration target in the center of the MFD. The rhino-boresight reset to the calibration target location at the instant of calibration, and subsequently became free to move in response to operator head movements. The rhino-boresight was visualized on the MFD with a small red cross. This differs from the eye-tracking implementation that did not have a visible symbol on the MFD. RP use is further described in sections 1.3.3.1 and 1.3.5.1.5.

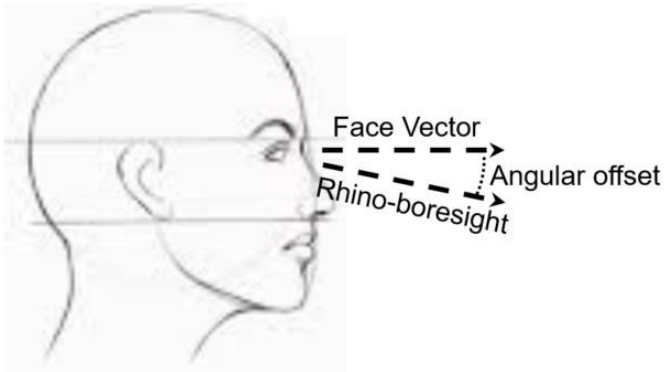


Figure 5. Rhino-boresight Definition

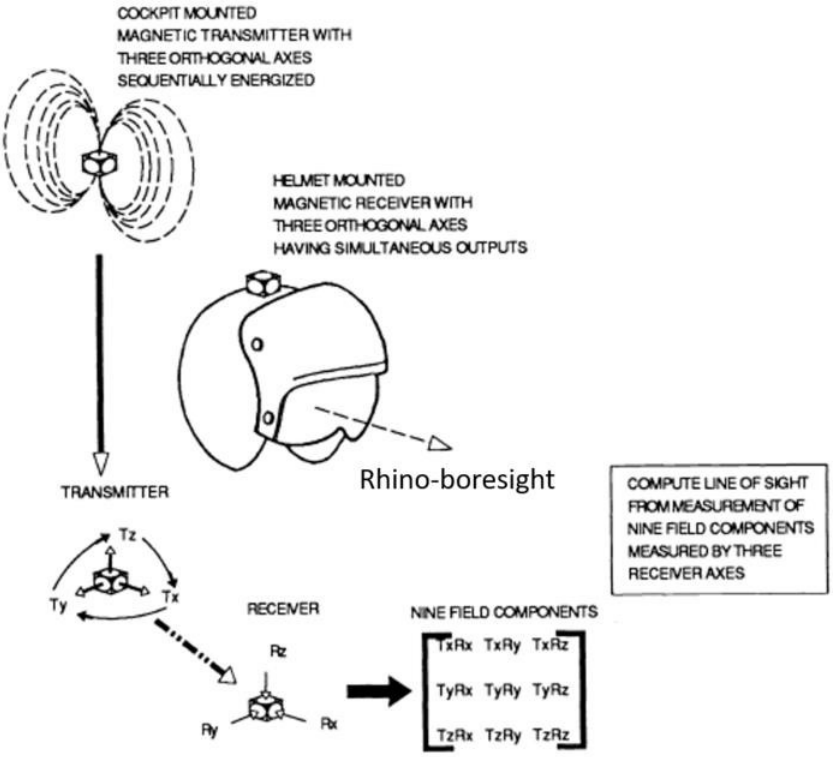


Figure 6. Helmet Tracking

1.3.3 Traditional Human System Interface (TH) Implementation

A combination of HOTAS and touchscreen inputs were used as the traditional HSI (TH). The operator interacted with the MFD using a HOTAS cursor slew or a touchscreen input depending on the predefined task. The TRL of all TH elements was assessed as a nine (see APPENDIX A – DETAILED TEST ITEM DESCRIPTION for details).

The primary flight control grips in the rear cockpit (RCP) of the test aircraft were modeled on the F-16 HOTAS. The Bugeye Technologies FC2000 stick and throttle retained their engine and flight control functionality to allow the operator to fly the test aircraft during specific test points. A functional speed brake switch was included on the throttle and a manual pitch trim lever was located to the left of the throttle; the trim switch on the control stick was not functional. All HOTAS controls not critical to this test were disabled so as not to impact flight or test operations.

The functionality of the switches modified and used during this test, known as "H2" and the "cursor slew" switches, are shown in Figure 7 and Figure 8.

1.3.3.1 Cursor Management Switch

The H2 switch on the stick as shown in Figure 7 was changed to the cursor management switch (CMS) and had the following functions: (1) designate, which selected an entity on an MFD; (2) snap, which caused the display-fixed cursor to jump to the location of the rhino-boresight or focus-boresight intercept point; and (3) expand, which increased the zoom level on an area of the display in the vicinity of the cursor's location.

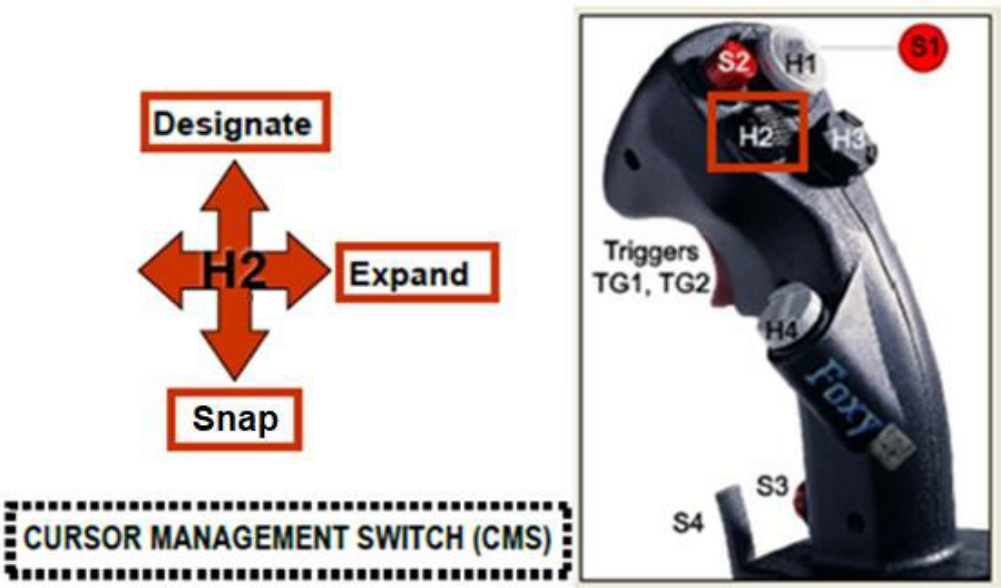


Figure 7. CMS Functionality

1.3.3.2 Cursor Slew Switch

The cursor slew switch, positioned at the thumb position on the throttle, was used for movement of the cursor on the MFD. Pushing the cursor slew switch in the forward direction moved the cursor vertically on the MFD as shown in Figure 8. Pushing the cursor slew switch in the down direction moved the cursor rightward on the MFD, also shown in Figure 8. Depressing the cursor slew switch on its Z-axis had no function.

Cursor slew as implemented in the test aircraft was non-proportional: any actuation of the cursor slew switch resulted in a constant maximum slew rate. This differed from contemporary proportional cursor slew systems, which adjusted slew rate as a function of the magnitude of the slew switch input. The cursor slew rate was set at 10 centimeters per second to allow for easier fine tracking. The full 50 centimeter wide screen could be traversed in five seconds. Using the originally planned 17.5 centimeters per second that better reflected the capabilities of modern fifth-generation fighter MFDs proved to make fine tracking in the terminal phase of designation too difficult.

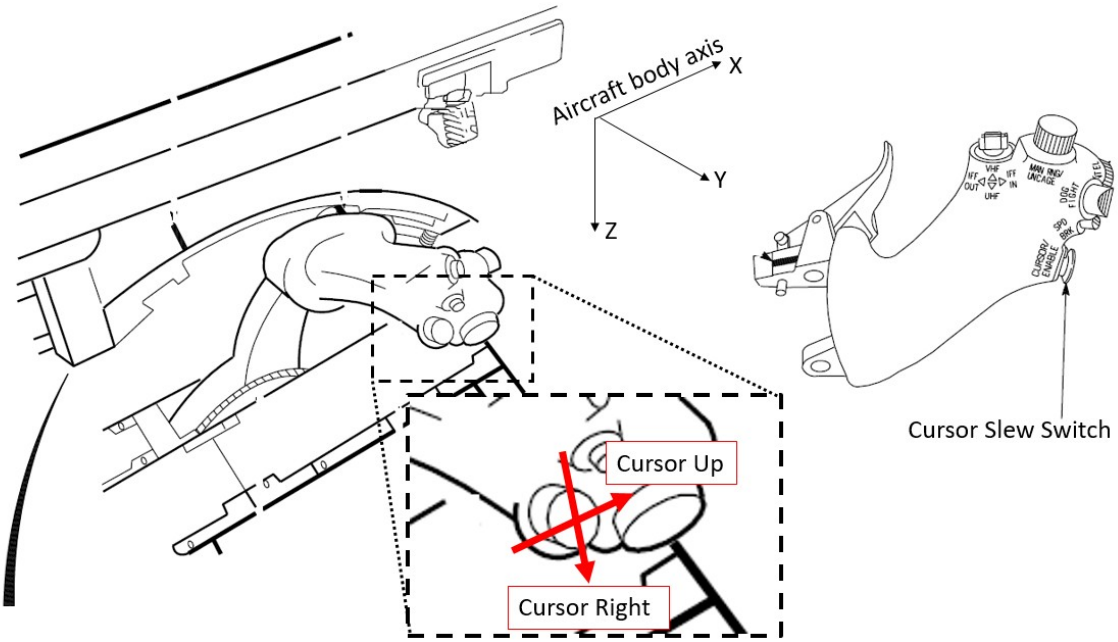


Figure 8. Cursor Slew Switch Functionality

1.3.3.3 Multi-Functional Display

The front panel of the RCP of the test aircraft had a large area LCD screen that served as the primary MFD display for this testing as shown in Figure 9. This LCD was an Advantech IDK-2121W touchscreen, installed in the RCP on pre-existing mounts. The LED backlight full-color display was 49.56 centimeters wide and 29.22 centimeters high, making its size comparable to the large format displays seen on fifth-generation fighter aircraft. The touchscreen used a 5-wire resistive solution to track the exact fingertip location during placement, slide, and release with a maximum response time of 25 milliseconds. The touchscreen was programmed to designate on release during all test-related tasks. Additional details on the MFD can be found in APPENDIX A – DETAILED TEST ITEM DESCRIPTION.

Figure 9. IDK-2121W Touchscreen, Installed

1.3.4 Operationally Flight Representative Helmet

The operationally flight representative helmet (OFRH), as seen in Figure 10, was funded by a Defense University Research Instrumentation Program in 2014. The asset was owned by the OPL and was not government furnished equipment (GFE). The Dikablis Professional eye-tracking system and Magnetic Polhemus Scout head tracker described previously were integrated into the OFRH without significantly affecting mass properties as seen in Figure 11. More information on the OFRH can be found in APPENDIX A – DETAILED TEST ITEM DESCRIPTION.

The OFRH used helmet mounted display optics focused at infinity to project a primary flight reference inside the visor. This flight reference was stabilized to the cockpit of the aircraft, and was referred to as a virtual HUD (VHUD), since it displayed the data of a traditional HUD but was actually projected in the helmet. An example of the VHUD symbology is shown in Figure 12, though the image projected was green and black.

Figure 10. OFRH

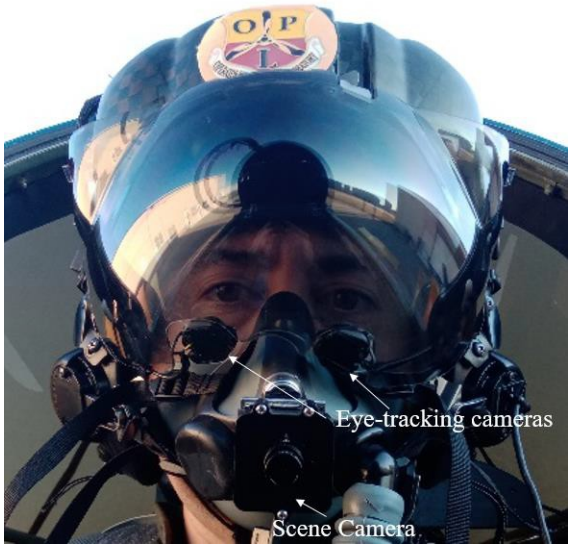


Figure 11. OFRH with Eye-Tracking Apparatus Installed

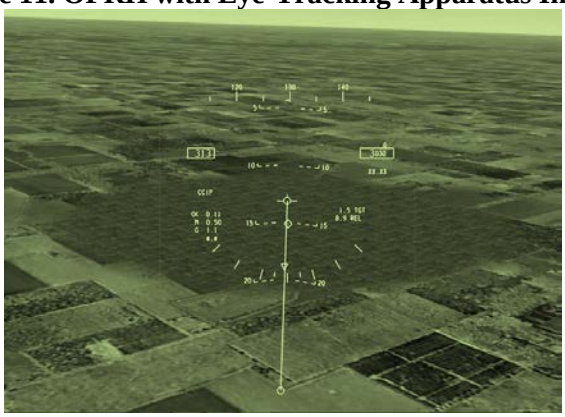


Figure 12. HMD VHUD Symbology

1.3.5 Research Vehicle

The OPL-owned Aero Vodochody L-29, tail number N429GC shown in Figure 13, was used as the platform for all HSI testing. The aircraft was maneuverable and capable of maintaining a 3.5G turn in a shallow dive, which made it a practical platform to emulate a fighter cockpit environment where system interaction would be necessary. The jet trainer aircraft could be flown from either the front cockpit (FCP) or rear cockpit (RCP) through mechanically linked flight controls. A qualified L-29 instructor pilot served as the pilot-in-command (PIC) and safety pilot (SP) from the FCP. The SP flew during all critical phases of flight, below 5,000 ft AGL, and as necessary during test maneuvers. The FCP did not contain any modified hardware relating to this test, with the exception of a small touch-screen display on the right console for in-flight ET calibrations. The FCP is shown in Figure 14. The HSI test operator (referred to as "operator") sat in the RCP and executed all test points and flew the aircraft during non-critical phases of flight to fly test-specific maneuvers. The OPL extensively modified the RCP to allow integration of test-specific cockpit hardware. The aircraft was instrumented with research displays, data collection equipment, and physiological measurement equipment as shown in Figure 15. More information on the OPL L-29 can be found in APPENDIX A – DETAILED TEST ITEM DESCRIPTION.

The OPL L-29 doubled as a fixed-base flight simulator when on the ground. The operator sat in the RCP and flew the aircraft through a simulated environment displayed on the VHUD, while executing test-related tasks on the MFD. The data acquisition system (DAS) on board the aircraft recorded the same data

that would be collected during an airborne sortie including simulated flight conditions. This simulator capability enabled familiarization of all three HSI methods and mission rehearsal prior to flight.

Figure 13. OPL L-29 Research Aircraft

Figure 14. N429GC Front Cockpit

Figure 15. N429GC Rear Cockpit

1.3.5.1 PREPAR3D REAR COCKPIT DISPLAY SOFTWARE

The operator in the RCP executed all test-related tasks using the HOTAS, touchscreen, and OFRH equipment, through the three HSI methods. A computer installed in the RCP ran the Lockheed Prepar3d software to emulate a representative fifth-generation cockpit display and show it on the MFD. The Prepar3d software was modified by OPL engineers to allow several test-related tasks to be performed by the operator while airborne using the provided HSI equipment. Prepar3d generated simulated sensor scope displays, data input menus, and air-to-air intercept portals, in addition to standard flight instruments for the purposes of this test. Flight data was fed to Prepar3d to create a realistic simulation, but no data from Prepar3d was fed back to the aircraft. Each display configuration corresponded to a test-specific task. Operator interaction technique varied based on HSI method.

1.3.5.1.1 Simulated Sensor Scope

Prepar3d generated a sensor scope display with simulated track symbols. These track symbols would either appear individually (Figure 16) or in a dense cluster of background clutter (Figure 17) depending on the software configuration. If the track symbol was surrounded by background clutter, the operator could magnify (x2) the area around the cursor location by commanding an expansion (CMS expand). A track symbol could be designated by moving the cursor within five centimeters and commanding a designation (CMS designate). The appearance of a white circle around the track provided designation confirmation. This symbology loosely reflected fifth-generation sensor scope displays.

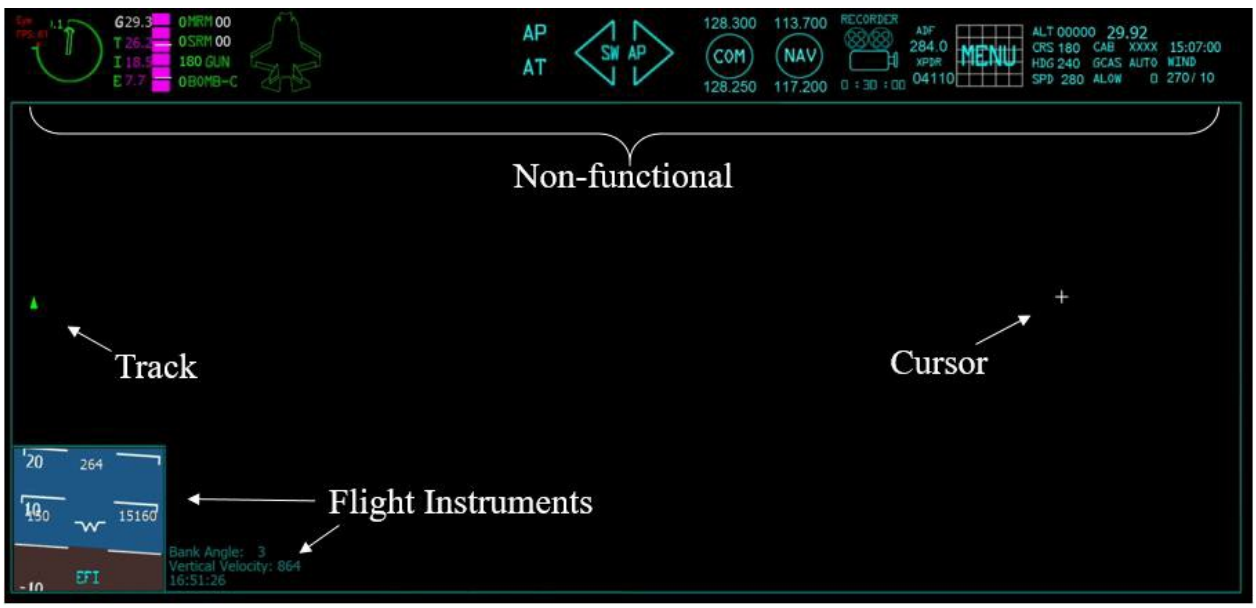


Figure 16: Simulated Sensor Scope, Without Background Clutter

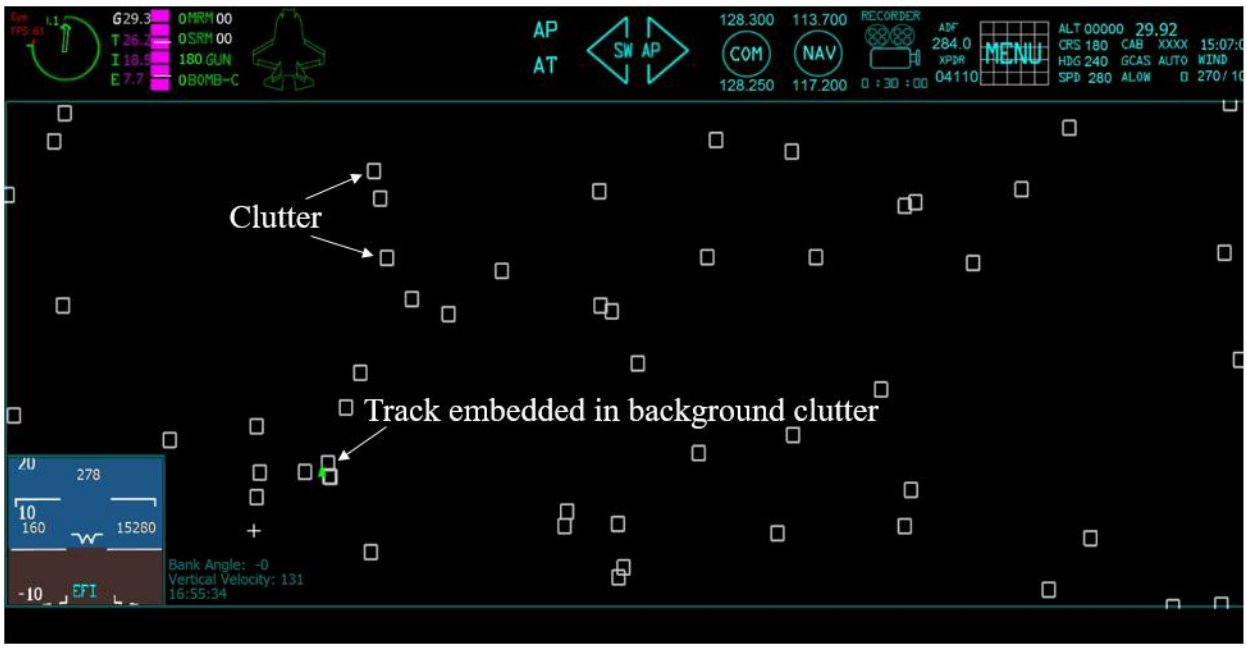


Figure 17: Simulated Sensor Scope, With Background Clutter

1.3.5.1.2 Data Entry Menu and Keyboard

Prepar3d generated a data entry menu and keyboard as shown in Figure 18. The menu keys and character keys were 2.54 cm by 2.54 cm (1” by 1”) to represent key dimensions typically seen on fifth-generation displays and the keyboard layout matched the QWERTY standard. The menu branch options represented cockpit functions that required multiple character inputs. A line of data could be entered by moving the cursor to each key and commanding a designation. The menu would change or text would appear at the bottom of the screen to confirm the input.

Figure 18: Data Entry Menu and Keyboard

1.3.5.1.3 Simulated Air-to-Air Portals

Prepar3d generated a set of air-to-air intercept portals to simulate an air-to-air engagement as shown in Figure 19. After task initiation, simulated enemy tracks would appear on the left primary situation display portal and right secondary situation display portal. The operator moved the cursor to an enemy track and designated it using the same technique described in in 1.3.5.1.1. After designation, the operator would select a simulated missile based on the range of the enemy track by moving the cursor to the missile selection keys and designating using the same technique described in 1.3.5.1.2.

Figure 19: Simulated Air-to-Air Portals

1.3.5.1.4 ET Task Execution

To execute an HSI task using ET, the operator looked at point of interest on the MFD. With his eyes focused on the point, he actuated the CMS snap to instantly move (or "snap") the cursor to his focus-boresight. The cursor on the MFD would snap from its original location to the calculated focus-boresight intercept location. The operator then used the cursor slew switch to slew the cursor as desired.

RP and TH both continuously displayed the rhino-boresight and the display-fixed cursor. However, the focus-boresight vector was removed from the display for test. Previous OPL research found that because state-of-the-art ET usually contained a bias error, display of the focus-boresight symbol caused operators to continuously drive the symbol off the display as they tried to track it with their eyes. Consequently, OPL recommended making the focus-boresight invisible to provide the closest comparison of all HSI methods given the current technological readiness of ET. The consequences of displayed and un-displayed boresight symbology is discussed at length in section 2 of this memorandum.

1.3.5.1.5 RP Task Execution

To execute an HSI task using RP, the operator moved his head to move the red rhino-boresight symbol (shown in Figure 20) to a point of interest on the MFD, to which he wanted to move the cursor. With the rhino-boresight hovering over the point, he snapped the cursor to the rhino-boresight location using the same CMS snap actuation used in ET. The cursor would snap from its original location to the instantaneous rhino-boresight location. The operator then used the cursor slew switch to further slew the cursor. As an additional functionality, an operator had the option to refine the rhino boresight directly onto a track and then rapidly actuate CMS snap (aft) followed by CMS designate (fwd). This rapid pair of inputs was referred to as "direct designate," but in fact always required both CMS actuations.

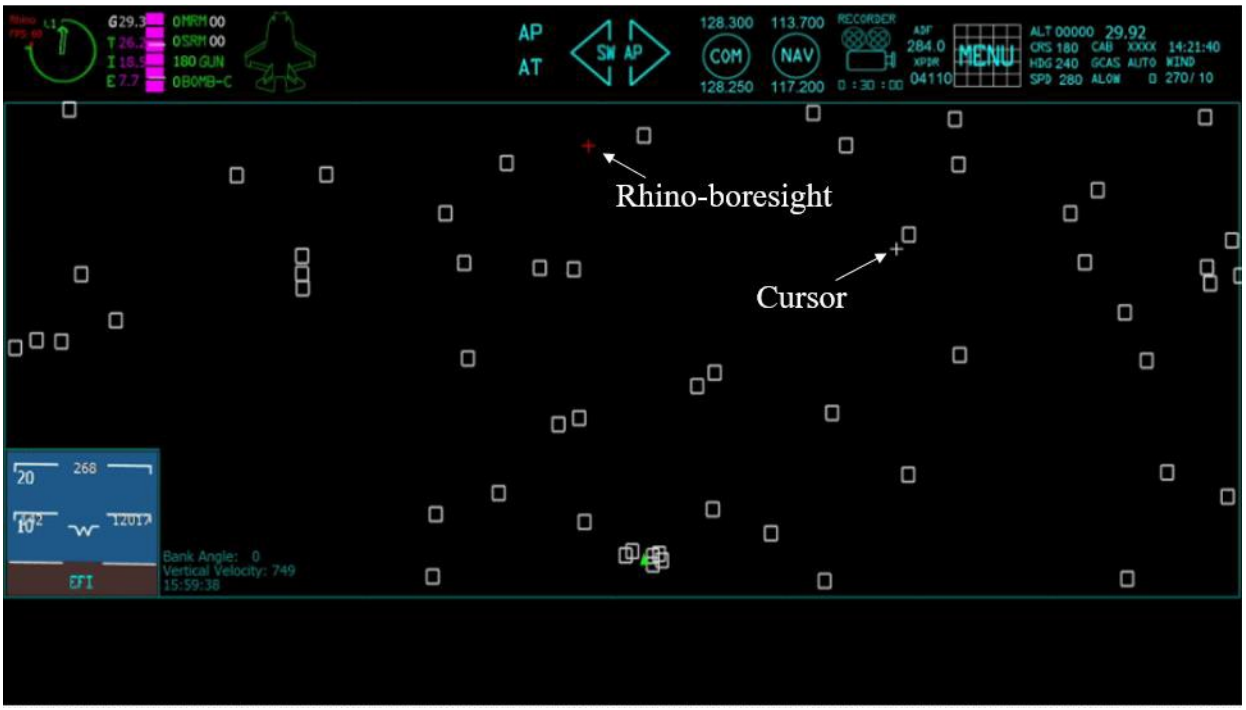


Figure 20: RP Task Execution Example

1.3.5.1.6 TH Task Execution

To execute an HSI task using TH, the operator moved the cursor using the cursor slew switch or pressed the touchscreen with a fingertip. A touchscreen press both snapped the cursor to the touch location and designated on touch release. The CMS snap function had no purpose when using TH and the operator could immediately designate or expand when the cursor was properly positioned.

1.3.6 Aircrew Flight Equipment (AFE)

Operators wore flight suits, flight-approved boots, and undergarments as specified in AFI 11-301 during all ground and flight tests as shown in Figure 21. The test team wore gloves that offered the necessary level of protection while optimizing touchscreen selection precision. The glove type was standardized across all operators. Additionally, aircrew wore personal G-suits, the OFRH, parachutes, and a harness, the latter three provided by OPL. A small Nexus-4 physiological monitoring system was stored in the left G-suit pocket.

Figure 21: Operator in Full AFE, with Nexus-4 stowed.

1.3.7 Workload Measurement Systems

Operator workload was measured using two different methods developed by OPL. The first method, Transition Probability Variance (TPV), calculated operator workload by tracking changes in operator cardiac activity. The second method, Normalized Aviation Performance (NAP), compared the operator's ability to accurately and precisely maintain specific flight parameters such as airspeed and altitude in a high workload and low workload situation.

1.3.7.1 Physiological Monitoring System

The Nexus-4 physiological monitoring system for Electrocardiogram (ECG) was worn by the operator under the Nomex flight suit and cotton T-shirt to provide the physiological component of the workload-related data. This system was wireless with no physical connections to the aircraft (see Figure 22). The wireless system minimized any concerns of safety of egress. The unit was stored in the left leg pocket during the test as shown in Figure 21. This system had been successfully used on several previous research studies (reference 4) prior to this test.

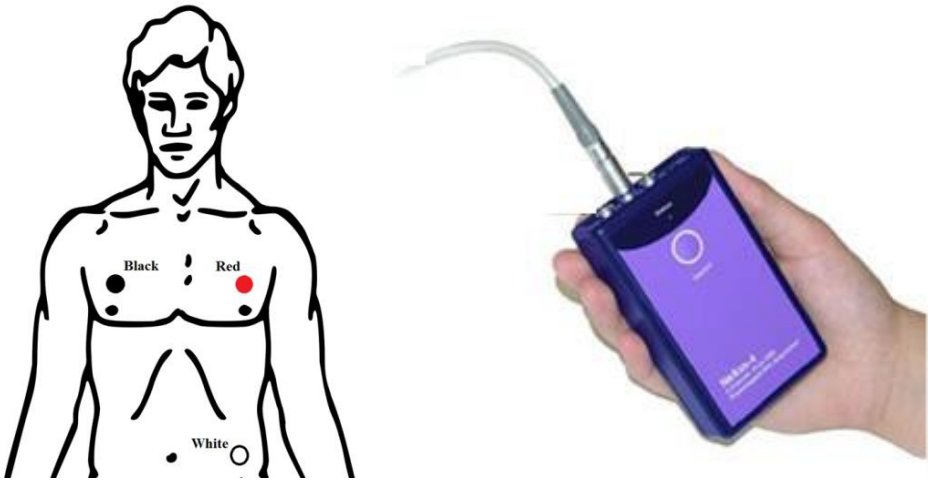


Figure 22: ECG Sensor Setup

1.3.7.1.1 Electrocardiogram (ECG) Workload Measurement

Electrical conductive systems of the heart responded strongly to changes in a person's attention and cognitive effort. The OPL developed a unique classifier to assess cognitive workload, referred to as the Chaotic Physiological Classifier (CPC) method. The OPL used the CPC method to transform raw ECG data into a cognitive workload metric, called the TPV-Based Workload metric.

ECG data was obtained from each operator throughout the flight during execution of test-specific tasks. The ECG time series data was then transformed into the TPV through a data reduction procedure run by the OPL. The TPV captured the variability in the dynamics of the ECG signal as the operator underwent different levels of cognitive loading. The final TPV value represented the quantitative operator workload for a given task and conditions, and was used in conjunction with operator comments and qualitative workload assessments. Final TPV values for each operator provided by OPL were not traceable back to the raw ECG data, protecting the privacy of the operator. More information on this type of quantitative workload measurement can be found in reference 5.

1.3.7.2 Normalized Aviation Performance

The meaningfulness of the TPV metric derived fundamentally from a parallel measurement of task performance (as seen in reference 1). For example, an operator could evince a nearly-constant TPV value, but show deteriorating task performance while executing maneuvers of increasing difficult or demand. This compensatory behavior is typically seen in flight tests, where actual workload expenditure is always relatively high, regardless of maneuver difficulty. For easier maneuvers, spare cognitive capacity gets applied toward maintaining exceptional task performance. As task difficulty increases, less spare capacity remains available; the main flight task increases cognitive demand. Therefore, a decrease in task performance is typically observed. Thus, to get a full picture of operator workload and spare capacity in flight tests, it is important to interpret quantitative workload data (TPV) in conjunction with quantitative maneuver performance data.

Maneuver performance in Have Rhino itself was explicitly measured during the increased workload maneuvers. For all such maneuvers, a normalized aviation performance (NAP) was calculated using an OPL algorithm during post-processing. NAP directly indicated the precision or quality of the maneuver. An operator with high TPV and high NAP was working hard but had enough spare cognitive capacity to perform the maneuver to a high degree. An operator with high TPV and low NAP was also working hard, but was overwhelmed and unable to perform the maneuver skillfully.

To calculate NAP, particular parameters for a pilot to achieve during a maneuver (airspeed, altitude, turn rate, descent rate, etc.) were assigned. During execution of the maneuver, the on-board system recorded the actual aircraft state. During post-processing, the recorded aircraft state was compared to the desired parameter by calculating the percent deviation for the duration of the maneuver. The percent deviations were normalized to achieve equal weighting for all measured parameters. The root-mean-square (RMS) of the complete set of normalized percent deviations was then computed and re-normalized to a reference maneuver flown by each operator while fully focused on flying without additional tasks. This single NAP value indicated aviation performance for the entire maneuver, providing a metric of total error relative to the reference maneuver. More information on NAP and TPV calculations is available in section 2.1.1.3, APPENDIX H – ABBREVIATED DATA ANALYSIS PLAN, and APPENDIX I – DIGITAL DATA APPENDIX.

1.3.8 Test Range

All flight testing was conducted in the Class E airspace in the vicinity of the Iowa City Municipal Airport (KIOW). Typical terrain in the airspace rose just over 1,000 ft MSL. The main hazards in the range consisted of routine VFR and IFR traffic.

1.3.9 Test Support Facilities and Equipment

The OPL operated a flight operations research hangar at KIOW. The hangar was large enough to conduct full ground tests of the systems under test prior to attempting during airborne tests. The L-29 test aircraft was interconnected to a ground station (shown in Figure 23) located in the OPL hangar via an encrypted L-band datalink that provided real-time data to test team members conducting and monitoring the test flights. A repeat of the MFD output, live video feed of the operator, and push-to-talk radio communication were available in the ground station. The ground station was manned by members of the Have Rhino test team who served as the test conductor (TC) and scribe for the mission and recorded relevant comments, thus freeing the aircrew of those responsibilities.

Figure 23: Ground Station

1.3.10 Instrumentation Requirements

A DAS installed in the L-29 provided both flight condition data and operator performance data. A list of DAS parameters and their purpose is shown in Table 1.

Table 1: Instrumentation Parameters

Parameter	Source	Description	Units
Subject ID	L-29 DAS	Operator	N/A
Airspeed	L-29 DAS	Calibrated Airspeed	knots
Altitude	L-29 DAS	Aircraft Altitude	feet MSL
Bank Angle	L-29 DAS	Aircraft Bank Angle	degrees
Load Factor	L-29 DAS	Load Factor Normal Body (ratio)	N/A
Heading	L-29 DAS	True Heading	degrees
Time	L-29 DAS	Task Initiation to Task Completion	seconds
Cursor Position	L-29 DAS	Centroid of HSI cursor	cm
Track Position	L-29 DAS	Centroid of Track Symbol	cm
Designation Button	L-29 DAS	Actuation of Designate Button	N/A
Expand Button	L-29 DAS	Actuation of Expand Button	N/A
Cursor Snap Button	L-29 DAS	Actuation of Cursor Snap Button	N/A
Pointing Type	L-29 DAS	RP, Eye Tracking, or Traditional	N/A
ECG	Nexus-4	Quantitative Workload Assessment	N/A
Hot-Mic Recording	L-29 DAS	All Intercom, Radio, and Hot Mic Audio	N/A

1.3.11 HSI Operators

Each member of the Have Rhino test execution team completed the full test matrix and some results varied based on individual operator metrics and experience. A summary of the background of each operator is given below:

- Operator 1
  - US Military Branch: USAF
  - Aircrew Rating: Senior Pilot
  - Total Flight Time: 930 hrs (880 hrs high-performance fighter PIC, 50 hrs GA PIC)
  - Height: 72”
  - Relevant HSI Experience:
    - 6 years of experience using HOTAS to control Radar and other sensors
    - Some experience using touchscreen devices airborne
- Operator 2
  - US Military Branch: USAF
  - Aircrew Rating: Senior Pilot
  - Total Flight Time: 840 hrs (750 high-performance fighter PIC, 40 hrs GA PIC)
  - Height: 67”
  - Relevant HSI Experience:
    - 6 years of experience using HOTAS to control Radar and other sensors
    - Some experience using touchscreen devices airborne
- Operator 3
  - US Military Branch: USMC
  - Aircrew Rating: Senior Pilot
  - Total Flight Time: 1,250 hrs (1,250 hrs high-performance fighter PIC)
  - Height: 70”
  - Relevant HSI Experience:
    - 8 years of experience using HOTAS to control targeting pods, Radar, and other sensors
    - Some experience using touchscreen devices airborne
- Operator 4
  - US Military Branch: USAF
  - Aircrew Rating: Combat Systems Officer
  - Total Flight Time: 1,469 hrs (1,400 hrs high-performance/fighter WSO, 69 hrs GA PIC)
  - Height: 65”
  - Relevant HSI Experience:
    - 6 years of experience using HOTAS actuations to control a variety of aircraft systems
    - Some experience using touchscreen devices airborne
- Operator 5
  - US Military Branch: USAF
  - Aircrew Rating: Flight Test Engineer (non-rated aircrew)
  - Total Flight Time: 100 hrs (33 hrs various military aircraft, 67 hrs GA PIC)
  - Height: 76”
  - Relevant HSI Experience:
    - Minimal experience using HOTAS to control aircraft systems
    - Some experience using touchscreen devices airborne

        - Led flight test program to upgrade legacy fighter aircraft with night vision device compatible cockpit instrumentation
- Operator 6
    - US Military Branch: USAF
    - Aircrew Rating: Flight Test Engineer (non-rated aircrew)
    - Total Flight Time: 37 hrs (33 hrs various military aircraft, 4 hrs GA PIC)
    - Height: 74”
    - Relevant HSI Experience:
        - Tropia eye condition – permanent, minor misalignment of left and right eye
        - Minimal experience using HOTAS to control aircraft systems
        - Some experience using touchscreen devices airborne
- Operator 7
    - US Military Branch: USAFR
    - Aircrew Rating: Senior Pilot
    - Total Flight Time: 2,400 hrs (2,000 hrs high-performance/fighter PIC)
    - Height: 70”
    - Relevant HSI Experience:
        - 600+ hrs of fifth-generation touchscreen aircraft experience
        - 1300+ hrs in aircraft requiring HOTAS to control targeting pods, Radar, and other sensors
        - 8 years of experience designing and testing aircraft HSI

The diversity in operator background gave unique perspectives and reflected the range of experience that would be present in a modern military.

1.4 TEST OBJECTIVES

The overall test objective was to compare eye-tracking, rhino-pointing, and traditional tactile input as components of a large area display HSI.

1.4.1 Specific Test Objectives

The specific test objectives supporting the overall test objective were:

1. Characterize operator performance while using eye-tracking, rhino-pointing, and HOTAS to designate a track symbol.
2. Characterize operator performance while using eye-tracking, rhino-pointing, and HOTAS to designate a track symbol embedded in a dense symbol environment.
3. Characterize operator performance while using eye-tracking, rhino-pointing, and touchscreen input to conduct a data entry task on a large area display.
4. Compare eye-tracking, rhino-pointing, and conventional tactile input during an operationally-representative task.

All test objectives were met. Of note, no evaluation criteria were used to make formal ratings during this testing. As an emerging technology, the current state of eye-tracking and rhino-pointing did not have a standard to measure against. As an HSI comparison, the intent of the test was to compare the effects of the three HSI methods, not to evaluate them against predetermined criteria.

2 TEST AND EVALUATION

This technical memorandum presents the results for Have Rhino, a United States Air Force (USAF) Test Pilot School (TPS) test management project (TMP). Have Rhino compared three different human-system interfaces (HSIs): eye-tracking, rhino-pointing (a form of head or nose tracking), and traditional HSI. Testing was requested by the 711th Human Performance Wing (HPW) of the Air Force Research Laboratory (AFRL) at Wright-Patterson Air Force Base (AFB), Ohio. The lead developmental test organization (LDTO) was the Air Force Test Center at Edwards AFB, California. The executing test organization was Class 19B of the USAF Test Pilot School, 412th Test Wing, Edwards AFB. Testing was conducted at Iowa City Municipal Airport in a modified L-29 Delfín, tail number N429GC, owned and operated by the Operator Performance Laboratory (OPL) of the University of Iowa. Testing was comprised of 14.5 ground test hours and 14 sorties totaling 21 flight test hours under the job order number MT19B800.

The test program was organized as a series of trials, each consisting of a task and a maneuver. There were three specific tasks and three sets of maneuvers. Operators from a variety of backgrounds flew all tasks and all maneuvers using each of the three HSI methods. These trials used the L-29 evaluator station and control room to gather detailed information about the performance of the HSI methods, which was augmented by operator comments and surveys. After HSI performance metrics were gathered, a separate task was flown to mimic an operational intercept.

This report first discusses test execution including conditions, the L-29 maneuvers, and the overall design of experiment. Test results are then presented, organized by task. For each task, quantitative then qualitative results are presented. While ground data was gathered to enhanced the dataset and familiarize the operator with the system under test and tasks, this report only presents flight data.

2.1 TEST CONDITIONS AND MANEUVERS

Table 2 presents a summary of the 14 flights flown. The “conditions” column provides the weather conditions in the test altitude band above 5,000ft AGL. The “data collected” column specifies which HSI methods and tasks the operators flew on the various sorties. All flights had sunshine and relatively calm air above any lower cloud layer. A few flights had occasional light turbulence that impacted RP performance; those effects are discussed in the results sections where applicable. No other environmental conditions impacted data quality. All operators performed all tasks and maneuvers for all HSI methods; a breakdown of test points conducted on each flight is available in APPENDIX C – FLIGHT LOG. Data collection tasks will be described in the following sections. Ground test data was collected but not reported on.

2.1.1 Overall Test Methods

All tests used a set of specific maneuvers intended to emulate a variety of aviation states aircrew experience when needing to interact with cockpit instruments. If the operator flew outside the boundaries of the databand, exceeded tolerances from starting conditions, or had to maneuver for airspace or traffic, the test maneuver was terminated and set up again.

2.1.1.1 Baseline Maneuver

The baseline maneuver provided a 1-G environment emulating a relaxed cruise state. In this maneuver, the operator flew at constant heading and constant flight path angle. The operator maintained sole control of the aircraft during this maneuver.

- Databand: 5,000 ft AGL to 25,000 ft MSL, 160 KCAS
- Tolerances: [Flight path angle] ±5° [Airspeed] ± 20 KCAS

Table 2. Flight Log

Flight	Operator	Date	Conditions	Data Collected
1	1	10 Mar	CAVU, LT OCN TURB	(HSI ET, RP | Task 1, 2, 3)
2	5	10 Mar	CAVU	(HSI RP, TH | Task 1, 2, 3)
3	7	10 Mar	CAVU	(HSI ET, RP, TH | Task 1, 2, 3)
4	2	11 Mar	Clear above BKN layer	(HSI ET, TH | Task 1, 2, 3)
5	4	11 Mar	Clear above BKN layer	(HSI ET, RP, TH | Task 1, 2, 3)
6	3	12 Mar	Clear above SCT layer	(HSI ET, RP | Task 1, 2, 3)
7	6	12 Mar	CAVU, LT OCN TURB	(HSI RP, TH | Task 1, 2, 3)
8	1	13 Mar	CAVU	(HSI TH | Task 1, 2, 3), (HSI ET, RP, TH | Task 4)
9	2	13 Mar	CAVU, LT OCN TURB	(HSI RP | Task 1, 2, 3), (HSI ET, RP, TH | Task 4)
10	5	13 Mar	CAVU	(HSI ET | Task 1, 2, 3), (HSI ET, RP, TH | Task 4)
11	4	15 Mar	Clear above BKN layer	(HSI ET, RP, TH | Task 4)
12	3	15 Mar	Clear above OVC layer	(HSI TH | Task 1, 2, 3), (HSI ET, RP, TH | Task 4)
13	6	15 Mar	Clear above OVC layer	(HSI ET | Task 1, 2, 3), (HSI ET, RP, TH | Task 4)
14	6	17 Mar	CAVU	(HSI ET | Task 1, 2, 3), (HSI ET, RP, TH | Task 4)
Note: Broken (BKN), Ceiling and Visibility Unrestricted (CAVU), Light Occasional Turbulence (LT OCN TURB), Overcast (OVC), Scattered (SCT)				

2.1.1.2 Varying Load Factor Maneuver

The varying load factor (G) maneuvers provided an increased-G environment emulating turning, higher energy flight states. Two levels of the load factor maneuver were flown. The medium load factor maneuver (2G) emulated a 1.5 to 2.5 G loading that might be experienced during an aggressive turn to intercept an instrument final. The high load factor maneuver (3.5G) emulated the 3 to 4 G loading that might be experienced when sharply turning to reposition during an air-to-air intercept or air-to-ground attack axis. In order to isolate the effect of load factor, the SP maintained sole control of the aircraft during this maneuver, with the operator only performing the HSI tasks.

- Databand: 10,000 to 25,000 ft MSL, 140 to 320 KCAS
- Tolerances: [G] ± 0.5 G

2.1.1.3 Varying Aviation Workload Maneuver

The varying aviation workload maneuver emulated task-saturating environments in which the precision or number of tasks increased. A bowtie pattern as shown in Figure 24 was planned. However, to increase the number of test points flown in a single sortie, operators only flew a 180° component of this maneuver, which never included the depicted turn reversal.

Two levels of the workload maneuver were flown. The low workload maneuver (LWL) required the operator to focus on maintaining constant bank angle, airspeed, and altitude in the turn. The high workload maneuver (HWL) added a requirement to set 500 vertical feet per minute, and changed the bank angle requirement to a turn rate requirement. Early in each sortie, the operator flew both the high and low workload maneuvers through 360° of heading change without any additional test-related tasks to establish a proficiency reference from which to grade future executions and establish a reference for calculating NAP. This report refers to this 360° turn as the “reference” maneuver for varying aviation workload.

- Databand: 5,000 ft AGL to 25,000 ft MSL, 160 KCAS

- Tolerances: None; operators instructed to fly as close to the prescribed parameters as possible

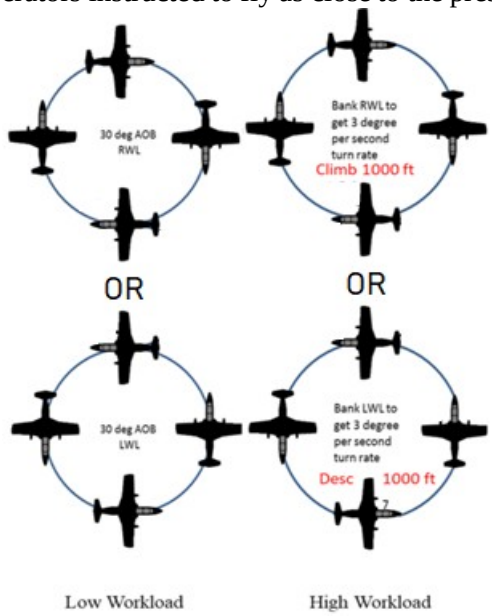


Figure 24: Varying Aviation Workload Reference Maneuver

2.1.2 Overall Statistical Methods

2.1.2.1 Experimental Design

The test program was designed as a set of randomized, blocked, single blind trials, controlled within-subject (i.e., each subject was his own control).

The individual tasks provided randomization: track locations and data strings were randomized during task executions. Blocking of operator, HSI method, and task order was accomplished within the constraints of available flight hours. Specifically, each operator was assigned a different order for using the HSI methods, and a different order for accomplishing the tasks. These "blocks" were assigned in a way to maximize the variation in the possible orders of HSI method and task, thereby minimizing the effect of task learning on resulting HSI performance. The experiment was single-blind in the sense that operators did not know their blocking until the day of execution, nor the randomization outcome of any trial until the moment of execution. Maneuver order was not blocked, and operators always accomplished maneuvers in order of increasing difficulty. Task 4 was completed last for every operator.

A full test matrix consisted of all HSI methods, all tasks, and all maneuvers, blocked differently for each operator. Each operator successfully accomplished the entire test matrix over the course of 2 flights. During a flight, the operator completed as many trials as possible of each given HSI method, task, and maneuver permutation within the available flight time. See APPENDIX D – STATISTICAL ANALYSIS for more clarification on experimental design and blocking.

2.1.2.2 Data Collection

The following types of quantitative data were collected by the aircraft during task execution: accuracy, time, and workload (in the form of TPV and NAP). Qualitative data, in the form of operator comments, were collected real-time by the TC and in HSI usability surveys administered to each pilot after each flight. An identical format survey was administered for each of the three HSI methods used. This survey may be found in APPENDIX F – HSI USABILITY SURVEY RESULTS.

Accuracy was always defined as the distance between track centroid and cursor location at some measured event. For data entry, a data entry key substituted for a track. For many test points, the measured event occurred at an "action." Action refers to the first actuation that an operator performed while using one of the HSI methods. For ET and RP, the first action was a CMS snap, since the snap preceded all designations and all expands. For TH, the first action was CMS designate or CMS expand, except in the case of data entry where the first action was the first press of the touchscreen.

A time measurement always indicated the elapsed seconds between task initiation onboard the L-29 and some measured event. The DAS recorded the initiation time when the initiate signal reached the aircraft and recorded completion time when the completion criteria was met. Each task had different completion criteria which are further explained in their respective sections. An alternate measure of time to complete a task, termed "pseudo-velocity," divided the initial distance between the cursor and the track (or data key) centroid by the time required to complete the task. This yielded a value in cm/sec indicating how fast the HSI method accomplished the task, standardized for starting distance. Operator workload was measured by the changes in TPV and NAP during the task, and was averaged over the entire task execution period.

To initiate data collection, first an operator established the assigned maneuver. The operator then communicated "ready" to the control station. Subsequently, the TC counted down from three over the radio and then transmitted a task start signal to the aircraft. Upon receiving the signal, the RCP MFD displayed the task as described in the individual task sections; data collection commenced simultaneous with task initiation. Accuracy, time and workload were all measured by systems on-board the aircraft, so any delays in the communication chain between the TC and the operator did not impact measurements.

Significantly, while workload in the form of TPV was collected continuously throughout a sortie, analysis using workload was only done if an increasing workload maneuver or operational maneuver was flown. The total meaningfulness of the TPV metric derived fundamentally from a parallel measurement of maneuver performance, which was not undertaken for the remaining maneuvers. For increasing workload maneuvers, NAP was the quantitative measure of task performance; for the operational maneuver, pilot comments were used as a qualitative measure of task performance.

Additionally, to scale TPV onto the range [0,1], TPV was normalized by reporting:

$$Normalized\ TPV = \frac{\log(TPV)}{\max(\log(TPV))}$$

Here, the max function returns the maximum value of log(TPV*)* evinced by an operator for a given task and maneuver, but any HSI method. Normalized TPV values were then binned by HSI method for comparison.

Refer to the data analysis found through APPENDIX I – DIGITAL DATA APPENDIX for more information on how each operator performance parameter was calculated.

2.1.2.3 Statistical Analysis

All collected data was indexed by operator, HSI method, task, and maneuver. In analysis, a "population" was defined as the complete set of data with any subset of these parameters fixed. The following is an example of a population description:

Operator – 1
HSI Method – ET
Task – 1, 2, and 3
Maneuver – LWL and HWL

A "distribution" was then defined as the complete set of a particular type of data for a given population. For example, in the above population, all the data for accuracy at designation would be a distribution

pertaining to that population. The design of the Have Rhino test provided a finite sample from each population. In the above example, operator 1 flew tasks 1, 2, and 3, in conjunction with the LWL or HWL maneuver, while using ET, a finite number of times. This finite sample provided a sample distribution, which was assumed to represent the overall population distribution. This assumption was validated by the t-tests described below.

For the following analysis, comparisons between distributions only with populations defined in this way were made. For any comparisons between means of sample distributions, a 2-sided t-test was performed using a 95% confidence interval and an assumption of unequal variances with Welch's modification to the degrees of freedom. This t-test inferred properties of the population distributions from the sample distributions. Of note, no assumption was made about the normality of the population distributions themselves; however, the distribution of the differences between the population means was assumed to be normal. Statements suggesting equality of means indicate an inability to reject the null hypothesis; statements suggesting inequality of means indicate a rejection of the null hypothesis. All t-test results are reproduced in APPENDIX D – STATISTICAL ANALYSIS. A graphical depiction of the box-and-whiskers format used in this memorandum also appear in that appendix.

The Circular Error Probable of 95% (CEP95) indicated that 95% of cursor actions will fall within the ellipse defined by the CEP value. CEP95 was calculated on the accuracy as defined above in Section 2.1.2.2 for a given population. The CEP assumed a circular bivariate normal distribution where the distribution clusters around the mean accuracy over a variety of azimuths, producing a circular confidence region.

2.2 COMPARISON OF EYE-TRACKING, RHINO-POINTING AND TRADITIONAL HSI

The overall test objective was to compare eye-tracking, rhino-pointing, and traditional tactile input as components of a large area display HSI. The overall test objective was met.

2.2.1 Single Track Designation Characterization

The first specific test objective was to characterize operator performance while using eye-tracking, rhino-pointing, and HOTAS to designate a track symbol. Operator performance was characterized by four metrics: (1) the accuracy of first action, (2) the time required to complete designation, (3) the operator's workload, and (4) HSI usability. This test objective was met.

Overall, ET was fast but inaccurate and unpredictable; operators found its performance highly desirable when it worked, and unacceptable when calibration drifted. RP was fast and accurate, but operators commented on the mental and physical effort required to attain that accuracy. TH was the slowest and most accurate, but operators thought it needed augmentation to accelerate cursor transit across the large display.

2.2.1.1 Single Track Designation Methods and Conditions

All single track designations were characterized using a common task, known as Task 1. In Task 1, a large blank portal appeared on the simulated sensor scope (described in section 1.3.5.1.1). When the TC transmitted the task, a small green triangle (a "track") appeared at a random location in the blank portion of the display. This situation is depicted in Figure 25.

After visually locating the track symbol, the operator moved his cursor to within the 5 cm designation radius of the track symbol, using the active HSI method. For TH, only the cursor slew switch was used for moving the cursor to the track. Once the cursor was close to the track, the operator designated the track, thereby completing the task and stopping the timer. The system indicated successful designation to the operator by putting a white circle around the track. This final situation is shown in Figure 26.

Task 1 time was limited to ten seconds. If the operator failed to designate the track after 7 seconds, the track began to flash; after 10 seconds the undesignated track disappeared, also stopping the timer. A more detailed description of Task 1 may be found in APPENDIX B – DETAILED TASK DESCRIPTION.

Figure 25. Task 1 at initiation

Figure 26. Task 1 at completion

During the task, the system continuously measured the locations of the cursor and the fixed track centroid, indexed to elapsed task time. All CMS actuations were also recorded and time-stamped. From these parameters, a data processing routine determined the accuracy at first action by calculating the magnitude of a vector between the cursor and the track when HOTAS indicated an action (see 1.3.3.1 for definition of "action"). Processing determined accuracy at designation with the same vector magnitude, but calculated at track designation. Workload and HSI usability were evaluated as described in section 1.3.7.

Task 1 was performed while operators executed the baseline maneuver, varying load factor maneuvers, and varying workload maneuvers. For the following Task 1 accuracy and time results, all operators, tasks, and maneuvers were pooled into a single data set, with a distinct population for each HSI method. A discrete distribution from these populations underlies any distribution plotted as continuous (to aid visualization). For such continuous curves, a kernel density estimation (KDE) of the underlying histogram is displayed; all KDEs used a Gaussian kernel and Silverman's rule of thumb for the bandwidth (reference 9).

2.2.1.2 Single Track Designation Results

2.2.1.2.1 Time and Accuracy

Figure 27 depicts the time to first action for each HSI method as a density distribution. The mean times to first action were 1.53 sec (ET), 2.12 sec (RP), and 4.9 sec (TH). These time differences appeared due to the strategies employed by operators when using the various HSI methods. When using ET, operators performed a snap as soon as their eyes found the track, which occurred very quickly. When using RP, operators spent additional time refining their rhino boresight prior to snap in order to immediately designate after snapping: a strategy referred to as "direct-designation." When using TH, operators began to slew their cursor in the general direction of the track as soon as they located the track. However, the chosen slew rate of 10 cm per second could never reposition the cursor faster than the snapping capability provided by ET and RP. Means, standard deviations, and t-tests between distributions such as these may all be found in APPENDIX D – STATISTICAL ANALYSIS.

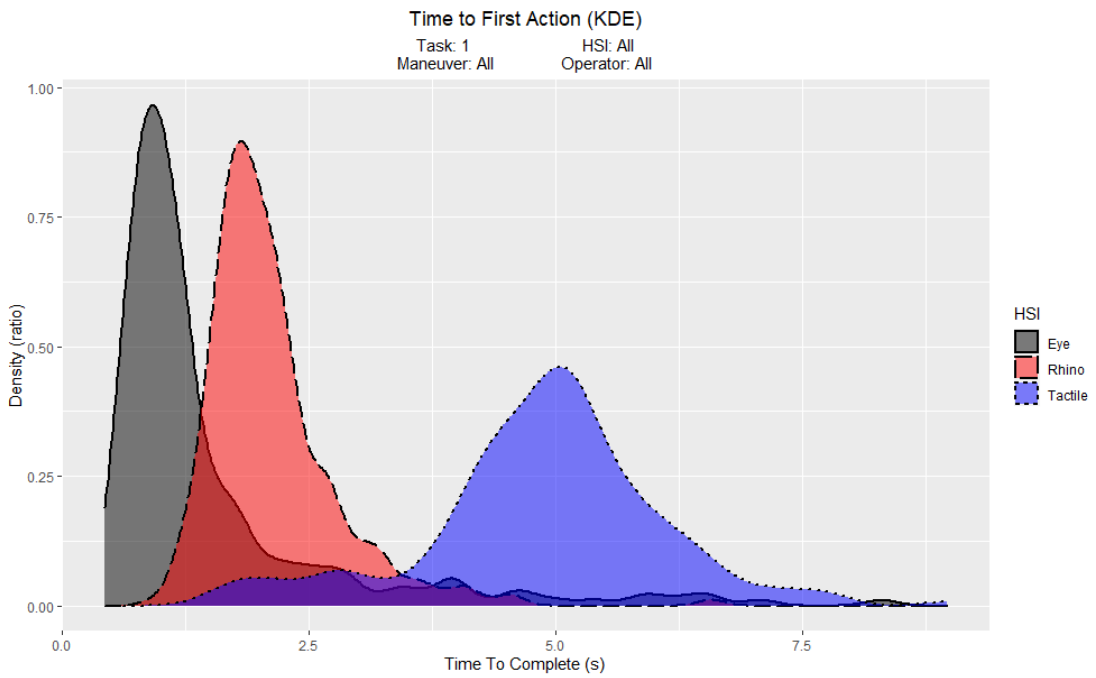


Figure 27. Task 1 time to complete first action.

A pseudo-velocity analysis (as described in section 2.1.2.2) somewhat complicates the straight-forward time to action analysis. Figure 28 shows that RP was indeed faster than TH; however, ET had a wide distribution of first action pseudo-velocities, and only had larger pseudo-velocities than RP (and TH) intermittently. This suggests that ET was not always faster to first action than RP and TH across similar distances, and will be discussed further in section 2.2.2.

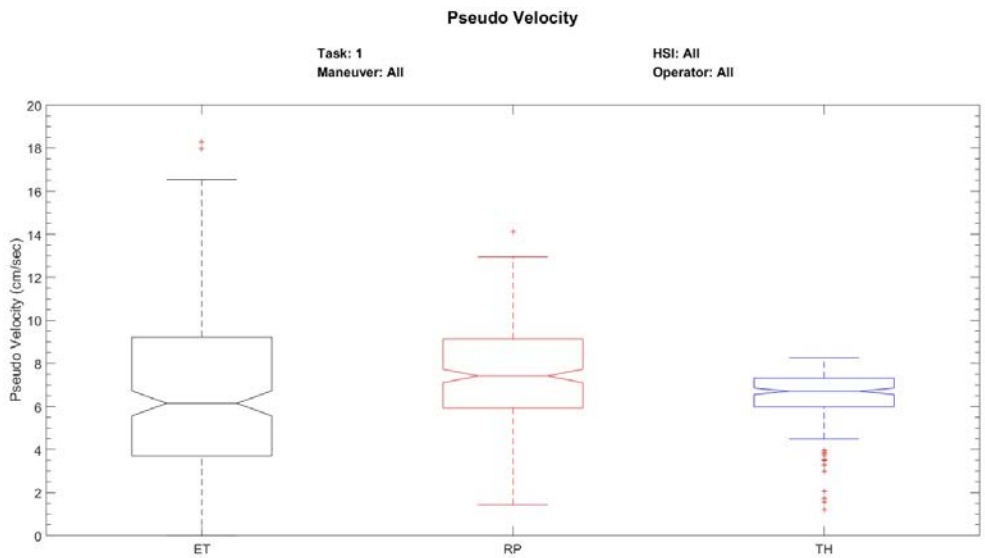


Figure 28. Task 1 pseudo-velocity during completion of first action.

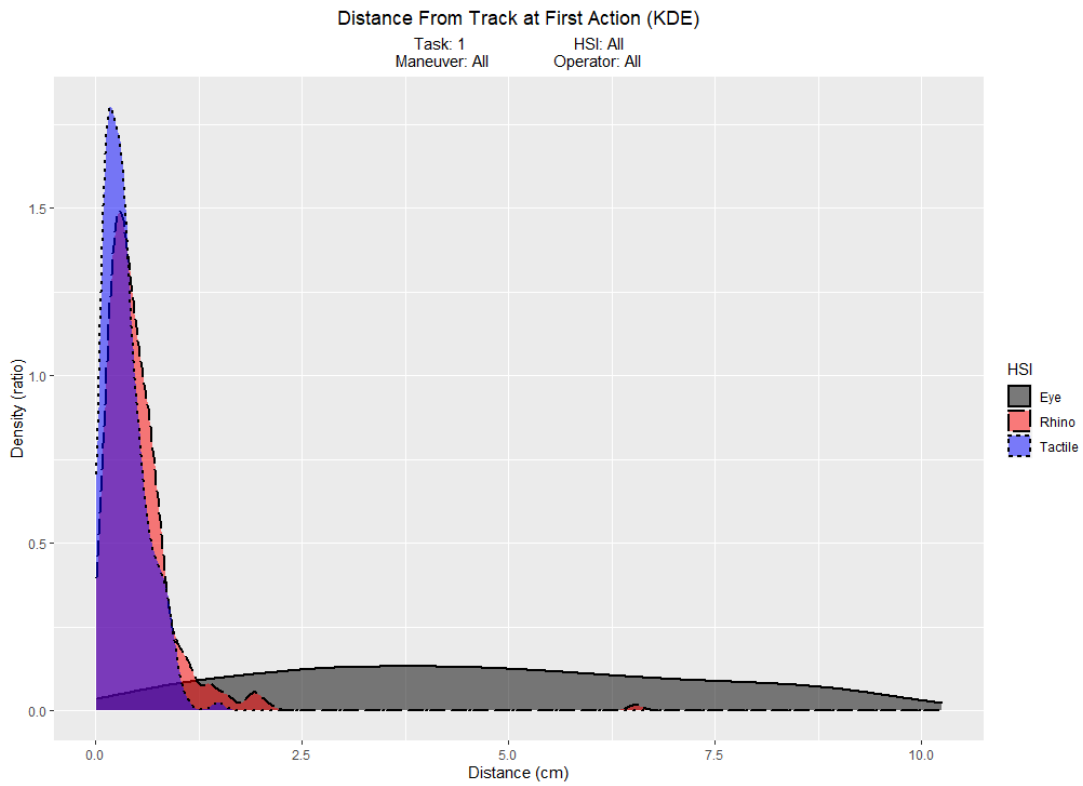


Figure 29. Task 1 Accuracy at First Action

Figure 29 depicts the accuracy at first action (measured in cm from the track centroid) for each HSI method. The mean accuracies at first action were 5.74 cm (ET), 0.51 cm (RP), and 0.36 cm (TH). Although a t-test indicated greater than 99% confidence that RP and TH had different accuracy means, the magnitude of the difference in sample means (0.15 cm), lacked operational relevance. From the operators' perspectives, RP and TH were equally accurate in placing a cursor at a given location on the display. By contrast, ET was unpredictable, snapping a cursor almost uniformly between 0 and 10 cm away from the track.

ET was also inaccurate relative to the centroid of the track. Figure 30 shows a two-dimensional error density plot for just ET. The origin of this plot, depicted with bold lines, represents the center of the track. The center of the ET error density distribution fell 1.75 cm from the center of the track for this distribution. This indicates a clear bias (inaccuracy) in ET relative to operator aim-point. This bias was not constant, and varied in size and magnitude for different populations with no discernible pattern. No such accuracy bias existed for RP or TH, whose two-dimensional error density distributions were always centered at the origin. Two-dimensional density plots for all HSI methods may be found in APPENDIX E – ADDITIONAL PLOTS.

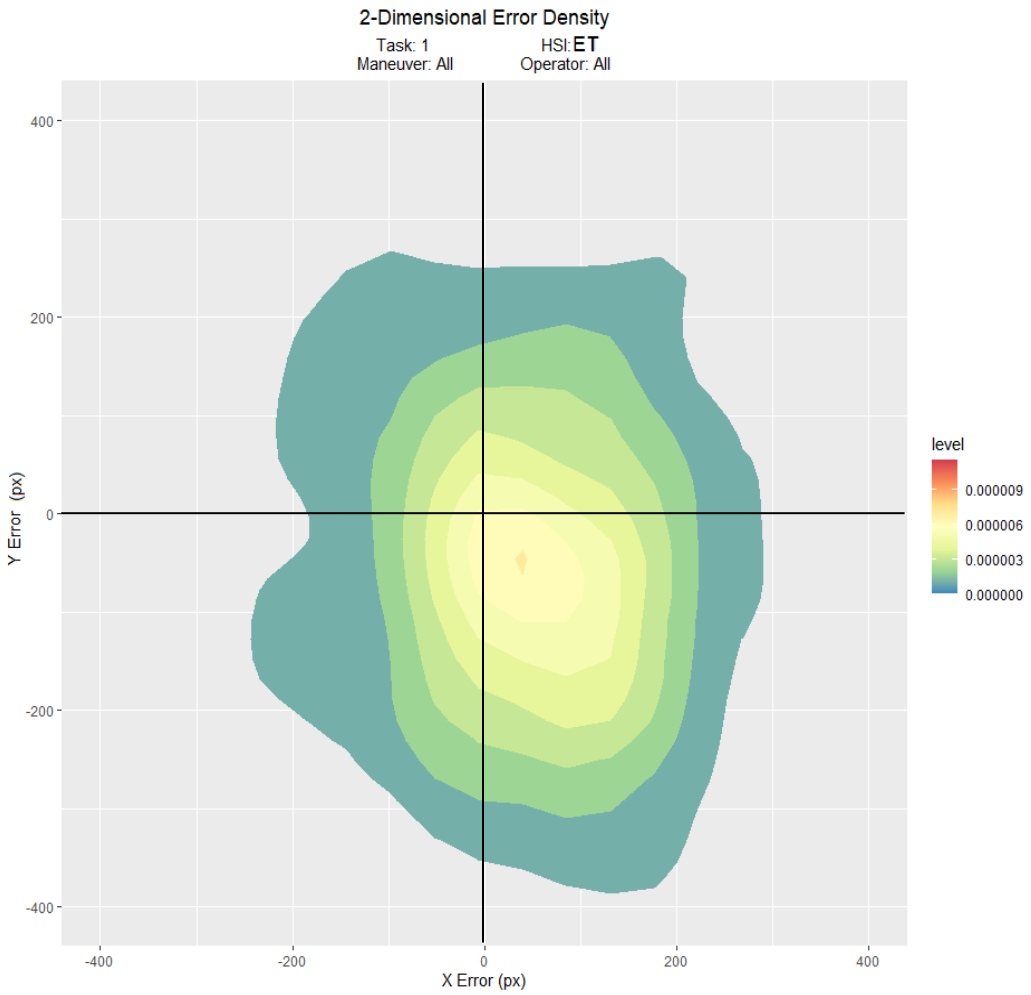


Figure 30. 2-Dimensional Error Plot, Task 1, ET only

Figure 31. Task 1 time to designate

In summary, relative to the other HSI methods, ET enabled quick, inaccurate, and unpredictable actions while RP enabled somewhat quick, accurate actions and TH enabled slow, highly accurate actions. Figure 31 places these results in the context of task success by depicting overall time to designate. Mean times to designate were 3.22 sec (ET), 2.69 sec (RP), and 4.9 sec (TH). Thus, RP was about a half-second faster to complete the task on average. Even though RP was slower to first action, the additional slew time required due to the inaccurate ET snap caused RP to be slower than ET overall. TH completed this task the slowest, as previously discussed.

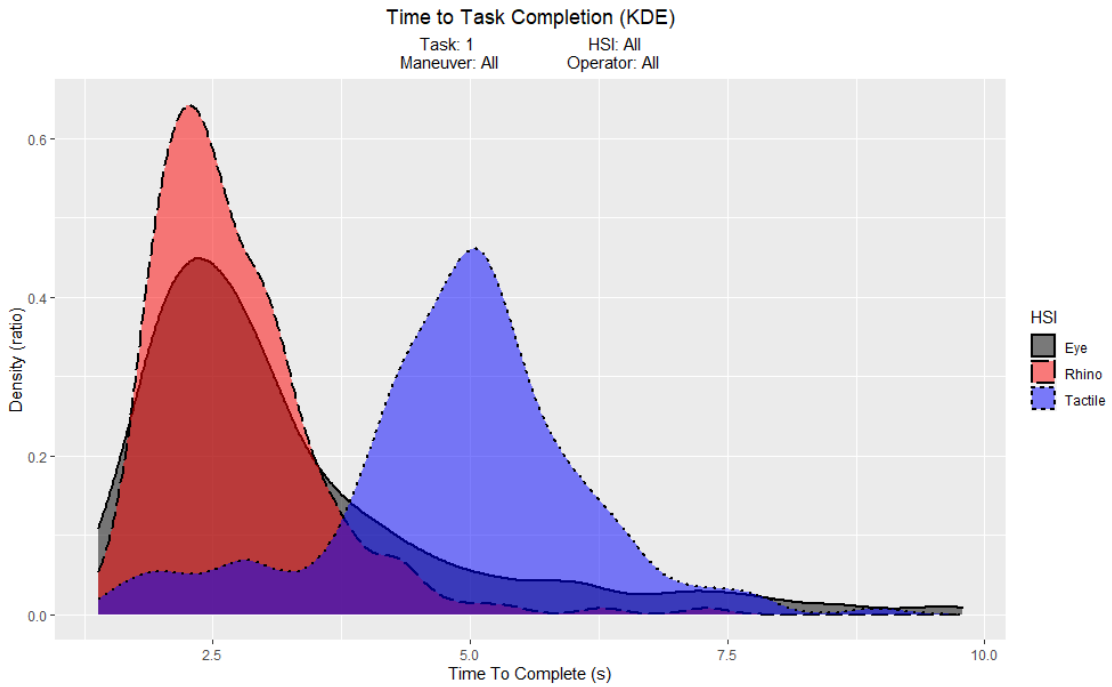
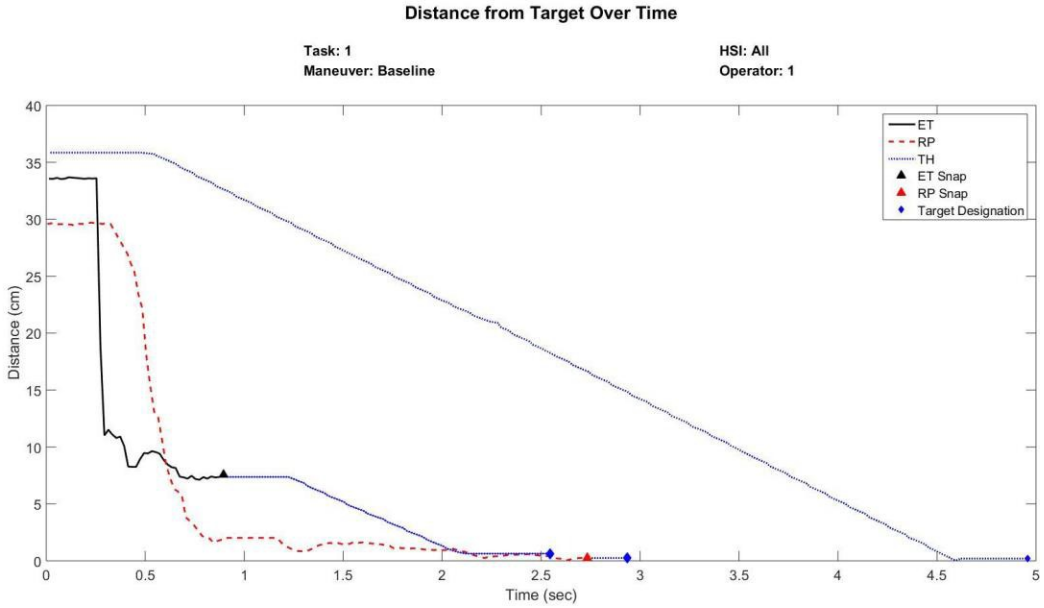


Figure 32. Example Task 1, cursor distance from track over time

Figure 32 exemplifies the previous discussion with three individual executions of Task 1 by showing the time-history path of the cursor toward the final track location for each HSI method.



The blue dotted line shows that for TH, a relatively long reaction time of over a half-second preceded a slow, linear approach to the track. Operators attributed the lengthy reaction time to a combination of locating the track and conceptualizing the cursor slew switch input that would yield the desired cursor direction.

The black solid line shows that for ET, a short reaction time of just under a quarter-second preceded an instantaneous jump of the focus-boresight, after which the eyes spent around a half-second focusing prior to an ET snap operation. After another short reaction time interval, tactile input took over and the cursor was slewed the remaining distance to the track.

The red dashed line shows that for RP, a short reaction time of just over a quarter-second preceded a rapid motion of the head towards the track. However, inside of about 2 cm, the operator spent nearly 2 full seconds refining the RP boresight in an attempt to get it close enough to the track to direct-designate. All operators were conscious of this mental and physical delay with RP.

Although the time history plot exemplifies typical behavior, the performance of ET specifically strongly depended on calibration. Figure 33 shows distributions of Task 1 trials for each HSI method. The origin represents an arbitrary track centroid, and the various dots indicate the relative location of the snap operation for that trial. There were several cases indicated by single black dots near the origin where ET was accurate enough to directly designate the target without updating the post-snap cursor with the cursor-slew switch. However, in most cases, large position updates were required with the cursor slew switch to achieve a position close enough to the track to designate. For the vast majority of RP cases, operators were able to direct-designate without updating the cursor position with the cursor slew switch. However this took time and effort. Actions for TH always took place within a designation radius, by definition. Cross plots for all operators can be found in APPENDIX E – ADDITIONAL PLOTS.

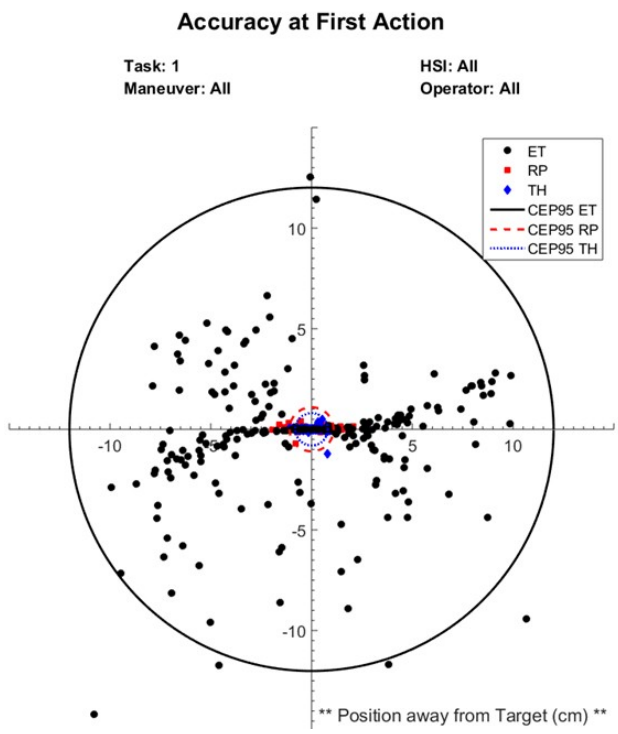


Figure 33: Task 1 Accuracy at first action

2.2.1.2.2 Workload and HSI Usability

Operators lauded ET and RP for the ease with which they navigated across large areas of the display when compared to TH. While TH was limited by its slew rate, both ET and RP were able to snap nearly instantly to areas of the screen that would take much longer with TH.

As seen in Figure 34, however, ET and TH both showed a slightly lower usability rating when compared to RP for survey question 1—Navigating Across Displays. (All survey questions may be found in APPENDIX F – HSI USABILITY SURVEY RESULTS). For ET, this lower rating on navigating across displays resulted from lack of predictability of the system. The cursor consistently appeared farther away from the track than expected, which led to a visual search for the cursor before further refinement was possible. This increased time and workload for the operator. For TH, the lower rating for display navigation resulted from the slowness of the fixed slew rate. RP was rated above both ET and TH; the slew rate was only slightly slower than ET, but predictability was vastly improved. Operators were able to direct-designate at snap for the majority of runs, with no further slew required.

However, this direct designation led to a tendency for operators to spend time refining the rhino-boresight prior to cursor snap in order to get the most accurate placement possible. The additional stabilization required for fine tracking and constant head movements resulted in RP being rated as the most uncomfortable to operate of the three HSI types evaluated. This can be seen in Figure 34, Question 10 - Overall Comfort. HSI Usability Ratings were on a 1 to 6 scale, with 1 being least favorable and 6 being most favorable.

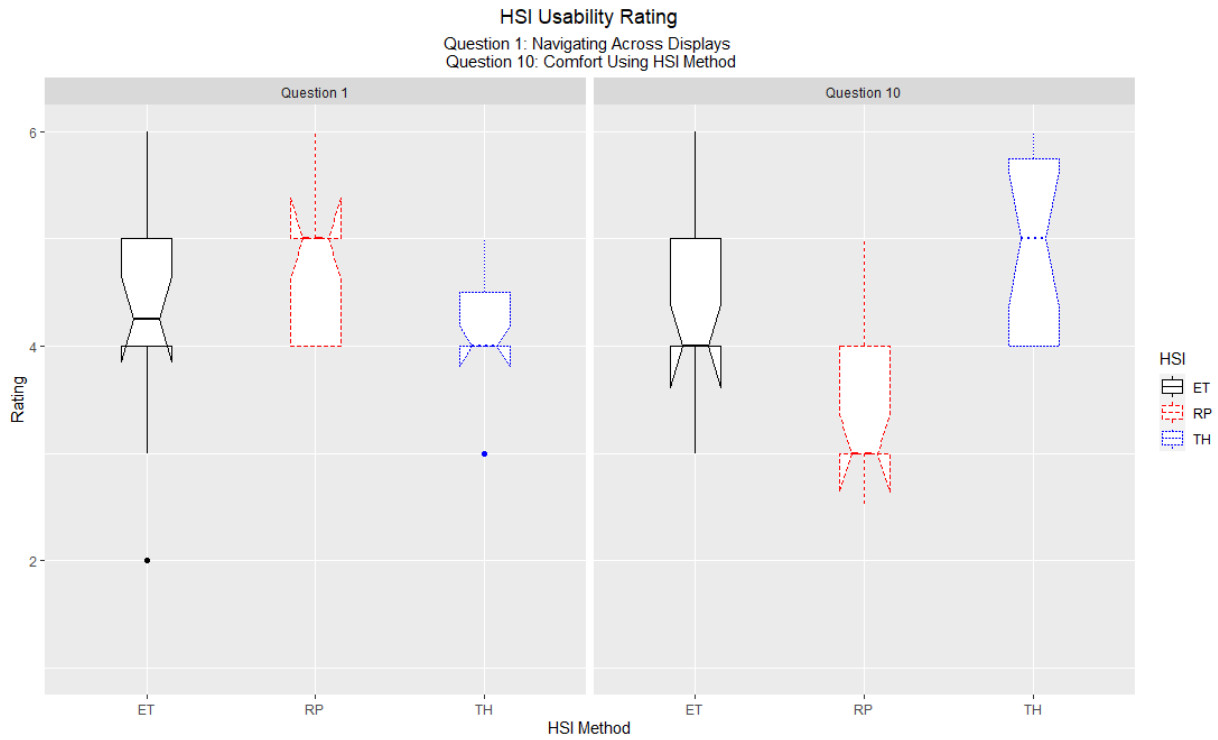


Figure 34. HSI usability survey results; question 1 & 10; Task 1

The requirement for the operator to move his head during maneuvering clearly detracted from RP comfort. Additionally, operators commented on the amount of mental effort required to consciously move their face to point at the extremities of a large format display. This led to distractions from primary flight tasks as well as added discomfort.

Overall, both ET and RP were found slightly more favorable than TH for navigating across the large format display, and had military utility as means to augment TH. As described, RP was typically able to directly designate a single track. Designating tracks with ET typically required additional slew, but this slew was not objectionable due to the speed at which ET acted. TH alone was simply too slow.

Normalized TPV while executing Task 1 is shown in Figure 35, for all three HSI methods.

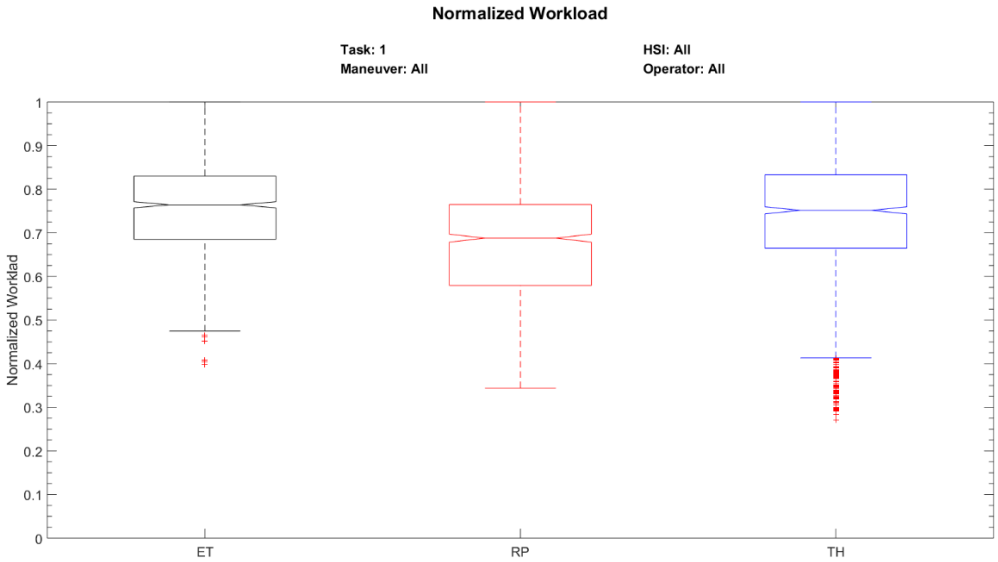


Figure 35: Normalized TPV (quantitative workload) for Task 1

Normalized TPV means were 0.755 (ET), 0.688 (RP), and 0.725 (TH). According to t-tests in APPENDIX D – STATISTICAL ANALYSIS, ET did generate higher TPVs than TH or RP; however RP and TH generated statistically equivalent TPVs. The additional workload when using ET resulted from the requirement to locate the cursor post-snap, as well as the deliberate slew operation. However, TH also required a deliberate slew operation, as well as monitoring and focus in order to direct the cursor to the track in the most efficient path. Operators also perceived additional workload from RP due to monitoring and control of head position. When considering normalized TPV in addition to operator comments, no difference in perceived workload was evident in task 1.

To fully characterize workload, NAP was also evaluated for Task 1. Results for the LWL and HWL maneuver NAP can be seen in Figure 36. This plot overlays the actual discrete sample distribution directly onto the box-and-whiskers summary, since NAP was highly operator dependent. A comparison of the means indicates that maneuver performance was slightly degraded when using ET as compared with RP and TH. Figure 36 in particular suggests that operators attained the best maneuver performance in the HWL maneuver when using Rhino pointing, although variance was high.

A comparison of the dramatic difference in NAP between the LWL and HWL maneuver while using eye tracking, in conjunction with pilot comments, highlights another principal finding from Task 1. As workload increased, the eye tracking maneuver performance distribution significantly narrowed, with operators all performing close to reference for HWL during Task 1. Moreover, the mean NAP for ET increased as workload increased for all tasks flown. In usability surveys, operators praised ET when it was working properly because it minimized the amount of time the operator was required to be “heads down” refining cursor placement. The operator simply had to glance down at the screen and cursor snap, which allowed for more time looking outside, flying the aircraft, and maintaining parameters. Although results from ET remained highly variable, the technology clearly indicated the possibility of allowing an operator to perform higher workload tasks with less time looking directly at the MFD.

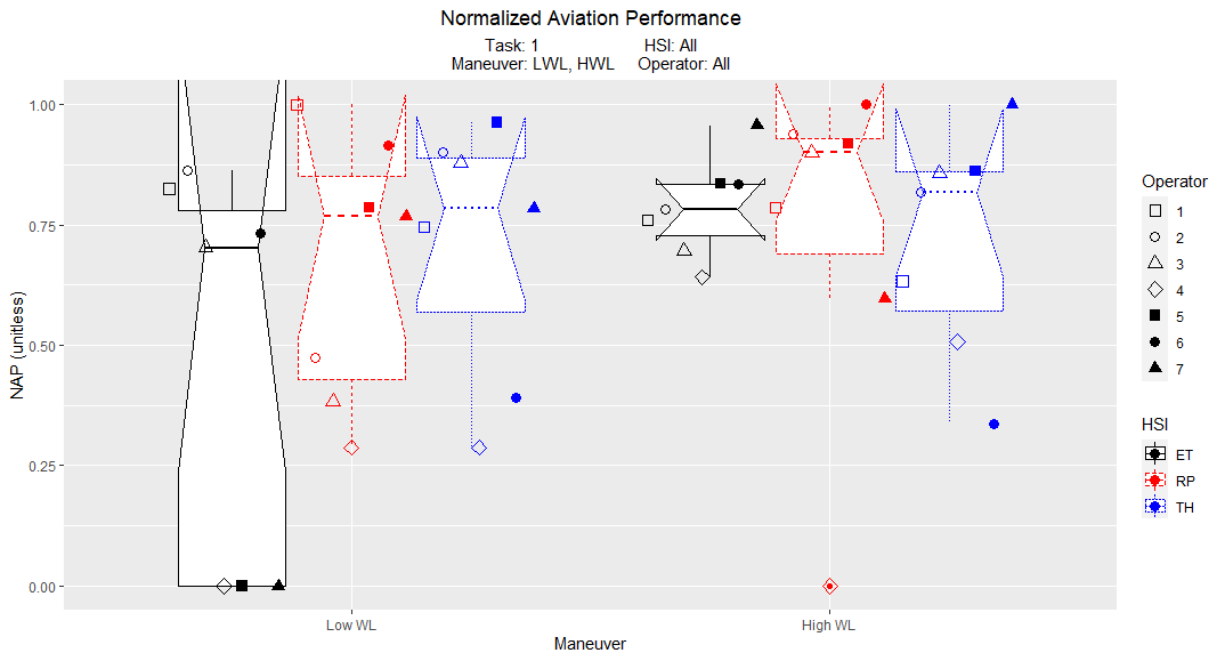


Figure 36: Normalized Aviation Performance, Task 1

2.2.2 Dense Symbol Environment Designation Characterization

The second specific test objective was to characterize operator performance while using eye-tracking, rhino-pointing, and HOTAS to designate a track symbol embedded in a dense symbol environment. Operator performance was characterized by five metrics: (1) the accuracy at expand, (2) the accuracy at designation, (3) the time to designate, (4) the operator's workload, and (5) HSI usability. This test objective was met.

The time and accuracy results were consistent with the results of the single track designation characterization. RP and TH provided the highest consistent accuracy for each user. However, Task 2 was unique in that operators used different strategies to complete the task, especially with TH. Specifically, each operator chose whether to sacrifice accuracy for time or whether to sacrifice time for accuracy. For ET, Task 2 amplified the disparity between time to act and time to complete that was seen in Task 1. The added step in Task 2 to expand before designating, combined with the added clutter, compounded the time delay intrinsic to ET's repeated visual search. Task 2 also amplified the discomfort experienced when using RP, since it doubled the head motion required in Task 1.

2.2.2.1 Dense Symbol Environment Designation Methods and Conditions

All dense symbol environment designation data were collected using a common task, known as Task 2. When the TC transmitted the task, a track symbol appeared on the display as in Task 1; however, additional square symbols appeared randomly across the display and in the vicinity of the green track symbol. 50 total (0.466 by 0.466 cm) square symbols appeared on the display and between 4 - 7 square symbols appeared centered between 0.420 and 0.885 cm from the track symbol. This situation is depicted in Figure 37.

The operator then attempted to designate the track using the assigned HSI method. Designation required two consecutive sub-tasks: cursor placement near the track cluster for expansion, then cursor placement on the track for final designation. A separate action was required for each of these sub-tasks. The display post-expansion is depicted in Figure 38. Observe that the cursor remained fixed relative to the MFD after the

expand operation, rather than repositioning to maintain a relative orientation to the track symbol. This was likely not representative of operational implementation and influenced operator strategies, as discussed below.

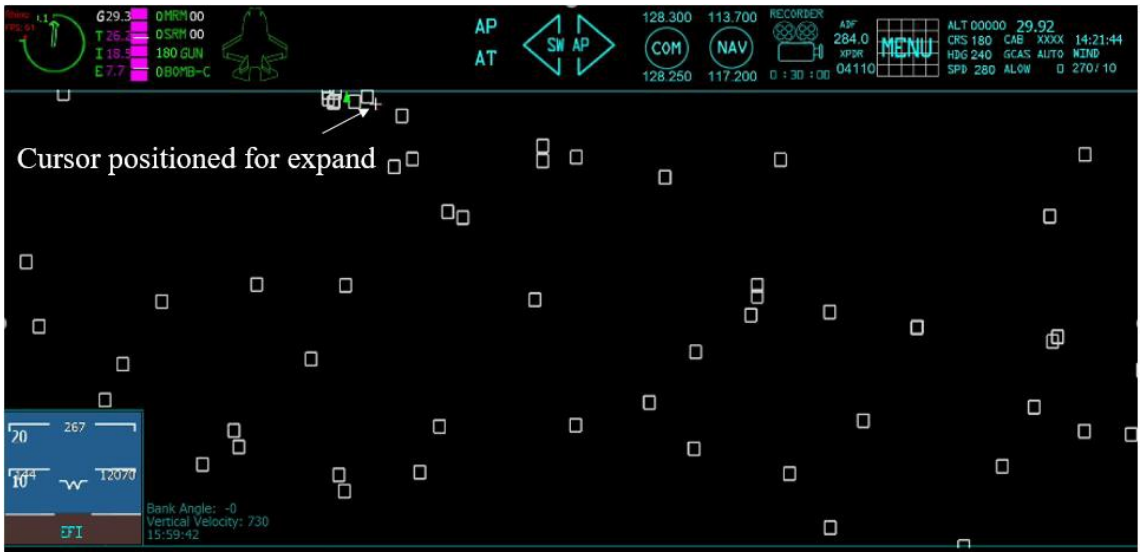


Figure 37: Task 2 at Pre-Expand

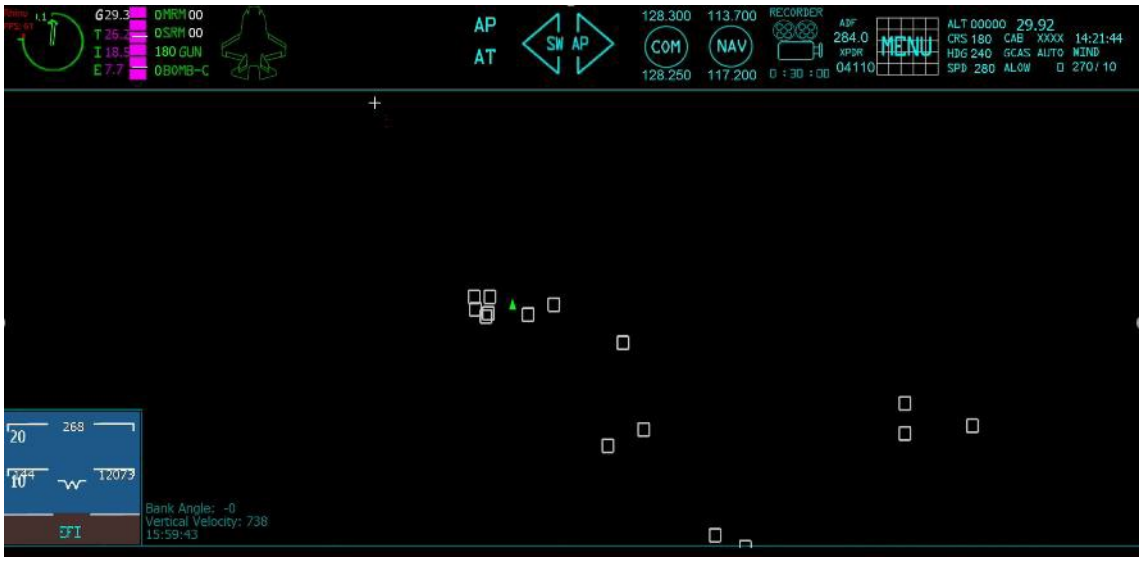


Figure 38: Task 2 at Post-Expand

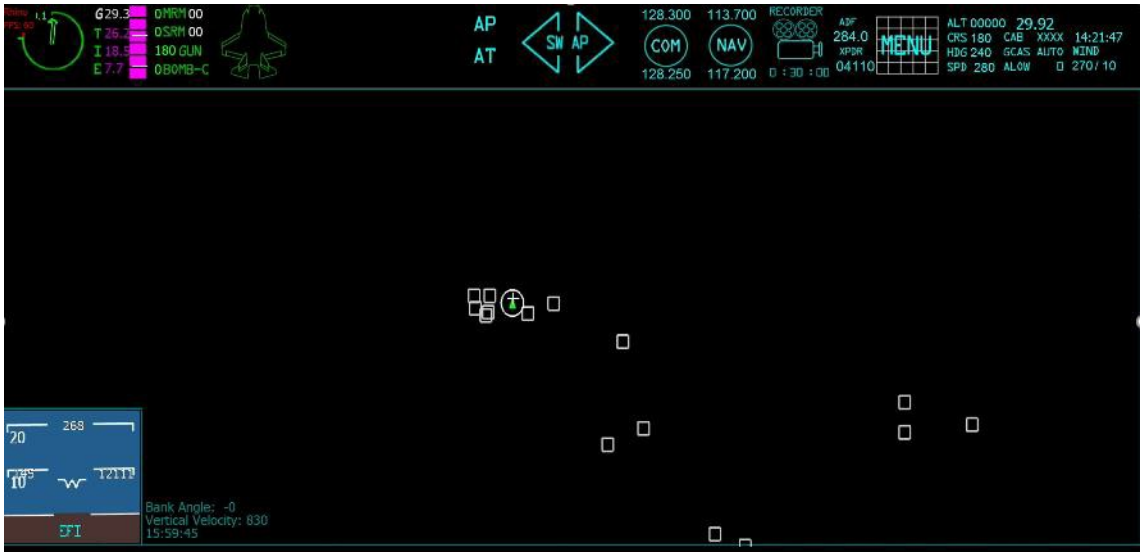


Figure 39: Task 2 at Completion

Task 2 time was limited to ten seconds, and system behavior at 7 and 10 seconds was as in Task 1. A more detailed description of Task 2 is provided in APPENDIX B – DETAILED TASK DESCRIPTION.

Task 2 was executed during the baseline, varying load factor and varying aviation workload maneuvers. Task 2 performance varied significantly by maneuver and operator. As such, two types of population were analyzed: the first binned all maneuvers as a function of operator; the second binned all operators as a function of maneuver.

2.2.2.2 Dense Symbol Environment Designation Results

2.2.2.2.1 Time and Accuracy

A time-history of cursor distances from the track for a typical Task 2 run is shown in Figure 40 and exemplifies the behavior of each HSI method. For all three HSI methods, the operator initially moved the cursor towards – but not necessarily onto – the track in order to expand. The expand operation caused a sharp discontinuity in distance as the expand location moved to the center of the MFD while the cursor remained fixed. The operator subsequently moved the cursor onto the track and designated.

For ET, approximately 0.7 sec of reaction time preceded a nearly instantaneous jump of the focus-boresight towards the track. A snap then occurred at 1.7 sec; as in Task 1, the operator spent this intervening second focusing on the track and commanding the snap. Another period of 0.7 sec in which the cursor did not close in distance toward the track followed the snap operation. Because the focus-boresight was not visible to the user, the operator spent this time visually locating the position of the cursor following the snap and then deciding upon the appropriate input to the cursor slew switch. At the snap, the cursor was only 1.5 cm away from the target, however, it took another 1.6 sec to move the cursor this remaining distance before an expand operation at 3.25 sec.

Post-expand, the same phenomena occurred, with half to full-second delays preceding and following snap. These delays disadvantaged ET in the time to complete metric for Task 2, since a minimum of two snaps were required to complete this task. Nevertheless, in this specific example in which ET calibration yielded an initial accuracy of 3 cm, ET was the fastest of the HSI methods.

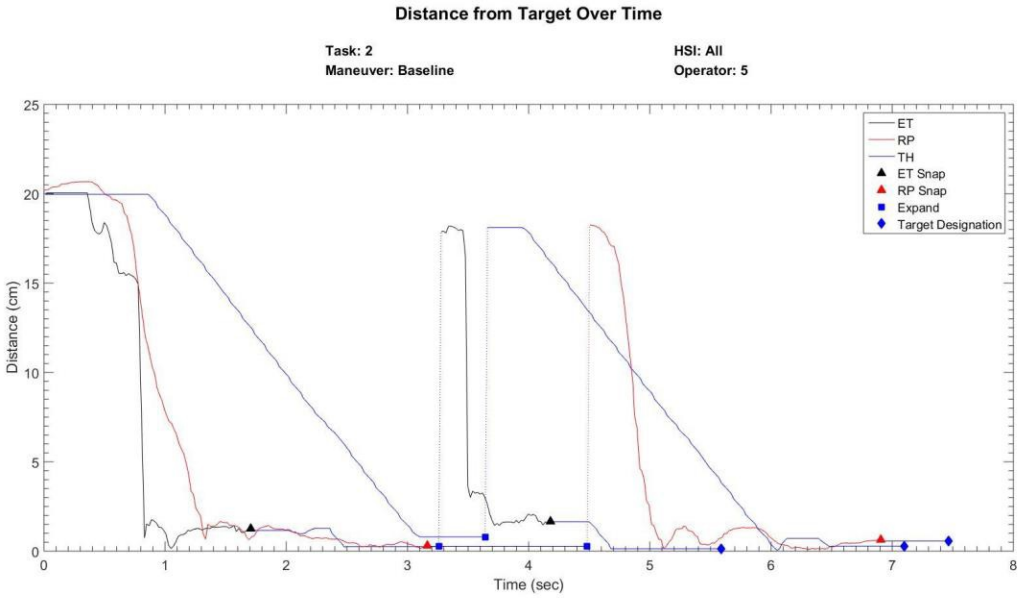


Figure 40: Task 2, Operator 5

For RP, the results resembled those from Task 1: the initial movement towards the target was rapid, closing the first 18 cm toward the track in 1.3 sec; however, this operator then spent nearly 2 sec refining the cursor position prior to expand. This 2 sec position update achieved only 2 cm of increased accuracy, and yielded the slowest expand time. A fundamental result, which held true across the entire test matrix, was that this operator could have successfully expanded prior to the 2 sec cursor refinement with no impact to task success. This operator chose not to because he could see the rhino-boresight and wanted to place it almost directly on the track.

Post-expand, the same phenomena occurred. The rhino-boresight stabilizations made RP the slowest HSI method for this specific example. A distinct oscillatory pattern is visible between 5 and 7 seconds, in which the operator actively used the neck musculature to stabilize the rhino boresight on the track. At any point in this interval, the operator could have snapped and refined the final 1-2 cm with the cursor-slew switch. However, all operators noted a compulsion to stabilize and refine due to the visual presence of the rhino boresight.

TH in this same example (Figure 40) operated as expected; distance from the target over time decreased linearly pre- and post-expand. Reaction times prior to cursor-slew inputs tended to be longer than reaction times prior to moving a rhino- or focus-boresight. In Figure 40, the blue line takes more than a full second to start to initially move toward the track, compared to about half a second before initial closure for ET or RP. Overall, TH was limited by slew rates and input reaction times, which compounded in Task 2 as compared to Task 1. Consequently, TH enabled higher accuracy, but with longer completion times.

The results from Figure 40 provided a heuristic to understand cursor behavior. However, the quantitative relations between the HSI methods in that figure did not always hold; in fact, time and accuracy metrics depended strongly on the quality of the ET calibration. In particular, for the task run in Figure 40, the quality of the ET calibration was high, as indicated by an initial snap to within 3 cm of the track location. ET accuracy was highly variable and did not typically achieve accuracies this high.

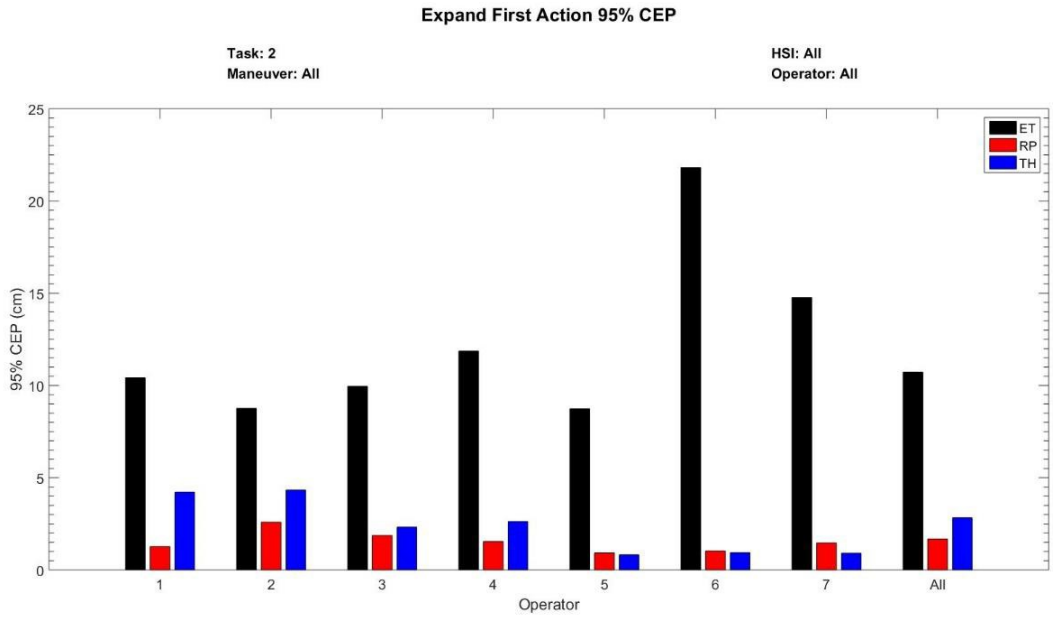


Figure 41: Task 2 CEP95 per Operator

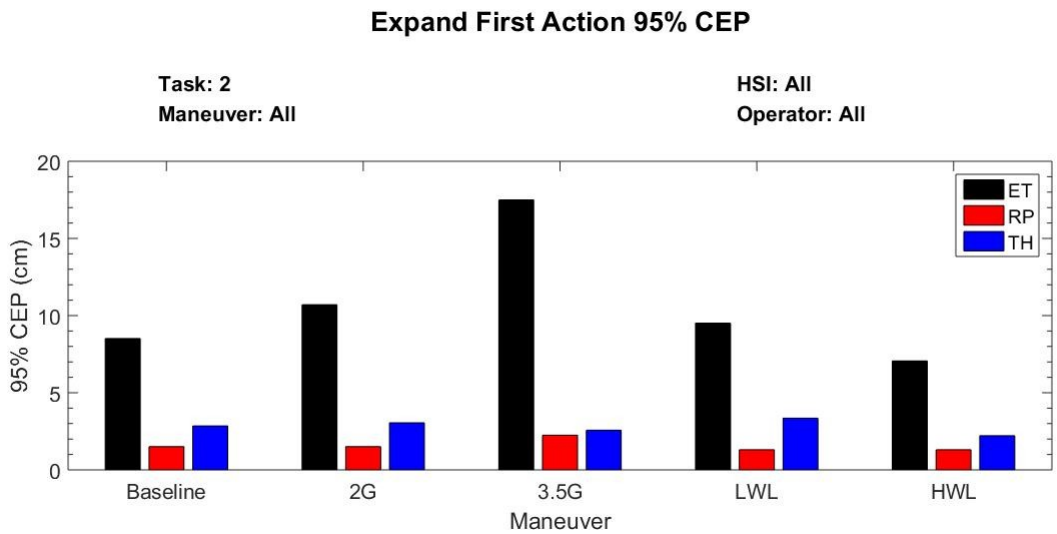


Figure 42: Task 2 CEP95 per Maneuver

A plot of CEP95 for all HSI methods and binned by operator is shown in Figure 41. ET provided the lowest accuracy, for all operators; RP usually provided the highest accuracy; TH provided an accuracy that fell in between, or roughly equaled that of RP. In general, operators 5, 6, and 7 reported the most deterioration in ET calibration throughout their test matrix. In particular operator 6, who reported having a minor eye tropia, experienced very poor ET calibration with an overall CEP95 at first action of 22 cm.

Figure 42 shows CEP95 binned by maneuver rather than operator. Again, ET provided the lowest accuracy; here, RP provided the highest accuracy across all maneuvers. There was a steep drop in accuracy for ET for the 3.5G maneuver. Load factors greater than 2.0 severely degraded the ET system, leading to a drop in accuracy and often a complete inability to perform any tasks. This specific effect, and its causes, will be discussed in more detail later.

Returning to Figure 41, observe that for operators 1-4, RP provided slightly better accuracy and for operators 5-7, TH provided slightly better accuracy. The differences between these two groups indicated different strategies for Task 2 accomplishment. The accuracy cross plots of Figure 43 exemplify the two different strategies. The unfilled symbols on these plots indicate a pre-expand first action and the filled circles indicate an expand action. Operator 2 (left plot) only moved the TH cursor the minimal distance required to successfully expand and have the track appear somewhere on the zoomed display, resulting in a CEP95 of 4.339 cm. Operator 5 (right plot) slewed the TH cursor as close to the target as possible before expanding, resulting in a CEP95 of 0.824 cm. The strategy used by operator 2 resulted in higher CEP95 but shorter completion times; operators 1, 3, and 4 shared this strategy. The strategy used by operator 5 resulted in low CEP95 but longer completion times; operators 5, 6, and 7 shared this strategy figure 46, mean time to expand, demonstrates the cost in time to utilize the higher accuracy strategy: operators 5, 6, and 7 used more time prior to expand when using RP or TH than did operators 1, 2, 3, and 4.

Post-expand, all operators used the same strategy, which resembled an accomplishment of task 1.

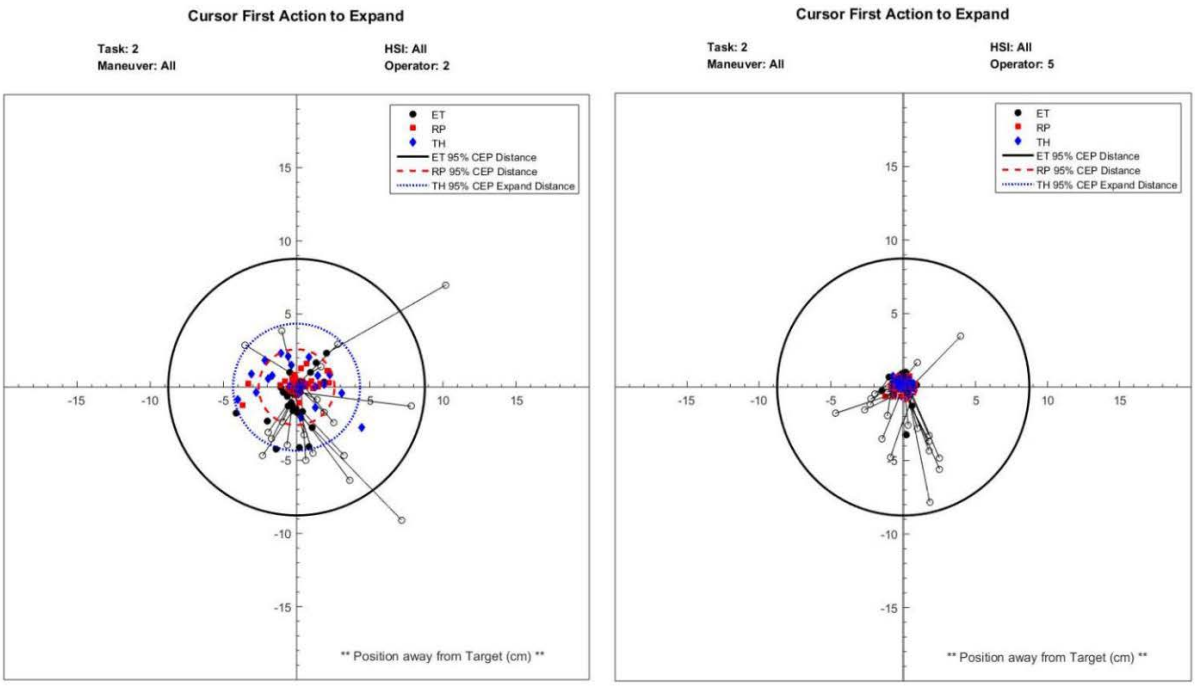


Figure 43. Operator strategies.

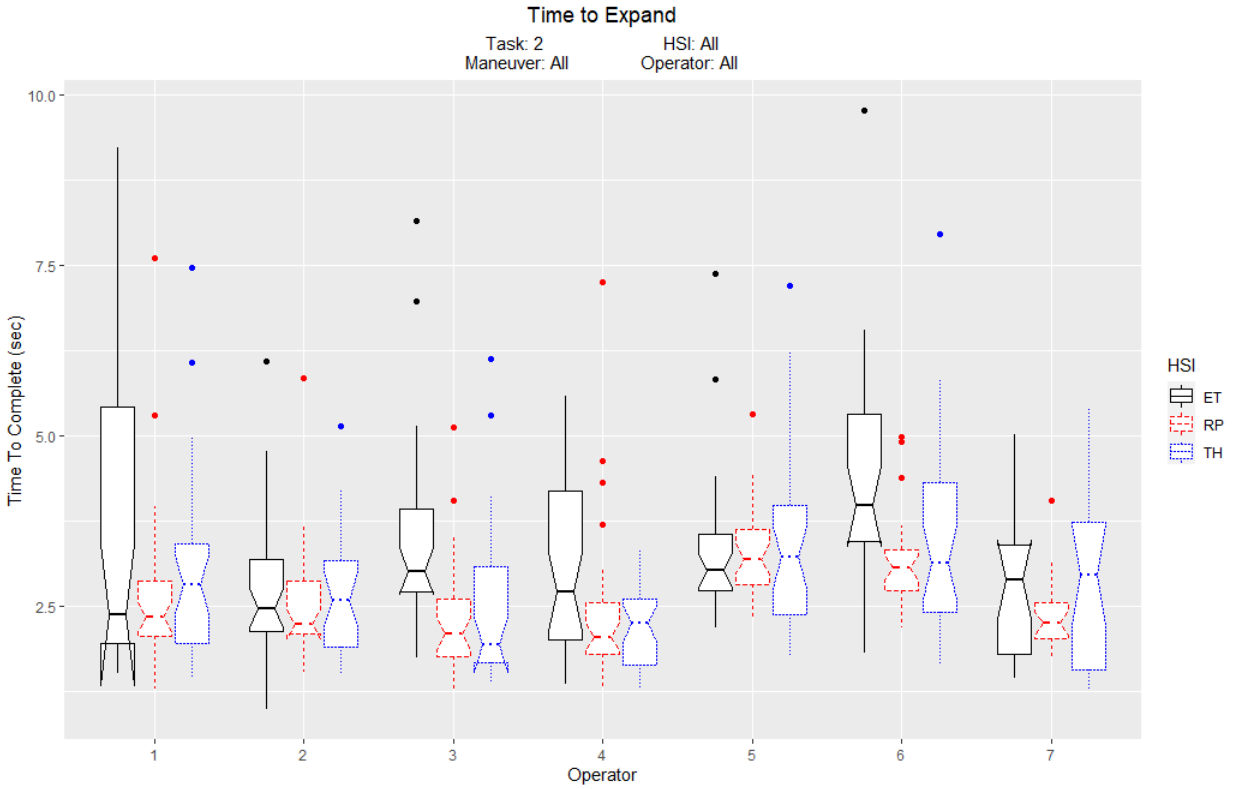


Figure 44. Task 2 Time to Expand, Binned by Operator

Figure 44 additionally shows that ET forced the longest mean time to expand for 6 of 7 operators. This remarkable result directly contradicts the results from Task 1, in which ET provided the fastest time to first action. Eye tracking delays came in two forms: eye focusing plus reaction time after a large shift in the focus-boresight, and the time required to locate the cursor after an inaccurate snap actuation. In Task 2, operators indicated that both of these effects increased due to the increased clutter on the screen; these increases sufficed to make ET the slowest method rather than the fastest.

When looking at total completion time in Figure 45 these two effects further compounded, since Task 2 required two snap operations. Furthermore, the inconsistent accuracy of ET sometimes forced the operator to perform even more than two snaps. Thus, ET forced the longest mean time to completions for all operators. In fact, the results for ET were worse than suggested by figure 46, because task 2 was limited to 10 seconds. When executing the 3.5 G maneuver, eye tracker performance sometimes deteriorated such that task timeout occurred without task success. The time to completion distributions would have shifted to larger values had time been unlimited. This effect is discussed further in 2.2.3.2.1.

The t-tests in APPENDIX D – STATISTICAL ANALYSIS indicate no meaningful statistical difference between operator performance when using RP and TH. Some operators acted slightly quicker with RP; some operators acted slightly quicker with TH. However, the distributions as a whole showed indistinguishable means both for time to expand and time to completion. Since the CEP95 results showed unnecessarily high accuracy at expand for RP due to time-intensive boresight refinement, these time results suggest that RP could complete this task significantly quicker than TH by obviating this boresight refinement effort for components of the task that did not necessitate high accuracy. **Develop a method to**

reduce the tendency for operators to fixate on making fine adjustments with the rhino-boresight (R4).[1] See APPENDIX J – FUTURE WORK for potential solutions for this issue.

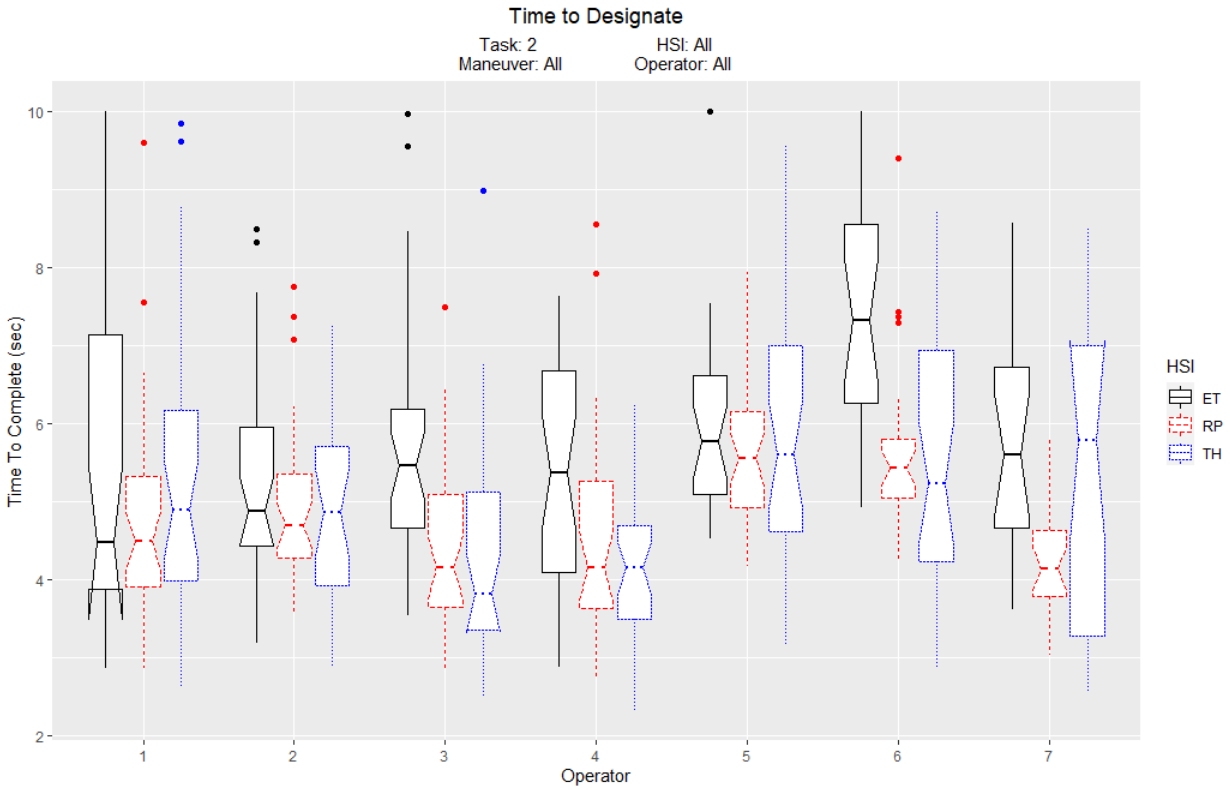


Figure 45. Average Task 2 Completion Times for All Operators

As in Task 1, the initial track position for Task 2 was randomized for each run. Therefore, time alone did not fully characterize operator speed using each HSI method. If the track appeared a long distance from the initial cursor location, that drove larger completion times. The effect of this larger (or smaller) distance varied between ET, RP, and TH. Figure 46 shows an example distribution of pseudo-velocities for each HSI method for all operators.

[1] Numerals following an R represent recommendation numbers, in priority order.

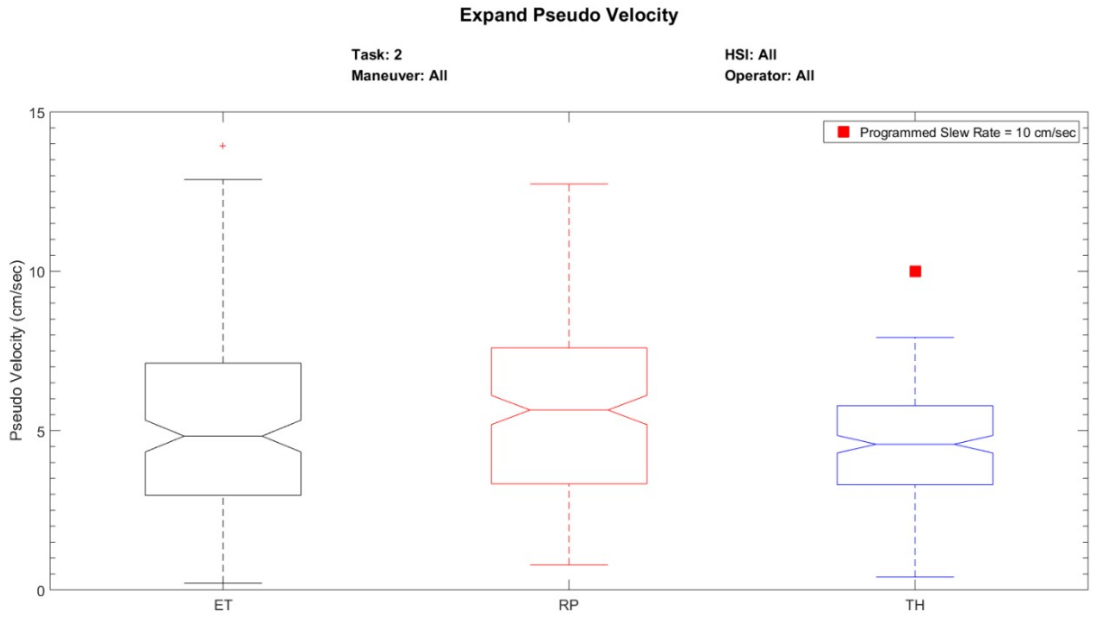


Figure 46. Task 2 Expand Pseudo Velocity

In pseudo-velocity, TH shows a relatively narrow distribution when compared to ET and RP due to the fact that TH had a fixed slew rate (i.e. pseudo velocity). The programmed TH maximum slew rate was 10 cm/sec, which was faster than the entire TH pseudo-velocity distribution due to delays caused by the operator locating the track, imperfect trajectories towards the track, and fine refinement prior to designation. These aspects combined to limit the task velocity to less than the maximum slew rate. ET and RP showed wide pseudo-velocity distributions, indicating inconsistent performance across all runs. However, the mean pseudo velocities for both ET and RP were both greater than that of TH, with no statistically discernible difference between their means.

Large format displays in particular require a cursor to travel larger distances. The box-and-whiskers plots of Figure 47 demonstrate the population results when only considering trials with large distances to travel. Specifically, this figure includes only trials with the largest 15% of initial cursor to track distances, but includes all operators and maneuvers.

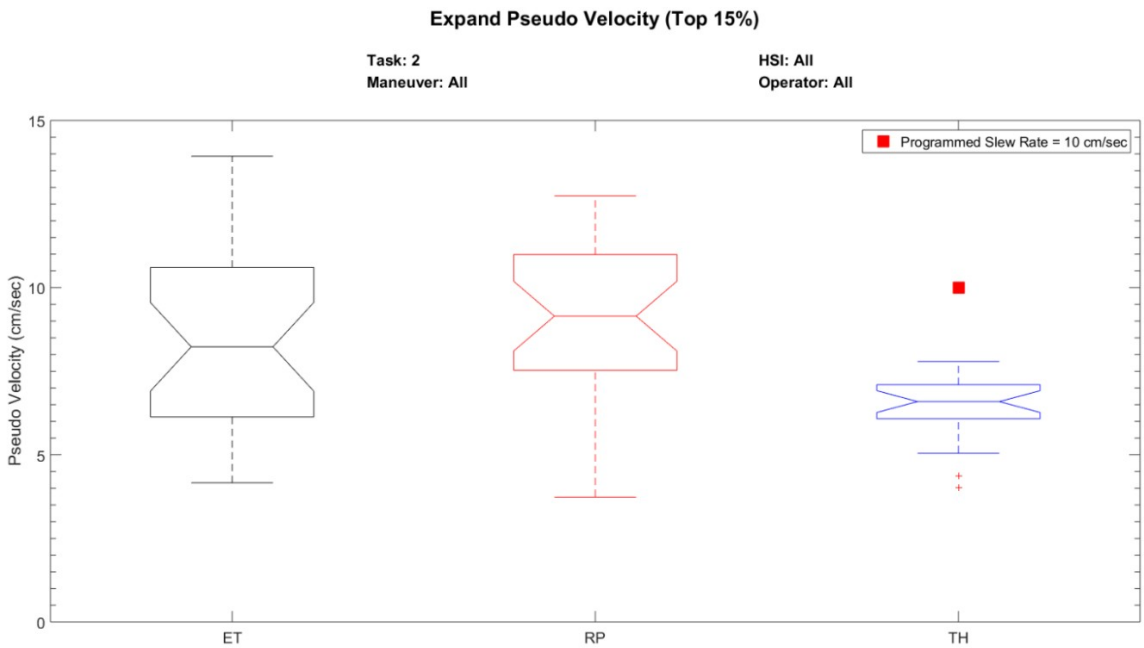


Figure 47. Expand pseudo velocities for the largest 15% of initial cursor to track distances.

This largest 15% population resulted in initial distances between the track and cursor ranging from 23 to 43 cm. The mean pseudo velocities for ET, RP, and TH resulted 8.40, 9.08, and 6.45 cm/sec respectively. This plot magnifies the high gross movement capability of both ET and RP, which had a clear and significant advantage in time to first action over TH. Furthermore, the results indicate no substantial difference in gross movement capability between ET and RP.

The smallest 15% of initial cursor and track distances were also evaluated and are depicted in Figure 48. The smallest 15% of initial distances encompassed 1 to 5 cm. The mean pseudo velocities for ET, RP and TH resulted 1.81, 2.07, and 2.35 cm/sec, respectively. Based on the statistical analysis, RP and ET were equivalent over small distances; ET, however, was slower than TH. Thus, TH had an advantage in time to first action over small distances.

In conclusion, TH was optimized for small displays only. An analysis of pseudo-velocity and screen size in APPENDIX D – STATISTICAL ANALYSIS indicates that for any MFD with cursor distances to travel greater than 15 cm, TH yielded distinctly inferior performance to that of RP or ET. **Augment traditional HSI with a mechanism to drive cursors rapidly across large-format displays (R1).**

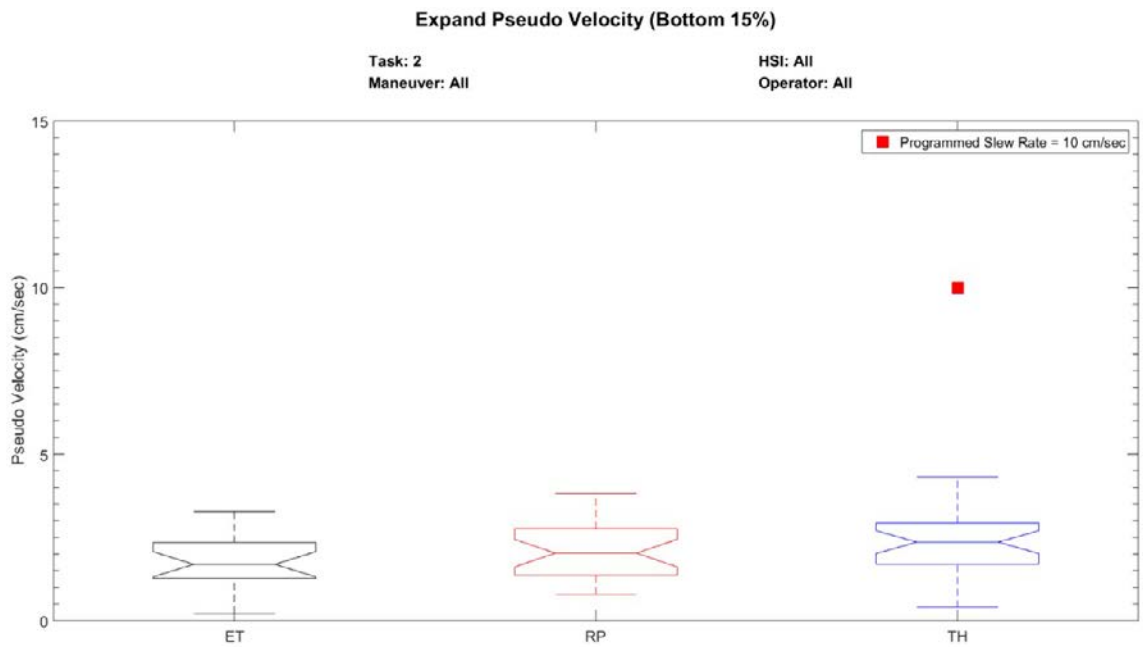


Figure 48. Expand pseudo-velocities for the smallest 15% of initial cursor to track distances.

2.2.2.2.2 Workload and Usability

Normalized TPV for Task 2 yielded mean normalized TPV values of 0.766 (ET), 0.705 (RP), and 0.729 (TH), as shown in Figure 49.

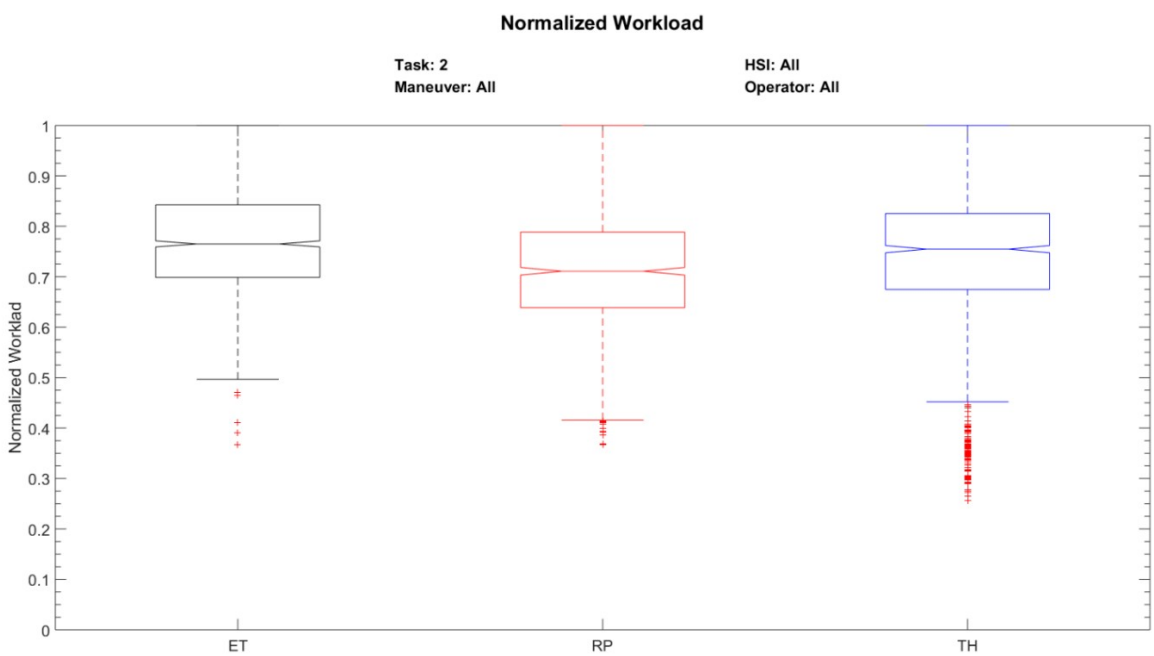


Figure 49: Normalized TPV (quantitative workload) for Task 2

T-tests indicated that ET resulted in higher normalized TPV when compared to RP and TH, just as in task 1. In Task 2, however, the additional step required also elevated the TH normalized TPV above that of RP. This analysis indicated that task 2 was performed with the lowest quantitative workload (indicated by normalized TPV) using RP.

NAP analysis for Task 2, shown in Figure 50 yielded variable data sets with results highly dependent upon operator. However, in general as workload increased between the LWL and HWL maneuver, the relative NAP for ET improved significantly. This hints that as workload increases, ET might additional extra capacity to maintain maneuver performance. Although the distributions were wide, TH also generally allowed higher NAP than RP. This corroborated an operator perception that the size of head movements in RP caused the VHUD to exit the peripheral field of view, and that those head movements increased time between maneuver cross-check.

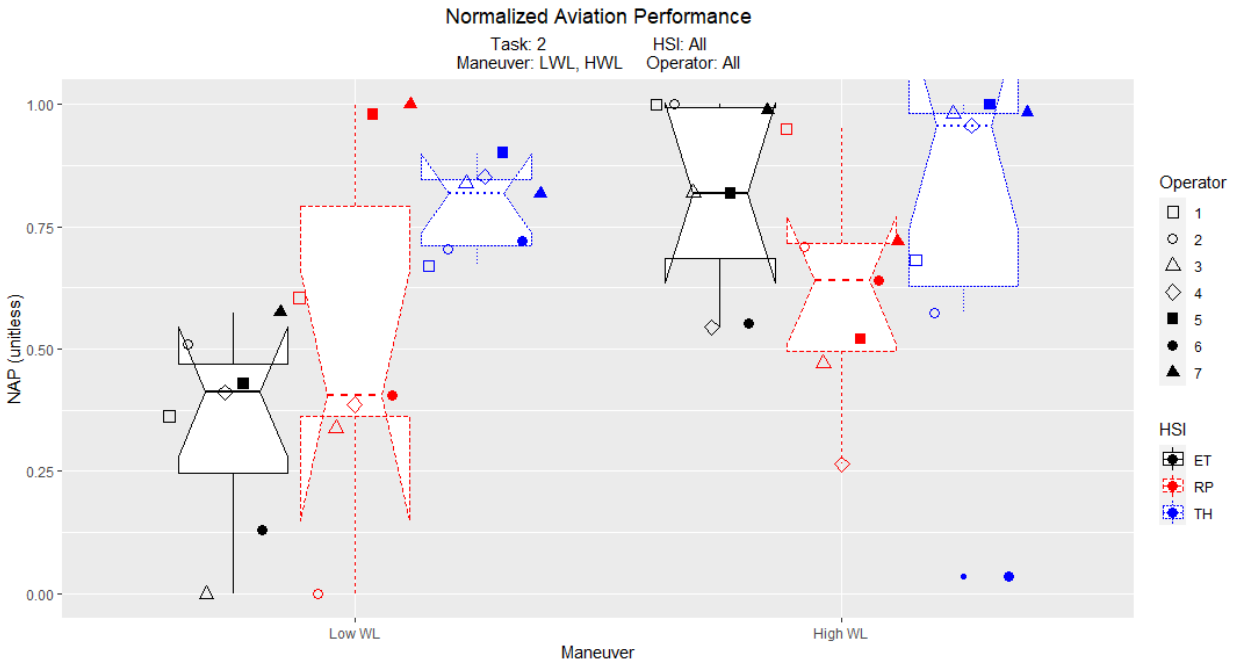


Figure 50. Normalized Aviation Performance, Task 2

In sum, RP had the lowest average workload, but also the lowest average NAP. These two results confound. As a whole, the workload results for Task 2 between the three HSI methods were inconclusive.

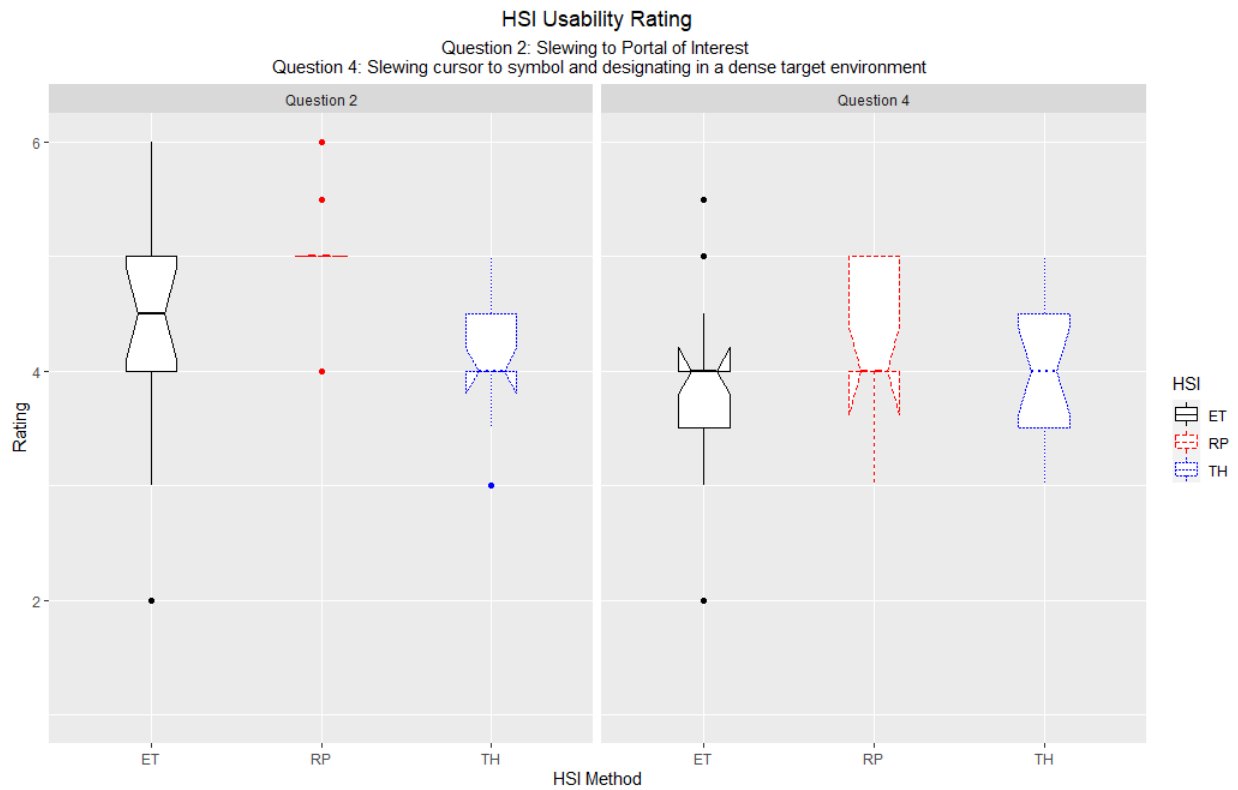


Figure 51: HSI usability survey results; questions 2 & 4; Task 2

Figure 51 shows the results of usability survey questions 2 and 4 for all operators. The mean ratings for question 4, which addressed the dense target environment specifically, were equal amongst all three HSI methods. However, operators commented upon the different strategies for the expand portion of the dense target environment task. Specifically for RP, some operators commented upon the lack of perceived need to get a precise cursor snap knowing that upon actuating the expand function, the cursor would still be in close proximity to the track. This technique led to faster expand operations, which culminated in faster overall designation times. This also proved beneficial for ET. As long as the ET snap was within a generally desired expand area, further refinement was not needed.

As reflected in the responses to question 2, which concerned slewing to different regions (or "portals" on the display), operators commented upon the desirability of rapid cursor snaps across entire large format display when using ET and RP. This effect was clearly amplified in the dense target environment, which required two snaps in order to achieve a designation.

Overall, operators highlighted that both ET and RP possessed many benefits compared to TH when manipulating a cursor over a large display in a dense target environment, and their military utility was high. TH was consistently able to achieve precise and repeatable results, especially when compared to ET. This led operators to rate TH highly in comparison to the other HSI methods due to their confidence and familiarity with TH. However, comments centered on the need for faster ways to transit the large format display when using TH.

2.2.3 Data Entry Characterization

The third specific test objective was to characterize operator performance while using eye-tracking, rhino-pointing and touchscreen input to conduct a data entry task on a large area display. Operator performance was characterized using five metrics: (1) the accuracy of each input relative to the desired key,

(2) overall data input error rate, (3) the time required to enter data, (4) the operator's workload, and (5) HSI usability. This test objective was met.

Overall, TH was the fastest and preferred method for entering data. RP was almost as fast as TH, but operators commented on the perceived increased workload required to achieve data input accuracy. The text data entry task, in which the keyboard extended over a large portion of the display, magnified discomfort with RP. ET was the slowest and the least preferred method for entering data. Operators noted large and random snap inaccuracy which led to the increased time to complete the task. All three HSI methods input data with nearly equal accuracy, resulting in a data entry error rate of approximately 3% independent of method.

2.2.3.1 Overall Testing Methodology and Conditions

All data entry characterization data was collected using a common task, known as Task 3. When the TC transmitted the task, the MFD displayed an array of data input cells, initialized to the first level of a multilevel menu; Figure 52 shows that start screen.

The operator referenced a two-sided data entry card with numbered lines of data. After sending the task, the TC read a line number to the operator, who located that line on the card. The operator then accessed the appropriate menu path, as depicted in Figure 53. Each menu path terminated in a data entry keyboard, as depicted in Figure 54. At the keyboard, the operator input the line of data as quickly as possible while minimizing errors, using the prescribed HSI method. Each data entry line required ten key actuations. For ET and RP, a snap was required prior to any key actuation, even if the cursor was already on that key. For TH, touchscreen was required for all key actuations.

Unlike in other tasks, which began timing after the TC sent the task, time to completion in Task 3 was measured starting from the first keystroke until the final "ENT" keystroke. Accuracy was determined as the distance between the center of the desired key and the cursor at first action. Error rate was calculated by taking the number of errors over total number of keystrokes. If an operator skipped a key on an input, or inadvertently input an extraneous key, it was counted as a single error. (This error rate was distinct from accuracy, a measure of the geometric miss distance between the center of each intended key and the point of actuation of that key.) Error rate was determined manually after each sortie, as clarification was occasionally required to determine which key the operator intended to press. The workload and HSI usability surveys were also used as discussed in previous tasks.

Figure 52. Data Input Screen, Level 1

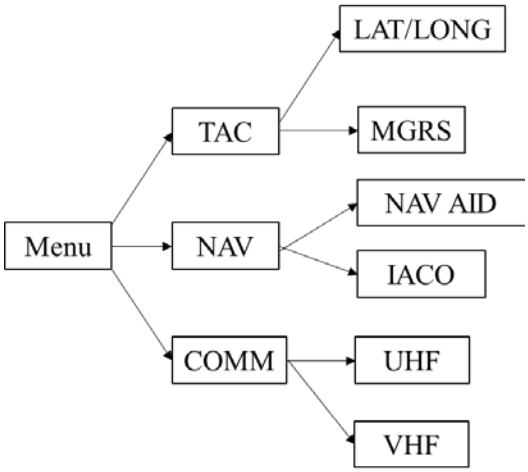


Figure 53: Data Menu Branching

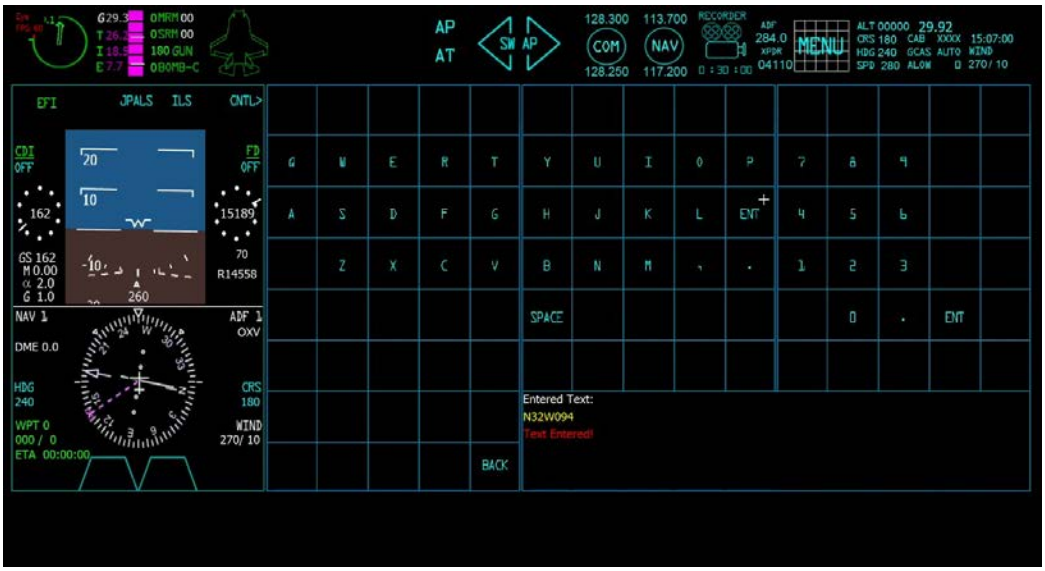


Figure 54. Data Entry Keyboard

Task 3 was executed during the baseline, varying load factor and varying aviation workload maneuvers. Task 3 performance varied somewhat by maneuver and operator. As such, two data sets were analyzed: the first binned all maneuvers as a function of operator; the second binned all operators as a function of maneuver. Where no dependence on operator or maneuver was noted, data for analysis was binned only by HSI method, though single-operator results also appear below for clarity and emphasis.

2.2.3.2 Overall Test Results

2.2.3.2.1 Time and Accuracy

The biggest delineator between the different HSI methods was the time to complete the task. Figure 55 below shows the distributions of time to complete the task for each operator and HSI method. TH was the fastest for all operators followed by RP and finally ET. An overall time to complete by HSI method is located in Figure 56. ET had the most variability followed by RP and the TH. As in previous tasks, the slowness of ET resulted from occasional cursor snaps to unpredictable locations, which required visual location of the cursor and additional slew time to the correct key. Mean times to input a full line of data were 31.4 sec (ET), 18.6 sec (RP), and 9.9 sec (TH), indicating unambiguous operator performance differences between the three methods.

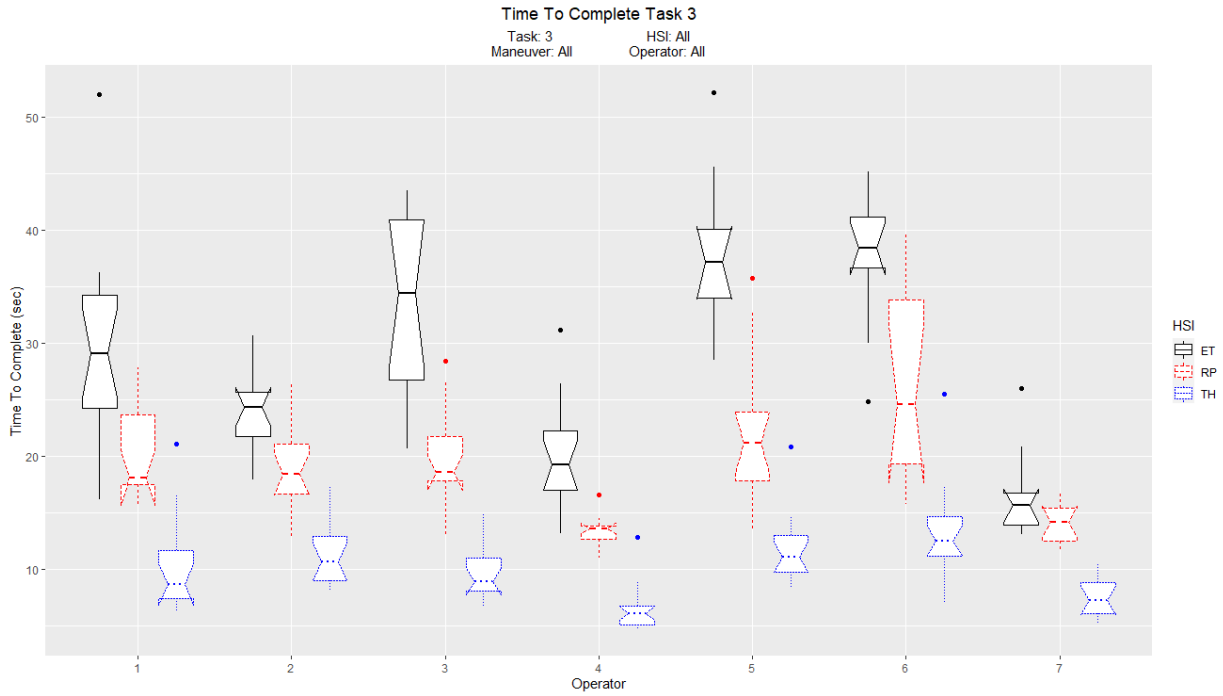


Figure 55: Data Entry Time to Complete by Operator

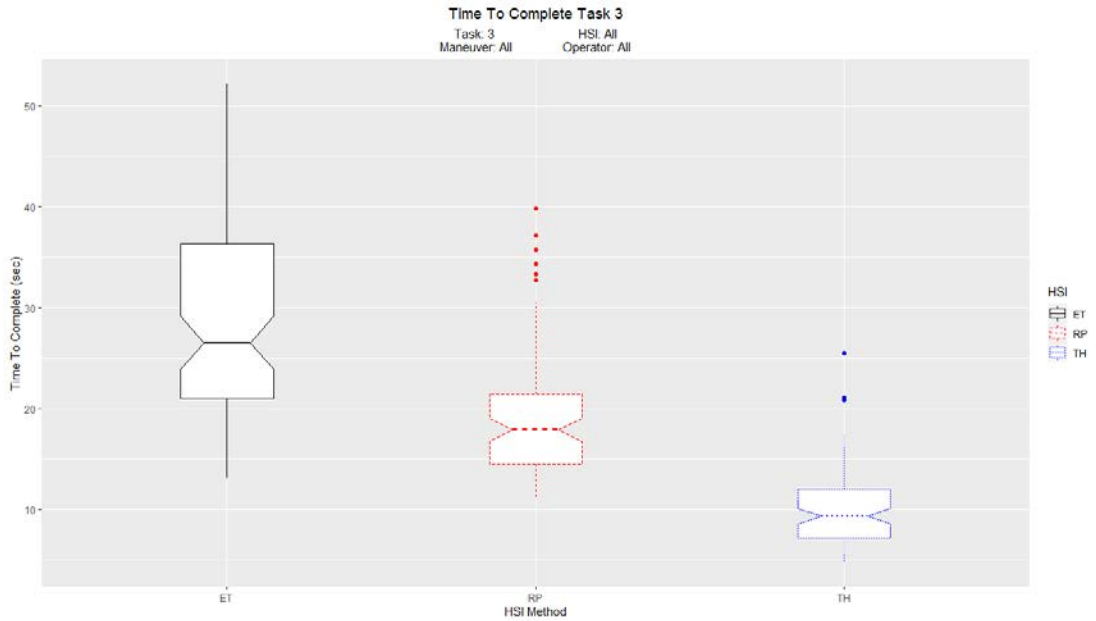


Figure 56: Data Entry Time to Complete by HSI method

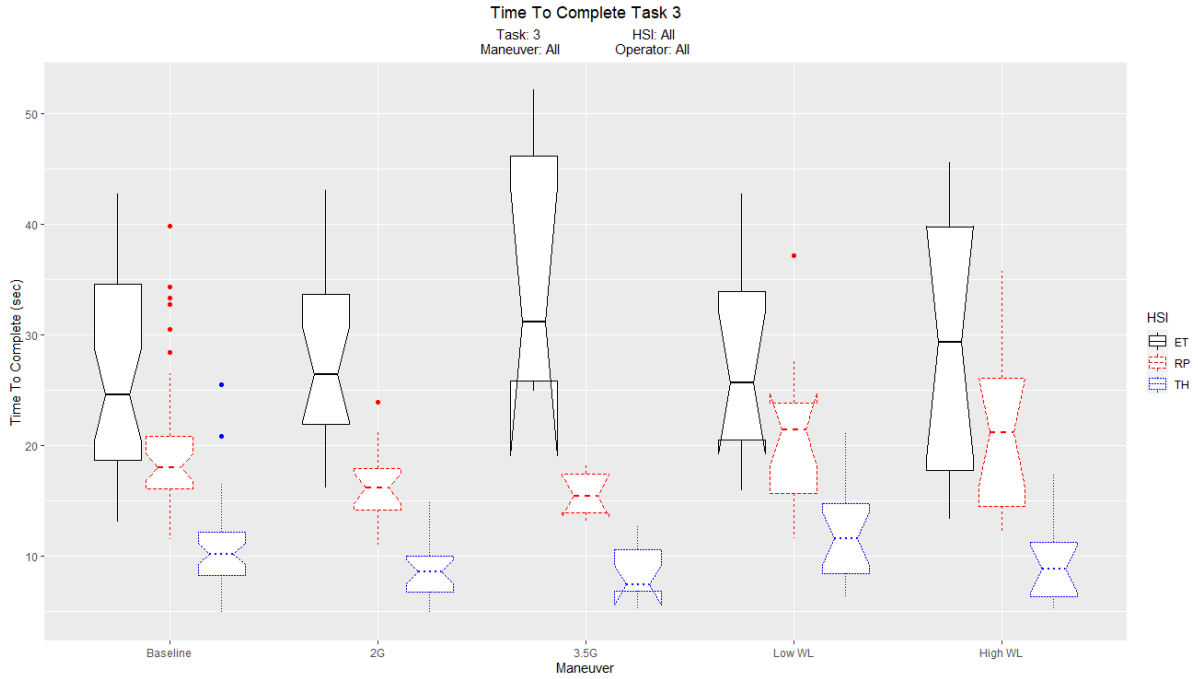


Figure 57. Data Entry Time to Complete by Maneuver

Figure 57 shows the breakdown of times to completion by maneuver. TH still had an advantage over RP and ET however the difference was most profound during the 3.5G maneuver. ET had a significant increase in the time to completion at 3.5G while the other HSI methods were somewhat consistent across all maneuvers. Minor movements of the helmet in relation to the head resulted in large degradations in ET accuracy. At 2.0 G, such helmet motion was not discernible through ET performance. At 3.5 G, however, multiple task 1 and 2 test points resulted in timeout, while at least one task 3 test point required over 180 seconds to complete (an outlier not depicted in Figure 57 for clarity of scale). During the test, multiple snap actuations were required to make the cursor appear anywhere on the screen. Eventually, all operators developed a compensatory technique in which at 3.5 Gs they looked 6 to 9 inches below – but on azimuth

with – the point to which they wished to snap. The success of this strategy indicated that the downward rotation of the helmet under G caused a large vertical bias error. In general, the effect of load factor above 2.0 made ET unusable. **Develop a method to compensate for the effect of load factor on eye-tracking accuracy (R3).** Future research ideas in order to address this issue are located in APPENDIX J – FUTURE WORK.

Accuracy at first action results are shown in Figure 58 for a single operator, and echoed the results from Tasks 1 and 2. The black square in the middle of this graphic shows the relative size of a key from the Task 3 keyboard (2.54 cm on each side). CEP95 for TH was 0.95 cm, and so it lay entirely within the key. CEP95 for RP was 2.32 cm (a radial measurement), and thus it lay just outside the key, although a significant portion of the RP distribution lay inside the key. CEP95 for ET was 10, or approximately 7 times the size of the key. These values allowed TH and RP to usually directly select the desired key, while ET usually required slewing with the cursor prior selecting a key.

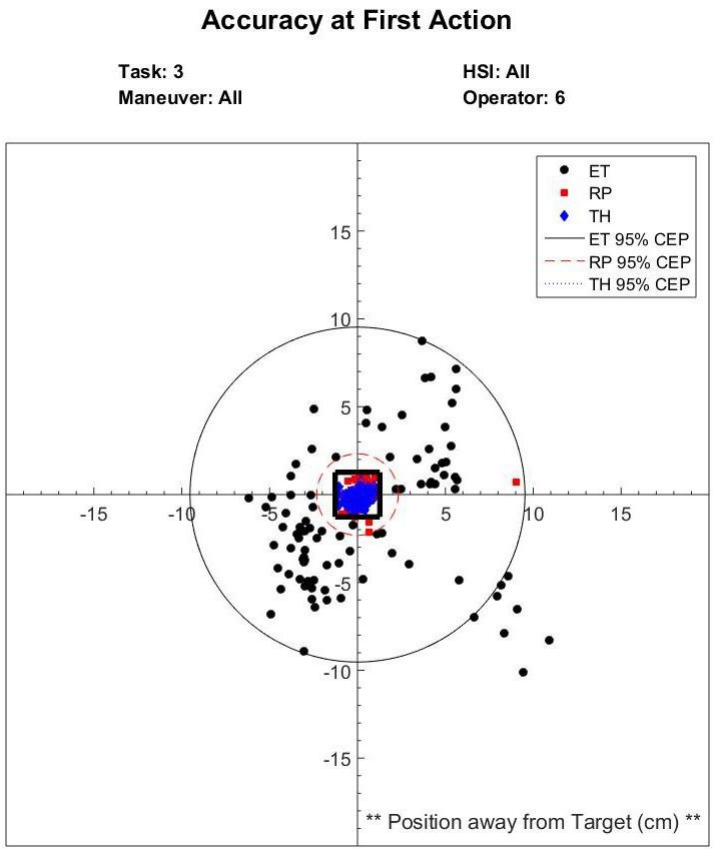


Figure 58: Data Entry Accuracy at First Action for Operator 6

The data entry error rate was consistent across all HSI methods. Across all operators, ET had an error rate of 2.1%, RP had an error rate of 3.1% and TH had an error rate of 2.7%. ET had the lowest error rate most likely due to operators having to slew using the cursor after first action since the cursor was rarely on the correct key post snap. Thus, the significantly longer time to completion for ET did yield slightly improved performance. RP had the highest error rate, due to operators attempting to direct-designate and sometimes being slightly off the correct key due to unrecognized rhino-boresight movement. This RP error source was strongly affected by turbulence. Multiple operators were surprised when encountering light turbulence that the rhino-boresight oscillated unpredictably, even with apparently minor aircraft motions. This effect also appeared in usability surveys for question 13, in which operators rated RP on average the worst HSI method for usability under G. The errors when using TH were mostly due to an operator inputting

the same key twice because of aircraft motion. Turbulence, G, and an inability to brace the arm during data entry combined to cause these occasional double entries when using touchscreen.

TH and RP were both accurate enough to direct designate using the key size that was representative of fifth generation aircraft implementations. TH was consistently faster which would aid the operator in returning back to other tasks quicker. In an operational scenario this would have led increased situational awareness. However, the accuracy of touchscreen with its TRL of nine nearly equaled the accuracy of RP and ET, which had much lower TRLs. Where data accuracy is critical, such as when entering weapon coordinates during close air support, the touchscreen error rate of 2.7 incorrect entries per 100 would not provide satisfactory performance without pilot compensation.

2.2.3.2.2 Workload and Usability

As in Tasks 1 and 2, normalized TPV and NAP data were noisy and highly operator dependent. Nonetheless, several trends were discernible.

Figure 59 shows an example comparison between normalized TPV and NAP for operator one executing Task 3 during the HWL maneuver. It shows that the TPV was higher for TH and there was little difference between ET and RP, though the mean normalized TPV for RP was slightly less. These results are only true for this operator and maneuver, and the full set of workload data can be found in APPENDIX E – ADDITIONAL PLOTS. However, for both LWL and HWL maneuvers, fully 86% of operators exhibited their highest TPV when using TH, and 71% exhibited their lowest TPV using RP, just as in this example.

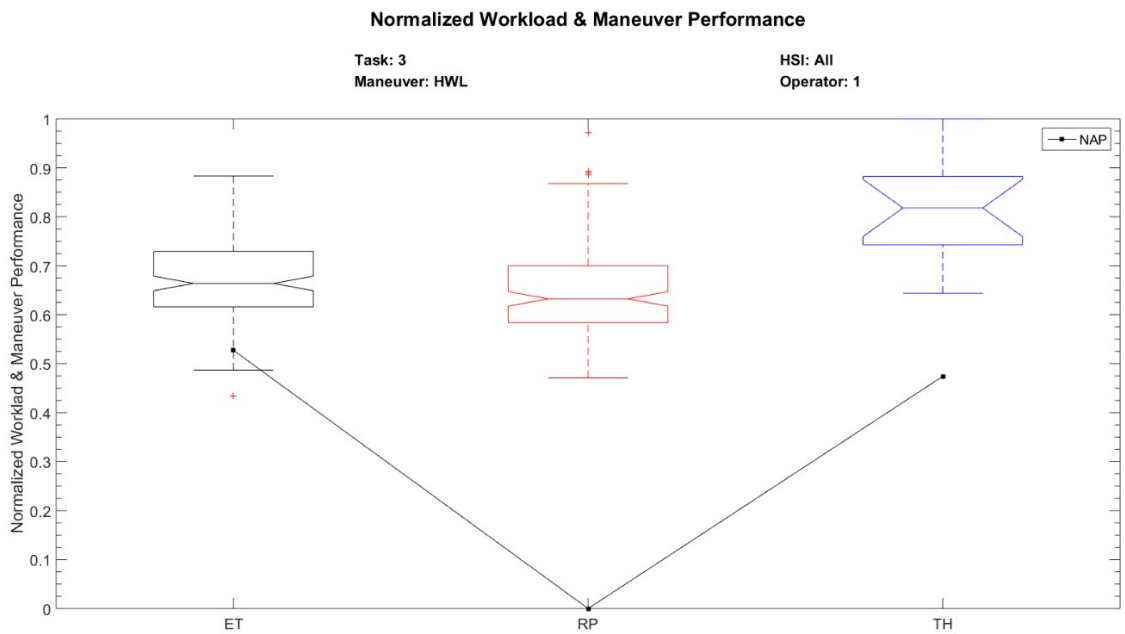


Figure 59: Normalized TPV (quantitative workload) and NAP (maneuver performance) for operator 1 during Task 3 HWL maneuver

In Figure 59, NAP for the HWL maneuver is superimposed on the normalized TPV plots as a black line. This line shows that operator one flew most precisely while using ET and most imprecisely while using RP. The remainder of NAP results appear in Figure 60. However, for the HWL maneuver, 71% of operators exhibited their highest NAP while using ET and 43% (a plurality) exhibited their lowest NAP while using RP, just as in this example. For the LWL maneuver, a small majority of operators (57%) exhibited their best NAP while using TH, rather than ET, but RP still exhibited the worst NAP.

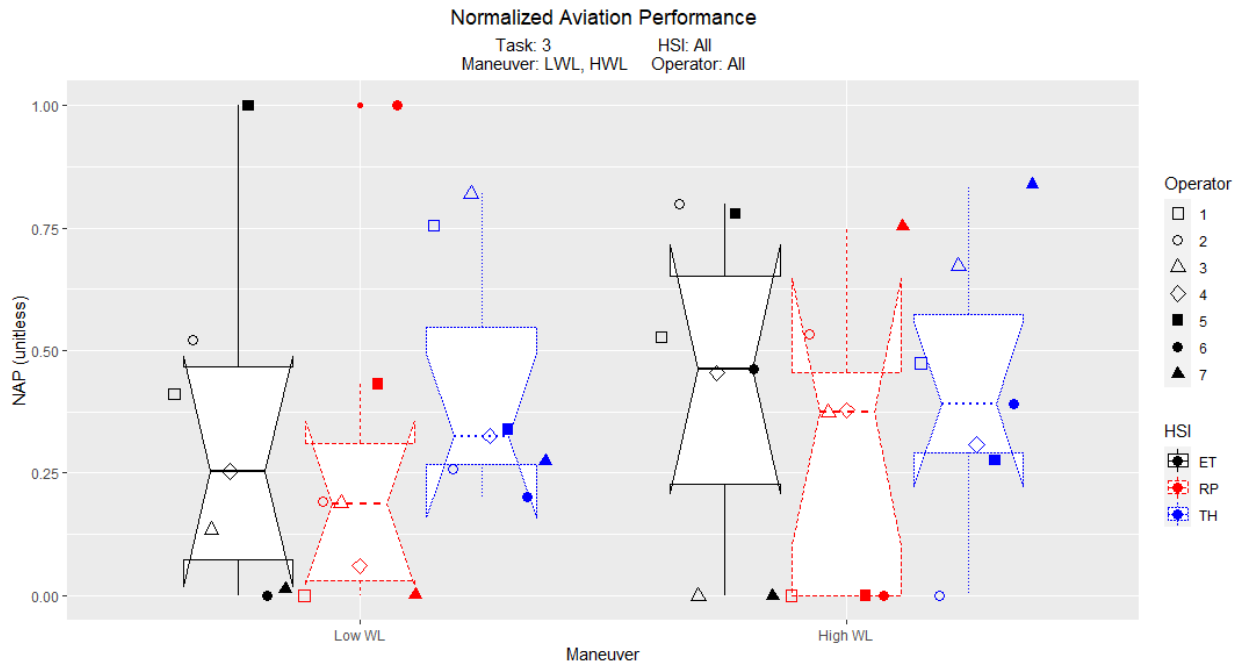


Figure 60. Normalized Aviation Performance, Task 3

Thus, RP tended to generate the lowest workload but also the lowest maneuver performance – the same result found in section 2.2.2.2.2. The lower workload when using RP likely resulted from the capability to enter data without removing hands from the flight controls, in combination with a reliable and accurate HSI method. However, a majority of operators flew worse aircraft parameters while using RP. This resulted from attempts to direct-designate, which forced longer head down time. During this head down time, operators were often unable to see the VHUD in their peripheral vision, due to the large head motions required. Increased discomfort also likely increased operator workload.

These data also suggest that TH drove the highest workload, albeit with minimal drop in performance. The high workload likely resulted from the requirement to remove hands from the controls in order to manipulate the data menus. This forced operators to work harder in order to maintain aircraft parameters while flying the aircraft with only one hand. Additionally, the large format display required pressing buttons on either side of the aircraft control stick. This required the operator to swap hands at times or reach across the aircraft controls in order to access the entire screen, which increased workload. NAP while using TH did not significantly suffer from these hand motions, likely due to the experience most operators had with TH. NAP scores when using TH for operators 5 and 6, who had the least experience with TH by a large margin, fell in the bottom 50th percentile of NAP scores overall.

Finally, ET tended to generate moderately high workloads, with the highest maneuver performance. Operators found it time-consuming and mentally taxing to execute repeated snap and slew operations with long slew times, which likely drove high workloads. However, the increased head up time with ET drove a quicker cross-check, improving maneuver performance. Additionally, operators were able to execute cursor snaps with minimal head movement; this allowed them to leave the VHUD in their peripheral vision while executing the task. The peripheral horizon, airspeed, and altitude tapes likely drove increased maneuver performance as well.

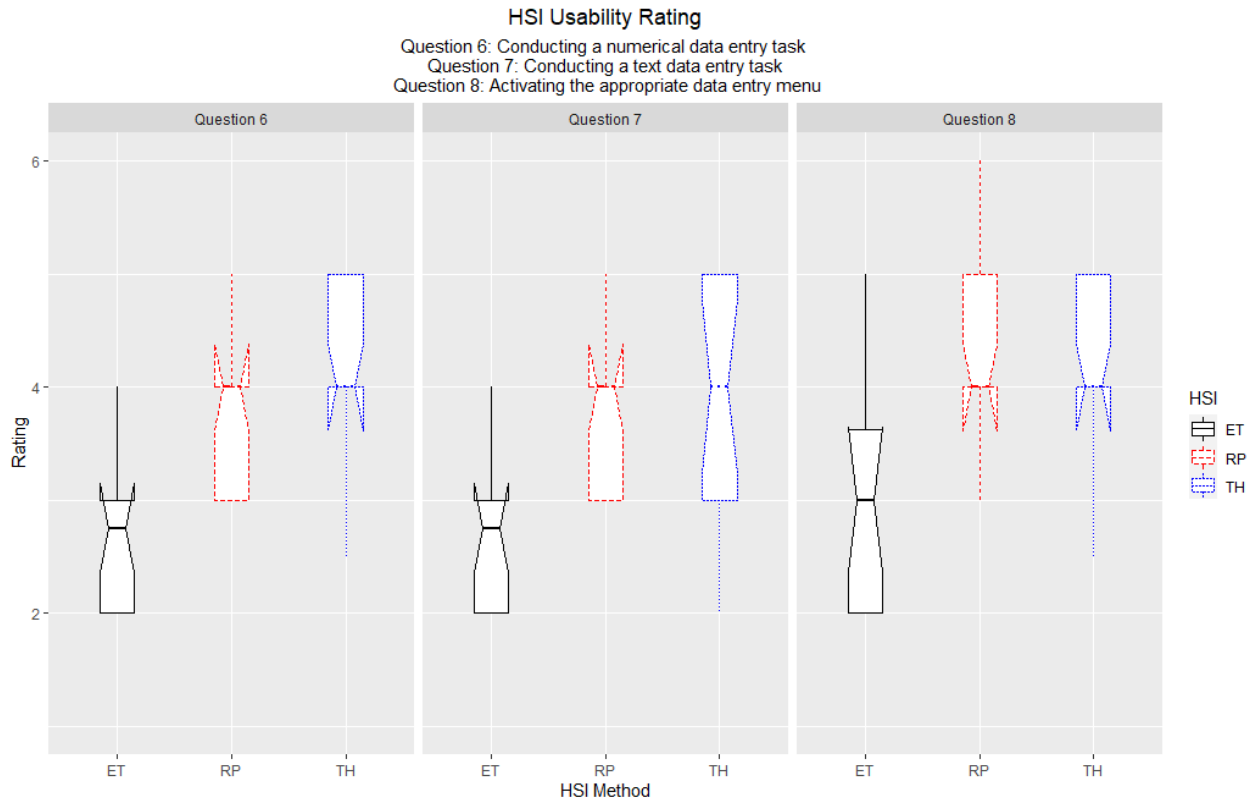


Figure 61. HSI Usability Results: questions 6, 7, and 8

The usability surveys for data entry demonstrated the strengths and weaknesses of each of the HSI methods for Task 3. As seen in Figure 61, ET was consistently rated lower than RP or TH due to the inaccuracy and unpredictability of ET. Additionally, operators ranked RP and TH closely for the numerical data entry and menu selection, but rated TH higher than RP for a text entry task. Operator comments suggest this is due to the QWERTY keyboard requiring large head movements during RP text entry. This led to greater operator discomfort, particularly under elevated G. Comments were also levied against RP during text entry due to the extended period of time the operator was required to be heads down looking at the display and not monitoring flight parameters. Conversely, it was observed from question 8 that RP was suited for single button operations (radio channel up/down or on/off type switches) in which head movements were quick and minimal, since RP did allow for continued hands-on flight controls.

TH was consistently rated highest among the three HSI types for usability and comfort during the data entry task. Operators did comment on the lack of bracing for their hand which led to a lack of precision when making an entry. This effect was exacerbated under elevated G. This led to errors with the operator primarily hitting the same key twice inadvertently. **Provide a method for operators to stabilize their hands during data entry when using touchscreen (R7).** See APPENDIX J – FUTURE WORK for further discussion of touchscreen bracing options.

Another recurring comment applied to all HSI methods. Operators objected that the ET cameras limited the operator's field of view out of the bottom of the visor. This blocked the data entry card on the operator's leg and the bottom portion of the display, and affected all HSI methods uniformly since the ET cameras were fixed on the helmet. The operator had to tilt their whole head down to view these areas which was very uncomfortable under G and disorienting in turns. **Place the ET cameras out of the normal field of view of the operator (R6).**

2.2.4 Operationally Representative Task Comparison

The final specific test objective was to compare eye-tracking, rhino-pointing, and conventional tactile input during an operationally representative task. Operator performance was characterized by three metrics: (1) overall HSI usability, (2) operator workload, and (3) military utility. This test objective was met.

The operational environment reinforced the laboratory results obtained in Tasks 1-3, magnifying both the strengths and weakness of each HSI method. All operators found TH to be functional, accurate, and familiar but in need of augmentation from one of the other HSI methods. RP was found to be fast and accurate but the large head movements required to reach the extremes of the display were deemed uncomfortable. Also, the visible feedback of the rhino-boresight was mentally distracting, particularly as the display became more cluttered in the operational task. ET was popular among the operators for whom the system and calibration worked well. They complimented ET as intuitive, comfortable, and less mentally taxing. Even when properly functioning, ET's accuracy was not as good as the other HSI methods but it was found that quickly moving the cursor to near the desired location was all that was needed. However, ET did have issues working consistently across different operators and needed frequent recalibration.

2.2.4.1 Operationally Representative Task Methods and Conditions

All operationally representative task data was collected using a common task, known as Task 4. Task 4 was flown in conjunction with the operational intercept. When the TC transmitted the task, the simulated air-to-air portals display appeared, empty of track symbols as depicted in Figure 62.

Figure 62: Operational Task at Initiation

Figure 63: Operational Task at Completion

As the operator maintained level flight between 150 and 250 knots, tracks began to appear left and right of the initial aircraft heading, as depicted in Figure 63. After the appearance of the first track, the operator maneuvered on a collision antenna train angle intercept and interacted with the MFD as he would in an air-to-air engagement. Every ten seconds, the TC requested the operator to read back a specific value from the VHUD. This interruption increased operator workload and simulated the regular external distractions that would normally occur in an operational environment. After two minutes, the Task 4 simulation ended and the operator provided real-time feedback and commentary. The TC repeated the process as desired by the operator. Each operator completed at least one instance of Task 4 per HSI method. A more detailed description of Task 4 is provided in APPENDIX B – DETAILED TASK DESCRIPTION.

Operator workload was defined by the changes in TPV during the task, and was averaged over the entire task execution period. No reference operational task was flown without the use of an HSI method, thus there was no reference performance to compare follow on operational tasks with. As a result, there was no NAP available and TPV data was used in conjunction with operator comments as a subjective measure of performance in order to characterize operator workload fully during Task 4. Overall HSI usability and military utility were defined by subjective ratings and comments given after each flight in a HSI usability survey, as in previous tasks. However HSI usability and military utility comments were also collected by the TC in real-time following any completion of Task 4.

2.2.4.2 Operationally Representative Task Results

2.2.4.2.1 HSI Usability

A sampling of notable operator comments for each HSI method is given below. A full listing of operator comments is available in APPENDIX F – HSI USABILITY SURVEY RESULTS.

Table 3: HSI Usability

ET Usability Comments
“The ability to just look and throw my cursor is great.”
“Felt like the future when it works.”
“The mental burden of using eye tracking during the operational task was very low.”
“On its own, it gets "close enough" usually.”
“When it worked, it was extremely usable … However, for the majority of the sortie, ET was not functional and the accuracy was dismal. It was completely unusable and made the tasks harder to complete than if I were to just slew the cursor with the cursor slew switch.”
RP Usability Comments
“Moving the cursor back and forth on the screen required substantial head motion.”
“The pointer was highly accurate and gross movements were quick.”
“RP allowed me to get the cursor back and forth between portals rapidly.”
“Holding the head still for precise designations was tiring, and required breaks every 30 seconds or so.”
“This system again worked very well for gross slewing across portals.”
“Gross movement is good. It gets you close. But you have to move your head more than I thought you would. Fine movement is difficult especially under g.”
TH Usability Comments
“This is the existing method and it works for what it is supposed to do but it could be improved with RP or ET.”
“Tactile input is a proven method for this and works great.”
“Unfortunately, it's extremely slow to get from one portal to another.”
“It was predictable and accurate.”
“Gross movements were pretty slow.”

All operators agreed that TH was functional, accurate, and familiar but interacting with a large area display needed improvement. Operators felt that ET was intuitive, comfortable, and less mentally taxing. ET exhibited poor accuracy; however, this often did not detract from overall performance, as just snapping the cursor across the large display quickly and finishing with TH was a favored strategy. Several comments were made about having to recalibrate ET, or that at times ET was so inaccurate that it was frustrating to use, particularly near the edges of the display. (See APPENDIX J – FUTURE WORK for detailed proposals on compensating for display edge effects with ET.) RP was found to be very accurate, fast, and good for large movements across the display. Operators often complained of discomfort making such deliberate head movements especially under elevated G loads. A common comment was that RP was accurate enough to direct-designate but it wasn’t required; just moving the large distance across the display was how it derived most of its value.

2.2.4.3 Operator Workload

For Task 4, normalized TPV distributions were compared to operator comments in order to draw conclusions. The TPV workload plots for all operators can be found in APPENDIX E – ADDITIONAL PLOTS.

Based on a statistical analysis of normalized TPV, for most operators TH generated the highest TPV during the execution of Task 4. RP most commonly generated the lowest TPV of the three HSI methods for Task 4, as shown in Figure 64. These results corroborate the results of section 2.2.3.2.2. However, they remain merely suggestive. The similarity in distributions between all three HSI methods in combination with the qualitative measurement of maneuver performance rendered overall workload results inconclusive for Task 4.

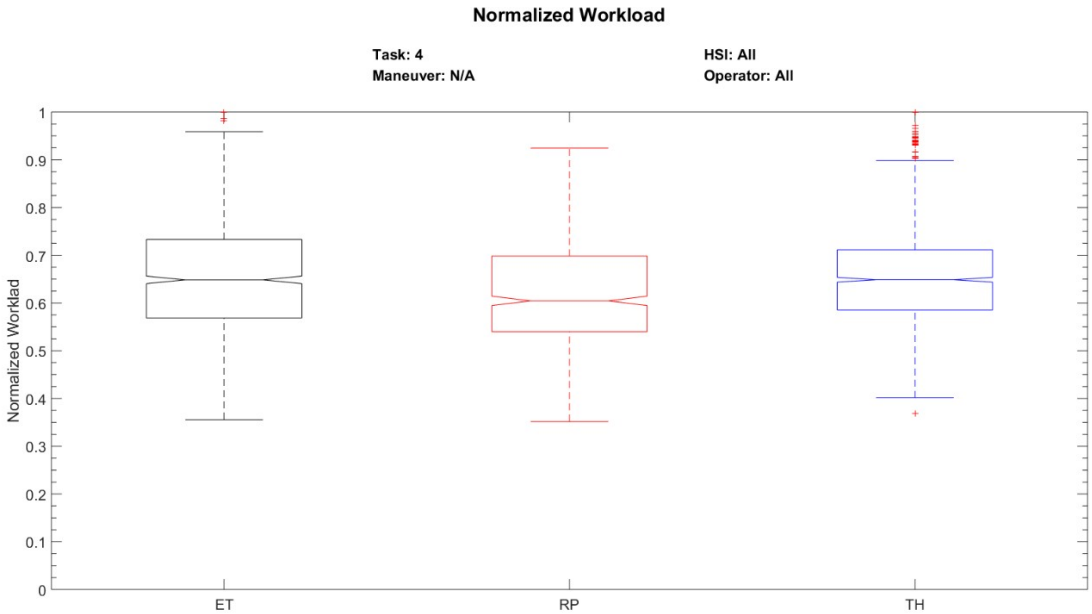


Figure 64. Task 4 Normalized Workload

2.2.4.4 Military Utility

A sampling of notable operator comments for each HSI method is given below. A full listing of operator comments is available in APPENDIX F – HSI USABILITY SURVEY RESULTS.

Overall, all operators saw the military utility in TH and acknowledged that ET and RP could not currently replace TH as the primary method of interacting with large format display because of the precise control of TH. However, the operators also stated the limitations of TH, and that it could be supplemented by ET and/or RP to increase the military utility of the HSI overall. For RP it was found that the visual feedback of the rhino-boresight that helped to make it so accurate became a distraction and mentally taxing when the display became more cluttered during the operational task.

During the test program, the team also executed some investigative trials both on-ground and in-flight with modifications to the specific implementations of RP and ET to try to address the negative comments operators gave both HSI methods. For RP, the rhino-boresight symbol was made invisible. For both RP and ET a software technique referred to as clamping was added; with clamping active, if a snap was commanded when the system thought the RP or ET cursor was off screen, the cursor would be placed at the edge of the display nearest to where the system thought the operator was pointing. See APPENDIX I – FUTURE WORK for significant detail on the motivation for these modifications, and proposals for future modifications responsive to the concerns of this report.

Table 4: Task 4 Military Utility

ET Military Utility Comments
"If technology allows eye tracking systems to be more robust and reliable, the accuracy it's capable of achieving when looking at a large area display would be extremely useful for making large cursor movement."
"When it is working feels like its magic or from the future. Had some hiccoughs and needed to be recalibrated frequently but has potential."
"Not having to move my head, it almost seemed like mind control. I looked where I wanted the cursor to go, commanded it to slave, and it went there."
"If technology allows eye tracking systems to be more robust and reliable, the accuracy it's capable of achieving when looking at a large area display would be extremely useful for making large cursor movement."
"This will easily speed up the kill chain once the tech matures a bit more."
RP Military Utility Comments
"RP has great potential. As an idea it works. It still has a few inherent issues involved with head movement under g and maneuvering."
"After snapping the cursor close to the target, I used the HOTAS for fine tracking, then quickly jumped back to the large missile selection buttons to complete the task."
"Direct designate was possible for stand-alone targets, but not for targets in tight groups."
"Moving the cursor back and forth on the screen required substantial head motion. Gross movements were quick and it was easy to designate."
"RP degraded performance in the operational environment. The tendency to try to direct designate (line #3) was a factor. However, it also became mentally taxing to track the two cursors."
"Tracking two cursors elevated mental workload significantly, and detracted from task performance."
TH Military Utility Comments
"Predictable, slow slew makes moving between portals undesirable."
"Designating tracks on the right-hand portal and then on the main portal took long enough that I would become slightly task overloaded by the number of tracks that had appeared in the interim."
"I think the combination of tactile HSI with another HSI such as the RP or ET would complement each other well. You could get the gross movement benefits of the ET/RP and the fine abilities of the tactile."
"It was predictable and accurate. However, slewing back and forth across the screen to designate and select a missile was objectionable. It was difficult to keep up with the targets that were popping up."

Operators thought clamping was extremely useful for ET and RP, and nearly eliminated issues observed with edge accuracy and lack of response to the snap command. With the Rhino boresight removed, operators stated that RP "felt almost like ET" and they felt that they were completing the tasks more quickly than with the cursor showing. However, RP was still reported as uncomfortable for large movements at the extremes of the display and ET still had the same accuracy issues.

For ET and RP, these investigative trials strongly affirmed the conclusions from pilot comments: that RP and ET both had clear military utility, but additional developments were required to take advantage of the strengths of each, while minimizing the weaknesses of each. To optimize operator performance with RP, **conduct further research with rhino-pointing in order to reduce mental and physical operator fatigue (R5)**. To optimize operator performance with ET, **conduct further research with eye-tracking in order to increase its accuracy, predictability, and reliability (R2).**

This page was intentionally left blank.

3 RECOMMENDATIONS

The Have Rhino team makes the following recommendations, listed here in priority order:

R1. Augment traditional human-system interfaces with a mechanism to drive cursors rapidly across large-format displays. (Page 53)

R2: Conduct further research with eye-tracking in order to increase its accuracy, predictability, and reliability. (Page 69)

R3: Develop a method to compensate for the effect of load factor on eye-tracking accuracy. (Page 61)

R4: Develop a method to reduce the tendency for operators to fixate on making fine adjustments with the rhino-boresight. (Page 51)

R5: Conduct further research with rhino-pointing in order to reduce mental and physical operator fatigue. (Page 69)

R6: Place the ET cameras out of the normal field of view of the operator. (Page 64)

R7: Provide a method for operators to stabilize their hands during data entry when using touchscreen. (Page 64)

This page was intentionally left blank.

4 REFERENCES

1. Schnell, T., Training Effectiveness Study. 2018, University of Iowa College of Engineering: Iowa City, Iowa.

2. Schnell, T., OPL FLIGHT OPERATIONS MANUAL. 2005, University of Iowa College of Engineering: Iowa City, Iowa. NAF T.O. 1T-L29-1 re-issued as OPL INST L29-1 07 JUL 2014.

3. OPL, Operators Flight Manual AERO VODOCHODY L-29 DELFIN (N429GC) U.o.I. State of Iowa, Editor. 2013: Iowa City. OPL-INST-L29-2.

4. OPL, Effects of Helmet Mounted Display Format and Spatial Audio Cueing on Pilot Performance and Spatial Disorientation Prevention, in support of Human Interface Research and Technology. University of Iowa College of Engineering: Iowa City, Iowa.

5. Reichlen, Christopher Patrick. "A comparison of linear and nonlinear ECG-based methods to assess pilot workload in a live-flight tactical setting." MS (Master of Science) thesis, University of Iowa, 2018. https://doi.org/10.17077/etd.hqx7edns

6. Geiselman, Eric. 'RE: HAVE RHINO - Eye tracking and RP TRL'. Email, April 06, 2020.

7. Schnell, T., Literature on Eye Gaze HMI Control. 2019, University of Iowa College of Engineering: Iowa City, Iowa.

8. D. Harp, R. Armstrong, H. Whiting, Z. Purser, S. Dold, S. Nevala, "*COMPARISON OF EYE-TRACKING, RHINO-POINTING, AND TRADITIONAL HUMAN SYSTEM INTERFACE, Test Plan*," 2020.

9. Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis. London: Chapman & Hall/CRC. p. 45. ISBN 978-0-412-24620-3.

This page was intentionally left blank.

APPENDIX A – DETAILED TEST ITEM DESCRIPTION

Further details and specifications on the following test hardware are provided in this appendix:

1. Eye-tracking system
2. Head-tracking system
3. Operationally Flight Representative Helmet (OFRH)
4. L-29 Aircraft
5. Technology Readiness Level (TRL)

Digital versions of user's manuals, spec sheets, and promotional brochures can be found in the respective file in the digital data appendix.

1. Eye-tracking system:

The Dikablis Professional ET eye camera recorded a black/white video of the eye with a resolution of 640 x 480 pixels and up to 60 frames per second. The scene camera recorded a high-resolution, color video of all the surroundings that were visible in the test subject's field of vision. The resolution of the scene camera was 1920 x 1080 pixels at 30 frames per second. The eye-tracking unit (separated from the helmet) is shown in Figure 65. ET hardware separated from OFRH.

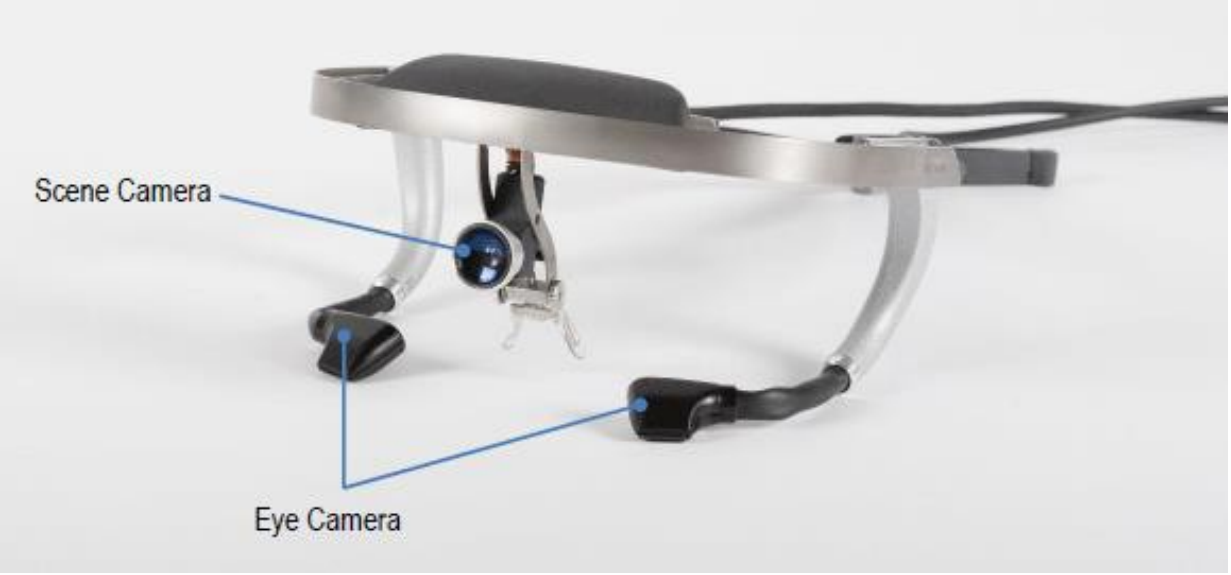


Figure 65. ET hardware separated from OFRH

As mentioned in section 1, the left camera IR illuminator was found to be defective a few days prior to the flight test. As a consequence, only the right eye camera was used for testing. It is possible that single eye-tracking degraded system performance, although this effect cannot be quantified without further research.

More importantly, this particular commercial-off-the-shelf eye tracker was chosen because it was available at OPL and therefore did not have to be purchased. The use of the Dikablis eye tracker was not the result of a hardware down-selection. In fact, this eye tracker, while relatively accurate and reasonably

functional in daylight and dark conditions, had a major drawback in that it did not use the glint from the IR illuminators to correct for headgear slippage. Some eye trackers use the glint spot to determine if the head gear slipped and automatically correct the tracking vector. The principle of glint correction is based on the fact that the eyeball is a sphere and therefore, the glint represents a point that will only move relative to the eye camera video coordinate system, if the headgear has slipped. Through a vector difference between the pupil and the glint spot, the error in the gaze vector due to slippage can thus be automatically corrected. As stated before, the Dikablis eye tracker did not have this correction mechanism, which would be an important requirement for a fielded system.

Eye-tracking calibration consisted of the following steps:

1. Operator looked at upper left corner of the MFD while the technician calibrated the pupil angle.

2. Step 1 was repeated on upper right corner, lower right corner, and lower left corner.

3. Operator looked at the center of the screen with the focus-boresight made visible to verify calibration accuracy.

4. Focus-boresight was made invisible and operator made several snaps around the screen to verify eye-tracking accuracy.

2. Head-tracking system:

The helmet worn during this test used a commercial off-the-shelf Magnetic Polhemus Scout head tracker, which was expected to perform at an accuracy of 3 arc minutes (0.05 degrees) and was considered comparable to trackers on modern fifth-generation fighter aircraft.

3. OFRH:

The OFRH used HMD optics with green high-brightness (10,000 cd/m^2) Organic LED imagers at a field of view (FOV) of 30 x 40 degrees at a resolution of 1280 x 1024. Therefore, the imagers required no backlight and they had effectively infinite contrast capability. The optics were focused at infinity.

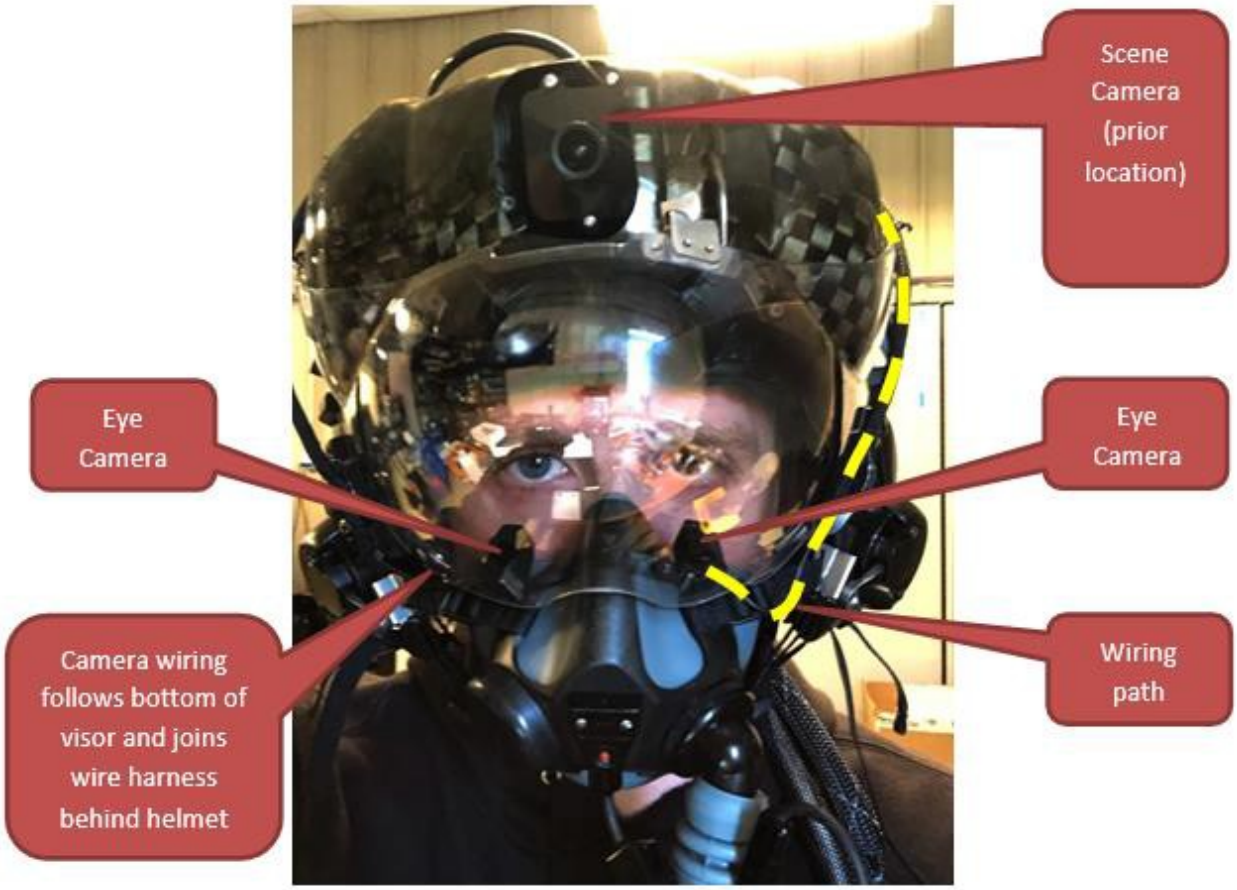


Figure 66. ET implementation in OFRH

Eye cameras were encased in machined Delrin cases, then further encased in polyolefin heat shrink. Each camera was adhesively attached to the inside of the helmet visor so as to be easily and quickly moved and adjusted for eye tracking calibration. From each camera, a thin wire ran along the bottom of the visor and then up and back along the curvature of the helmet (figure 1) before meeting up with the wires from the other eye camera and scene camera. These wires were joined into a cable harness (figure 2) and followed the existing helmet cabling where they terminate in a USB and HDMI output to be plugged into the onboard Dikablis eye tracker box.

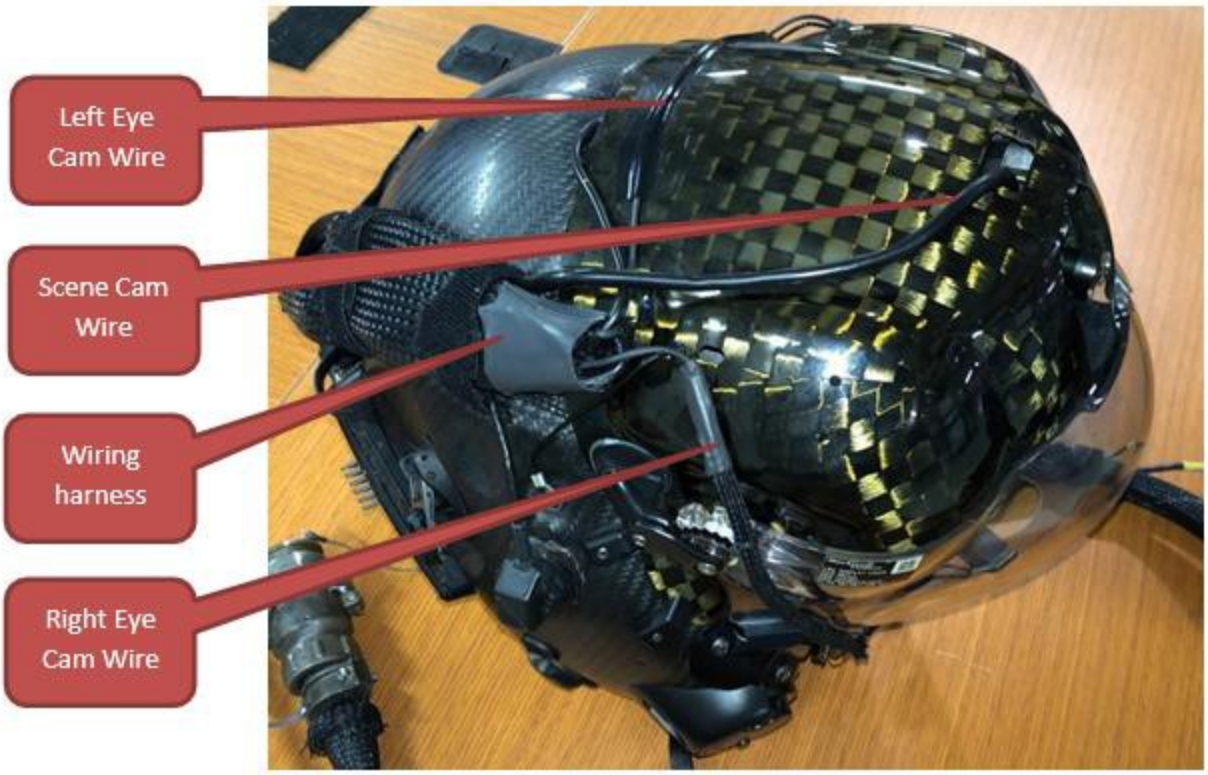


Figure 67. Upper OFRH wiring

4. LCD Screen:

IDK-2121W series is a Color Active Matrix Liquid Crystal Display composed of a TFTLCD panel, a driver circuit, and backlight system. The screen format is intended to support the FHD (1920(H) x 1080(V)) screen and 16.7M colors (RGB 8-bits data). All input signals are dual LVDS interface. Driver board for the backlight is included. The IDK-2121W touchscreen delivered the fingertip contact location with 1,920 x 1,080 resolution.

The following are characteristics summary under 25°C condition:

Table 1.1: Display Characteristics

Item	Unit	Description
Screen Diagonal	[mm]	546.86(21.53")
Active Area	[mm]	476.64 (H) x 268.11 (V)
Pixels H x V		1920 (x3) x 1080
Pixel Pitch		248.25 (per one triad) × 248.25
Pixel Arrangement	[um]	R.G.B. Vertical Stripe
Display Mode		VA Mode, Normally Black
White Luminance (Center)	[cd/m^2]	1200 (Typ.)
Contrast Ratio		3500:1 (Typ.)
Optical Response Time	[msec]	16 ms (Typ., on/off)
Input Voltage VDD	[Volt]	+5.0 V
Backlight Input Voltage	[Volt]	+12.0 V
Power Consumption (VDD line + LED line)	[Watt]	VDD: 4.4 + PLED: 43.56 = 48 (Typ.) (with LED driver board, all white pattern)
Weight	[Grams]	R series: 3200 (Typ.) N series: 2400 (Typ.)
Physical Size	[mm]	R series: 495.6(W) × 292.2(H) × 21.65(D) Typ. N series: 495.6(W) × 292.2(H) × 17.7(D) Typ.
Electrical Interface		Dual channel LVDS
Support Color		16.7M colors (RGB 8 bits)
Surface Treatment		Anti-Glare, 3H
Temperature Range		N-series/R series
Operating	[°C]	0 to 55
Storage (Shipping)	[°C]	-20 to 60
RoHS Compliance		RoHS Compliance

L-29 Test Aircraft:

The OPL used an L-29 flight test aircraft, tail number N429GC, as seen in Figure 68 and Figure 69 for all tests. The following were basic specifications of the aircraft:

- Aero Vodochody L-29, US Registration N429GC
- Crew: 1 safety pilot (front cockpit), 1 test operator (rear cockpit operator station)
- Length: 10.81 m (35 ft 6 in)
- Wingspan: 10.29 m (33 ft 9 in)
- Height: 3.13 m (10 ft 3 in)
- Wing area: 19.8 m² (213 ft²)
- Empty weight: 2,280 kg (5,030 lb) Loaded weight: 3,286 kg (7,244 lb)
- Max takeoff weight: 3,540 kg (7,800 lb)
- Powerplant: 1× Motorlet M-701C , 8.7 kN (1,960 lbf)
- Maximum speed: 820 km/hr (443 knots, 510 mph)
- Range: 900 km (486 nm, 560 mi)
- Service ceiling 11,500 m (37,700 ft)
- Rate of climb: 14 m/s (2,800 ft/min)

- Wing loading: 166 kg/m² (34.3 lb/ft²)
- Thrust/weight: 0.25
- G-loading Limits: G +7.5/-3.5 with external fuel tanks, +8/-4 without external fuel tanks
- USAF G-suit compatible
- Cold ejection seats

The test only required the aircraft to be subjected to positive-g loads, not exceeding 4 G. Two wing-mounted external fuel tanks were loaded on the aircraft to increase sortie duration. These were considered standard equipment by the flight manual.

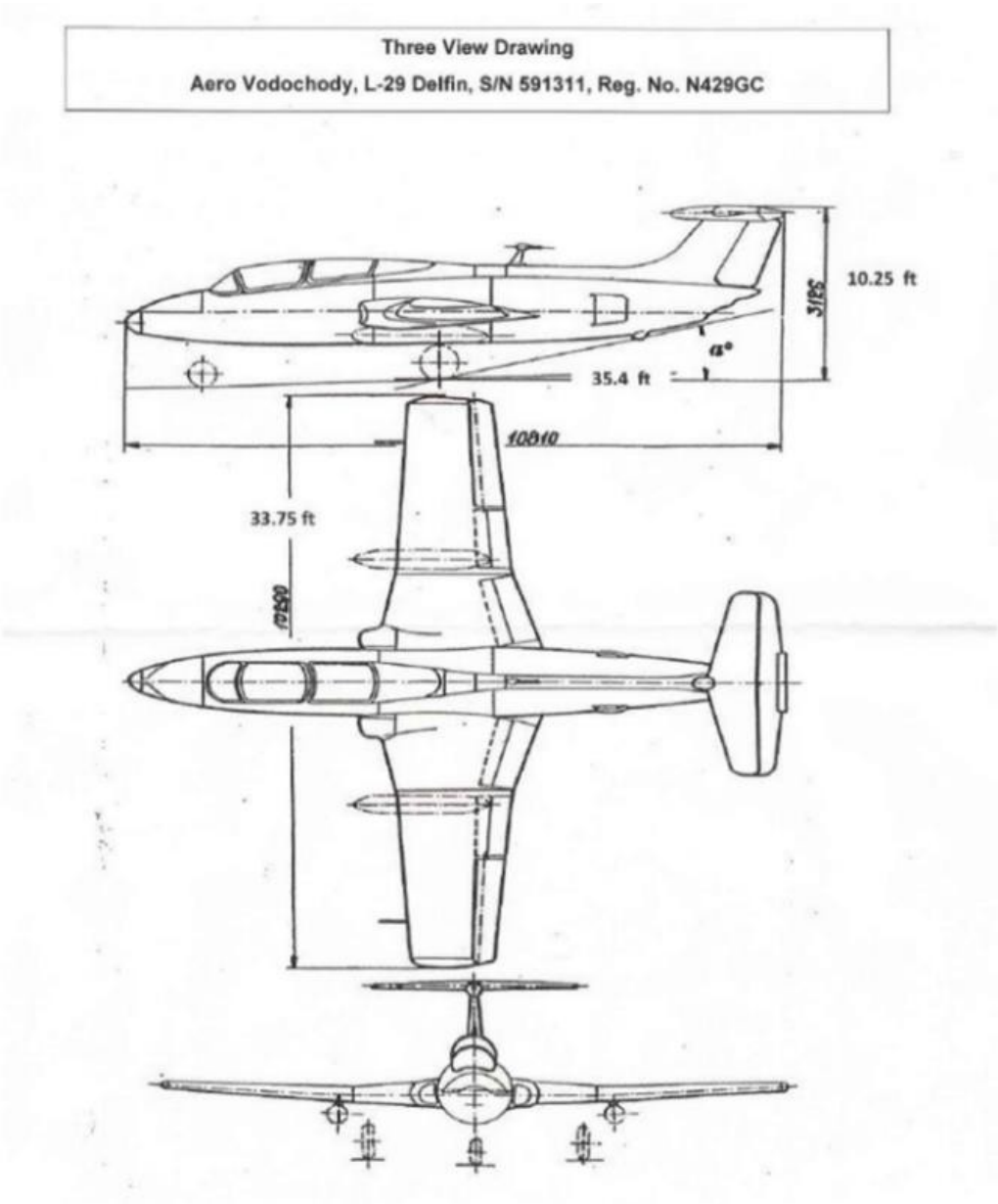


Figure 68: N429GC Three View Diagram

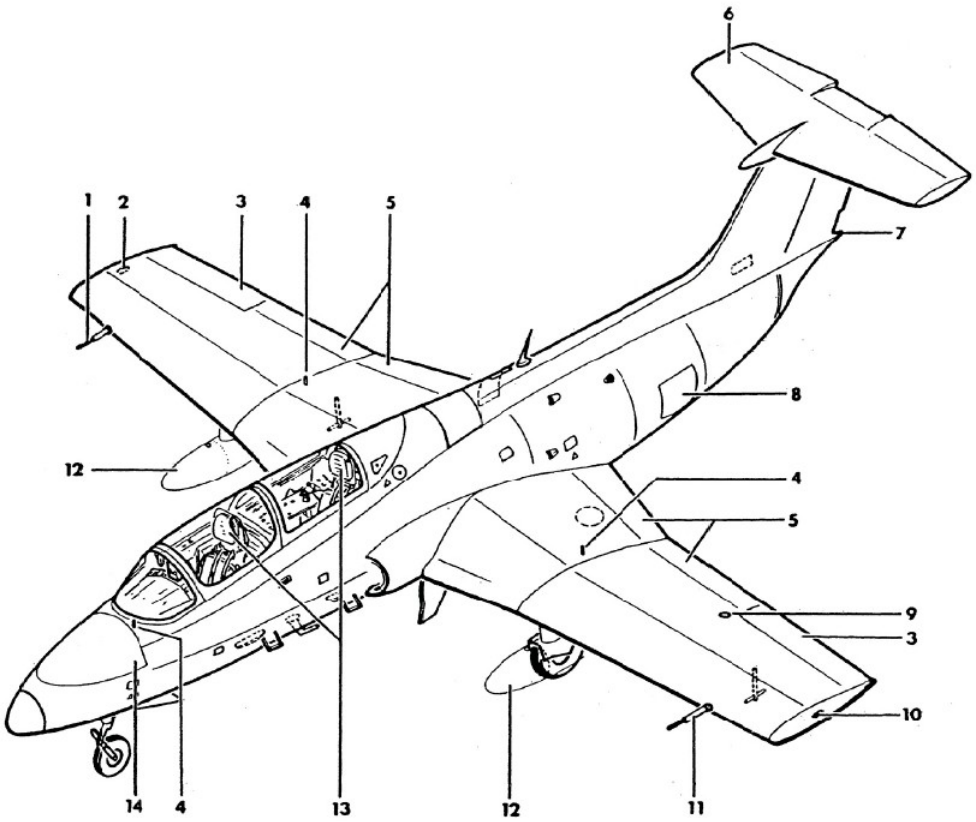


Figure 69: N429GC, General Arrangement of Aircraft Components

No structural or aerodynamic modifications were made to the OPL L-29 aircraft. The OPL purchased the aircraft in 2008 and had been used for over 600 incident-free flight hours at the time of this test.

The test was conducted using the aircraft with its existing research instrumentation. The only modifications specific to this test were the hardware and software changes to the RCP display system, to accommodate RP. The FAA approved all other existing instrumentation, which was documented in the aircraft logs and L-29 Operators Flight Manual (reference 2) issued under OPL Instruction OPL-INST 2 (reference 3). The installation of the experimental computers and displays used the factory instrument panel, factory avionics box trays, and factory hard points with an equivalent or better structural safety factor because the modern computers that were installed in lieu of the obsolete factory avionics boxes were lighter. All equipment was owned by the OPL and no GFE was onboard for this flight test. The RCP is shown in Figure 15.

The safety pilot (SP) operated from the front cockpit (FCP). The FCP did not contain any modified hardware relating to this test, with the exception of a small experimental touch-screen display on the right console for in-flight ET calibrations. The FCP is shown in Figure 14.

In the RCP, the aircraft was instrumented with research displays, data collection equipment, and physiological measurement equipment as shown in Figure 15. Two experimental displays were installed in the rear cockpit:

- A 54.6 cm (21.5-inch) diagonal daylight-readable touchscreen in the head-down position (described previously in 1.3.4).
- A 38.1 cm (15-inch) diagonal daylight-readable touchscreen display was also available in the head-up position showing a virtual view of the outside visuals with overlaid HUD symbology. The image on this display was repeated in the VHUD of the HMD, so this display was covered during each flight.

Technology Readiness Level (TRL):

Technology readiness levels (TRLs) were a means of estimating the maturity of technologies during the acquisition phase of a program, developed at NASA in the 1970s. The use of TRLs as shown in Figure 70 enabled consistent, uniform discussions of technical maturity across different types of technology. The higher the TRL value, the more capable the technology was and ready for operational use. A lower TRL value implied that technology was still emerging and only usable in a laboratory setting.

The nine-point NASA scale was:

TRL 9 – Actual system "flight proven" through successful mission operations
TRL 8 – Actual system completed and "flight qualified" through test and demonstration (ground or space)
TRL 7 – System prototype demonstration in a space environment
TRL 6 – System/subsystem model or prototype demonstration in a relevant environment (ground or space)
TRL 5 – Component and/or breadboard validation in relevant environment
TRL 4 – Component and/or breadboard validation in laboratory environment
TRL 3 – Analytical and experimental critical function and/or characteristic proof-of concept
TRL 2 – Technology concept and/or application formulated
TRL 1 – Basic principles observed and reported

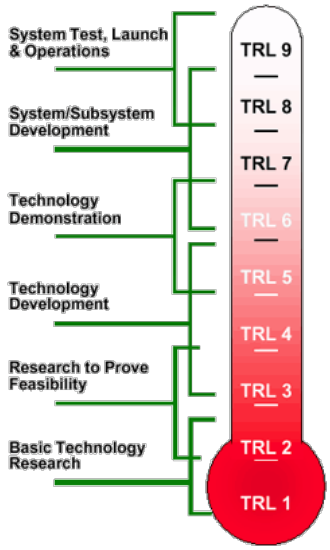


Figure 70. TRL Scale

APPENDIX B – DETAILED TASK DESCRIPTION

A more detailed version of the steps required for each HSI task is given in this appendix.

Task 1:

Step	Event	Result
1	Operator/SP initiated maneuver.	Overall operator workload increased and flight environment became more dynamic.
2	Operator cross-check and interaction allowed to vary as desired to sustain the maneuver.	The operator's gaze and rhino-boresight location varied. However, the conventional cursor remained fixed to the center of the display.
3	After a random time interval, a distinct track symbol appeared on the MFD at a random location, accompanied by a notification from the control room.	This event interrupted the operator's cross-check, requiring additional input. The system recorded track symbol location and notification time.
4	Operator directed attention to MFD and began searching for track symbol.	Operator used designated HSI method to manipulate displayed cursor(s).
5	Operator slewed the cursor over the track symbol and pressed the designate button. For rhino-pointing and eye tracking, this event required an intermediate cursor snap operation.	The system recorded the times at cursor snap (if applicable) and designation. The system also recorded the cursor location at cursor snap (if applicable) and designation.
6	After 7 seconds, if no designation had occurred the track began to flash. After 10 seconds, if no designation had occurred, the track disappeared.	If the operator could not complete the task, the task ended after 10 seconds. If successful designation occurred, the task ended.
7	Operator/SP continued maneuver until next test point.	The operator's cross check again varied as required to continue the maneuver.

Only one cursor snap operation in step (5) was allowed. No *a priori* indication was given to the operator that a given designation operation in step (5) would be successful. However, an unsuccessful designation resulted in no graphical feedback while a successful designation results in positive graphical feedback and task termination. Following any unsuccessful designation in step (7), the operator could re-attempt designation as many times as desired until task timeout.

Task 2:

Step	Event	Result
1	Operator/SP initiated maneuver.	Overall operator workload increased and flight environment became more dynamic.
2	Operator cross-check and interaction allowed to vary as desired to sustain the maneuver.	The operator's gaze and rhino-boresight location varied. However, the conventional cursor remained fixed to the center of the display.
3	After a random time interval, a randomized track cluster with one distinct track appeared on a portal at a random location accompanied by a notification from the control room.	This event interrupted the operator's cross-check, requiring additional input. The system recorded the distinct track symbol location and notification time.

4	Operator directed attention to MFD and began searching for track symbol.	Operator used the designated HSI method to manipulate displayed cursor(s).
5	Operator slewed the cursor near the track symbol and pressed the expand button.	A predefined region of the display around the expand coordinate appeared at a higher zoom level on the display. System recorded the time elapsed since notification, and the location of the expand operation.
6	Operator used the current HSI method to designate a track on the expanded region. For rhino-pointing and eye-tracking, this operation could have required an intermediate cursor snap operation.	If a snap operation occurs, the system recorded the location of the snap and time since notification. If designation occurs, the system recorded the location of the designation and time since notification.
6	After 7 seconds, if no designation had occurred the distinct track began to flash. After 10 seconds, if no designation had occurred, the track disappeared.	If the operator could not complete the task, the task ended after 10 seconds. If successful designation occurred, the task ended.
7	Operator/SP continued maneuver until next test point.	The operator's cross check again varied as required to continue the maneuver.

Only one expand operation was allowed in step (5). Following the expand operation, the display-fixed cursor appeared on the expanded region of the MFD at the same relative orientation to the track cluster as the orientation presented at the expand operation. All non-unique tracks in the cluster were the same symbol and a designation operation on the non-unique tracks was not possible. Overall, operator strived to successfully complete the entire task by designating the unique track in the minimum time; however this guidance imposed no requirements on the proximity of the expand operation to the distinct track.

Task 3:

Step	Event	Result
1	Operator/SP initiated maneuver.	Overall operator workload increased and flight environment became more dynamic.
2	Operator cross-check and interaction allowed to vary as desired to sustain the maneuver.	The operator's gaze and rhino-boresight location varied. However, the conventional cursor remained fixed to the center of the display.
3	After a random time interval, the ground station prompted the operator to input a random line of data entry from the predetermined data card.	This event interrupted the operator's cross-check, requiring additional input. The system or test conductor recorded the predetermined data and notification time.
4	Operator activated the appropriate data entry menu.	Operator used designated HSI method to complete task. System recorded the time at the first interaction with the data entry menu.
5	Operator inputted the data specified, reading that data off a card placed on the operator’s leg.	System recorded the time elapsed between notification, individual input coordinates, and data input completion.
6	Operator/SP continued maneuver until next test point.	The operator's cross check again varied as required to continue the maneuver.

The data input was a combination of branching menu selections and alphanumerical keypad inputs. Each keypad key (whether menu or alphanumerical) was 1” by 1” on the MFD. Figure 18 gives a notional example of the submenu layout and the keypad interface.

Each instance of Task 3 required ten actuations to navigate the menu branches and provide alphanumerical inputs. If an incorrect menu branch was accessed, the operator used a “Back” button to return to the correct menu and it was noted by the test team. If an incorrect alphanumerical input was made, the test team recorded the error but allowed the operator to continue data entry with the next required character, without error notification. A list of all possible data entry lines is given below:

1. **TAC – LAT/LONG - N 3 4 W 1 1 8**
2. **TAC – LAT/LONG - N 4 1 W 1 1 9**
3. **TAC – LAT/LONG - N 3 2 W 0 9 4**
4. **TAC – LAT/LONG - N 4 1 W 1 2 3**
5. **TAC – LAT/LONG - N 4 4 W 1 1 8**
6. **NAV – NAVAID – T A C 1 1 1 X**
7. **NAV – NAVAID – T A C 0 9 4 X**
8. **NAV – NAVAID – 1 5 7 . 3 2 5**
9. **NAV – NAVAID – 1 2 8 . 5 2 0**
10. **AV – ICAO – C L U T C H 2**
11. **NAV – ICAO – K N O T S 0 8**
12. **NAV – ICAO – H A V O C I V**
13. **NAV – ICAO – G O L D E N 1**
14. **NAV – ICAO – F U Z Z Y 6 9**
15. **NAV – ICAO – B M W 3 3 0 I**
16. **COMM – UHF – 2 4 6 . 1 2 5**
17. **COMM – UHF – 3 6 9 . 3 2 0**
18. **COMM – UHF – 2 0 5 . 5 7 3**
19. **COMM – UHF – 2 5 1 . 8 0 2**
20. **COMM – UHF – 3 7 2 . 2 5 0**
21. **COMM – VHF – 1 4 6 . 1 2 5**
22. **COMM – VHF – 1 2 8 . 5 6 0**
23. **COMM – VHF – 1 4 5 . 6 2 0**
24. **COMM – VHF – 1 3 2 . 9 2 5**
25. **COMM – VHF – 1 0 9 . 4 3 0**

Task 4:

Step	Event	Result
1	Operator held straight and level flight at assigned altitude and airspeed.	Simulated loiter in a combat environment.
2	Operator cross-check and interaction allowed to vary as desired to sustain the maneuver.	The operator's gaze and rhino-boresight location varied. However,

		the conventional cursor remained fixed to the center of the display.
3	After a random time interval, a number of distinct track symbols appeared on the MFD at pseudo-random locations. Each track was offset no more than 5 nautical miles laterally from a bearing line drawn off the nose of the test aircraft.	Operational scenario was posed; additional operator input was required. System recorded notification time.
4	Operator directed attention to MFD to react accordingly.	Operator used designated HSI method to complete task.
5	Operator slewed cursors in such a way to designate all tasks in minimum time.	Operator received graphical feedback that the tracks had been designated.
6	Operator pointed aircraft at the closest designated track and continued to complete a collision antenna train angle (CATA) intercept to fly directly to the track.	Operator workload increased.
7	Operator selected a missile based on the range to the closest target on the display from a menu situated in another region of the display.	System recorded the time elapsed between notification and correct missile selection.
8	Tracks continued to appear on the display between 10 and 50 NM slant range on a time-random basis. At each appearance, the operator designated and the system reacted as in steps 5-7. Operator updated the CATA intercept and re-selected a missile only if the new target is closer.	Operator continued task in an operationally relevant manner.
9	Tracks on a separate portal also appeared on a time-random basis and needed to be designated as in steps 5-7.	The additional tracks populated in order to force additional iterations of HSI use as well as force a cross-check of the various portals on the MFD in order to evaluate the HSI modes in an operationally representative environment.
10	During the event, the TC made a call every 10 seconds requesting a random VHUD parameter to add an additional task to the event.	Additional tasks resulted in additional workload to the aircrew and provided greater opportunities to evaluate HSI method as the operator was not solely focused on the portal of interest.

The MFD displayed several “portals” that presented different data. The final portal geometry was reported in the results; however, notionally, they consisted of: (1) representative 180 degree radar scope which will be populated with notional targets at various ranges and azimuths from the ownship radar (depicted as the blue fighter aircraft shape); (2) a missile select menu for selecting notional missiles which correlate to target ranges; and (3) an additional notional 360 degree defensive sensor portal for additional target cueing.

APPENDIX C – FLIGHT LOG

The following table documents the flight each test point was flown on.

Table C-1 - Test Points Flown

Test Point	# of Runs	Parameter	Maneuver	Flight #
1	65	Single Target Track Designation (T)	Baseline	1-14
2	65	Single Target Track Designation (RP)	Baseline	1-14
3	61	Single Target Track Designation (ET)	Baseline	1-14
4	40	Single Target Track Designation (T)	Constant G - 2G	1-14
5	31	Single Target Track Designation (T)	Constant G - 3.5 G	1-14
6	40	Single Target Track Designation (RP)	Constant G - 2G	1-14
7	34	Single Target Track Designation (RP)	Constant G - 3.5 G	1-14
8	37	Single Target Track Designation (ET)	Constant G - 2G	1-14
9	22	Single Target Track Designation (ET)	Constant G - 3.5 G	1-14
10	11	Baseline Aviation Workload	Low Bowtie	1-14
11	11	Baseline Aviation Workload	High Bowtie	1-14
12	42	Single Target Track Designation (T)	Low Bowtie	1-14
13	47	Single Target Track Designation (T)	High Bowtie	1-14
14	51	Single Target Track Designation (RP)	Low Bowtie	1-14
15	57	Single Target Track Designation (RP)	High Bowtie	1-14
16	47	Single Target Track Designation (ET)	Low Bowtie	1-14
17	49	Single Target Track Designation (ET)	High Bowtie	1-14
18	40	Single Target Track in Dense Environment (T)	Baseline	1-14
19	41	Single Target Track in Dense Environment (RP)	Baseline	1-14
20	45	Single Target Track in Dense Environment (ET)	Baseline	1-14
21	35	Single Target Track in Dense Environment (T)	Constant G - 2G	1-14
22	16	Single Target Track in Dense Environment (T)	Constant G - 3.5 G	1-14
23	35	Single Target Track in Dense Environment (RP)	Constant G - 2G	1-14
24	25	Single Target Track in Dense Environment (RP)	Constant G - 3.5 G	1-14

25	55	Single Target Track in Dense Environment (ET)	Constant G - 2G	1-14
26	25	Single Target Track in Dense Environment (ET)	Constant G - 3.5 G	1-14
27	37	Single Target Track in Dense Environment (T)	Low Bowtie	1-14
28	49	Single Target Track in Dense Environment (T)	High Bowtie	1-14
29	33	Single Target Track in Dense Environment (RP)	Low Bowtie	1-14
30	44	Single Target Track in Dense Environment (RP)	High Bowtie	1-14
31	40	Single Target Track in Dense Environment (ET)	Low Bowtie	1-14
32	42	Single Target Track in Dense Environment (ET)	High Bowtie	1-14
33	26	Data Input (T)	Baseline	1-14
34	27	Data Input (RP)	Baseline	1-14
35	24	Data Input (ET)	Baseline	1-14
36	19	Data Input (T)	Constant G - 2G	1-14
37	8	Data Input (T)	Constant G - 3.5 G	1-14
38	12	Data Input (RP)	Constant G - 2G	1-14
39	8	Data Input (RP)	Constant G - 3.5 G	1-14
40	18	Data Input (ET)	Constant G - 2G	1-14
41	13	Data Input (ET)	Constant G - 3.5 G	1-14
42	14	Data Input (T)	Low Bowtie	1-14
43	10	Data Input (T)	High Bowtie	1-14
44	15	Data Input (RP)	Low Bowtie	1-14
45	16	Data Input (RP)	High Bowtie	1-14
46	14	Data Input (ET)	Low Bowtie	1-14
47	14	Data Input (ET)	High Bowtie	1-14
48	6	Operational Scenario (T)	N/A	7-14
49	6	Operational Scenario (RP)	N/A	7-14
50	6	Operational Scenario (ET)	N/A	7-14

APPENDIX D – STATISTICAL ANALYSIS

Experimental Blocking

Consider the three HSI methods, three tasks (omitting the operational task), and five total maneuvers.

As a compromise between experimental rigor and test matrix size, all operators were considered roughly equivalent, but the experiment sought to maximize the repeated numbers of HSI method and task pairs. HSI method and task pairs were generated by generating permutations of HSI methods and permutations of task separately, and then combining them. For example, HSI methods could be permutated as: ET, TH, RP. (In the succeeding, this will be abbreviated as 1-3-2.) Also as example, tasks could be permutated as: 3-2-1. This would generate the following HSI-task pairs: 13, 32, 21. In execution, this operator would fly HSI method one first, cycling through all three tasks in the order 3-2-1. Subsequently, he would fly HSI method three, cycling through all three tasks in the order 3-2-1 before going on to the final HSI method (two).

As stated in the body of the memorandum, all maneuvers were flown in order (1 through 5), and Task 4 was flown after the entire remainder of the test matrix was complete. Operators were flown in a semi-random manner according to schedule availability.

The total number of such HSI method and task pairs was just $3 \cdot 3 = 9$. The number of operators was 7. Every operator was assigned one HSI method permutation (e.g. 3-2-1), and one task permutation (e.g. 1-2-3); thus he tested a total of 3 HSI method and task pairs (in the parenthetical examples, 3-1, 2-2, and 1-3).

With 6 operators, 18 total pairs were tested. The 7th operator was assigned HSI method and task permutations which mirrored the operator with the closest operational background. Each HSI method and task pair was thus tested two or more times.

It is important to observe that in this blocking, not every operator tested every HSI-Task permutation, but at least two operators tested each HSI-Task permutation. Paper records of HSI, Task, and Maneuvers flown by the Have Rhino team in execution provide the exact details of which operator flew which blocking. However, this experimental design intended to obscure the covariate of operator-blocking pairing.

CEP95

The CEP95 was calculated by multiplying the root-mean-square (RMS) and the Y value of the population. The Y value was determined by the geometric mean (GM) and RMS, producing an R value. Using the R value, linear interpolation was used to calculate the Y value for the population in the table below.

R = GM / RMS	Y
0.02	2.41
0.05	2.375
0.53	1.955
0.745	1.735
0.9	1.49
0.95	1.35
0.975	1.25
1.0	1.0

T-Tests

The following tables present the full results of the Welch's T-tests performed in the population analyses performed, as described in section 2.1.2.3. Any statements regarding differences between means in this memorandum are supported by the population inferences formed using these T-tests.

TASK / STO: 1

	ET vs RP				RP vs TH				ET vs TH					ET		RP		TH		
Distribution	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result		Mean	StDev	Mean	StDev	Mean	StDev	unit
Accuracyat Action	15.298	225	2.20E-16	reject	4.2282	383	0.0000295	reject	15.803	222	2.2E-16	reject		5.743	5.067	0.513	0.518	0.357	0.249	cm
Time at Action	-6.1147	319	2.82E-09	reject	-29.319	303	2.2E-16	reject	-27.632	426	2.2E-16	reject	Sample	1.529	1.292	2.117	0.664	4.866	1.207	s
Accuracyat Designate	-0.42398	455	6.72E-01	accept	1.5535	257	0.1215	accept	1.3572	226	0.1761	accept	Parameters	0.576	2.388	0.689	3.399	0.357	0.249	cm
Time to Designate	4.5996	313	6.15E-06	reject	-22.409	337	2.2E-16	reject	-12.152	411	2.2E-16	reject		3.222	1.579	2.685	0.784	4.866	1.207	s
Velocity	-1.74	339.02	8.20E-02	accept	5.61	394.4	3.88E-08	reject	1.42	262.52	0.1556	accept		6.885	4.065	7.426	2.312	6.477	1.226	cm/s

TASK / STO: 2

	ET vs RP				RP vs TH				ET vs TH					ET		RP		TH		
Distribution	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result		Mean	StDev	Mean	StDev	Mean	StDev	unit
Pre-Expand Accuracyat Action	3.61	380.65	3.45E-04	reject	2.91	416.71	3.85E-03	reject	6.35	371.73	6.11E-10	reject		14.248	6.143	12.045	5.967	10.380	5.756	cm
Pre-Expand Time at Action	-2.64	322.92	8.60E-03	reject	-5.54	372.02	5.80E-08	reject	-6.96	375.96	1.53E-11	reject		1.958	1.327	2.270	0.960	2.888	1.291	s
Post-Expand Accuracyat Action	12.63	198.66	3.20E-27	reject	3.03	235.44	2.76E-03	reject	13.23	190.44	9.12E-29	reject		5.672	5.581	0.515	0.894	0.326	0.193	cm
Post-Expand Time at Action	-0.23	321.26	8.14E-01	accept	-5.66	335.30	3.20E-08	reject	-4.90	388.73	1.41E-06	reject		4.186	1.664	4.219	1.091	5.019	1.698	s
Accuracyat Designate	2.39	185.25	1.78E-02	reject	-0.13	412.52	8.97E-01	accept	2.38	185.34	1.85E-02	reject		0.733	2.318	0.324	0.197	0.326	0.193	cm
Time to Complete	5.90	328.57	9.00E-09	reject	-0.59	361.79	5.53E-01	accept	4.63	379.31	5.02E-06	reject	Sample	5.822	1.699	4.932	1.237	5.019	1.698	s
Pre-Expand High 15% Time to Action	1.97	52.91	5.42E-02	accept	-4.78	49.00	1.63E-05	reject	-2.60	49.92	2.12E-02	reject	Parameters	3.740	1.292	3.080	1.194	4.589	1.060	s
Pre-Expand Low 15% Time to Action	1.12	34.28	2.71E-01	accept	1.25	57.65	2.15E-01	accept	2.07	39.14	4.55E-02	reject		2.446	0.857	2.189	0.663	1.949	0.821	s
Pre-Expand High 15% Pseudo Velocity	-0.96	52.43	3.42E-01	accept	5.18	34.17	9.92E-06	reject	3.44	33.69	1.56E-03	reject		8.399	2.826	9.084	2.452	6.448	0.935	cm/s
Pre-Expand Low 15 % Pseudo Velocity	-1.09	42.86	2.80E-01	accept	-1.22	59.91	2.27E-01	accept	-2.25	45.58	2.96E-02	reject		1.810	0.780	2.068	0.841	2.347	0.927	cm/s
Pre-Expand 20cm or More at Task Initiation Veclocity	-1.16	95.46	2.48E-01	accept	5.52	74.59	1.37E-07	reject	3.70	65.10	4.49E-04	reject		7.808	2.548	8.362	2.228	6.339	0.994	cm/s
Pre-Expand 17.5cm or Less at Task Initiation Velocity	-2.15	227.21	3.30E-02	reject	3.03	235.30	2.74E-03	reject	0.27	164.50	7.90E-01	accept		3.905	2.068	4.489	2.158	3.844	1.334	cm/s
Pre-Expand 15cm or Less at Task Initiation Velocity	-0.58	178.94	5.60E-01	accept	0.01	197.63	9.92E-01	accept	-0.65	144.16	5.18E-01	accept		3.569	1.877	3.721	1.724	3.719	1.317	cm/s

TASK / STO: 3

	ET vs RP				RP vs TH				ET vs TH					ET		RP		TH		
Distribution	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result		Mean	StDev	Mean	StDev	Mean	StDev	unit
Action (by key) Accuracy	42.85	933.36	1.22E-222	reject	4.28	991.41	2.02E-05	reject	46.20	786.76	2.45E-226	reject	Sample	5.101	2.788	0.612	0.944	0.473	0.247	cm
Designate (by key) Accuracy	3.83	1801.29	1.32E-04	reject	3.28	1969.03	1.04E-03	reject	7.17	1725.73	1.14E-12	reject	Parameters	0.561	0.288	0.511	0.274	0.473	0.247	s
Total Data Entry Error Rate	-0.48196	150.663	0.6305367	accept	0.39	195.23	6.98E-01	accept	-0.16	142.09	8.72E-01	accept		2.469	8.250	3.054	7.465	2.659	6.896	cm
Time to Complete	5.16	107.97	1.13E-06	reject	11.26	140.12	2.44E-21	reject	9.68	91.89	1.08E-15	reject		31.670	20.662	19.633	7.596	10.050	3.798	s

TASK / STO (Workload Only): 1

	ET vs RP				RP vs TH				ET vs TH					ET		RP		TH	
Distribution	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result		Mean	StDev	Mean	StDev	Mean	StDev
Operator 1 (LWL)	1.21	90.92	2.29E-01	accept	-5.49	95.59	3.35E-07	reject	-4.01	104.59	1.16E-04	reject		0.650	0.073	0.634	0.056	0.712	0.089
Operator 1 (HWL)	6.77	111.54	6.22E-10	reject	-10.88	118.85	1.52E-19	reject	-4.14	114.97	6.75E-05	reject		0.745	0.068	0.661	0.066	0.800	0.075
Operator 2 (LWL)	13.84	87.88	8.44E-24	reject	-18.02	85.92	6.26E-31	reject	-5.34	91.83	6.70E-07	reject		0.887	0.041	0.733	0.069	0.931	0.039
Operator 2 (HWL)	17.67	109.35	8.18E-34	reject	-36.58	89.00	2.08E-55	reject	-13.74	81.05	7.13E-23	reject		0.855	0.044	0.707	0.044	0.949	0.026
Operator 3 (LWL)	37.19	68.69	3.07E-47	reject	22.92	100.20	1.23E-41	reject	50.95	87.22	1.11E-66	reject		0.911	0.053	0.572	0.030	0.401	0.046
Operator 3 (HWL	-3.72	96.42	3.37E-04	reject	37.56	70.44	2.49E-48	reject	31.41	64.64	7.35E-41	reject		0.796	0.086	0.860	0.085	0.370	0.045
Operator 4 (LWL)	-0.85	140.86	3.98E-01	accept	-2.63	108.77	9.68E-03	reject	-3.37	109.99	1.05E-03	reject	Sample	0.746	0.062	0.755	0.061	0.785	0.065
Operator 4 (HWL)	6.72	90.96	1.57E-09	reject	-4.34	114.53	3.11E-05	reject	3.54	79.72	6.81E-04	reject	Parameters	0.893	0.064	0.812	0.057	0.853	0.050
Operator 5 (LWL)	11.17	91.82	8.32E-19	reject	-7.67	94.83	-1.49E+00	reject	2.33	90.7	2.22E-02	reject		0.840	0.076	0.653	0.091	0.800	0.097
Operator 5 (HWL)	6.33	107.7	5.78E-09	reject	-5.87	117.75	4.04E-08	reject	-0.53	97.39	5.94E-01	accept		0.816	0.069	0.713	0.108	0.824	0.099
Operator 6 (LWL)	-0.27	91.68	7.88E-01	accept	0.50	93.92	6.19E-01	accept	0.18	93.94	8.60E-01	accept		0.783	0.101	0.788	0.087	0.780	0.086
Operator 6 (HWL)	-1.17	94.99	2.44E-01	accept	-3.40	105.28	9.62E-04	reject	-4.77	98.02	6.41E-06	reject		0.713	0.091	0.737	0.104	0.802	0.097
Operator 7 (LWL)	0.10	88.05	9.19E-01	accept	-3.18	105.21	1.94E-03	reject	-2.72	76.41	8.01E-03	reject		0.760	0.109	0.757	0.103	0.814	0.086
Operator 7 (HWL)	1.11	49.32	2.73E-01	accept	1.54	112.99	1.26E-01	accept	2.28	50.17	2.71E-02	reject		0.753	0.100	0.728	0.091	0.702	0.093
All Operators (All Maneuvers)	4.93	1963.00	8.80E-07	reject	-1.73	2062.80	8.36E-02	accept	3.48	1956.70	5.05E-04	reject		0.755	0.105	0.668	0.129	0.725	0.166

TASK / STO (Workload Only): 2

	ET vs RP				RP vs TH				ET vs TH					ET		RP		TH	
Distribution	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result		Mean	StDev	Mean	StDev	Mean	StDev
Operator 1 (LWL)	0.69	94.21	4.93E-01	accept	-5.72	95.94	1.19E-07	reject	-5.03	105.57	2.04E-06	reject		0.723	0.087	0.711	0.076	0.813	0.100
Operator 1 (HWL)	4.32	107.09	3.55E-05	reject	-8.34	122.69	1.29E-13	reject	-3.42	116.62	8.69E-04	reject		0.764	0.084	0.703	0.069	0.817	0.084
Operator 2 (LWL)	15.34	97.86	8.98E-28	reject	24.56	86.34	1.03E-40	reject	-6.60	85.09	3.31E-09	reject		0.874	51.000	0.721	0.049	0.932	0.033
Operator 2 (HWL)	18.16	107.03	1.81E-34	reject	-36.03	78.67	1.17E-50	reject	-15.88	106.29	8.24E-30	reject		0.845	0.047	0.687	0.045	0.954	0.028
Operator 3 (LWL)	0.48	74.01	6.30E-01	accept	47.75	102.56	7.05E-72	reject	29.51	71.62	8.20E-42	reject		0.857	0.108	0.085	0.055	0.371	0.048
Operator 3 (HWL	1.27	111.39	2.07E-01	accept	43.21	89.43	1.05E-61	reject	36.86	92.25	5.47E-57	reject		0.843	0.093	0.824	0.065	0.363	0.045
Operator 4 (LWL)	5.86	84.15	8.68E-08	reject	-2.74	93.02	7.31E-03	reject	3.48	88.36	7.70E-04	reject	Sample	0.787	0.075	0.713	0.049	0.741	0.053
Operator 4 (HWL)	3.55	105.76	5.79E-04	reject	-2.04	93.47	4.41E-02	reject	1.27	96.89	2.06E-01	accept	Parameters	0.833	0.058	0.795	0.057	0.819	0.057
Operator 5 (LWL)	12.76	92.28	2.10E-22	reject	-5.52	85.82	3.47E-07	reject	4.50	70.54	2.62E-05	reject		0.879	0.068	0.708	0.078	0.807	0.094
Operator 5 (HWL)	9.85	100.89	1.94E-16	reject	-9.57	101.32	7.55E-16	reject	0.77	89.79	4.43E-01	accept		0.854	0.084	0.675	0.101	0.841	0.076
Operator 6 (LWL)	2.14	115.54	3.48E-02	reject	-3.77	79.97	3.13E-04	reject	-1.31	148.07	1.91E-01	accept		0.774	0.120	0.737	0.081	0.794	0.067
Operator 6 (HWL)	-1.65	91.72	1.03E-01	accept	0.14	87.84	8.85E-01	accept	-1.90	64.13	6.16E-02	accept		0.700	0.097	0.737	0.117	0.734	0.075
Operator 7 (LWL)	2.36	43.70	2.28E-02	reject	.2.28	64.92	2.57E-02	reject	0.51	26.55	6.14E-01	accept		0.847	0.071	0.784	0.124	0.836	0.079
Operator 7 (HWL)	-1.43	84.22	1.56E-01	accept	-0.23	99.77	8.20E-01	accept	-1.75	70.48	8.40E-02	accept		0.752	0.069	0.778	0.102	0.782	0.078
All Operators (All Maneuvers)	3.63	2080.7	2.88E-04	reject	2.04	1880.5	4.14E-02	reject	6.37	2520.6	2.26E-10	reject		0.766	0.103	0.705	0.121	0.729	0.158

TASK / STO (Workload Only): 3

	ET vs RP				RP vs TH				ET vs TH					ET		RP		TH	
Distribution	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result		Mean	StDev	Mean	StDev	Mean	StDev
Operator 1 (LWL)	6.89	249.98	4.55E-11	reject	-20.36	94.99	2.03E-36	reject	-15.44	96.74	7.49E-28	reject		0.597	0.070	0.535	0.075	0.799	0.087
Operator 1 (HWL)	2.01	290.29	4.55E-02	reject	-6.17	15.35	1.63E-05	reject	-5.42	15.02	7.06E-05	reject		0.671	0.087	0.649	0.096	0.821	0.100
Operator 2 (LWL)	14.32	79.68	9.60E-24	reject	-23.3	109.5	3.06E-44	reject	-7.84	#######	8.29E-11	reject		0.830	0.038	0.687	0.060	0.903	0.039
Operator 2 (HWL)	13.65	59.10	6.39E-20	reject	-27.23	51.63	2.70E-32	reject	-9.22	41.99	1.21E-11	reject		0.847	0.047	0.697	0.060	0.948	0.026
Operator 3 (LWL)	38.60	82.03	2.27E-54	reject	6.72	68.11	4.52E-09	reject	36.89	116.48	4.29E-66	reject		0.856	0.080	0.469	0.032	0.418	0.050
Operator 3 (HWL	0.24	59.30	8.14E-01	accept	37.67	117.77	1.58E-67	reject	23.25	60.63	6.80E-32	reject		0.846	0.103	0.842	0.060	0.435	0.059
Operator 4 (LWL)	-0.36	111.72	7.21E-01	accept	-5.81	31.59	1.96E-06	reject	-5.83	35.26	1.27E-06	reject	Sample	0.652	0.080	0.656	0.063	0.799	0.118
Operator 4 (HWL)	4.22	92.72	5.63E-05	reject	-10.17	72.97	1.22E-15	reject	-7.88	66.14	4.28E-11	reject	Parameters	0.660	0.053	0.613	0.059	0.800	0.114
Operator 5 (LWL)	5.77	126.58	5.83E-08	reject	-13.23	164.13	1.17E-27	reject	-9.50	125.58	1.86E-16	reject		0.602	0.045	0.536	0.088	0.737	0.117
Operator 5 (HWL)	6.92	47.31	1.04E-08	reject	-11.02	67.67	9.60E-17	reject	-6.99	67.38	1.57E-09	reject		0.702	0.052	0.574	0.098	0.812	0.089
Operator 6 (LWL)	0.55	68.80	5.85E-01	accept	-5.3	101.40	6.72E-07	reject	-5.54	11.79	2.01E-07	reject		0.603	0.072	0.596	0.065	0.686	0.109
Operator 6 (HWL)	0.19	77.15	8.49E-01	accept	-6.13	63.72	6.26E-08	reject	-5.33	79.81	8.93E-07	reject		0.552	0.088	0.549	0.065	0.670	0.117
Operator 7 (LWL)	1.04	91.48	2.99E-01	accept	-6.3	97.78	8.42E-09	reject	-5.29	106.42	6.86E-07	reject		0.615	0.097	0.596	0.084	0.726	0.123
Operator 7 (HWL)	6.96	91.89	5.01E-10	reject	-5.48	33.56	4.20E-06	reject	-0.20	42.97	8.46E-01	accept		0.749	0.097	0.627	0.091	0.754	0.099
All Operators (All Maneuvers)	9.53	3878.90	2.69E-21	reject	-15.63	2871.80	6.36E-53	reject	-6.22	3498.40	5.41E-10	reject		0.650	0.132	0.584	0.125	0.673	0.158

TASK / STO (Workload Only): 4

	ET vs RP				RP vs TH				ET vs TH					ET		RP		TH	
Distribution	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result	T-Stat	DoF	P Value	Result		Mean	StDev	Mean	StDev	Mean	StDev
Operator 1	1.15	251.02	2.52E-01	reject	-5.41	293.19	1.28E-07	reject	-4.45	296.58	1.20E-05	reject		0.620	0.072	0.609	0.079	0.660	0.089
Operator 2	2.39	296.95	-1.76E-02	reject	-8.73	229.98	5.14E-16	reject	-6.81	253.88	-7.05E-11	reject		0.768	0.071	0.751	0.057	0.822	0.069
Operator 3	no data	no data	no data	no data	no data	no data	no data	no data	-1.66	380.22	9.68E-02	accept	Sample	0.562	0.073	no data	no data	0.572	0.090
Operator 4	-2.31	282.18	2.12E-02	reject	-5.15	399.33	4.13E-07	reject	-7.25	389.39	2.33E-12	reject	Parameters	0.651	0.058	0.664	0.058	0.697	0.067
Operator 5	5.60	227.27	6.07E-08	reject	-11.02	264.85	1.65E-23	reject	-7.03	472.64	7.41E-12	reject		0.616	0.086	0.561	0.088	0.676	0.101
Operator 6	3.25	172.57	1.40E-03	reject	-2.75	157.14	6.75E-03	reject	0.77	222.87	4.44E-01	accept		0.683	0.110	0.635	0.087	0.672	0.095
All Operators	0.58	1435.10	5.64E-01	accept	0.97	1419.50	3.32E-01	accept	1.74	2454.10	8.29E-02	accept		0.650	0.104	0.647	0.095	0.642	0.114

Box-and-Whiskers

The box-and-whiskers plots used in this memorandum comply with the format depicted below.

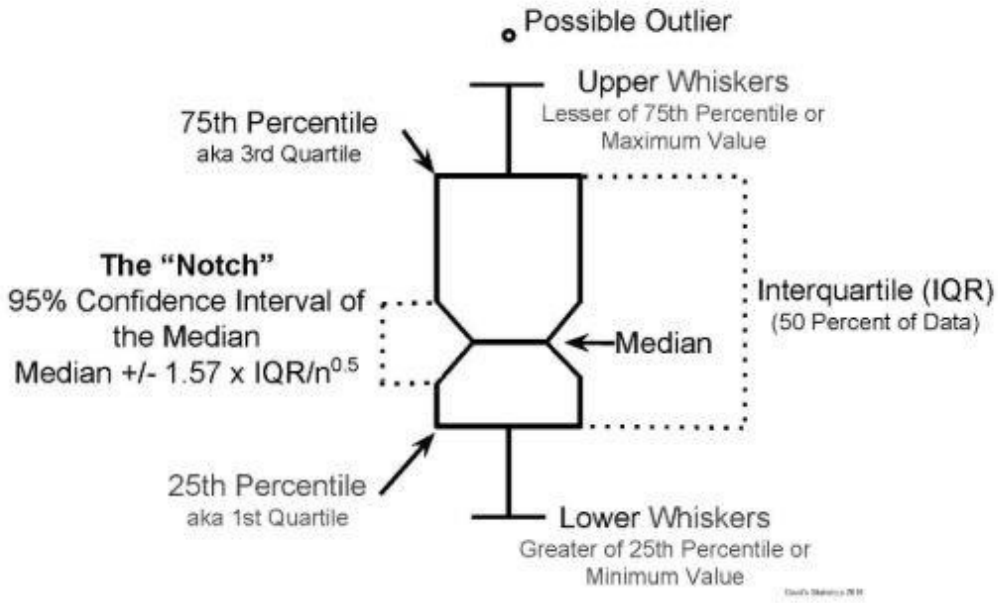
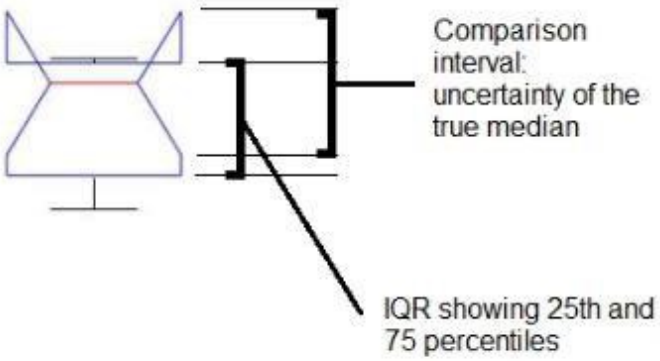


image above: [2]

To apply that to your box plots:



[1] https://www.mathworks.com/help/stats/boxplot.html#bu180jd-Notch

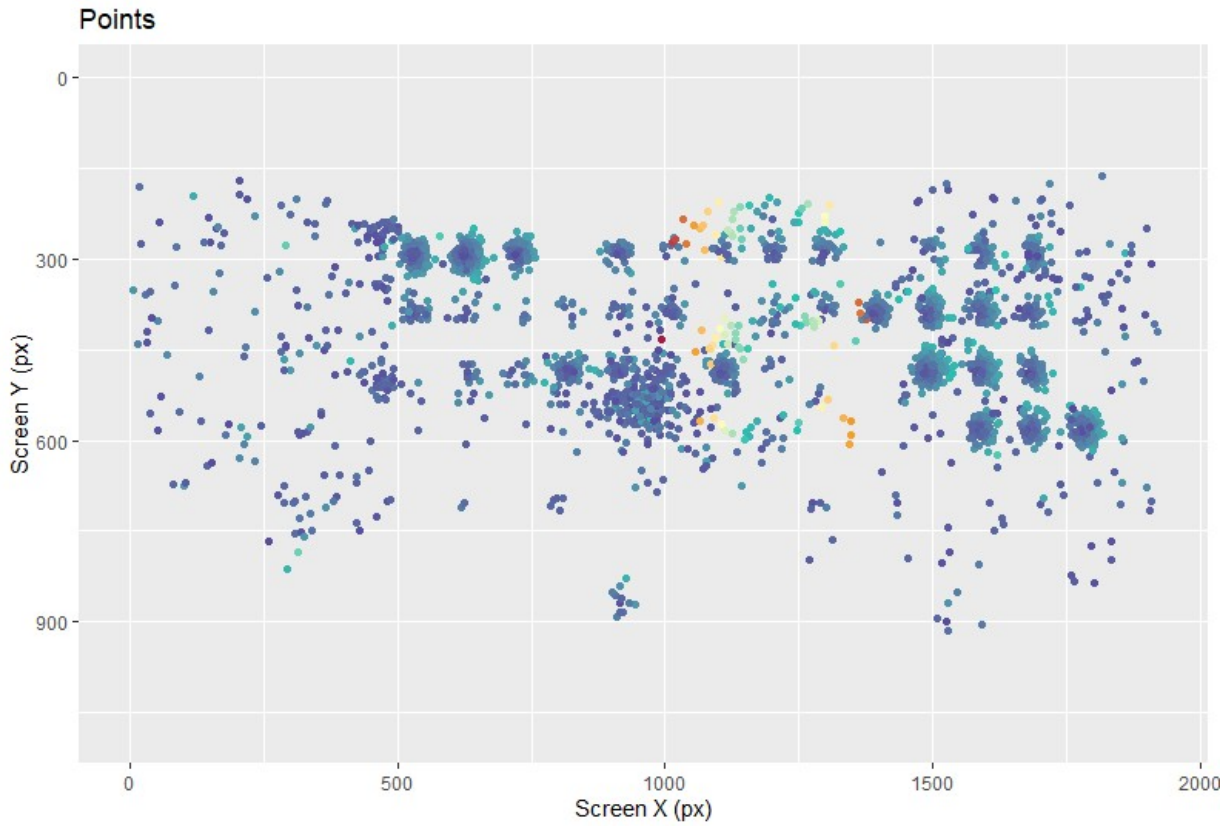
[2] https://sites.google.com/site/davidsstatistics/home/notched-box-plots

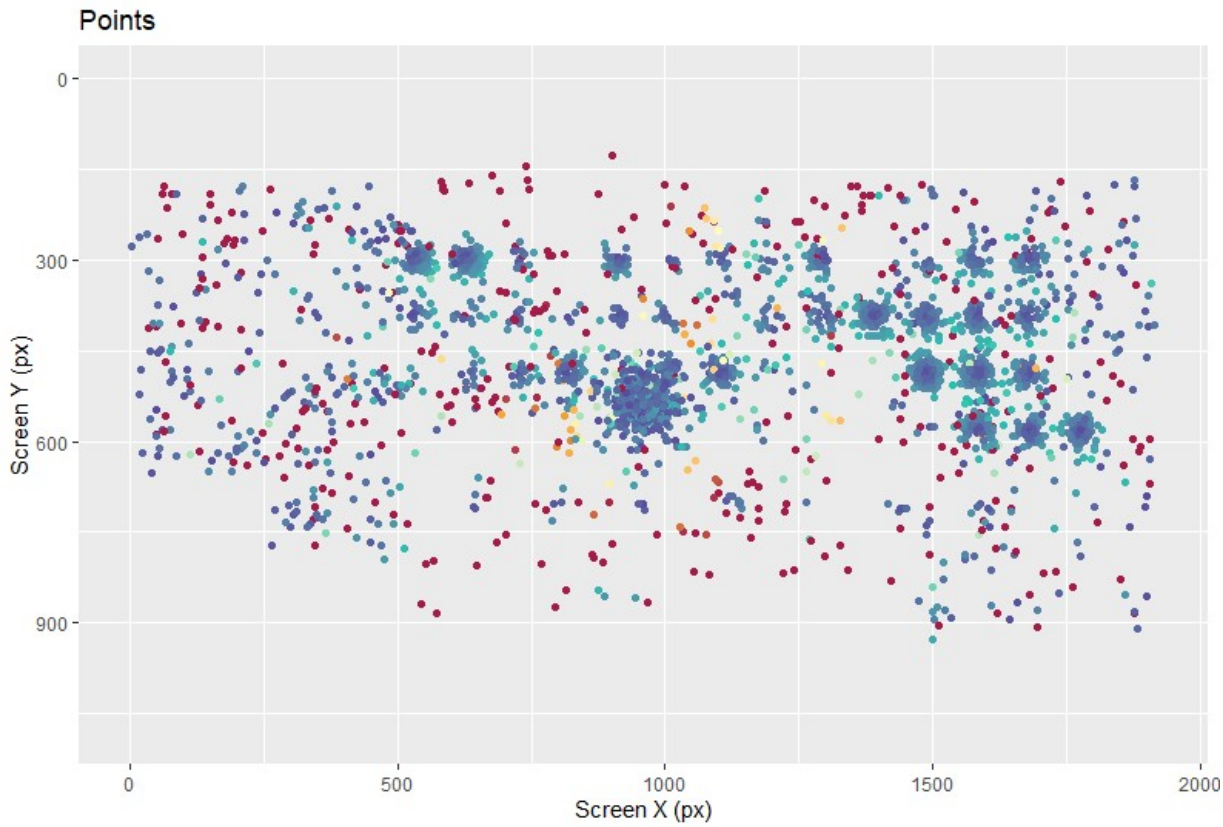
[3] https://www.nature.com/articles/d41586-019-00857-9

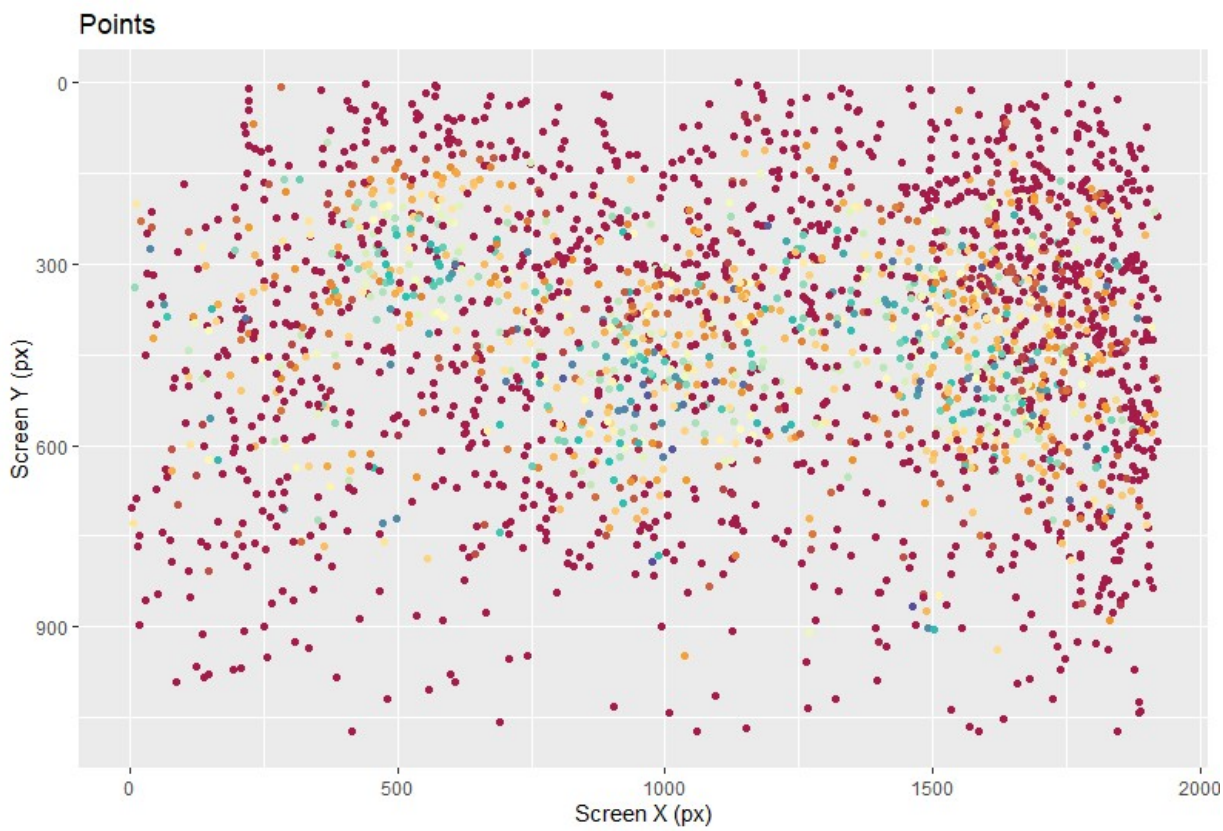
APPENDIX E – ADDITIONAL PLOTS

Heat Maps

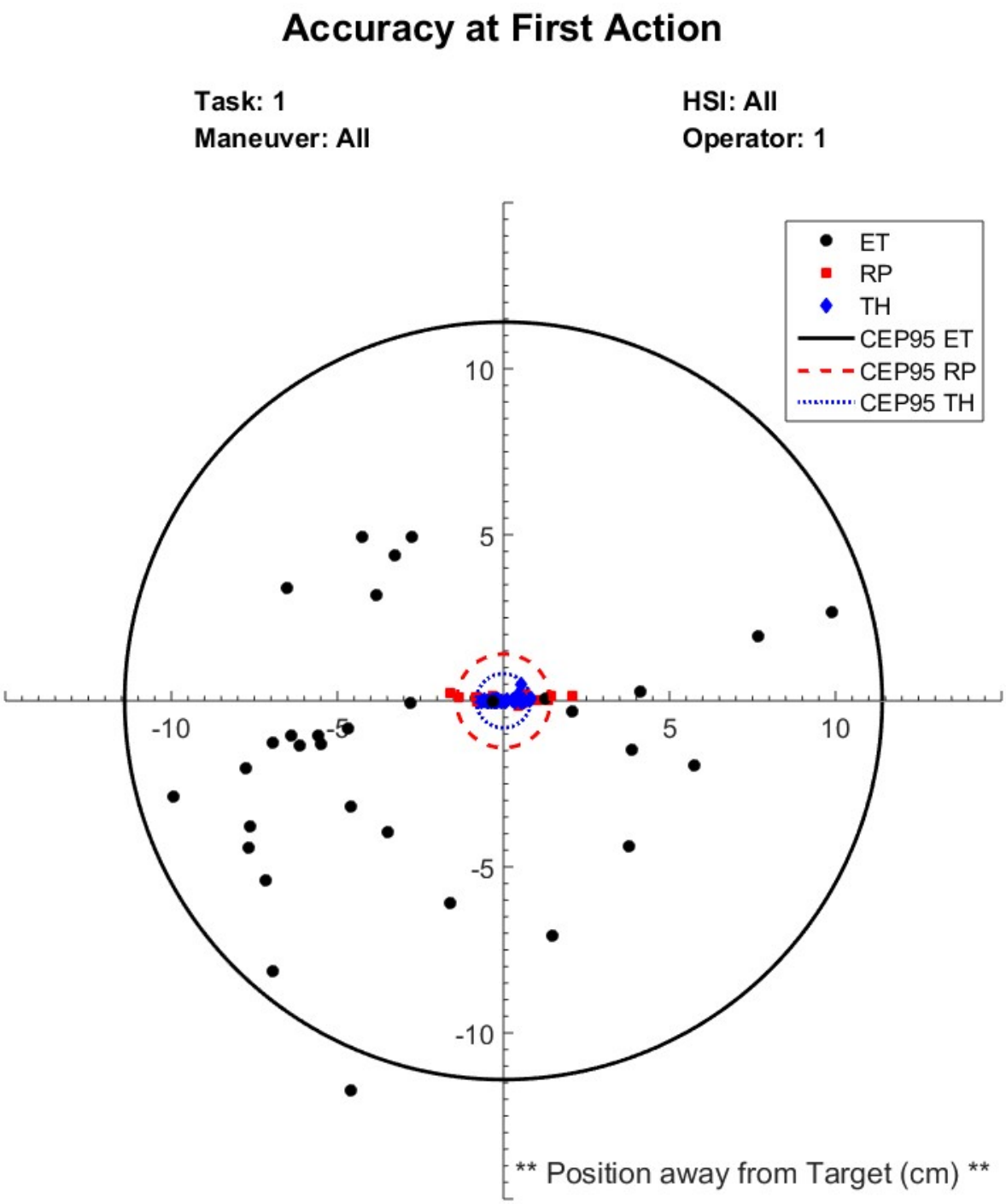
Below are shown heat maps for TH, RP, and ET (top-to-bottom). Population included all tasks, all maneuvers, and all operators. Each plot x- and y-axis spans the display, in pixels. Each point represents the point at which an operator successfully executed an action. The color of each point, scaled by a "Spectral" palette in R programming language, indicates the distance from the intended point (e.g., track) when the action was performed. Blues indicate an action very close to the intended point; reds indicate an action very far from the intended point




Points
Screen Y (px)
0
300
600
900
0
500
1000
1500
2000
Screen X (px)


Points
Screen Y (px)
0
300
600
900
0
500
1000
1500
2000
Screen X (px)

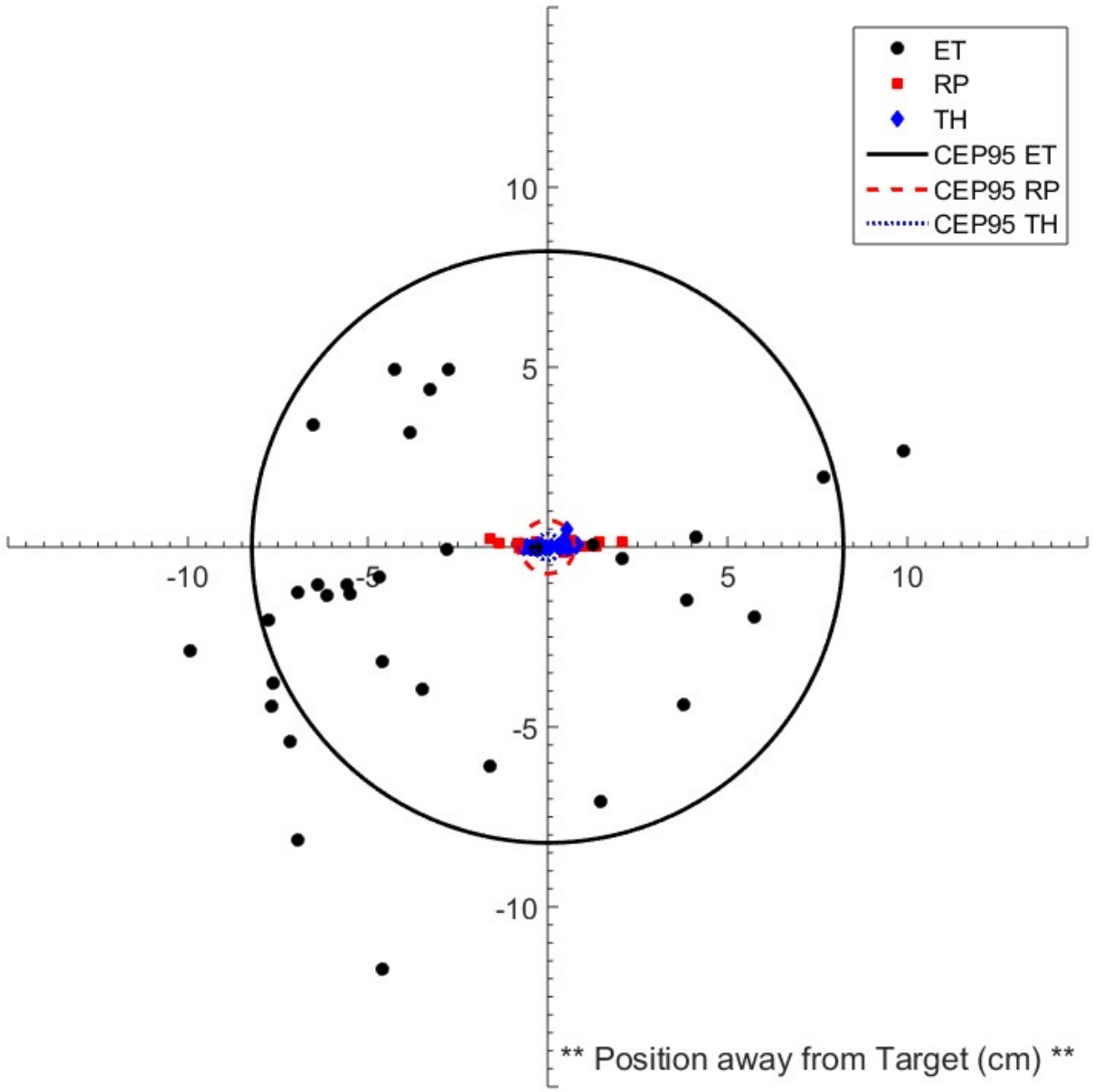
STO 1 – Accuracy



Accuracy at First Action

Task: 1
Maneuver: All

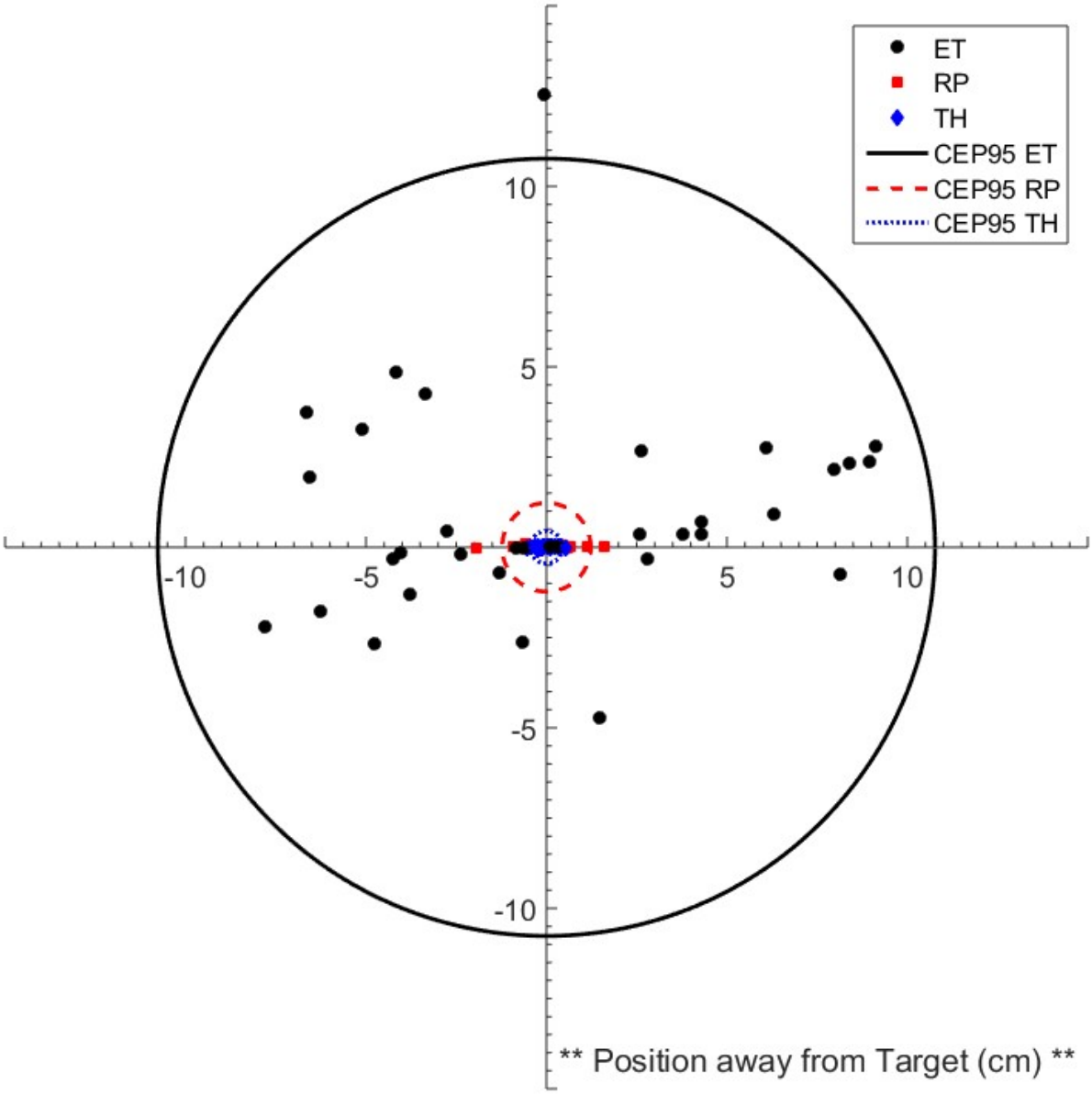
HSI: All
Operator: 2



Accuracy at First Action

Task: 1
Maneuver: All

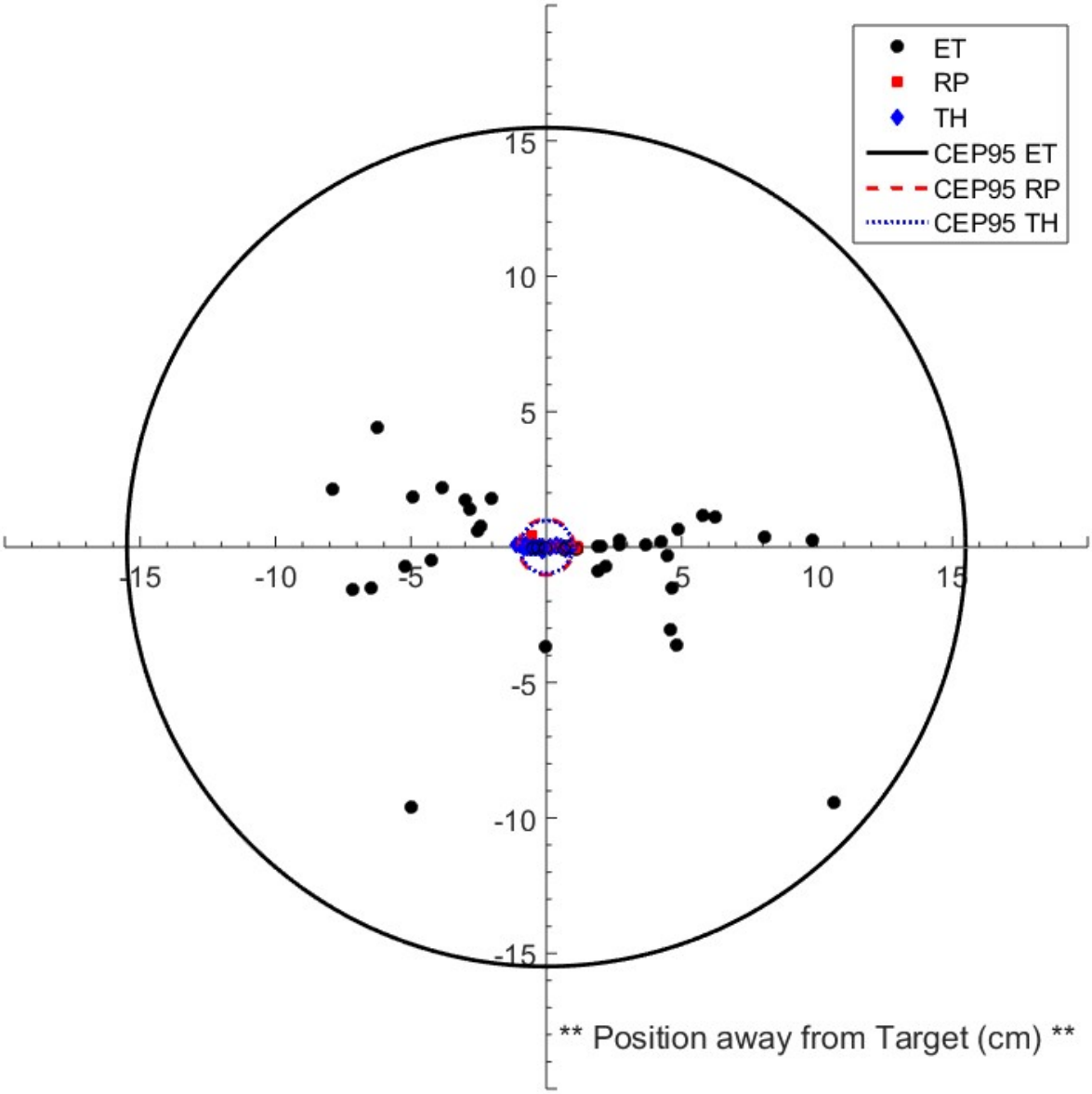
HSI: All
Operator: 3



Accuracy at First Action

Task: 1
Maneuver: All

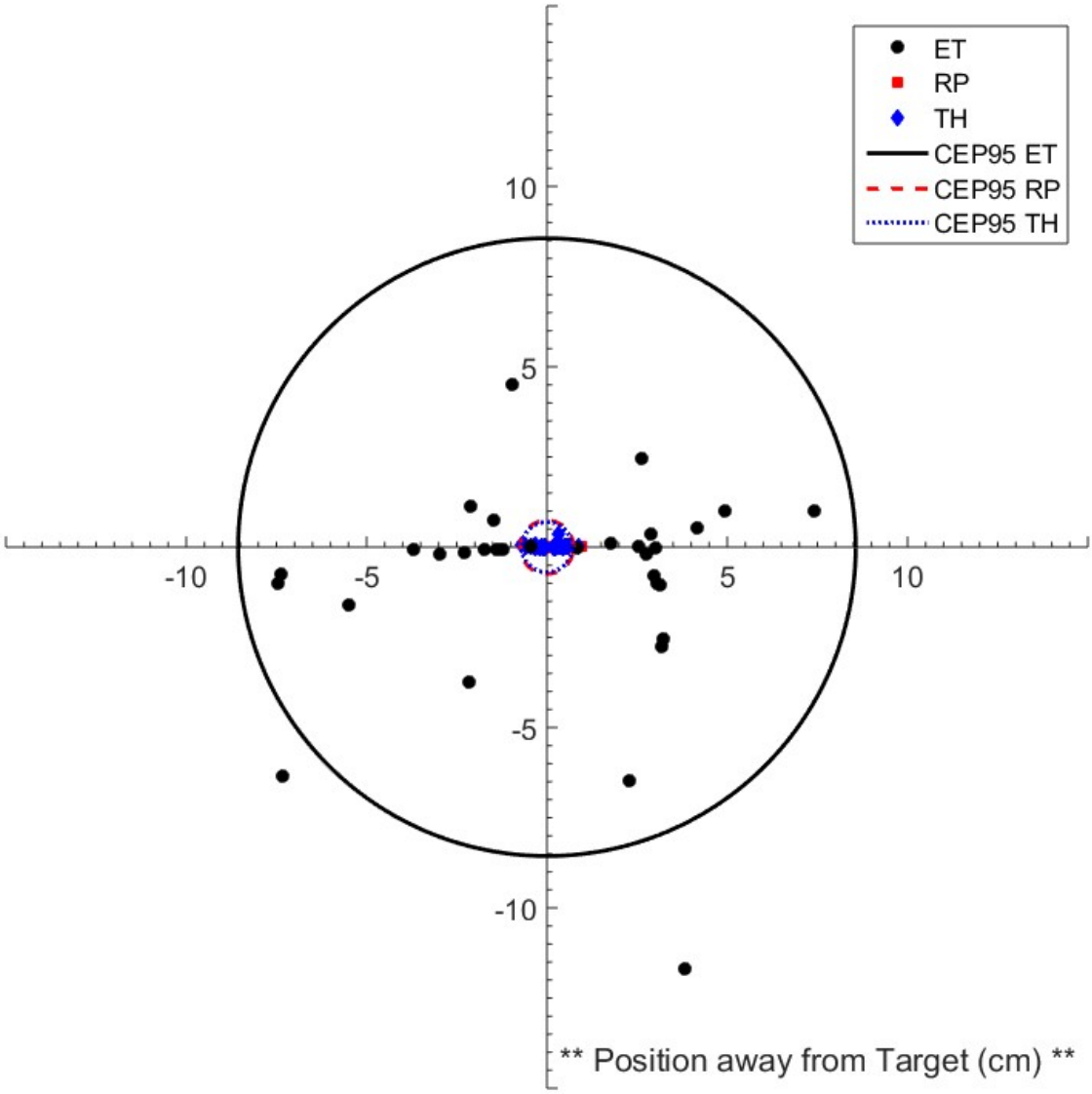
HSI: All
Operator: 4



Accuracy at First Action

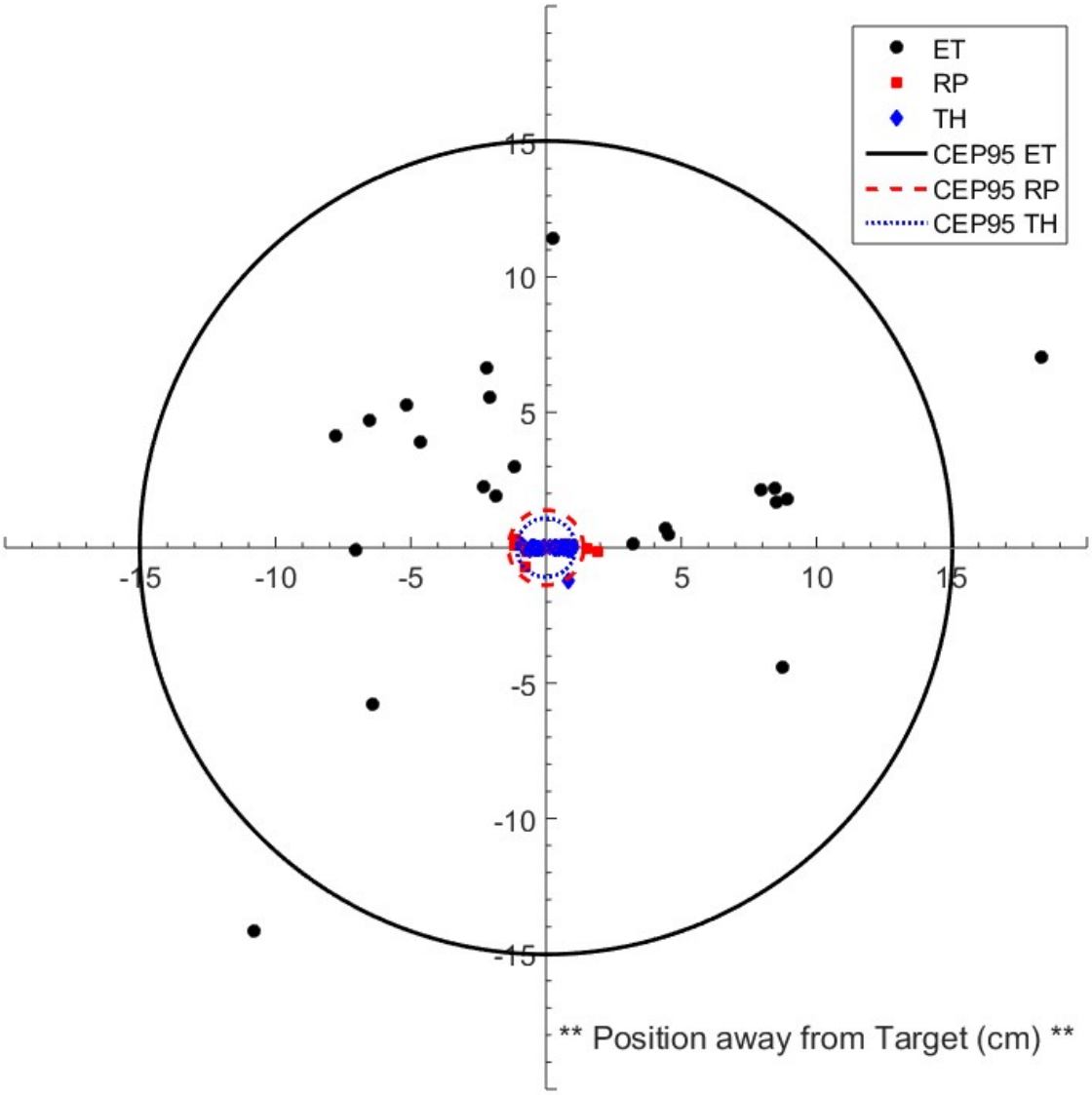
Task: 1
Maneuver: All

HSI: All
Operator: 5



Accuracy at First Action

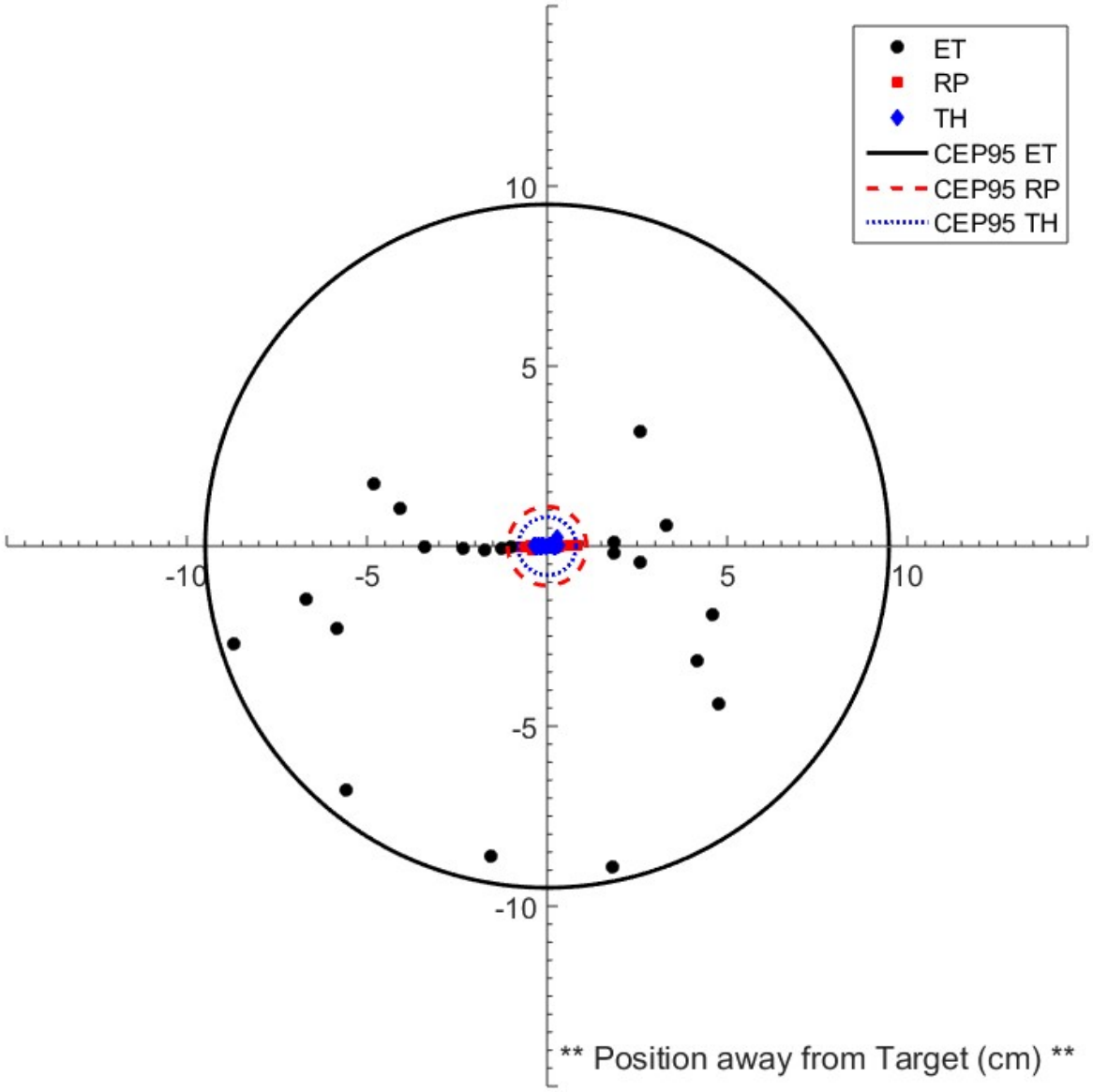
Task: 1 **HSI: All**
Maneuver: All **Operator: 6**

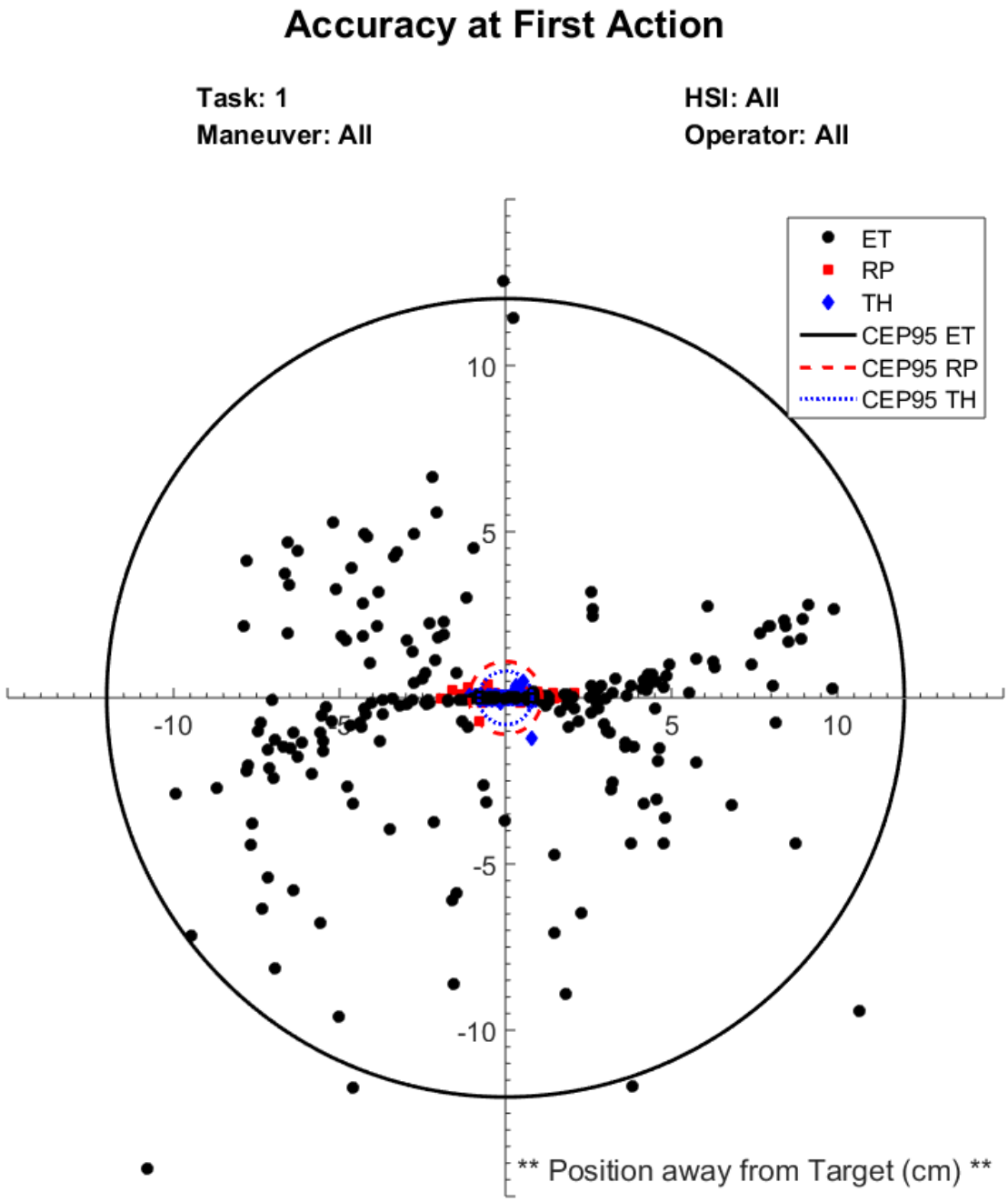


Accuracy at First Action

Task: 1
Maneuver: All

HSI: All
Operator: 7



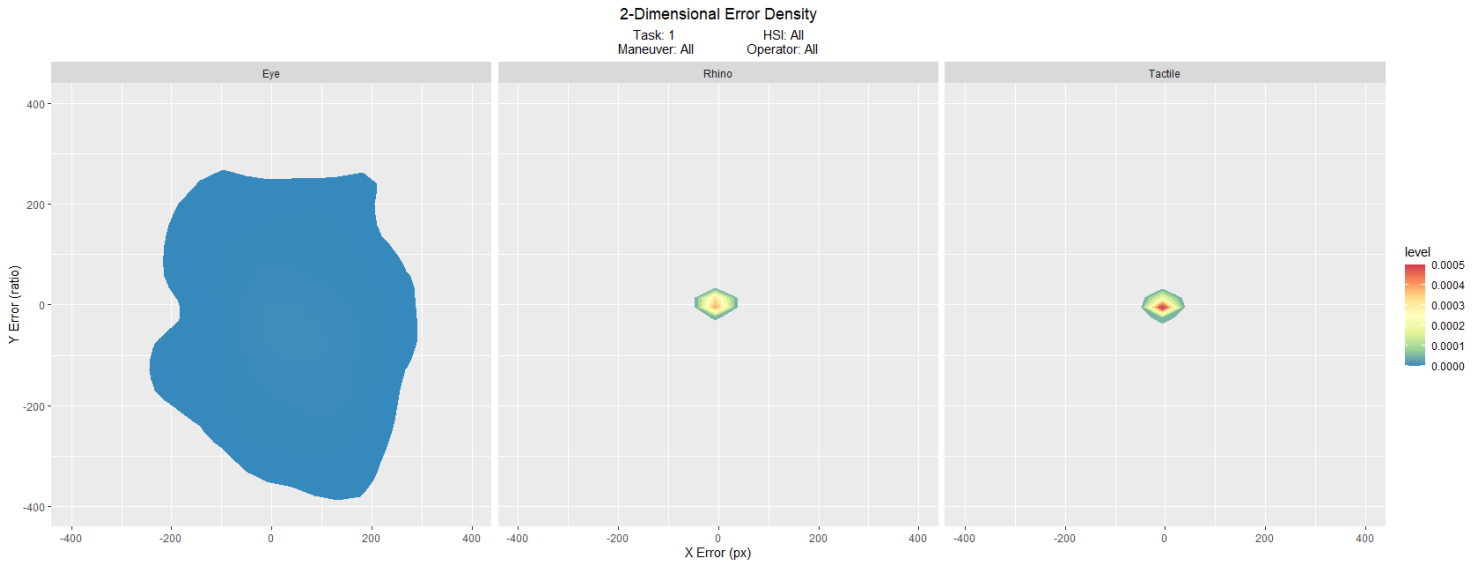

Accuracy at First Action
Task: 1
Maneuver: All
HSI: All
Operator: All
ET
RP
TH
CEP95 ET
CEP95 RP
CEP95 TH
10
5
-5
-10
-10
-5
5
10
** Position away from Target (cm) **

CEP95 Results

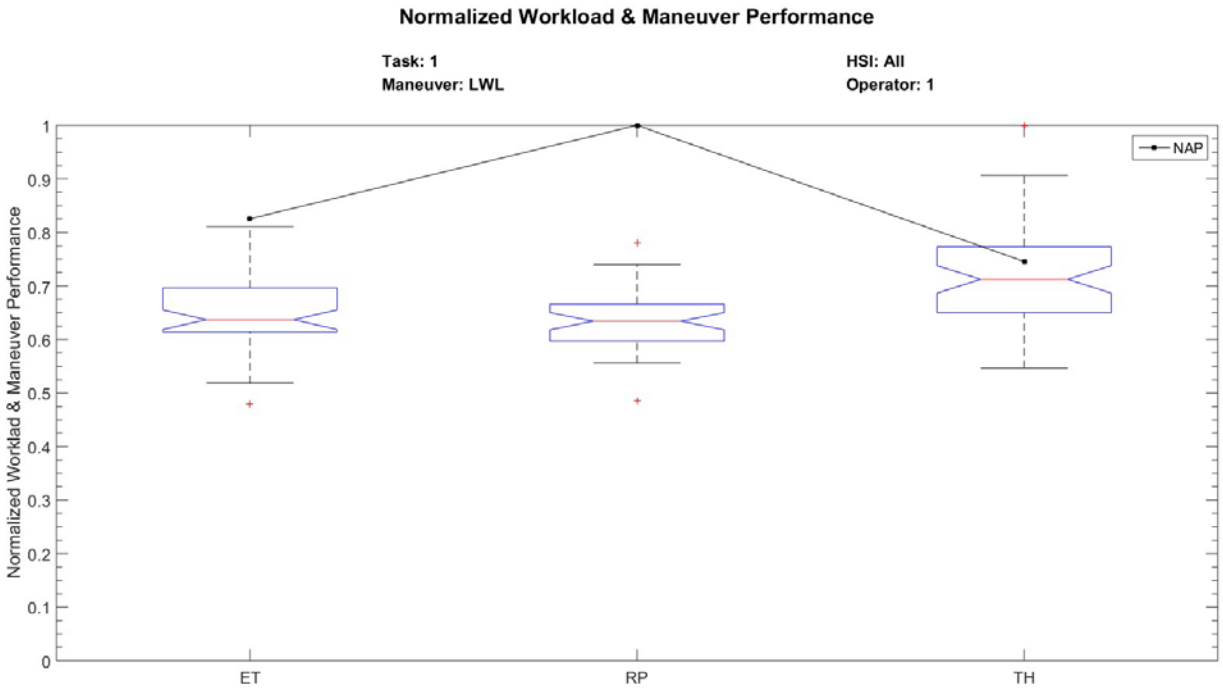
Op	Man	ET_R	ET_Y	ET_RMS	**ET_CEP**	RP_R	RP_Y	RP_RMS	**RP_CEP_95**	T_R	T_Y	T_RMS	**T_CEP_95**
1	All	0.794	1.658	6.879	**11.405**	0.697	1.784	0.791	**1.411**	0.728	1.752	0.466	**0.816**
2	All	0.5	1.981	4.15	**8.221**	0.672	1.81	0.409	**0.740**	0.814	1.626	0.225	**0.366**
3	All	0.639	1.843	5.842	**10.767**	0.602	1.881	0.653	**1.228**	0.621	1.862	0.242	**0.451**
4	All	0.43	2.0425	7.581	**15.484**	0.766	1.702	0.601	**1.023**	0.742	1.738	0.561	**0.975**
5	All	0.666	1.816	4.715	**8.562**	0.682	1.799	0.41	**0.738**	0.707	1.774	0.394	**0.699**
6	All	0.778	1.683	8.926	**15.022**	0.632	1.851	0.723	**1.338**	0.756	1.718	0.632	**1.086**
7	All	0.718	1.763	5.382	**9.488**	0.648	1.834	0.396	**0.726**	0.767	1.7	0.228	**0.388**
All	All	0.574	1.91	6.289	**12.012**	0.626	1.857	0.589	**1.094**	0.628	1.855	0.434	**0.805**

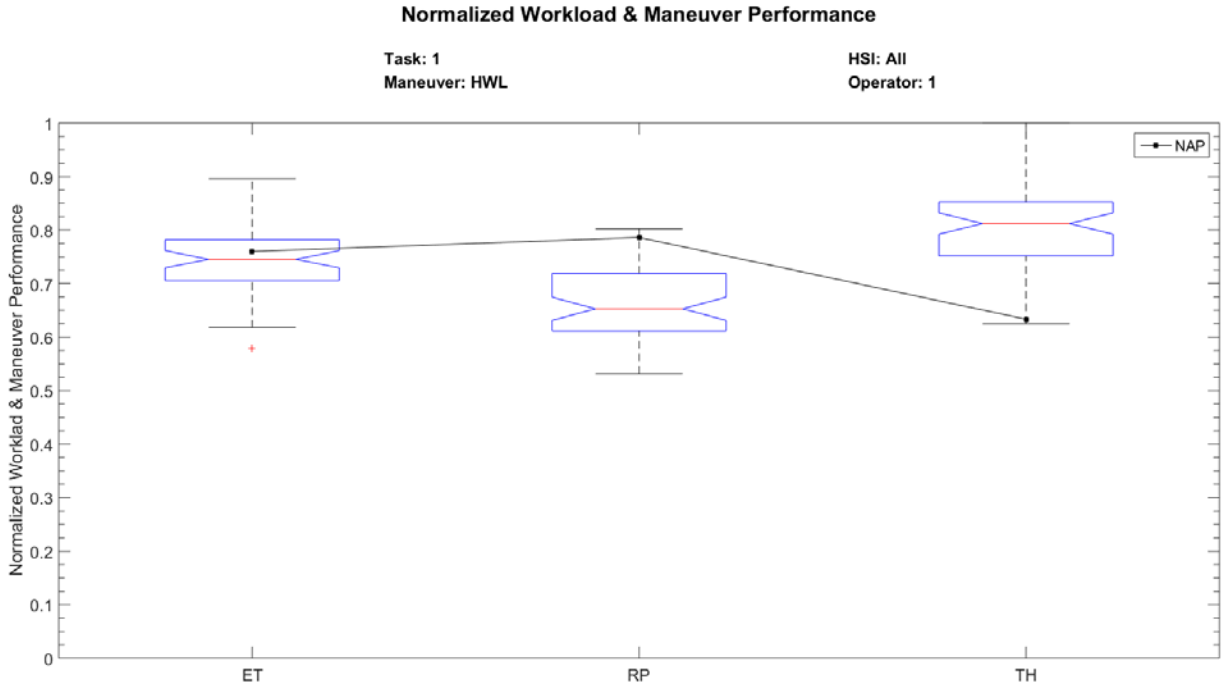
<u>2-dimensional Error Densities</u>

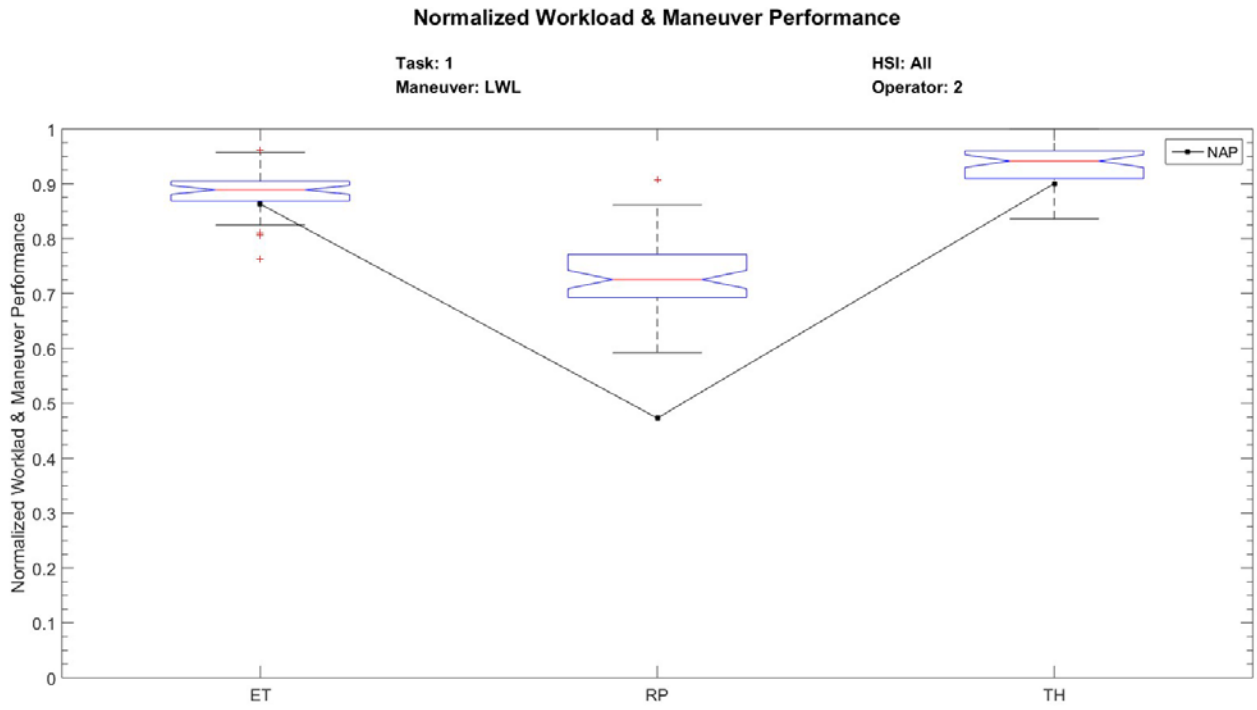
All HSI Methods depicted on a common density level scale.

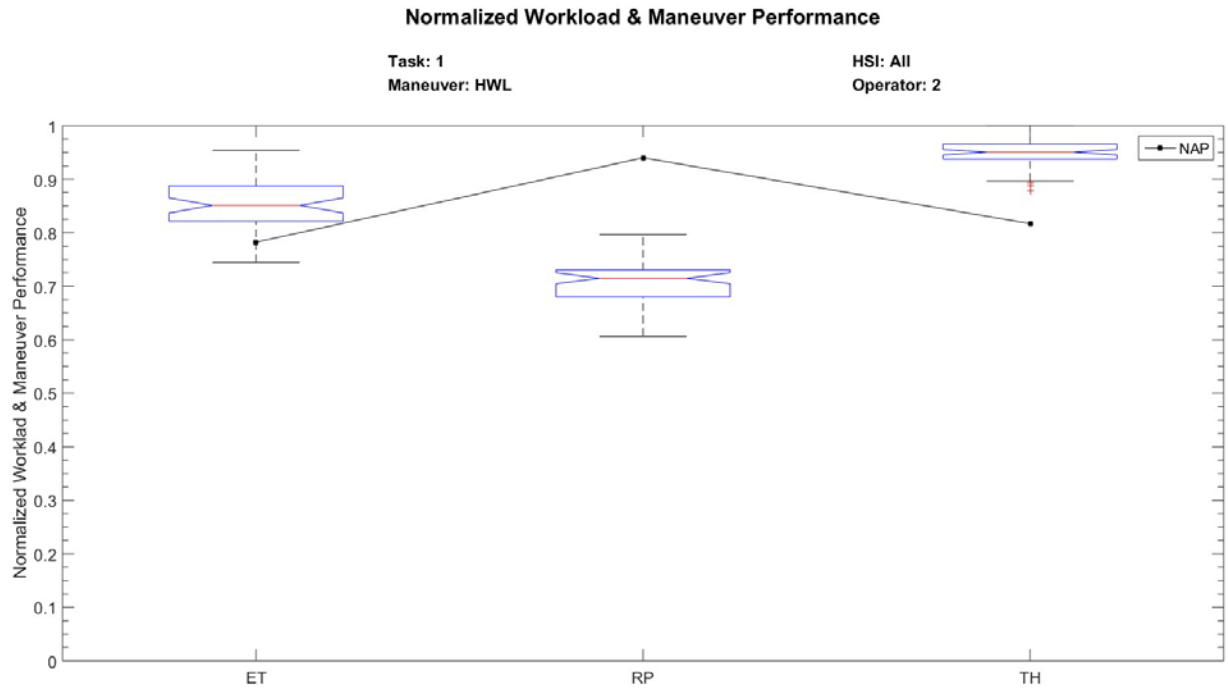


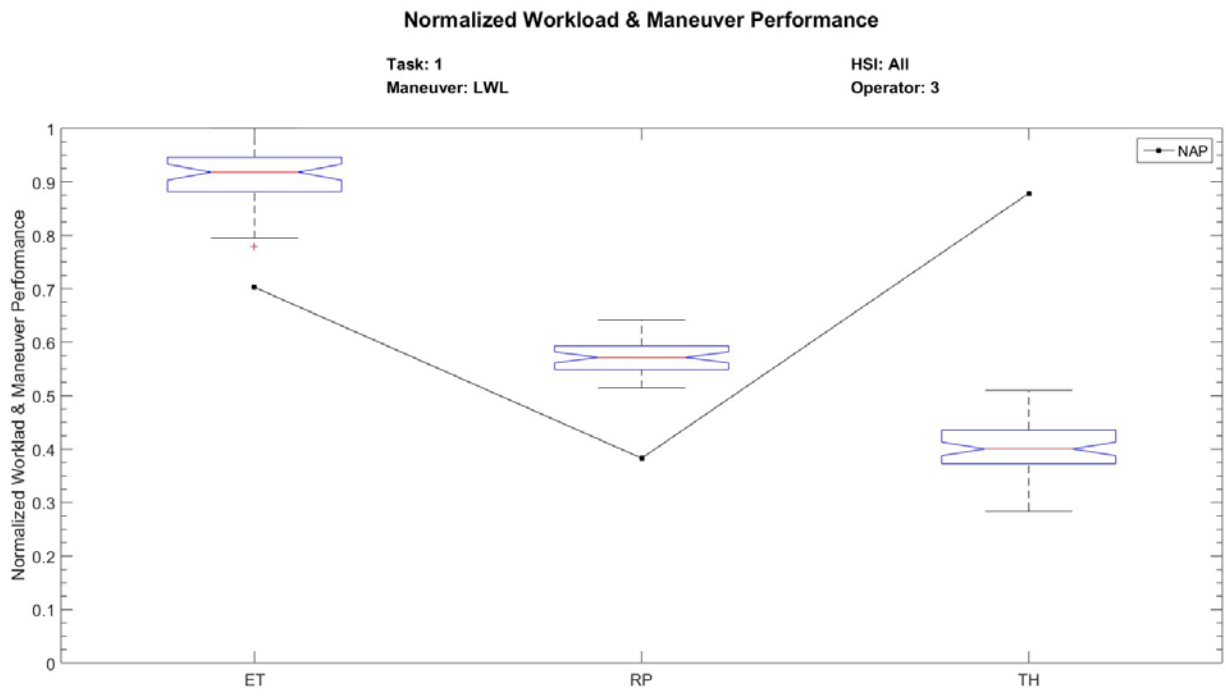
Task 1 Workload

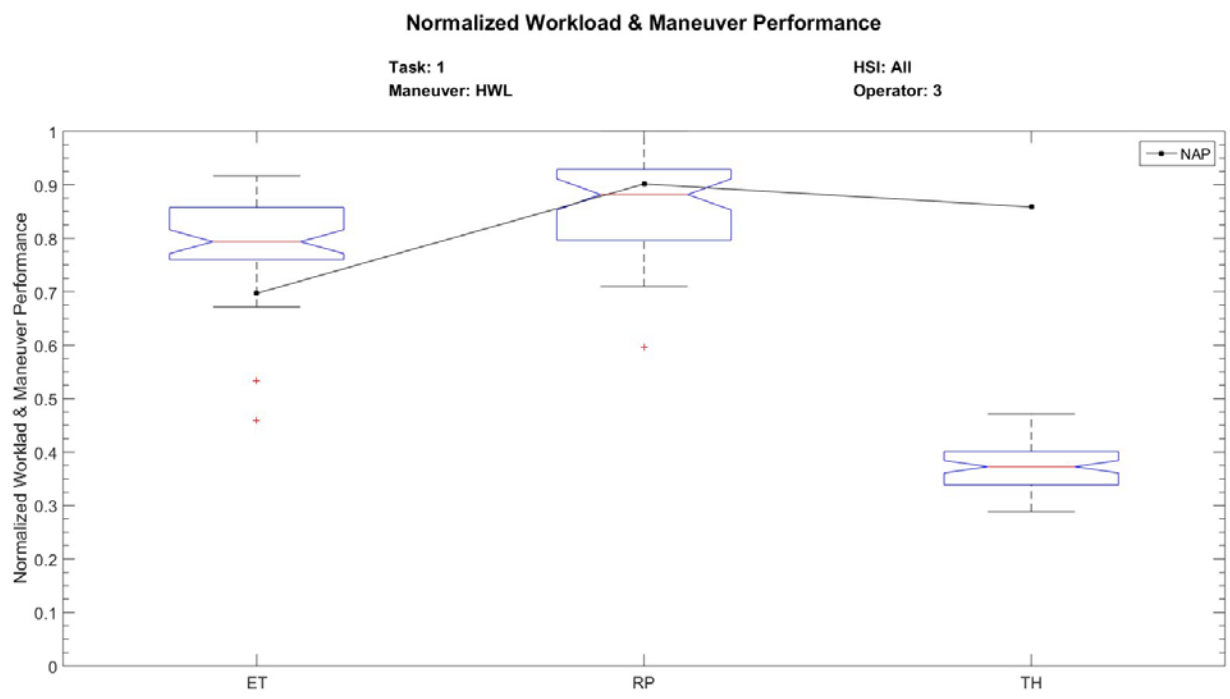


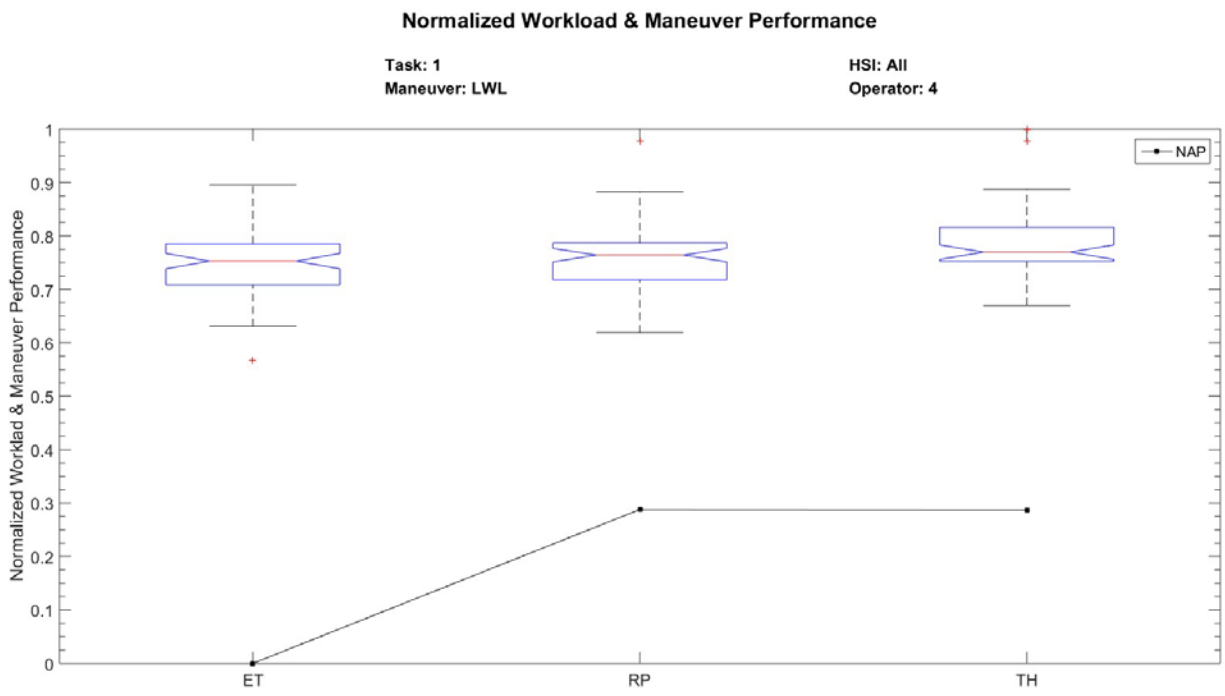


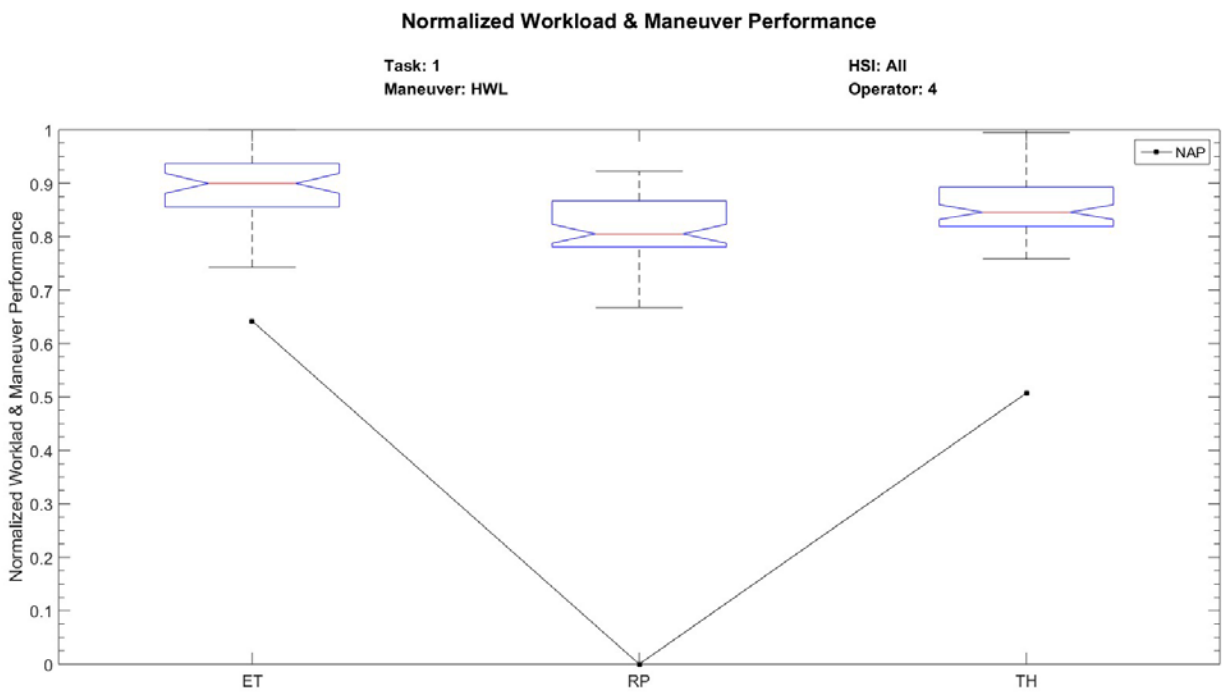

Normalized Workload & Maneuver Performance
Task: 1
Maneuver: LWL
HSI: All
Operator: 2
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

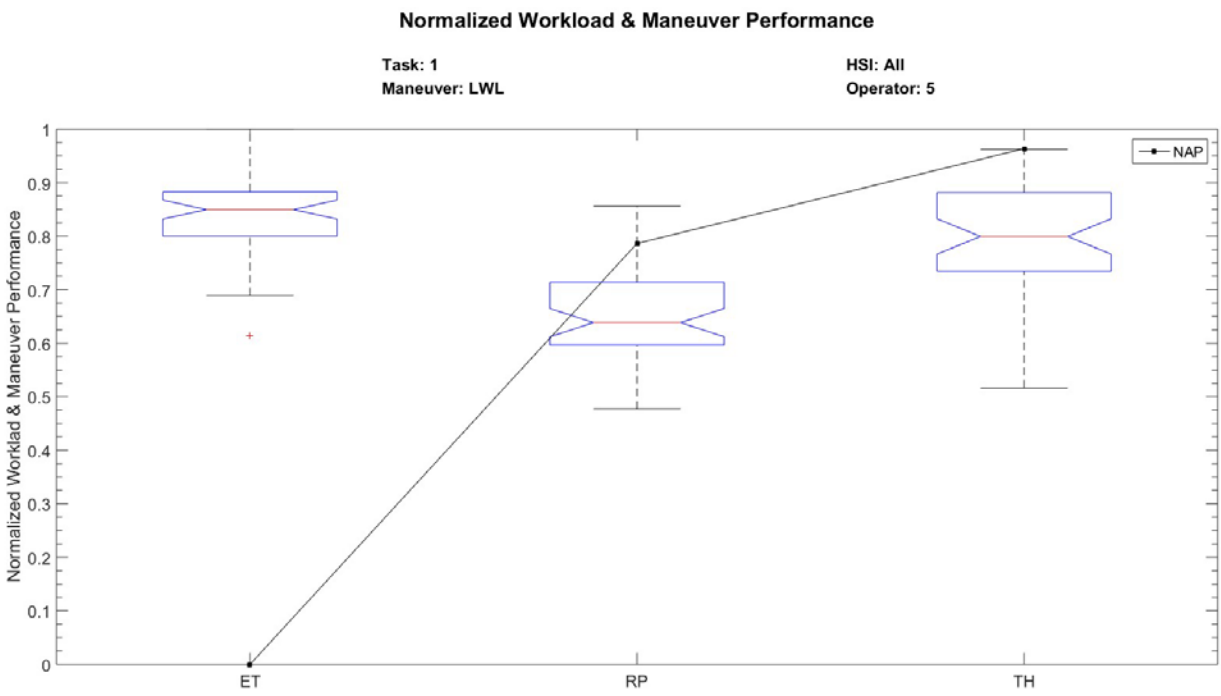

Normalized Workload & Maneuver Performance
Task: 1
Maneuver: HWL
HSI: All
Operator: 2
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

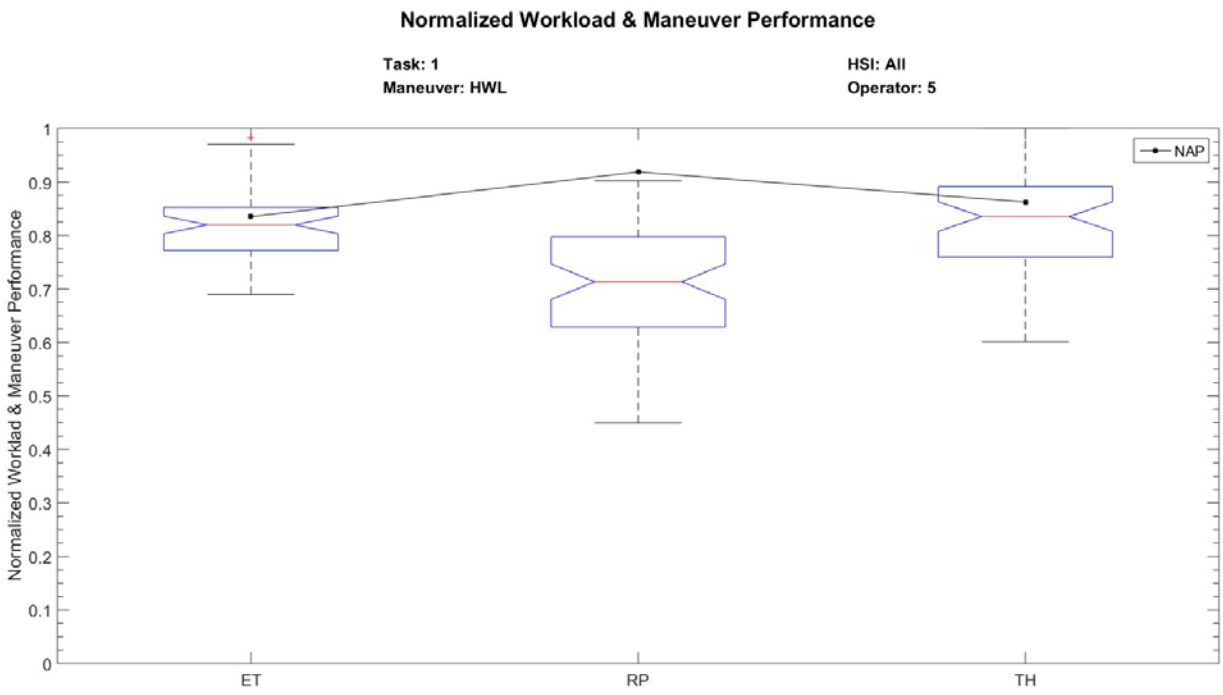

Normalized Workload & Maneuver Performance
Task: 1
Maneuver: LWL
HSI: All
Operator: 3
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

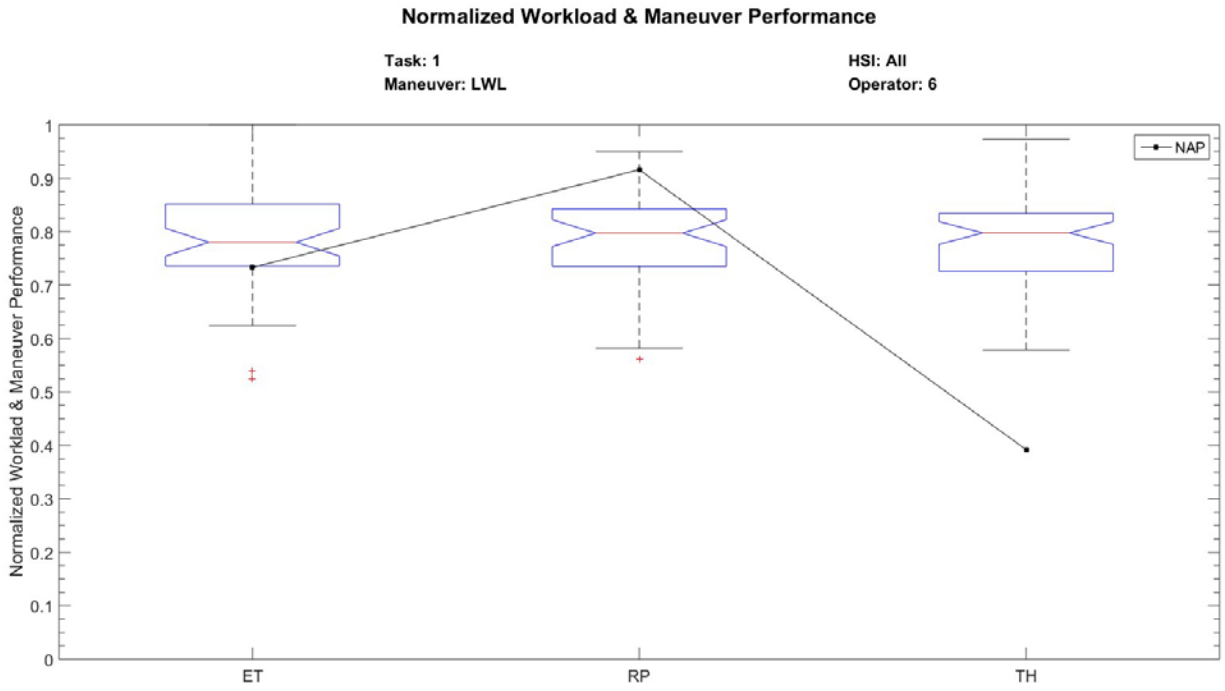

Normalized Workload & Maneuver Performance
Task: 1
Maneuver: HWL
HSI: All
Operator: 3
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

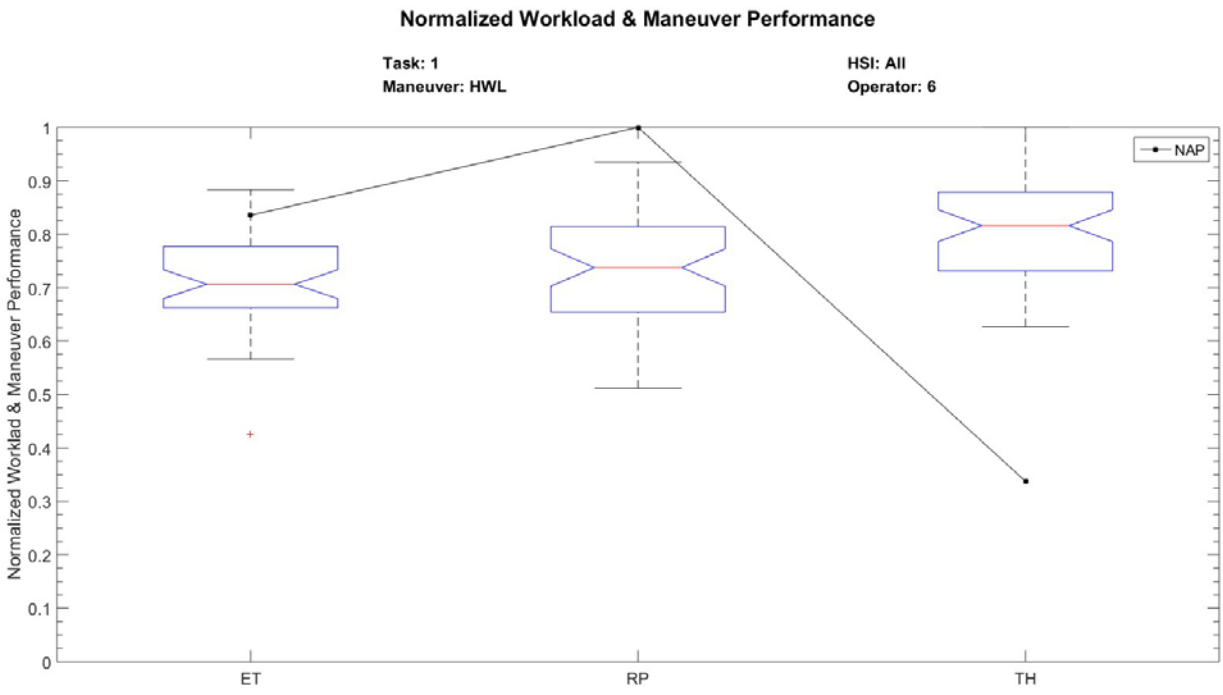

Normalized Workload & Maneuver Performance
Task: 1
Maneuver: LWL
HSI: All
Operator: 4
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

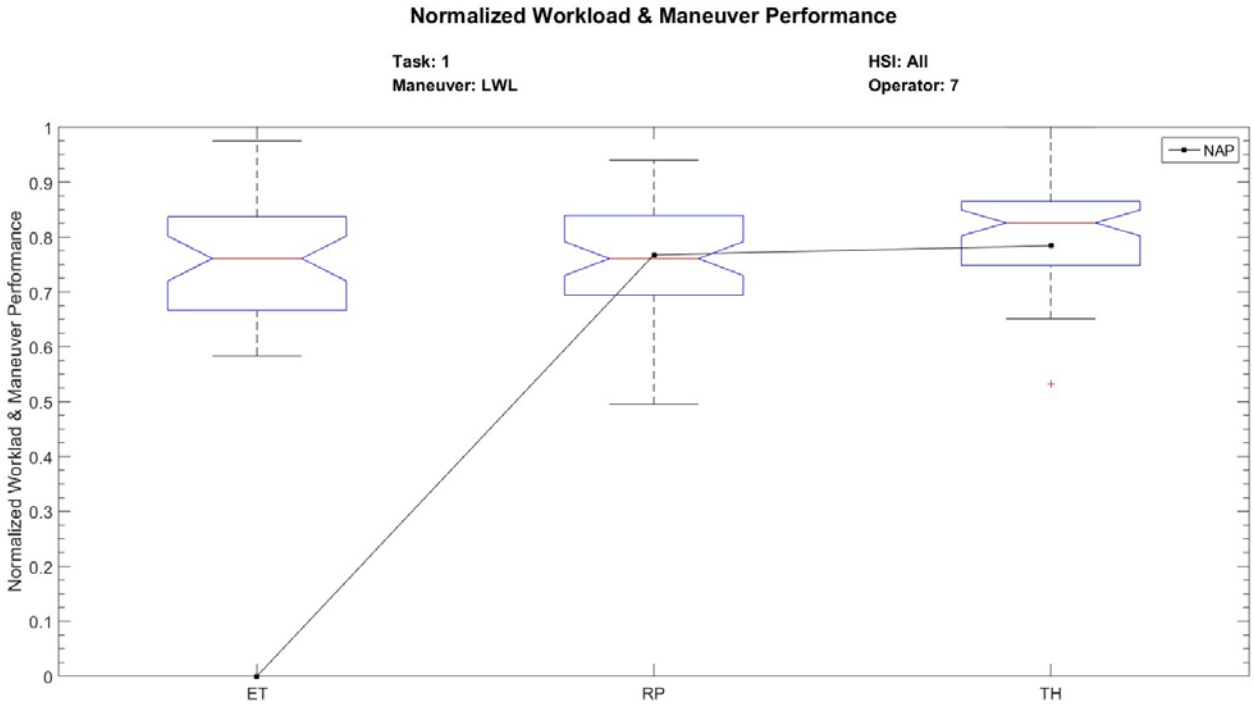

Normalized Workload & Maneuver Performance
Task: 1
Maneuver: HWL
HSI: All
Operator: 4
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

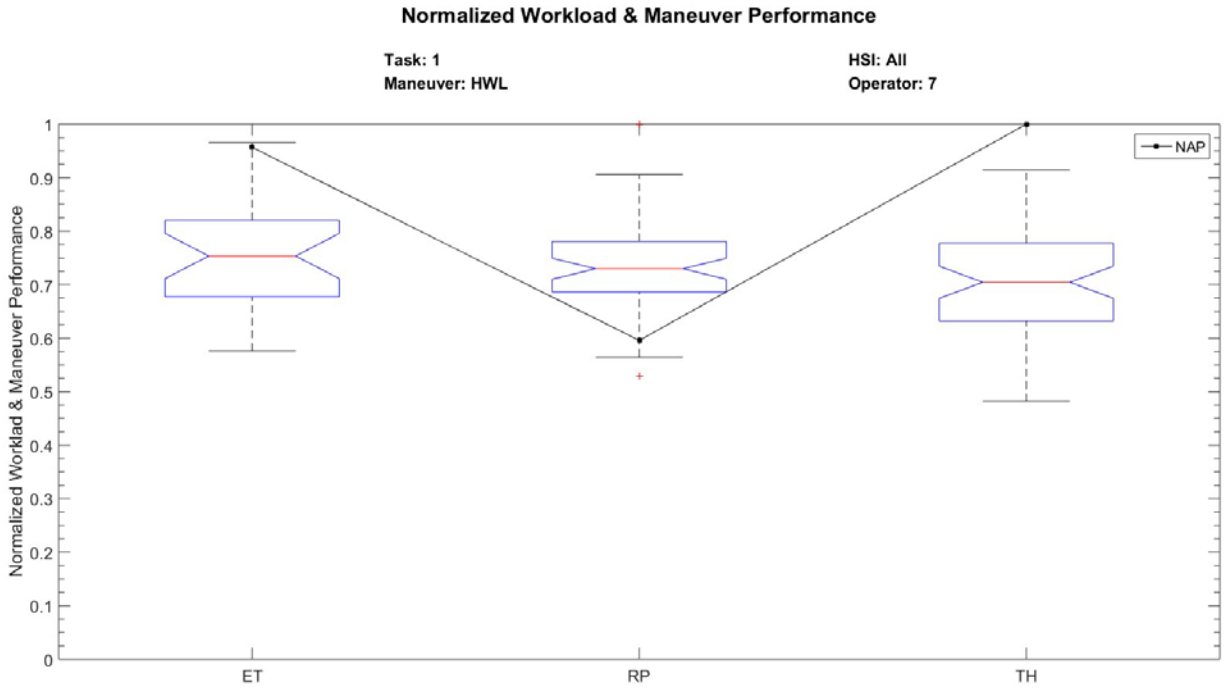

Normalized Workload & Maneuver Performance
Task: 1
Maneuver: LWL
HSI: All
Operator: 5
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH


Normalized Workload & Maneuver Performance
Task: 1
Maneuver: HWL
HSI: All
Operator: 5
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

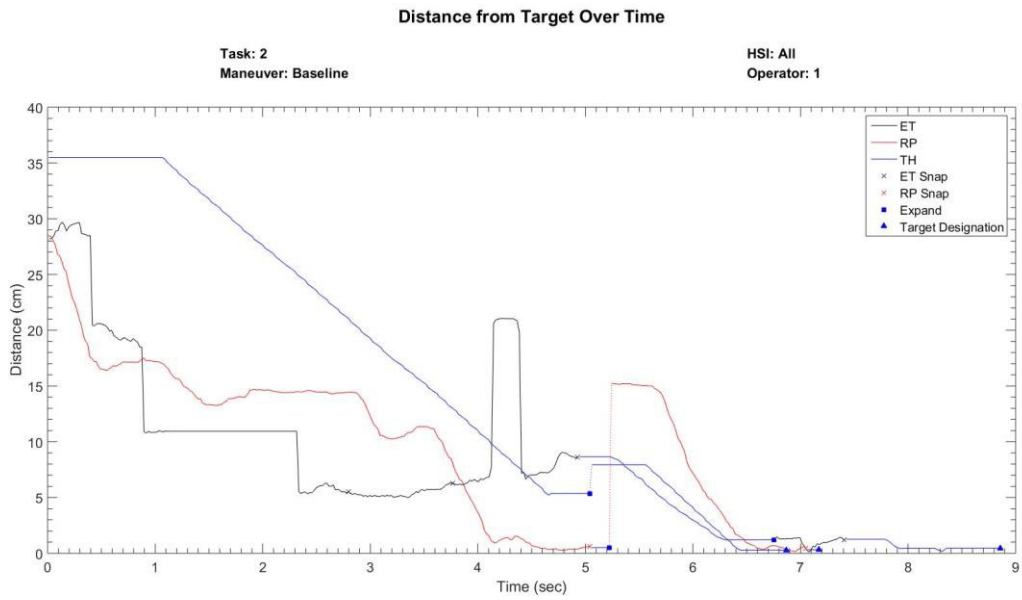

Normalized Workload & Maneuver Performance
Task: 1
Maneuver: LWL
HSI: All
Operator: 6
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH


Normalized Workload & Maneuver Performance
Task: 1
Maneuver: HWL
HSI: All
Operator: 6
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH


Normalized Workload & Maneuver Performance
Task: 1
Maneuver: LWL
HSI: All
Operator: 7
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH


Normalized Workload & Maneuver Performance
Task: 1
Maneuver: HWL
HSI: All
Operator: 7
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

Task 2 HSI Time-Histories, Poor ET Calibration



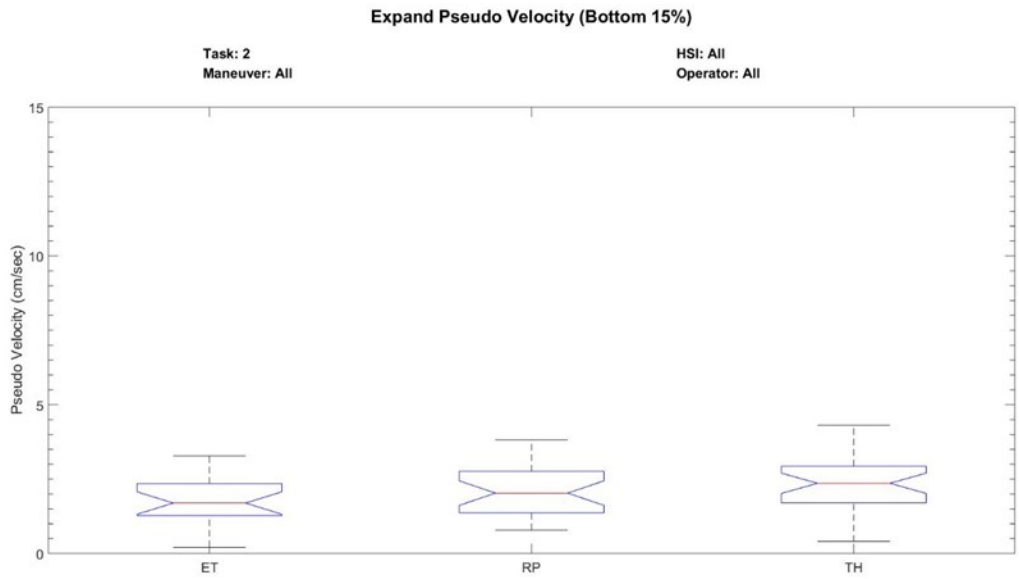
This figure shows the same information described in the similar time-history plot of section 2.2.2.2.1., but with a low quality ET calibration.

For this run, initial movement of the focus-boresight occurred 9 cm away from the track. Again, there was time where the operator must locate the cursor before updating the position with the slew switch, approximately 0.3 sec. Because of the long distance away from the track at first action, the operator was forced to spend two seconds to update the position to get close enough to the track to expand. It is also important to note here, that two other pre-expand snap attempts were made on this run, the first at 2.8 sec and the second at 3.8 sec. For this particular ET run, one of the main limitations of ET was observed. The initial pre-expand track location was near the far left edge of the screen. Because the ET cursor was not visible, the cursor snaps based off where the ET system thinks the operator is looking. For cases where the track was on the edge of the screen and the calibration was off, the cursor would not snap because the systems thought the operator was looking off the screen. For this case, after two snap attempts, the operator looked away from the track and was finally able to get the cursor to snap 9 cm away from the track.

Due to the post-expand predictability, the operator in this case was able to anticipate the future location of the track following the expand. Therefore, just prior to expanding, the operator focused his eyes on the center of the screen. Following the expand operation, the track appeared close to the ET boresight, which allowed the operator to quickly snap, locate the display-fixed cursor, update the position, and designate the track. Again, there was a delay for the operator to locate the cursor and update the position prior to designation. The final designation occurred at 8.9 sec.

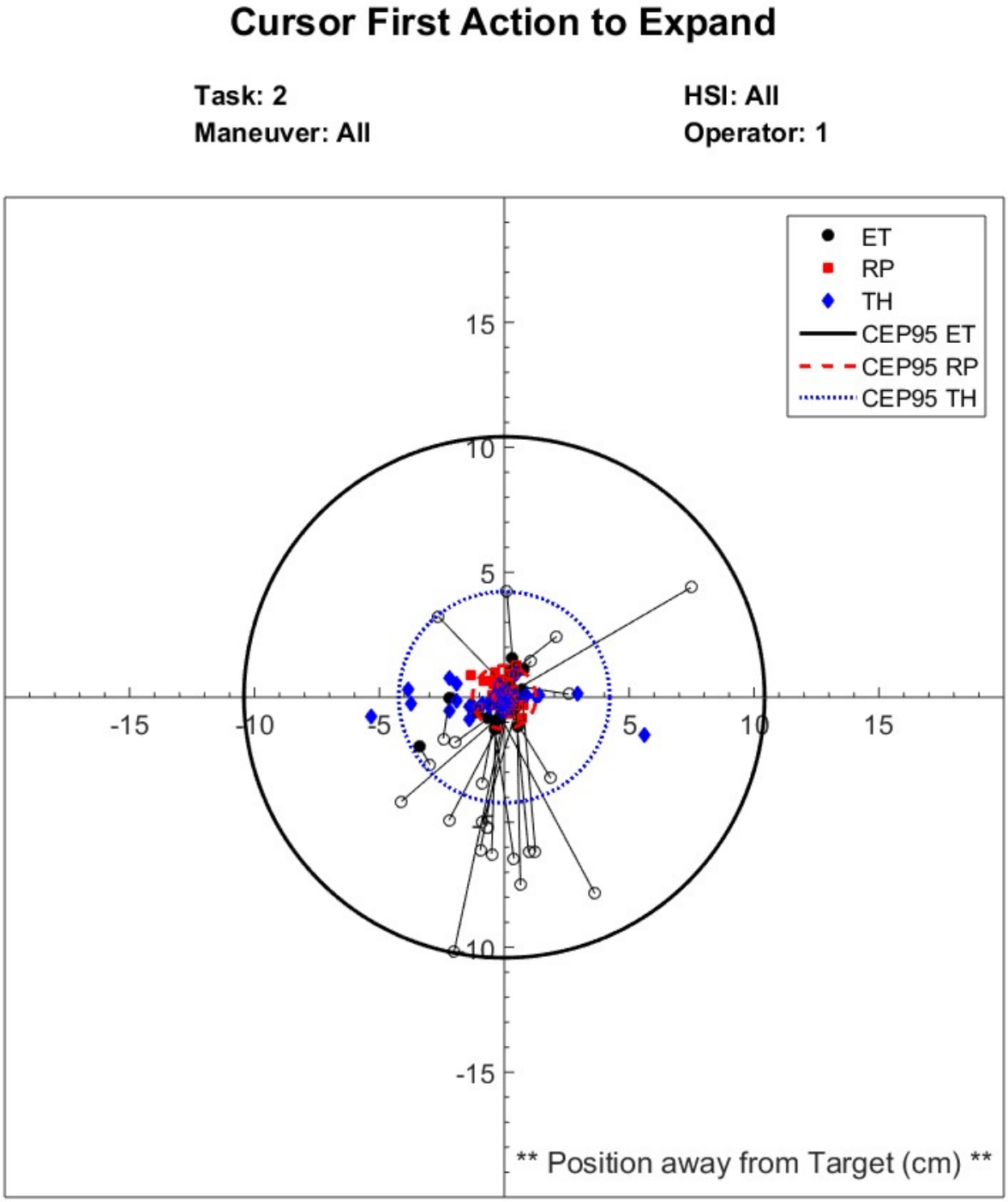
For RP and TH, similar observations can be made as stated above. RP showed quick gross movements towards the track, but time was spent by the operator refining the rhino-boresight before snapping and expanding/designating. TH again showed constant slew rates towards the track both pre- and post-expand. These characteristics of RP and TH held for all operators and all maneuvers; however, the accuracies and times measured for ET depended strongly on the quality of calibration. Since the ET system had no internal mechanism to report calibration quality, pilot comments binned by operator and maneuver were used to capture this covariate.

Task 2 Pseudo-Velocity and Screen Size Analysis



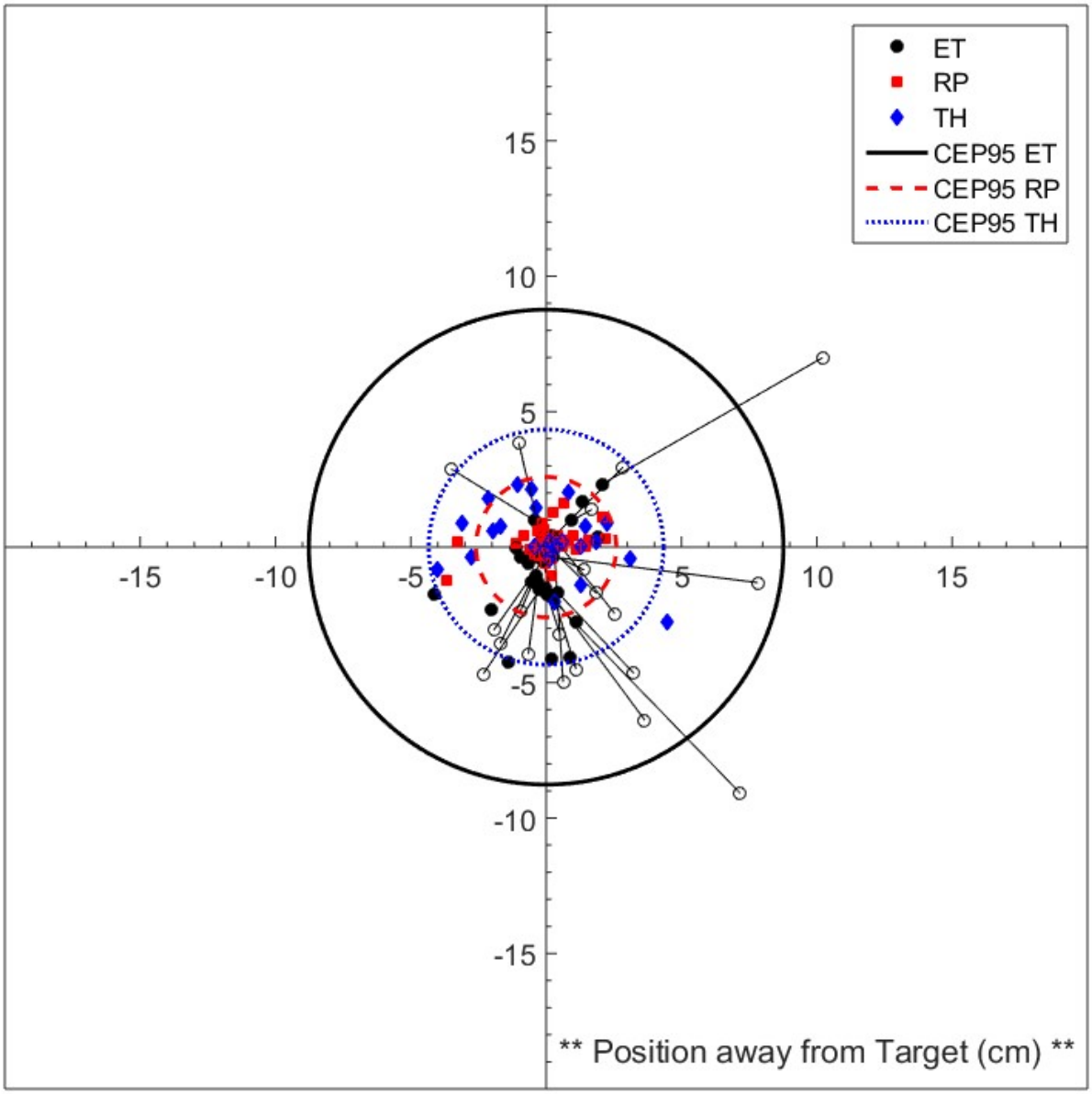
Additional analysis was also performed in order to determine at what initial distance is there a difference in the mean pseudo velocities. The analysis looked an initial distances prior to expand of less than 15 cm, less than 17.5 cm, and greater than 20 cm. For the first case, the mean velocities were 3.569, 3.721, and 3.719 cm/sec for ET, RP, and TH respectfully. A T-test was performed and showed no significant difference in the means. For the second case, the means were 3.905, 4.489, and 3.844 cm/sec for ET, RP, and TH respectfully. The T-test indicated that there was difference in the means between RP and both ET and TH. For the final scenario, the means were 4.808, 8.362, and 6.339 cm/sec for ET, RP, and TH respectfully. The T-test indicated difference between all the mean. The results showed that when the track was less than 15 cm away at initiation, it did not matter which HSI method was used as all three displayed similar results. When the initial distance is expanded to 17.5 cm away from the track, the results showed that RP began to show favorable results over both ET and TH. When the distance at initiation was greater than 20 cm, the results showed that both ET and RP had improved pseudo velocities as compared to TH, although RP still displayed a higher average velocity over ET.

Task 2 Accuracy at Expand



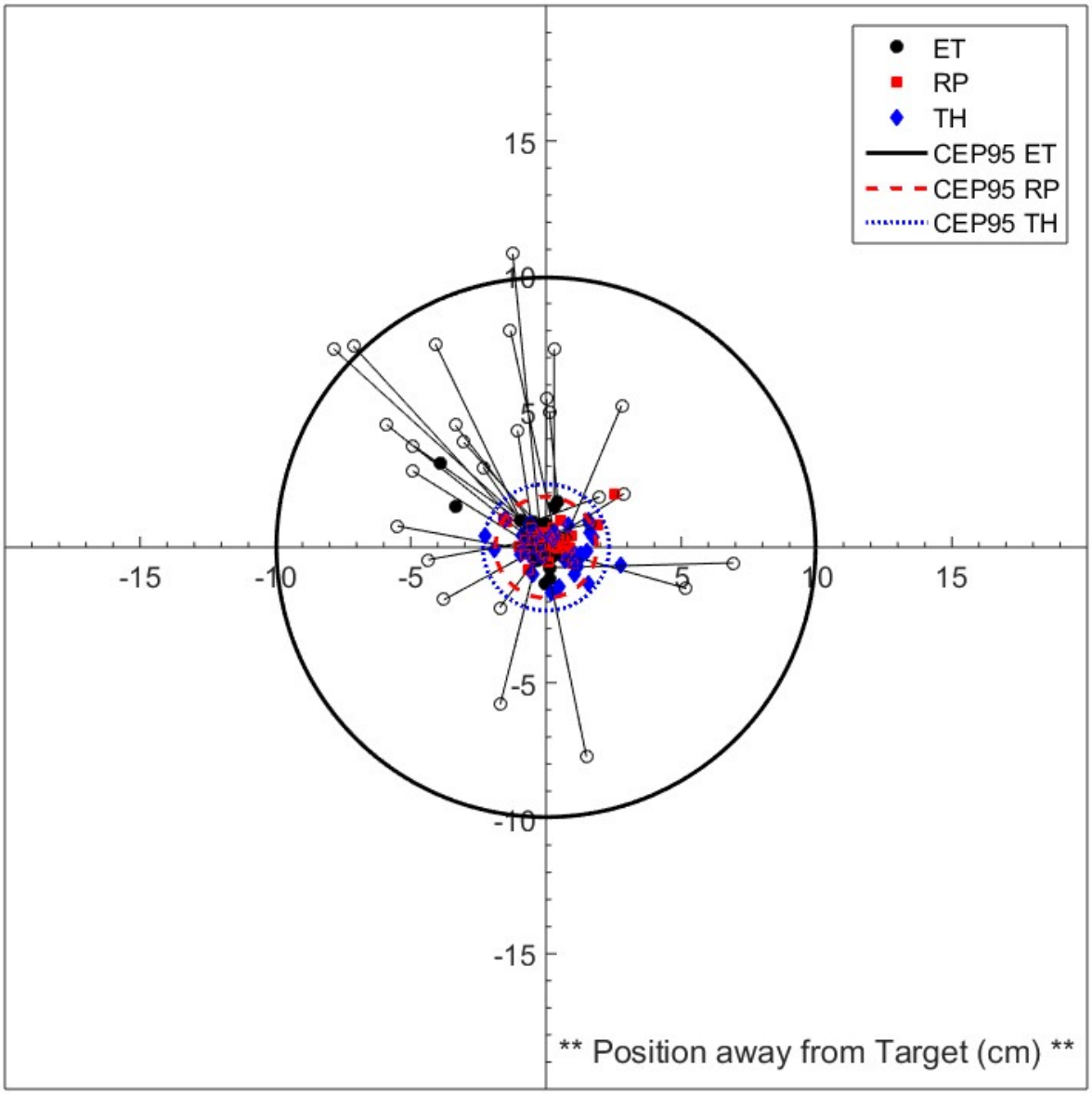
Cursor First Action to Expand

Task: 2 **HSI: All**
Maneuver: All **Operator: 2**



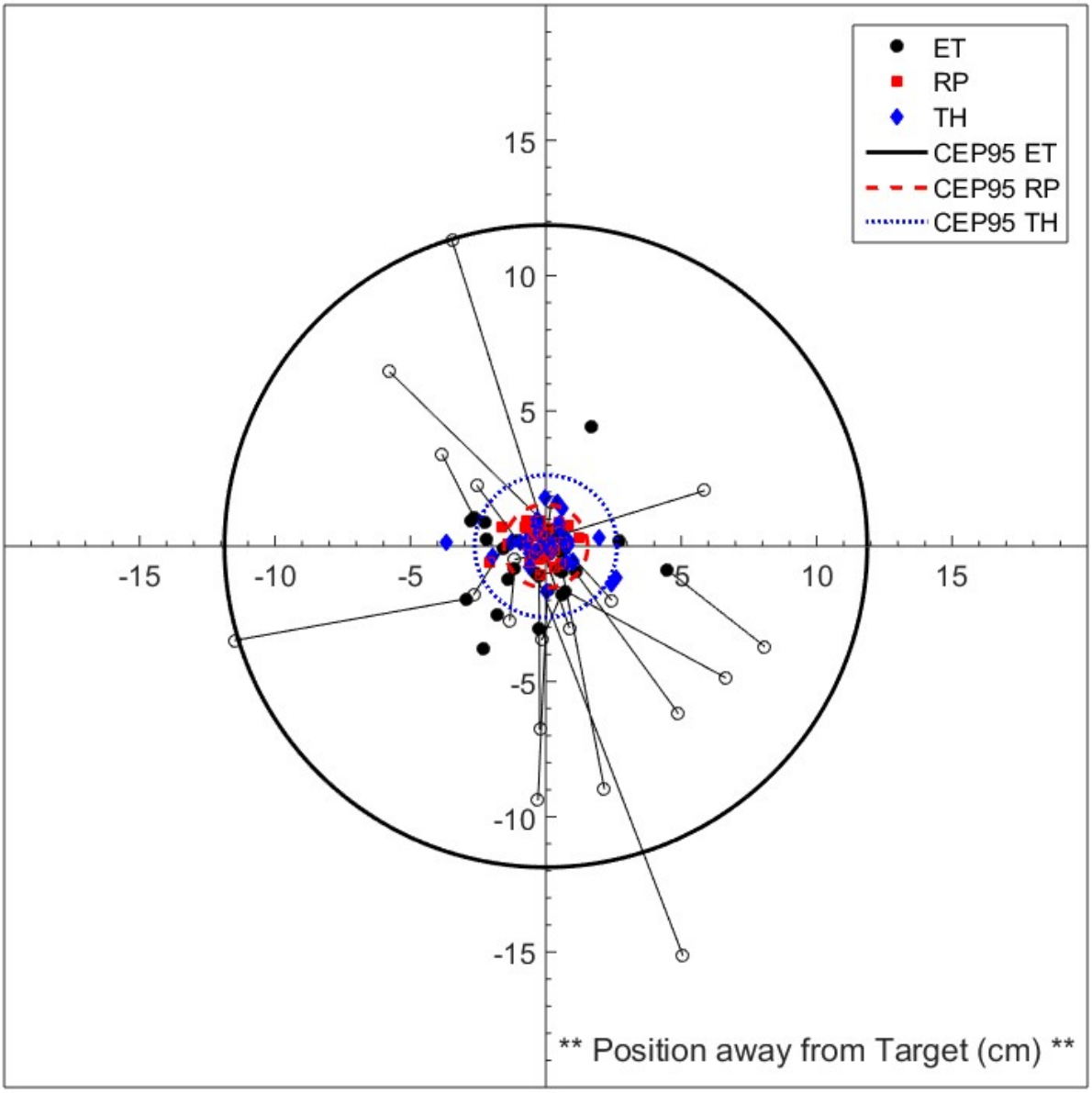
Cursor First Action to Expand

Task: 2 **HSI: All**
Maneuver: All **Operator: 3**



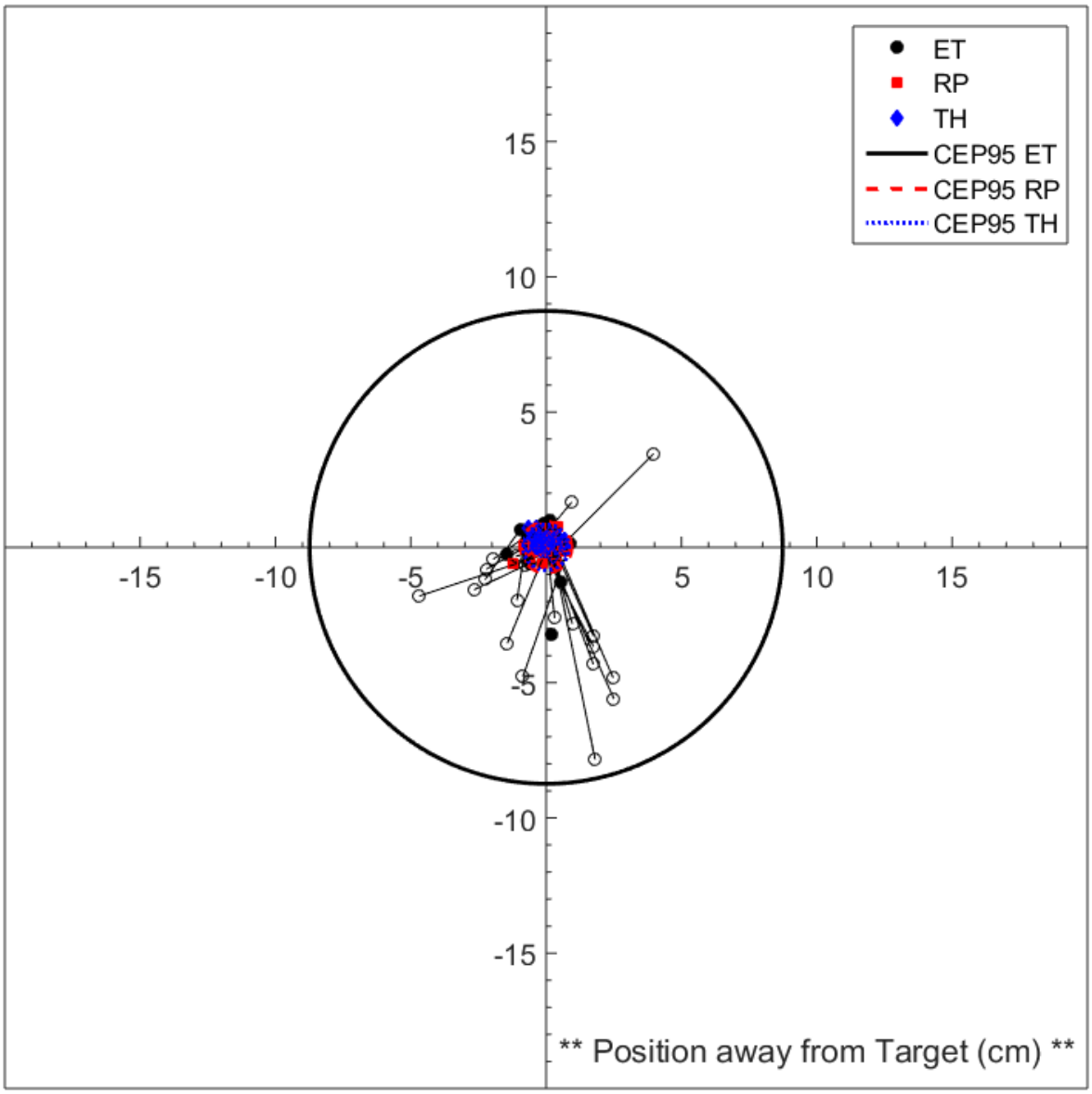
Cursor First Action to Expand

Task: 2 **HSI: All**
Maneuver: All **Operator: 4**



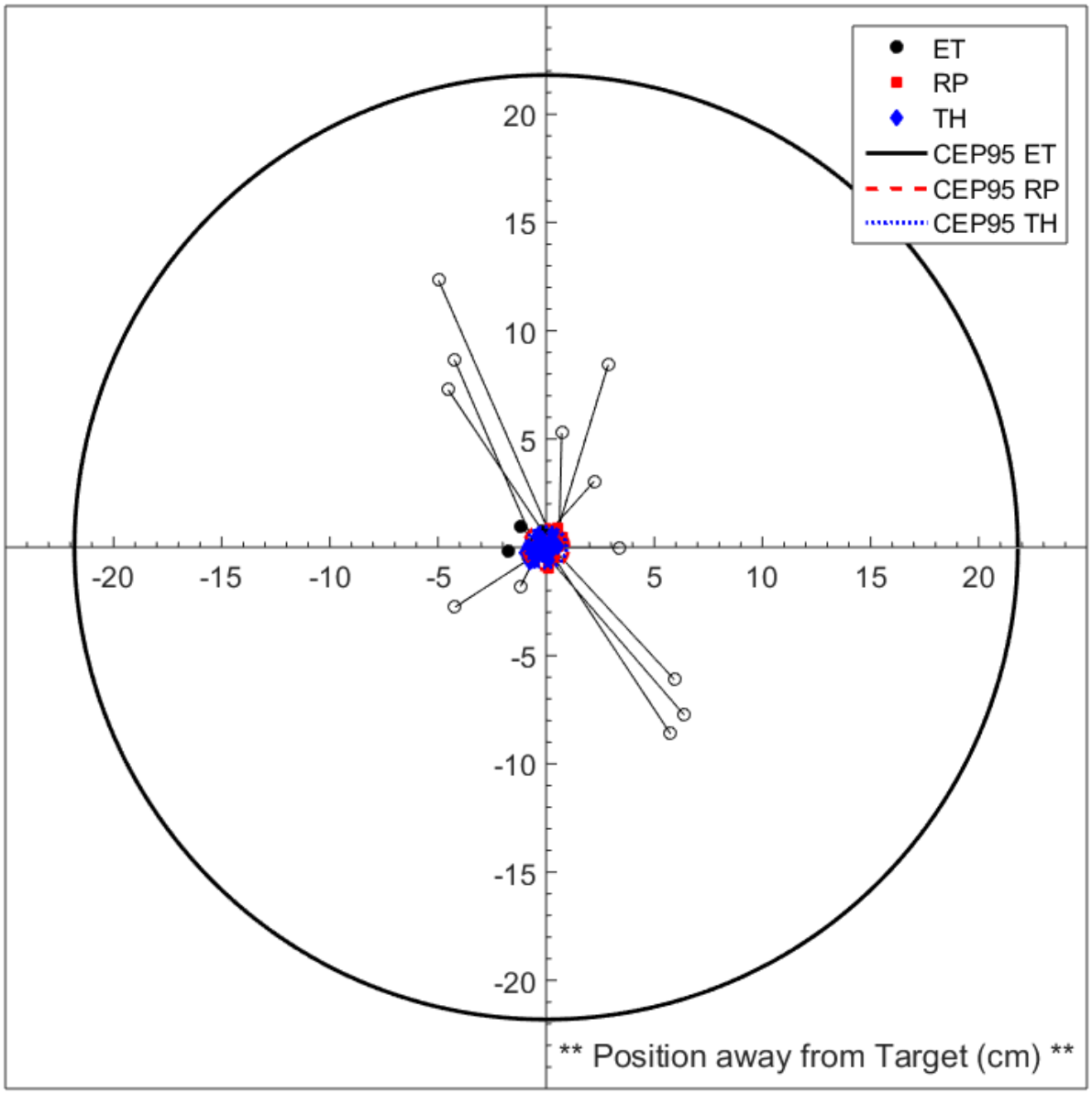
Cursor First Action to Expand

Task: 2 **HSI: All**
Maneuver: All **Operator: 5**



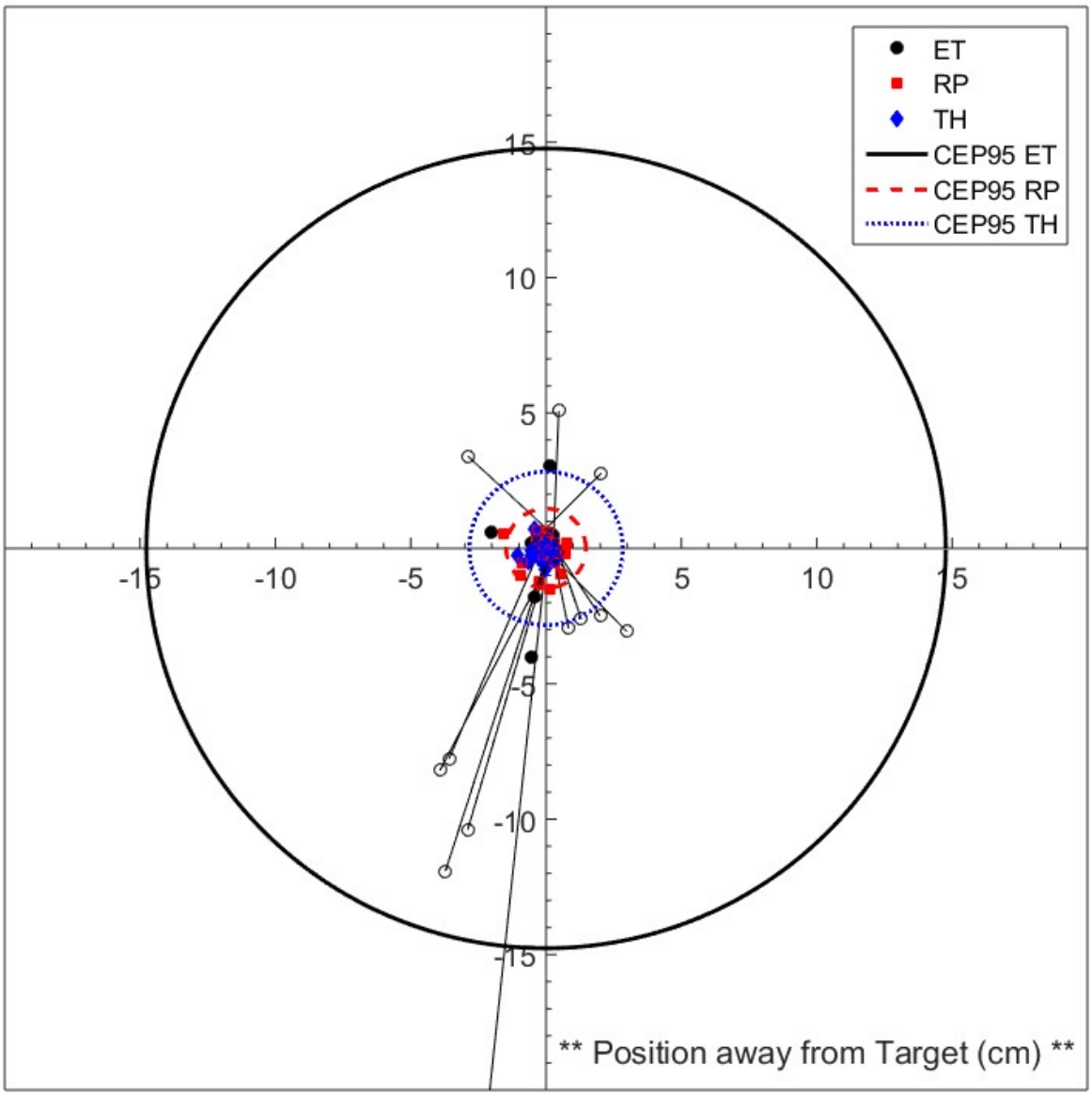
Cursor First Action to Expand

Task: 2 **HSI: All**
Maneuver: All **Operator: 6**



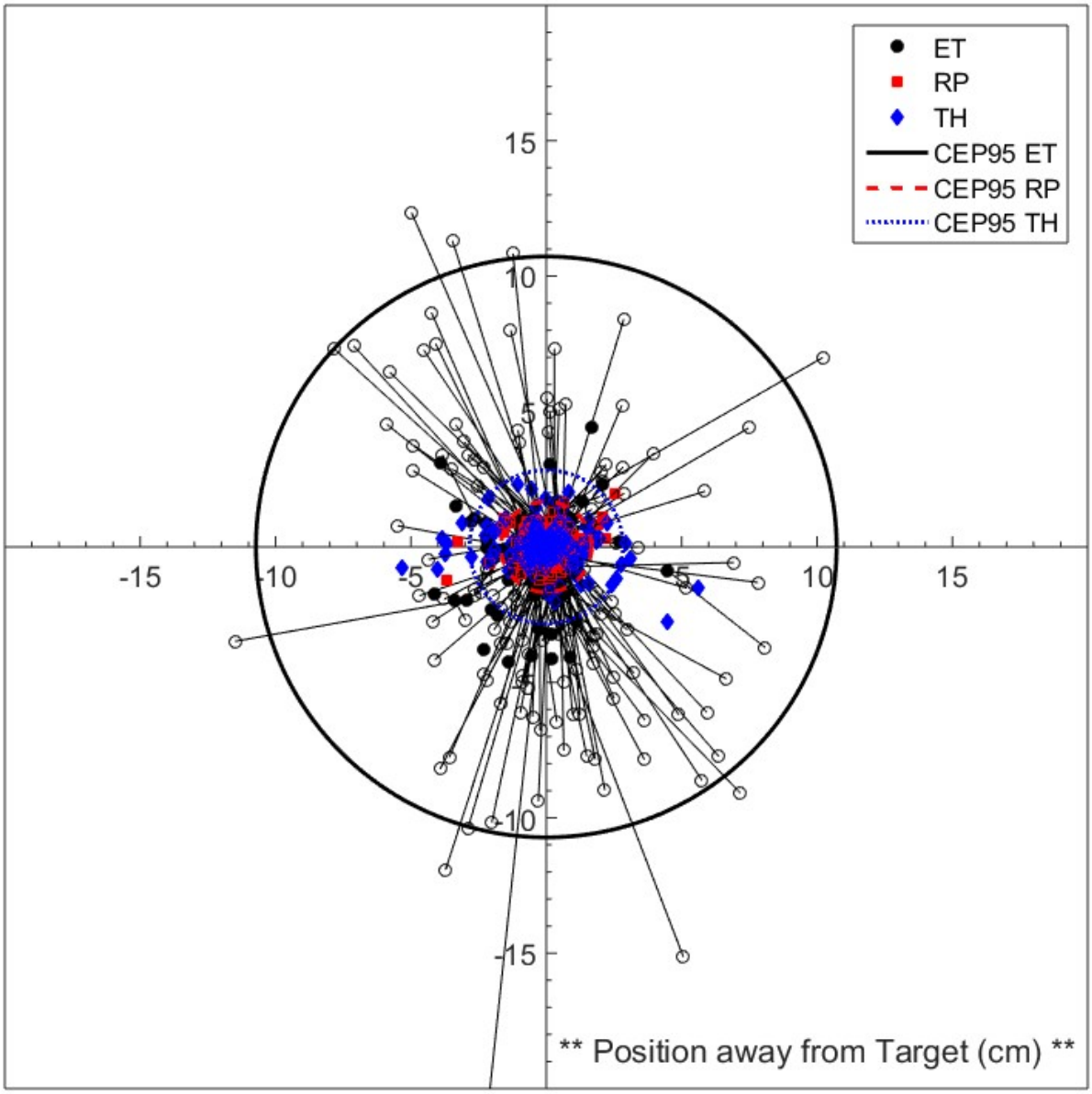
Cursor First Action to Expand

Task: 2 **HSI: All**
Maneuver: All **Operator: 7**



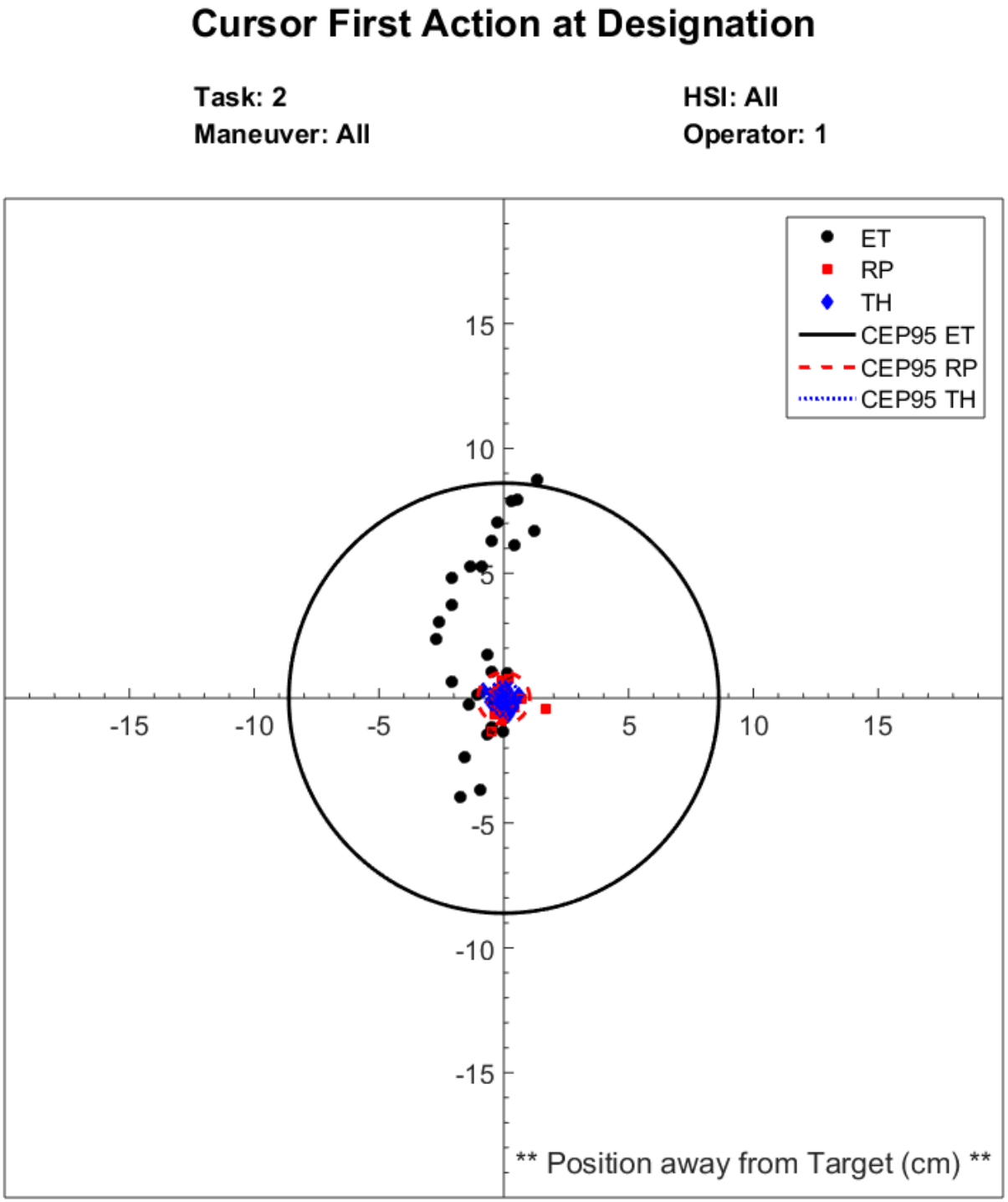
Cursor First Action to Expand

Task: 2 **HSI: All**
Maneuver: All **Operator: All**



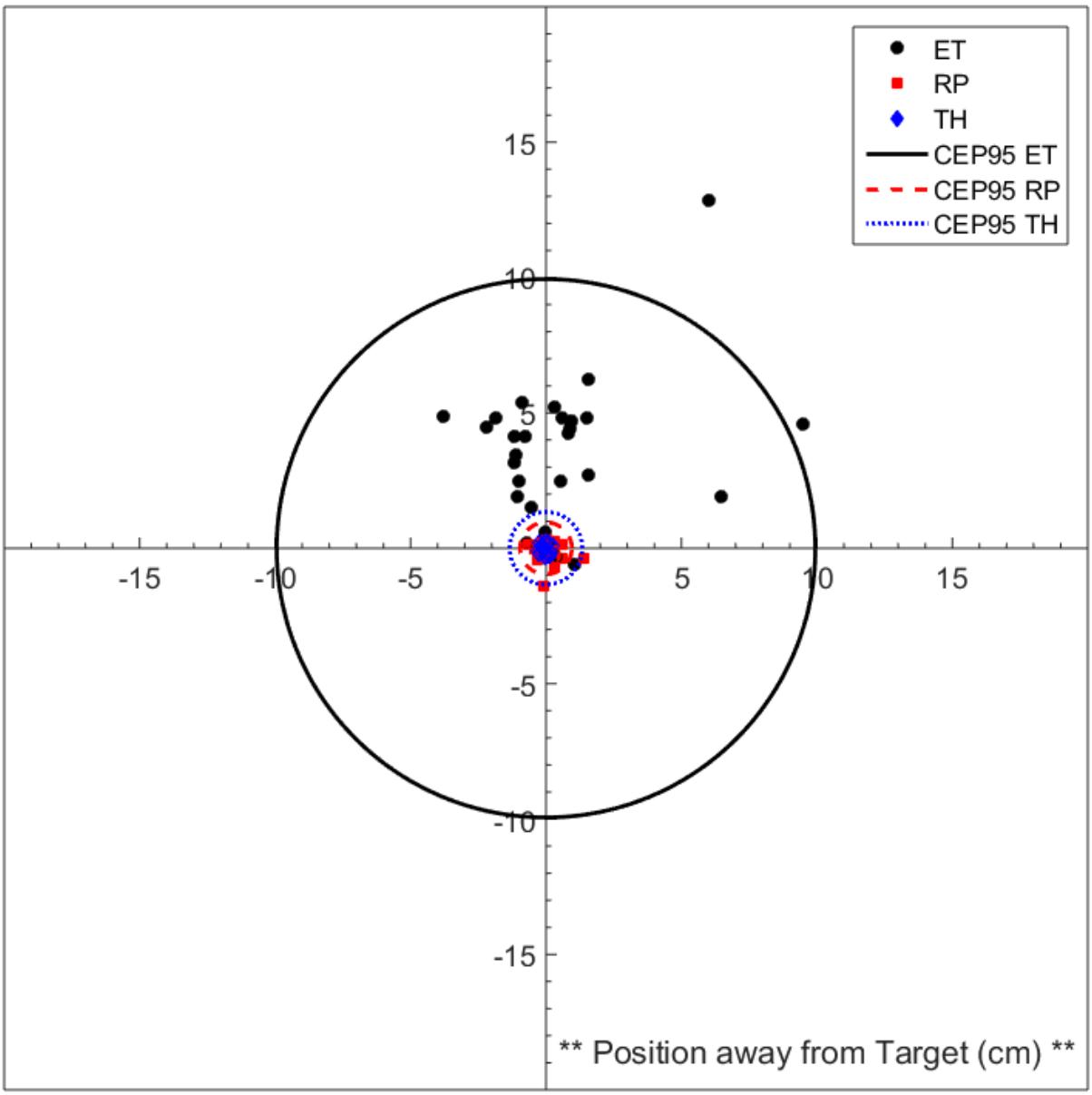
Op	Man	ET_R	ET_Y	ET_RMS	**ET_CEP_95**	RP_R	RP_Y	RP_RMS	**RP_CEP_95**	T_R	T_Y	T_RMS	**T_CEP_95**
1	All	0.6768	1.805	5.773	**10.421**	0.639	1.844	0.6889	**1.270**	0.433	2.04	2.070	**4.223**
2	All	0.7319	1.748	5.015	**8.766**	0.558	1.927	1.3443	**2.590**	0.571	1.912	2.269	**4.339**
3	All	0.8091	1.633	6.099	**9.960**	0.677	1.805	1.0351	**1.868**	0.779	1.682	1.387	**2.333**
4	All	0.6559	1.826	6.499	**11.866**	0.704	1.777	0.8672	**1.541**	0.609	1.874	1.403	**2.628**
5	All	0.651	1.831	4.773	**8.738**	0.663	1.818	0.5126	**0.932**	0.738	1.742	0.473	**0.824**
6	All	0.882	1.518	14.365	**21.806**	0.663	1.818	0.5678	**1.032**	0.781	1.679	0.560	**0.940**
7	All	0.6156	1.867	7.909	**14.765**	0.633	1.849	0.7953	**1.471**	0.706	1.775	0.514	**0.913**
All	All	0.679	1.802	5.951	**10.723**	0.584	1.899	0.881	**1.673**	0.476	2.002	1.415	**2.833**

Task 2 – Accuracy at Designation



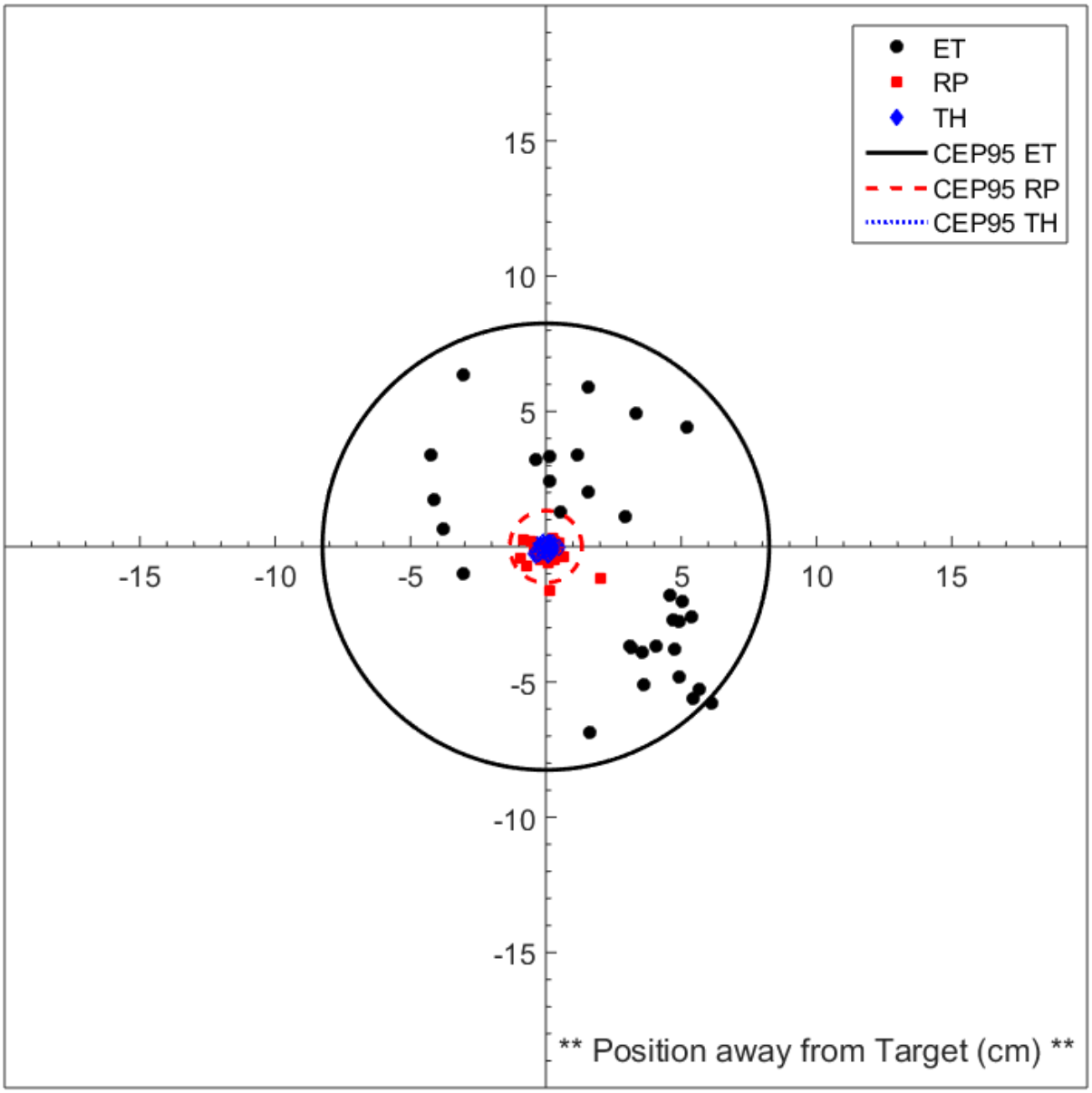
Cursor First Action at Designation

Task: 2 **HSI: All**
Maneuver: All **Operator: 2**



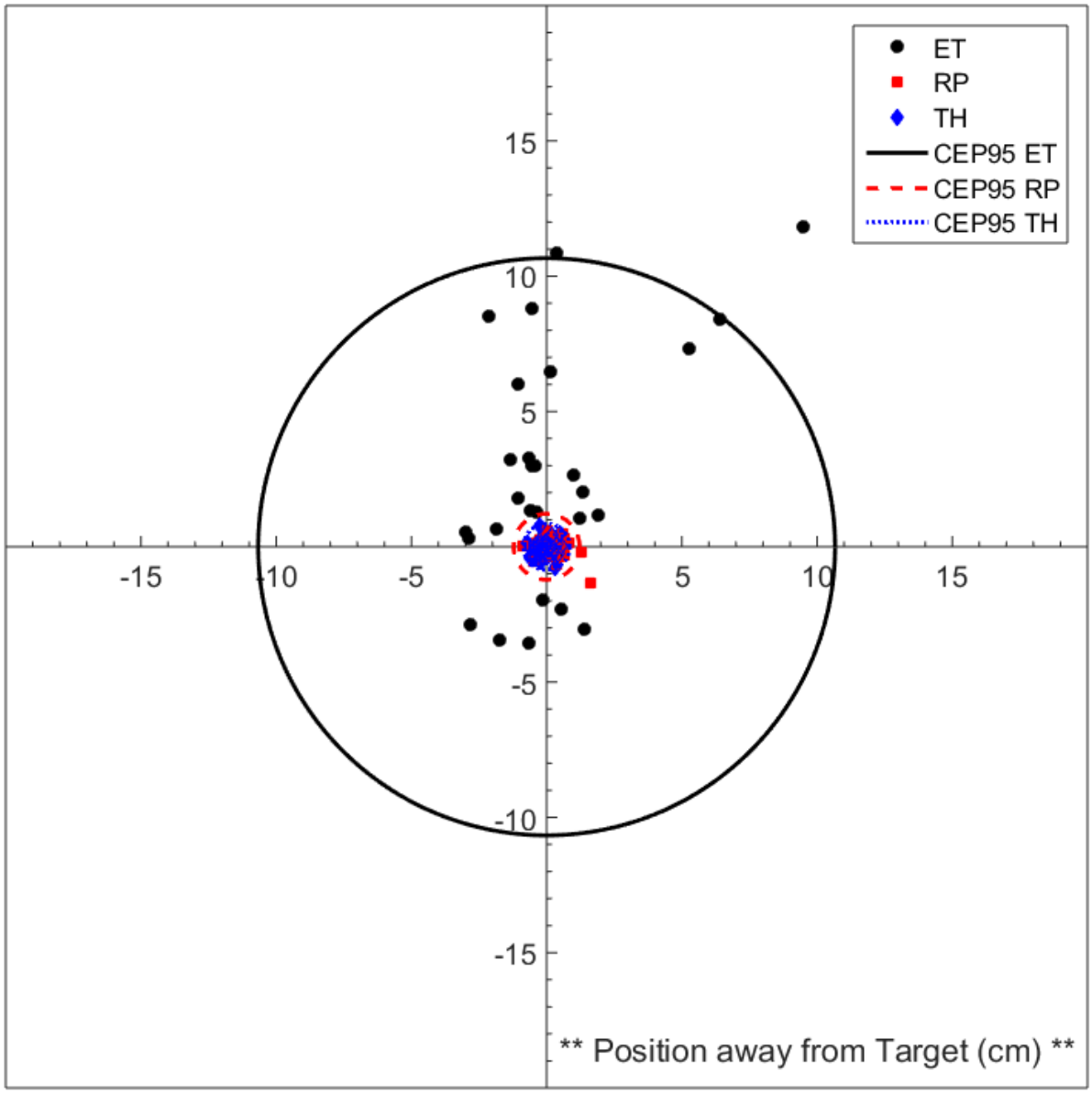
Cursor First Action at Designation

Task: 2 **HSI: All**
Maneuver: All **Operator: 3**



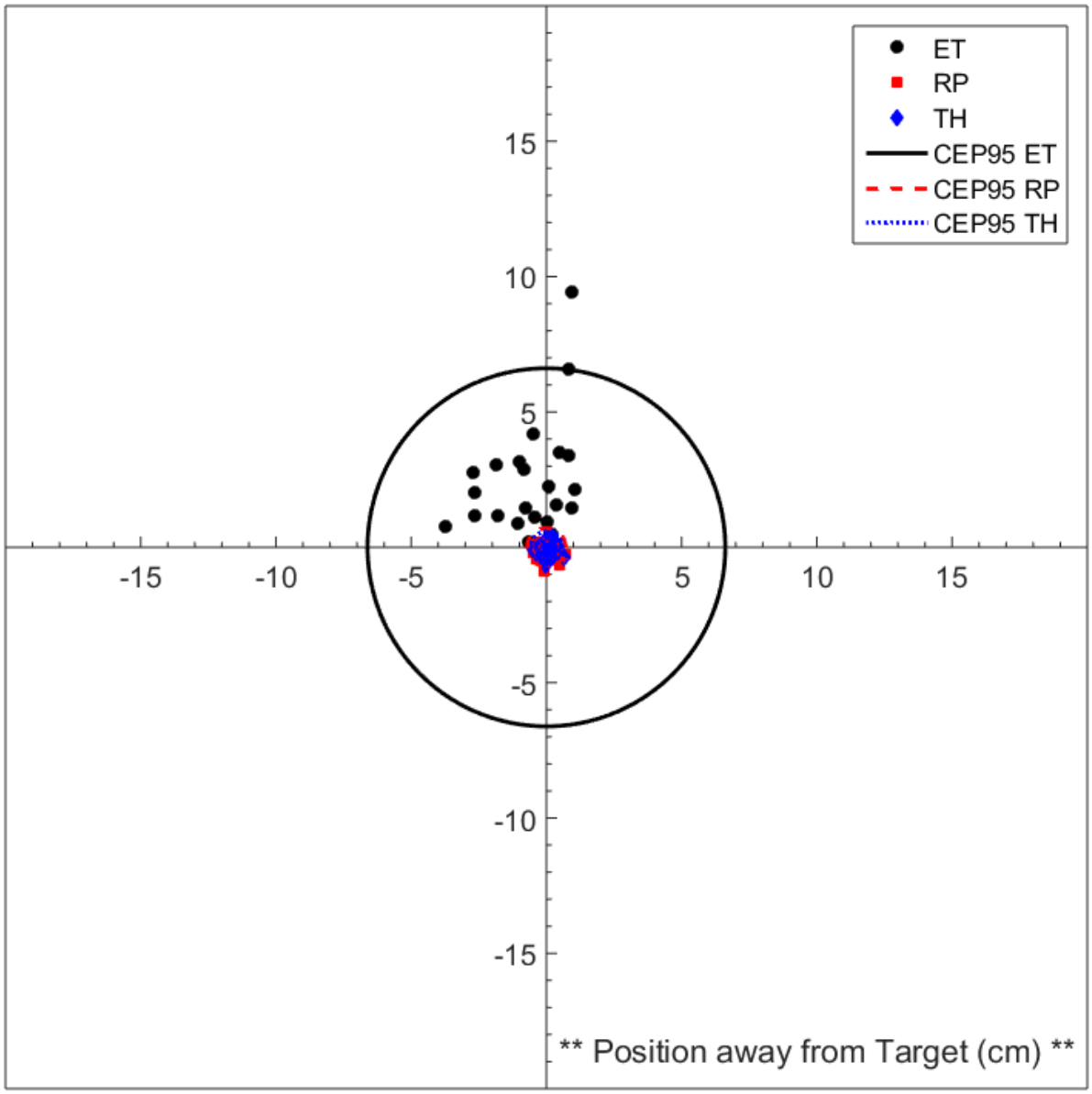
Cursor First Action at Designation

Task: 2 **HSI: All**
Maneuver: All **Operator: 4**



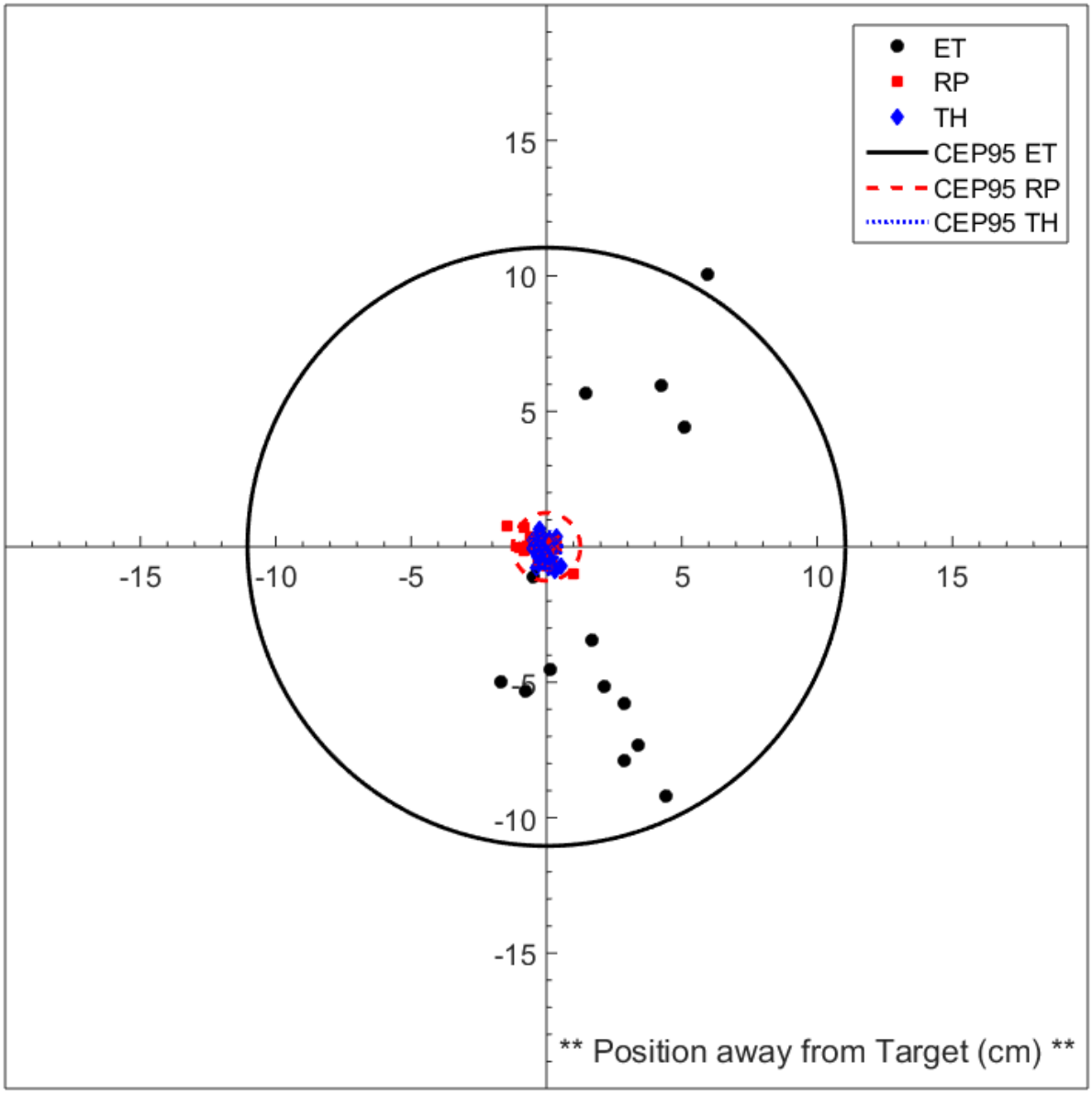
Cursor First Action at Designation

Task: 2 **HSI: All**
Maneuver: All **Operator: 5**



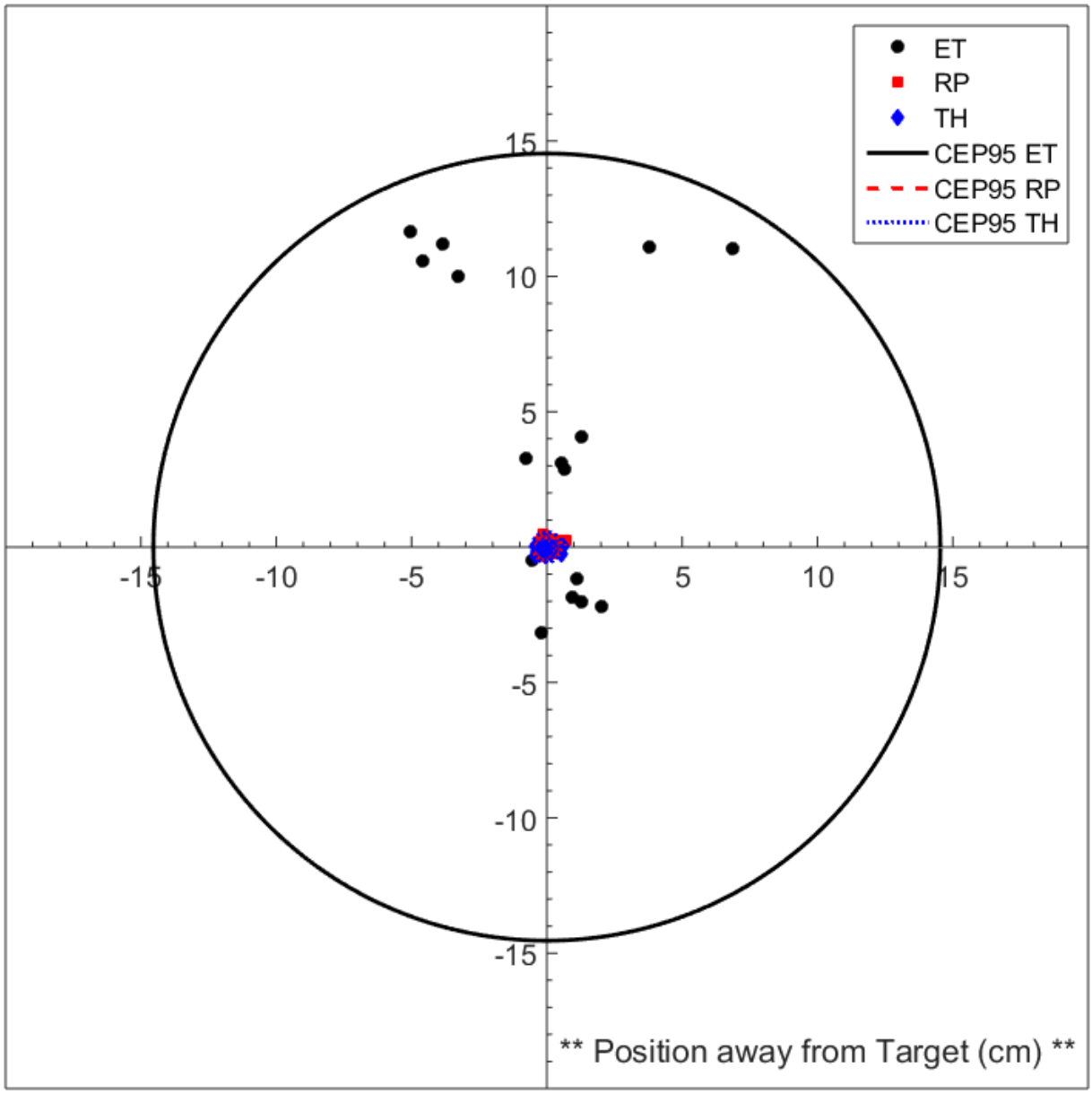
Cursor First Action at Designation

Task: 2 **HSI: All**
Maneuver: All **Operator: 6**



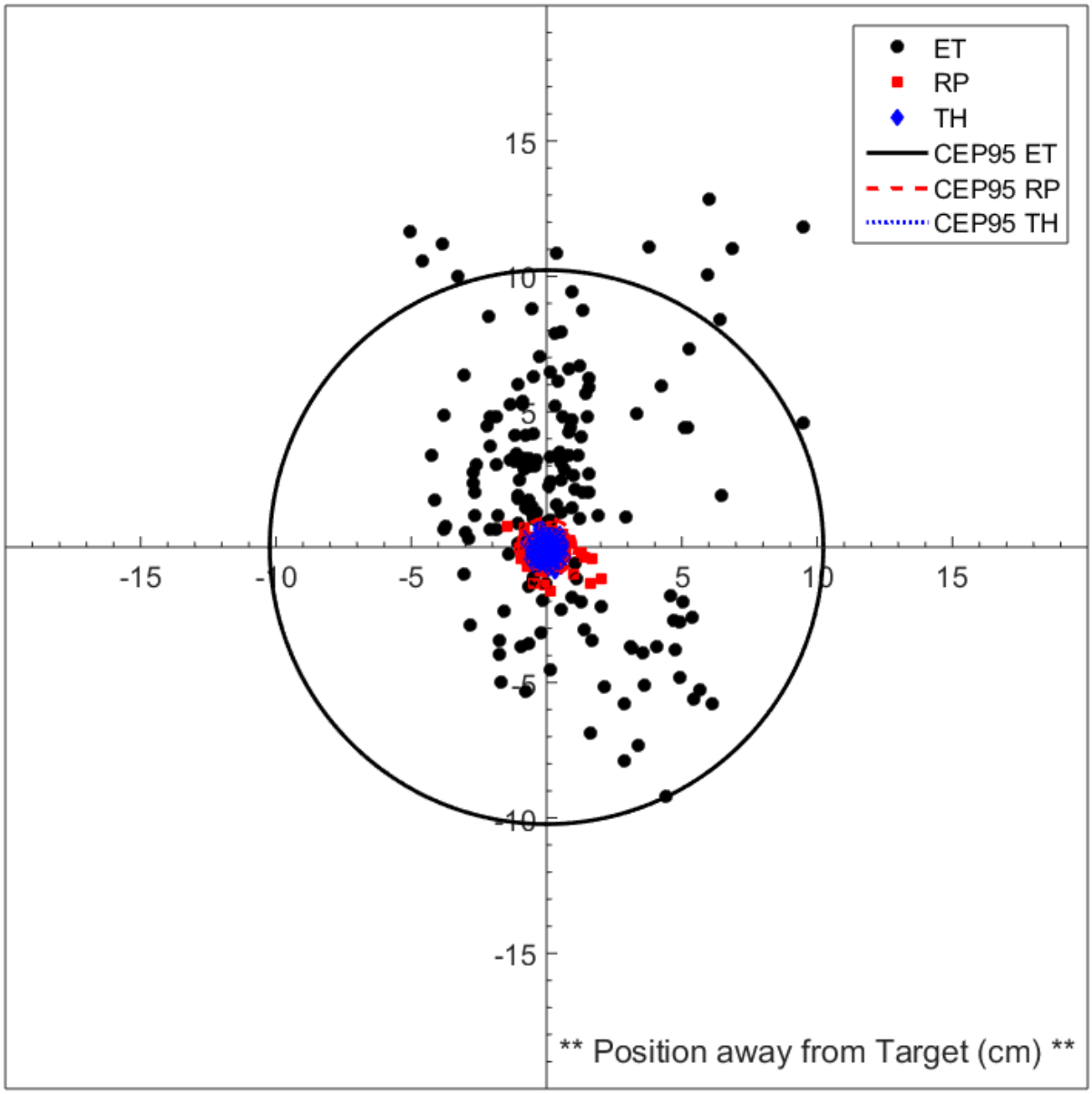
Cursor First Action at Designation

Task: 2 **HSI: All**
Maneuver: All **Operator: 7**



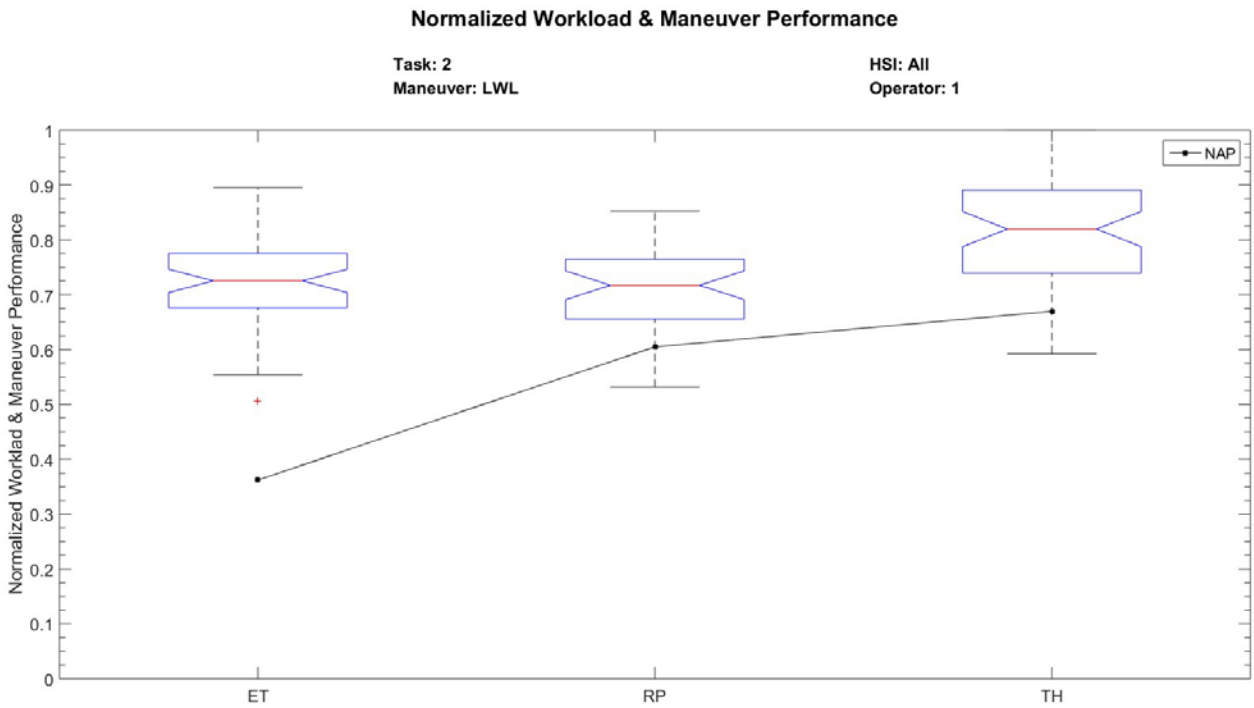
Cursor First Action at Designation

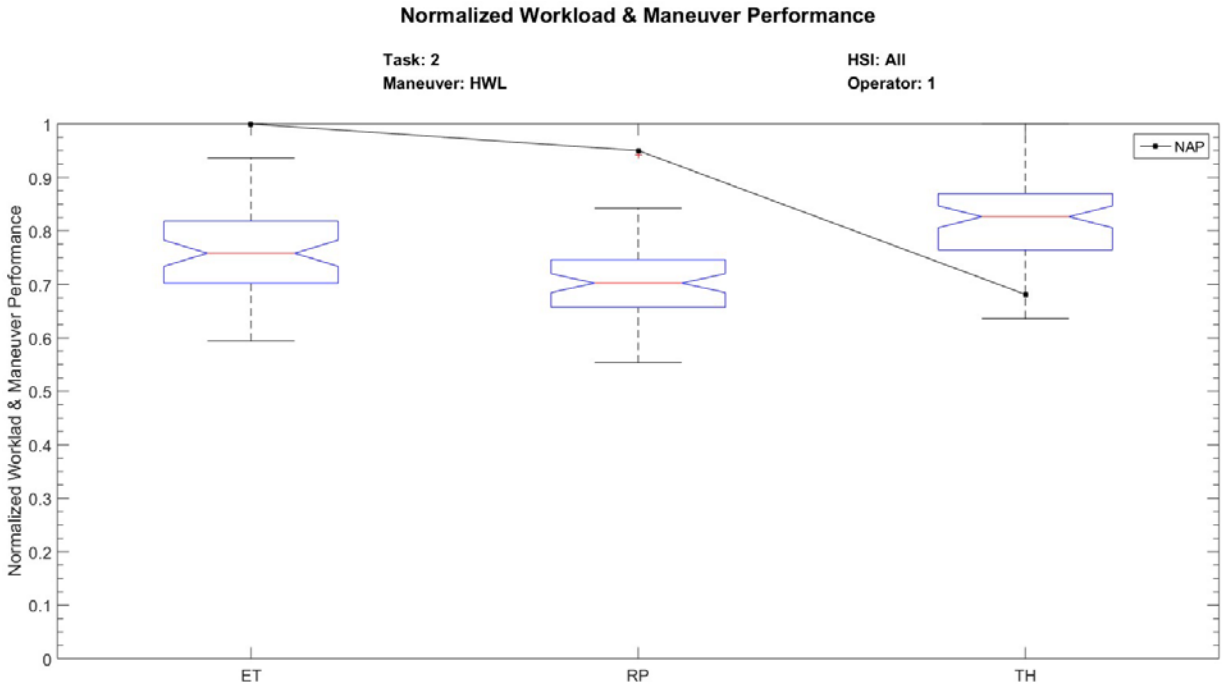
Task: 2 **HSI: All**
Maneuver: All **Operator: All**

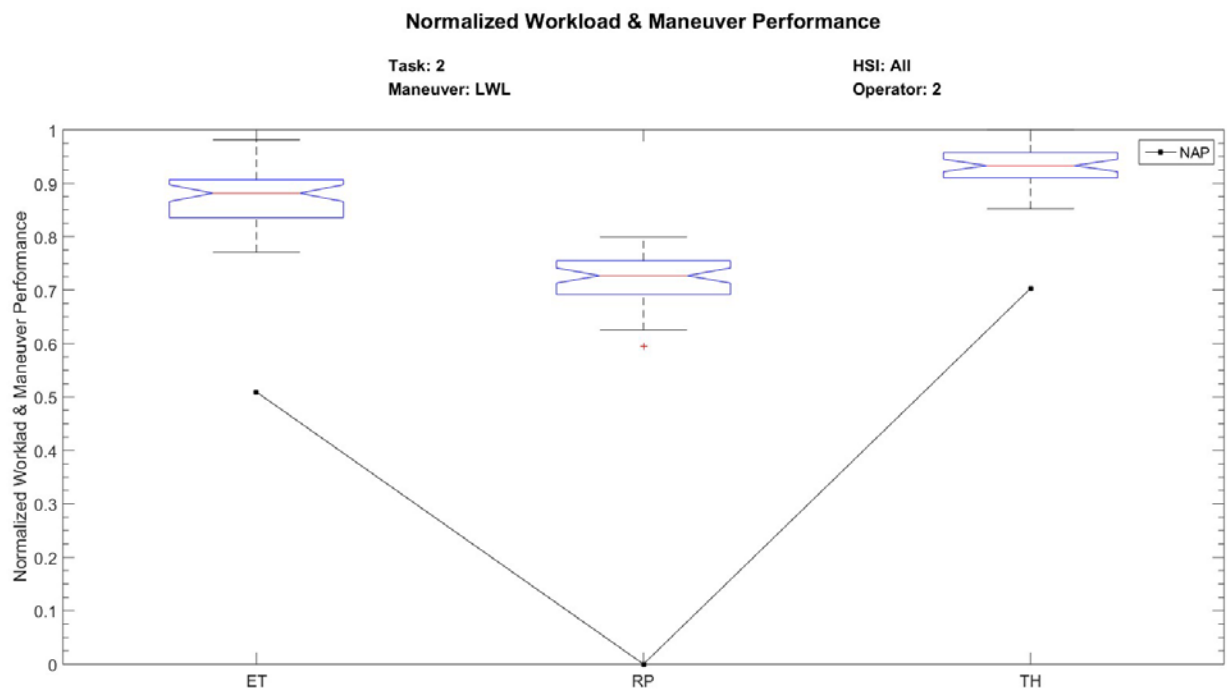


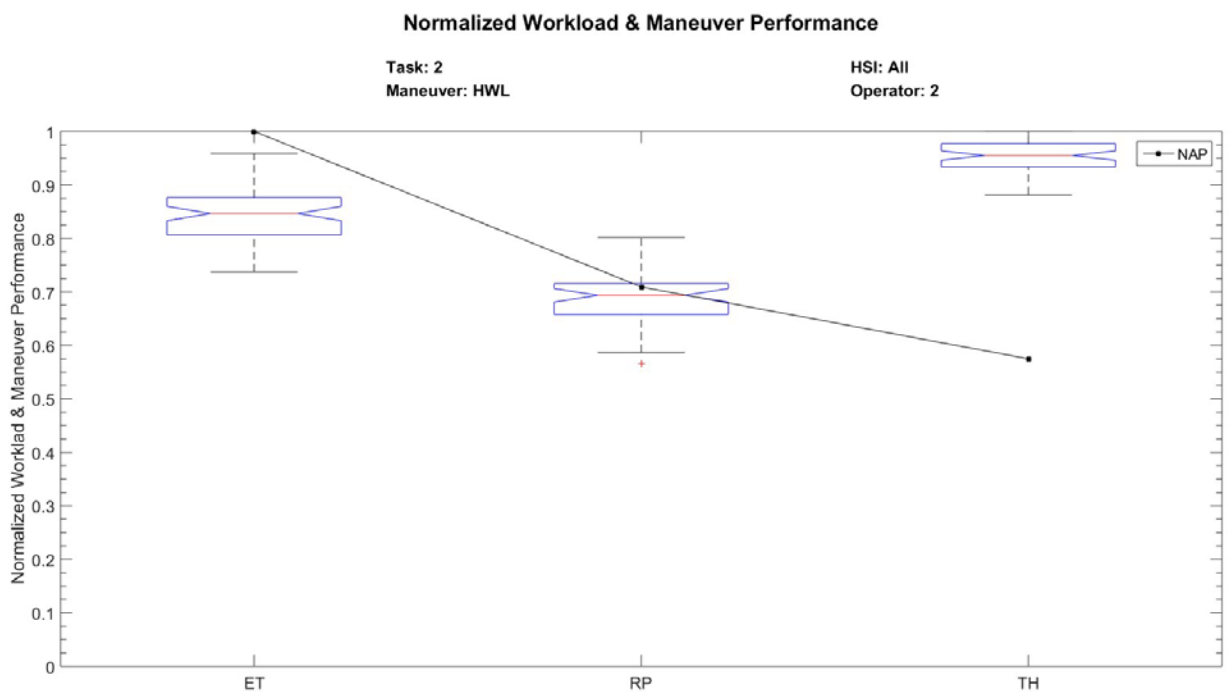
Op	Man	ET_R	ET_Y	ET_RMS	ET_CEP_95	RP_R	RP_Y	RP_RMS	RP_CEP_95	T_R	T_Y	T_RMS	T_CEP_95
1	All	0.646	1.836	4.690	8.611	0.634	1.849	0.561	1.037	0.744	1.736	0.377	0.654
2	All	0.492	1.988	5.001	9.942	0.590	1.894	0.504	0.955	0.896	1.496	0.896	1.340
3	All	0.887	1.511	5.461	8.252	0.540	1.945	0.685	1.332	0.730	1.750	0.238	0.417
4	All	0.572	1.912	5.578	10.665	0.641	1.841	0.662	1.219	0.706	1.775	0.481	0.854
5	All	0.567	1.917	3.450	6.614	0.692	1.789	0.397	0.710	0.841	1.583	0.371	0.587
6	All	0.837	1.590	6.945	11.043	0.674	1.808	0.692	1.251	0.838	1.588	0.509	0.808
7	All	0.577	1.907	7.623	14.537	0.742	1.738	0.294	0.511	0.704	1.777	0.310	0.551
All	All	0.605	1.878	5.446	10.228	0.596	1.887	0.563	1.062	0.715	1.766	0.379	0.669

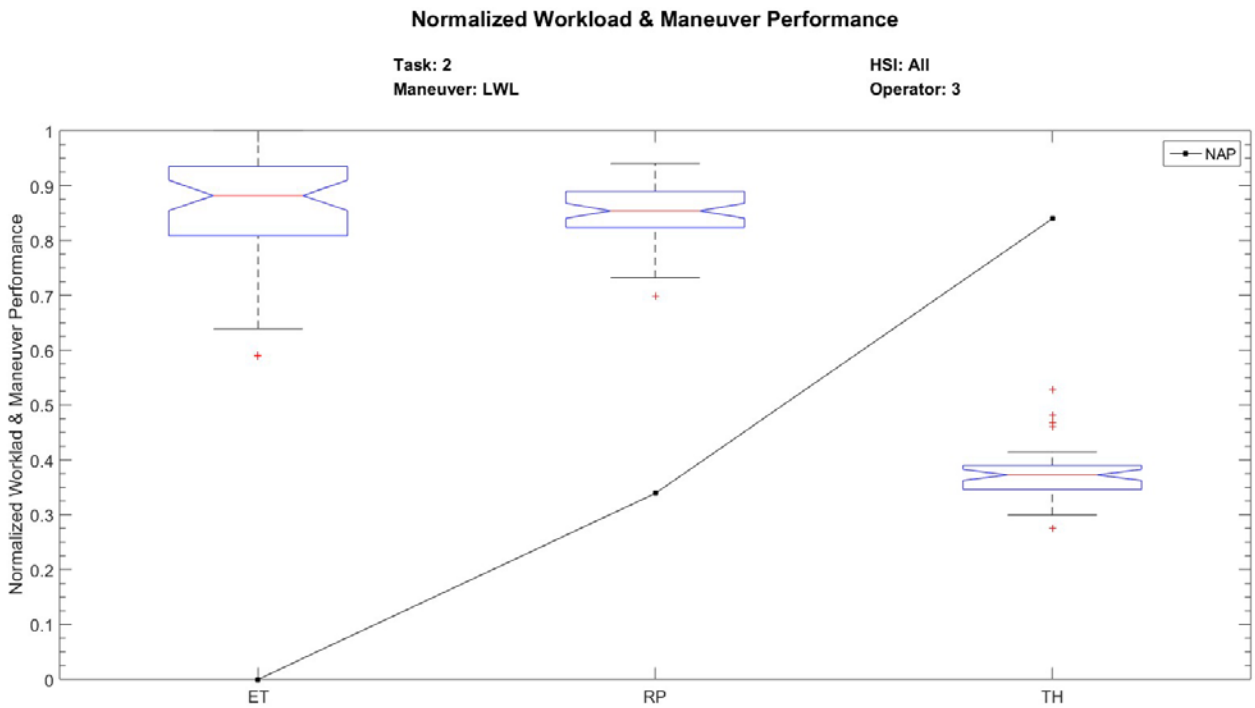
Task 2 – Workload

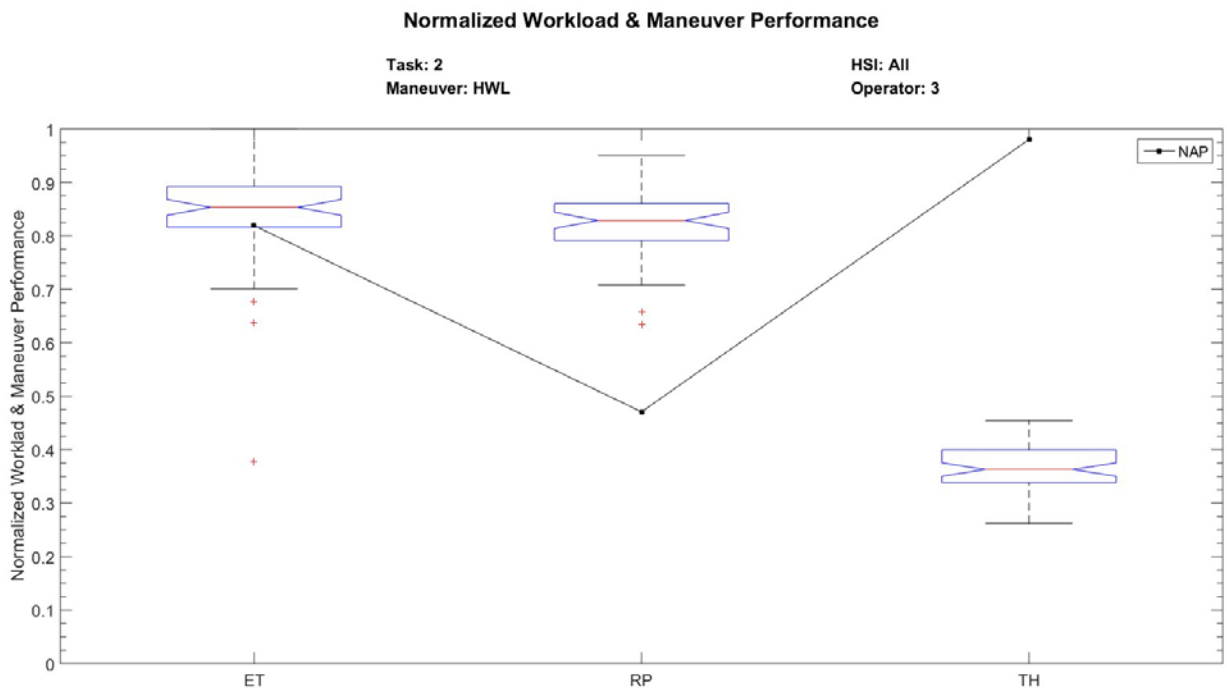


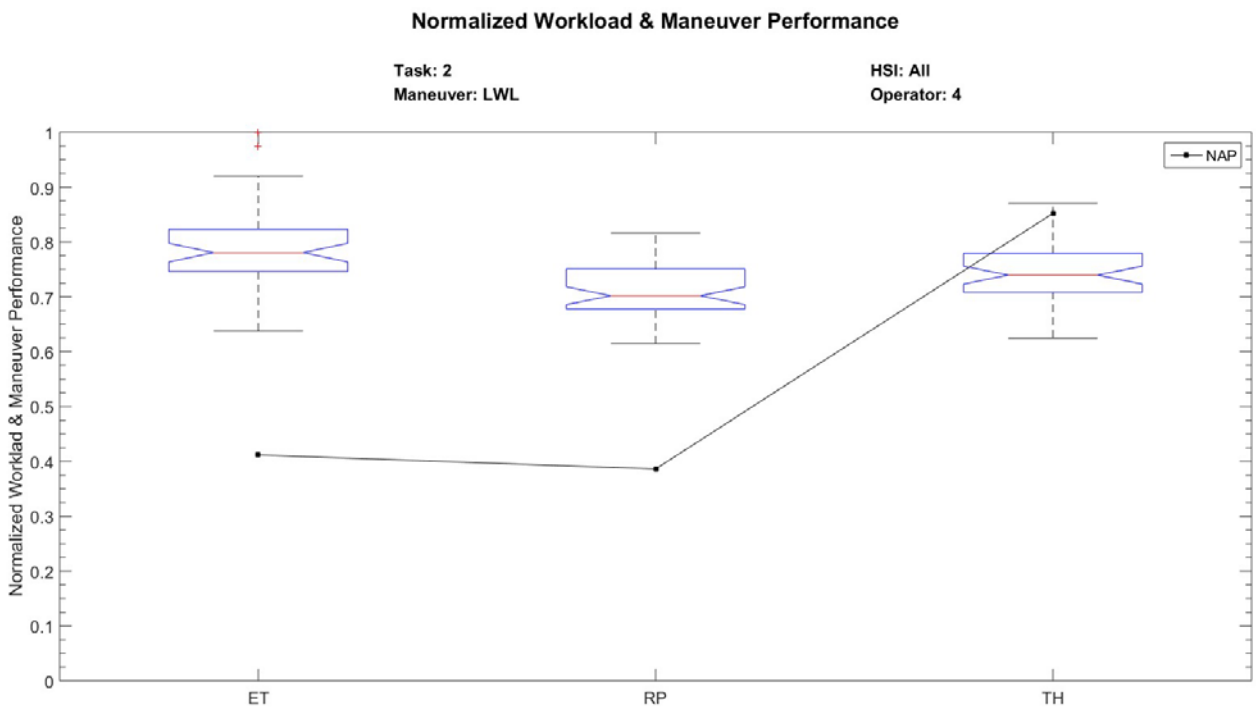


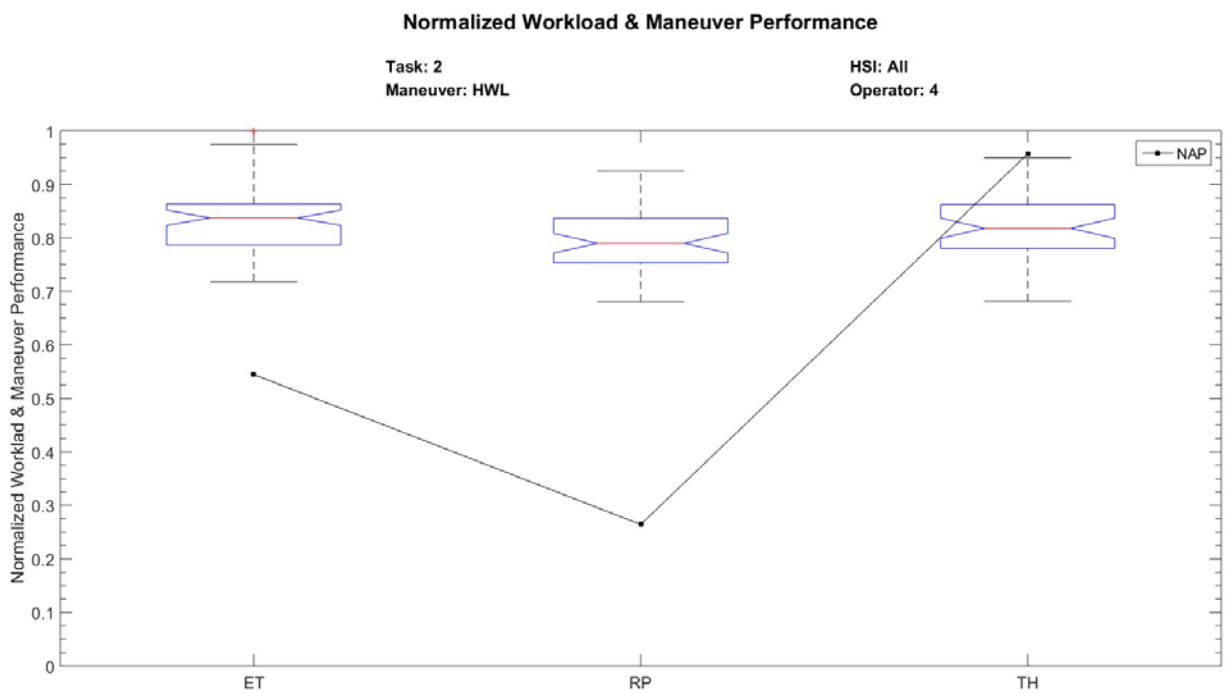

Normalized Workload & Maneuver Performance
Task: 2
Maneuver: LWL
HSI: All
Operator: 2
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

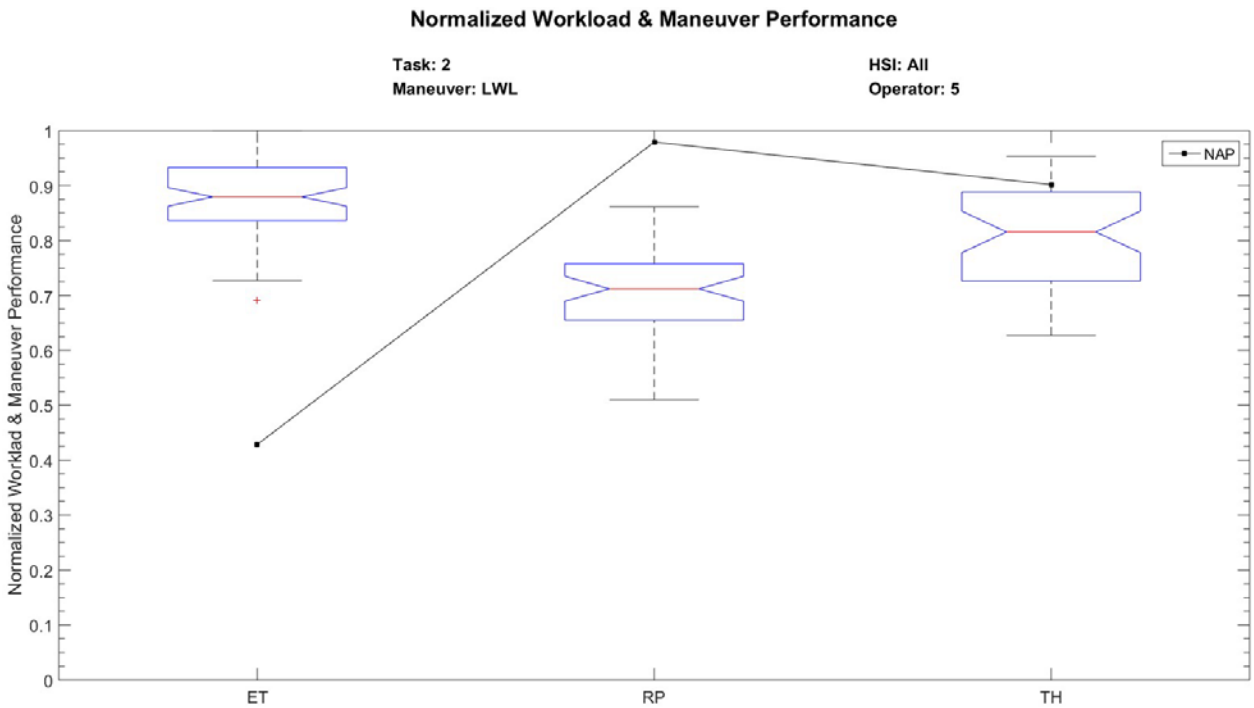

Normalized Workload & Maneuver Performance
Task: 2
Maneuver: HWL
HSI: All
Operator: 2
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

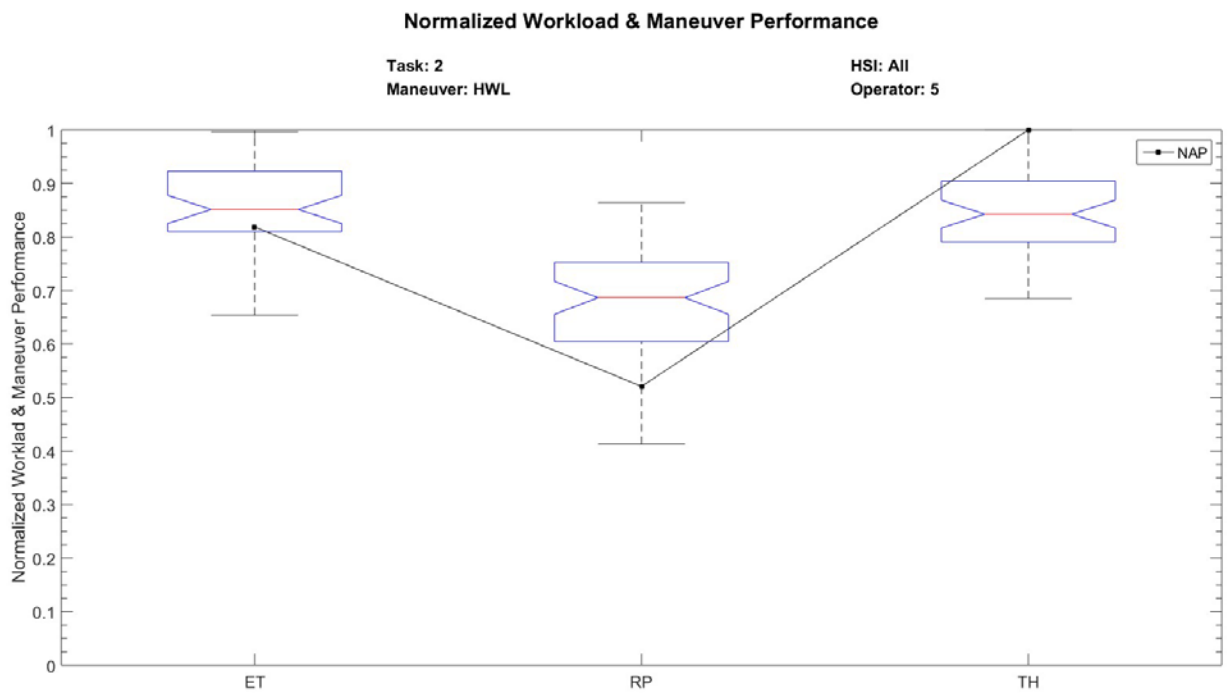

Normalized Workload & Maneuver Performance
Task: 2
Maneuver: LWL
HSI: All
Operator: 3
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

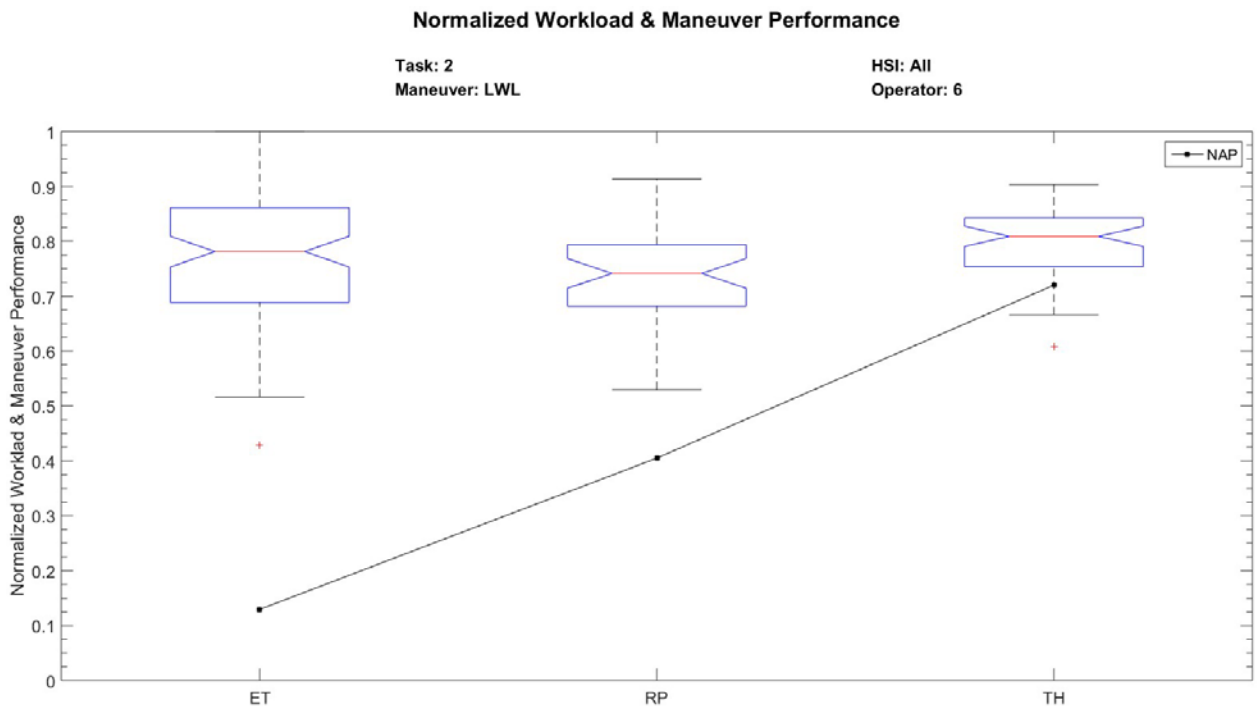

Normalized Workload & Maneuver Performance
Task: 2
Maneuver: HWL
HSI: All
Operator: 3
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

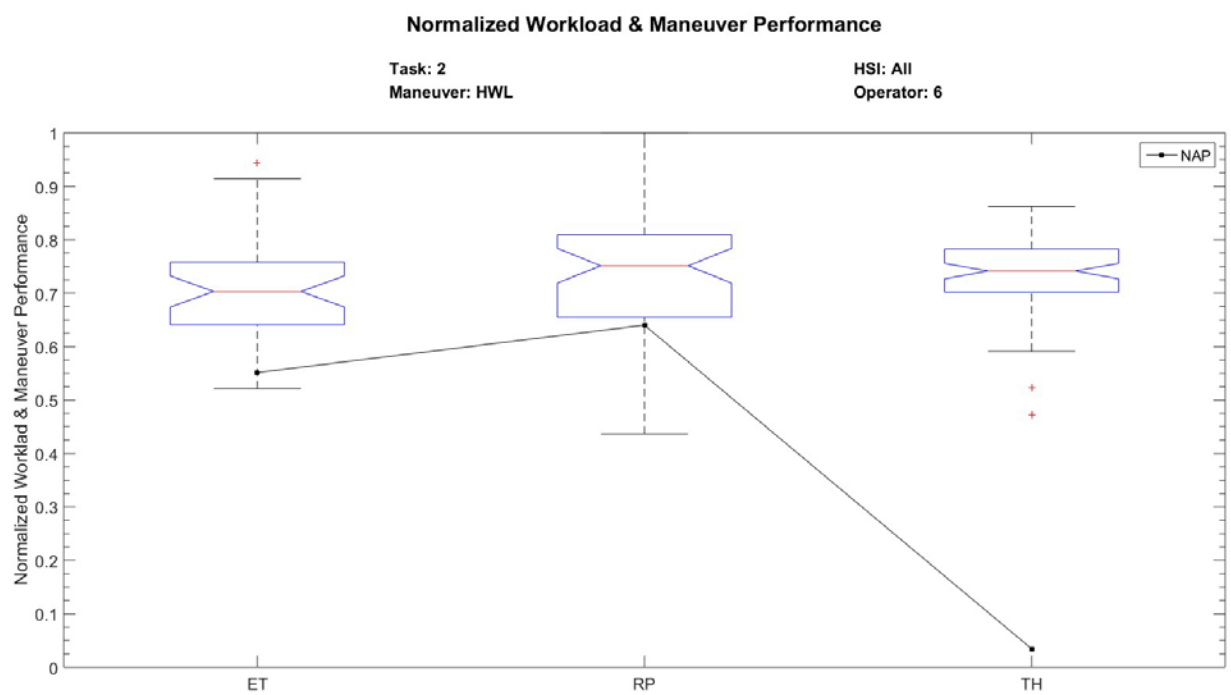

Normalized Workload & Maneuver Performance
Task: 2
Maneuver: LWL
HSI: All
Operator: 4
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

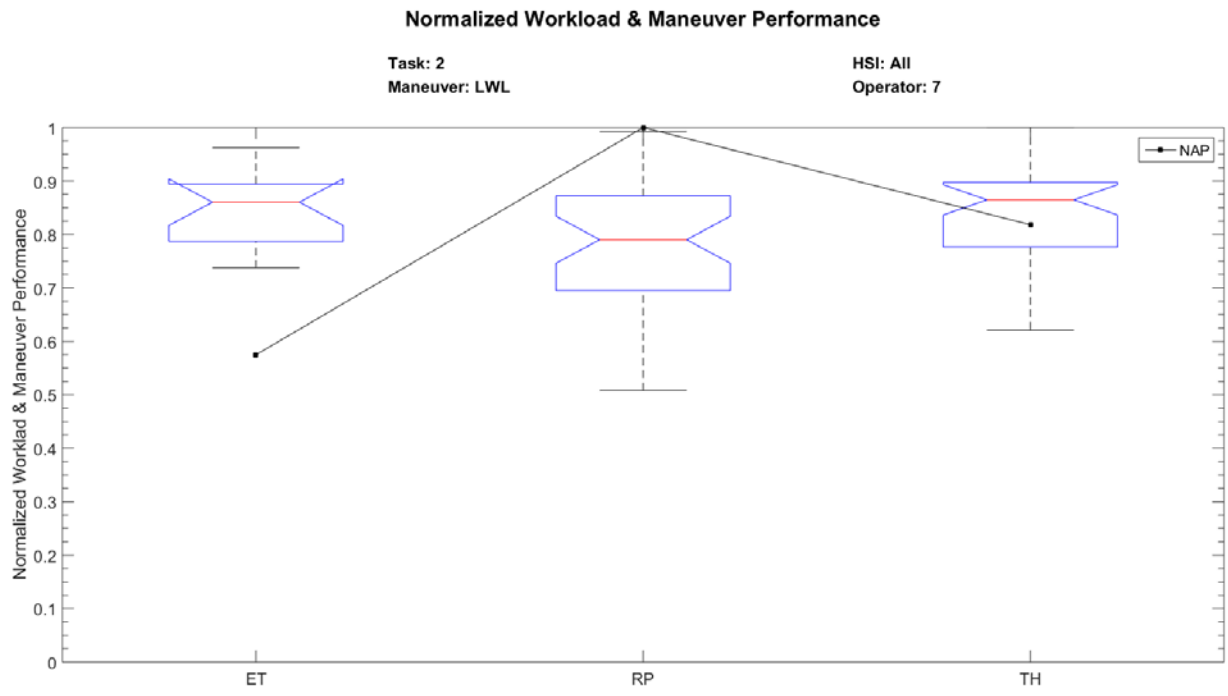

Normalized Workload & Maneuver Performance
Task: 2
Maneuver: HWL
HSI: All
Operator: 4
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

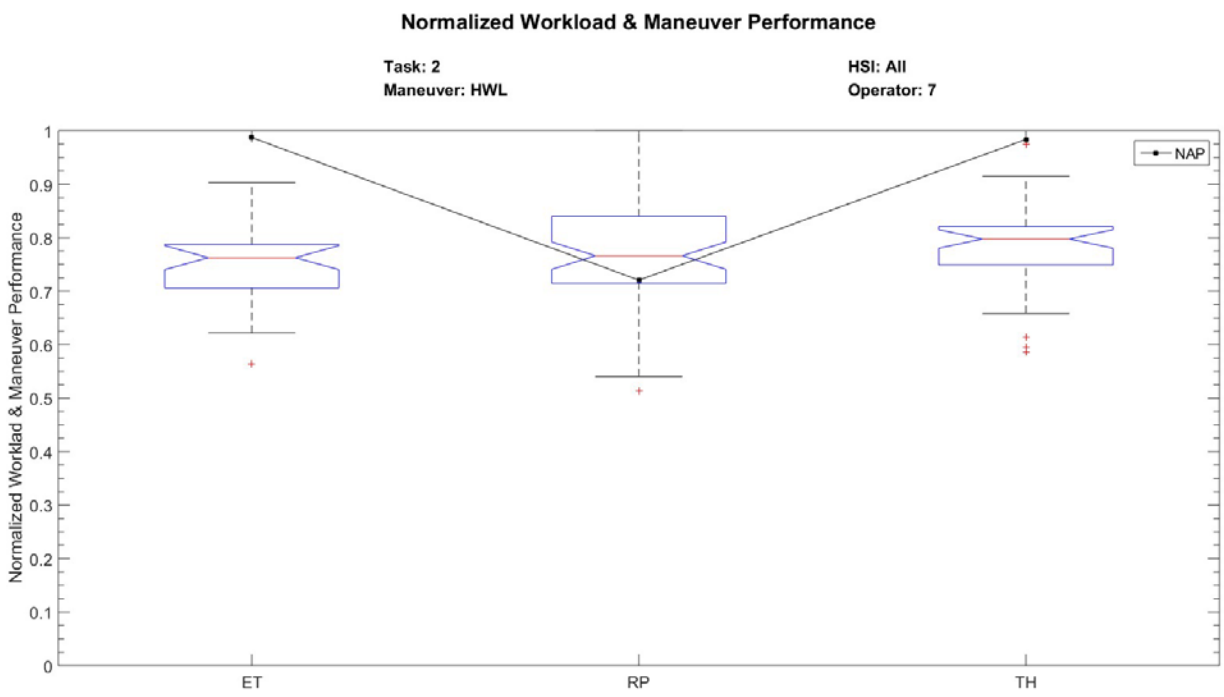

Normalized Workload & Maneuver Performance
Task: 2
Maneuver: LWL
HSI: All
Operator: 5
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH


Normalized Workload & Maneuver Performance
Task: 2
Maneuver: HWL
HSI: All
Operator: 5
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

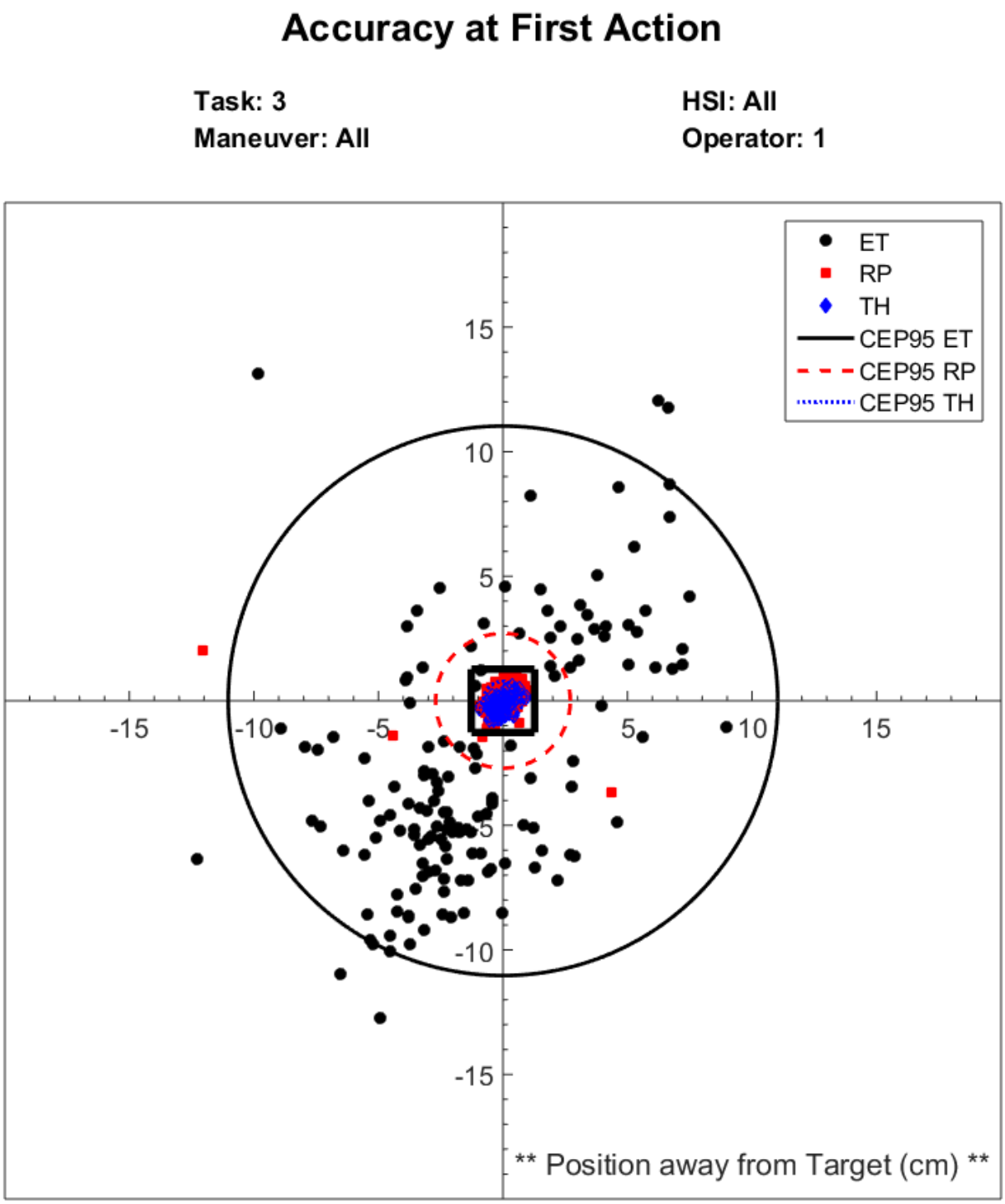

Normalized Workload & Maneuver Performance
Task: 2
Maneuver: LWL
HSI: All
Operator: 6
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH


Normalized Workload & Maneuver Performance
Task: 2
Maneuver: HWL
HSI: All
Operator: 6
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH


Normalized Workload & Maneuver Performance
Task: 2
Maneuver: LWL
HSI: All
Operator: 7
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

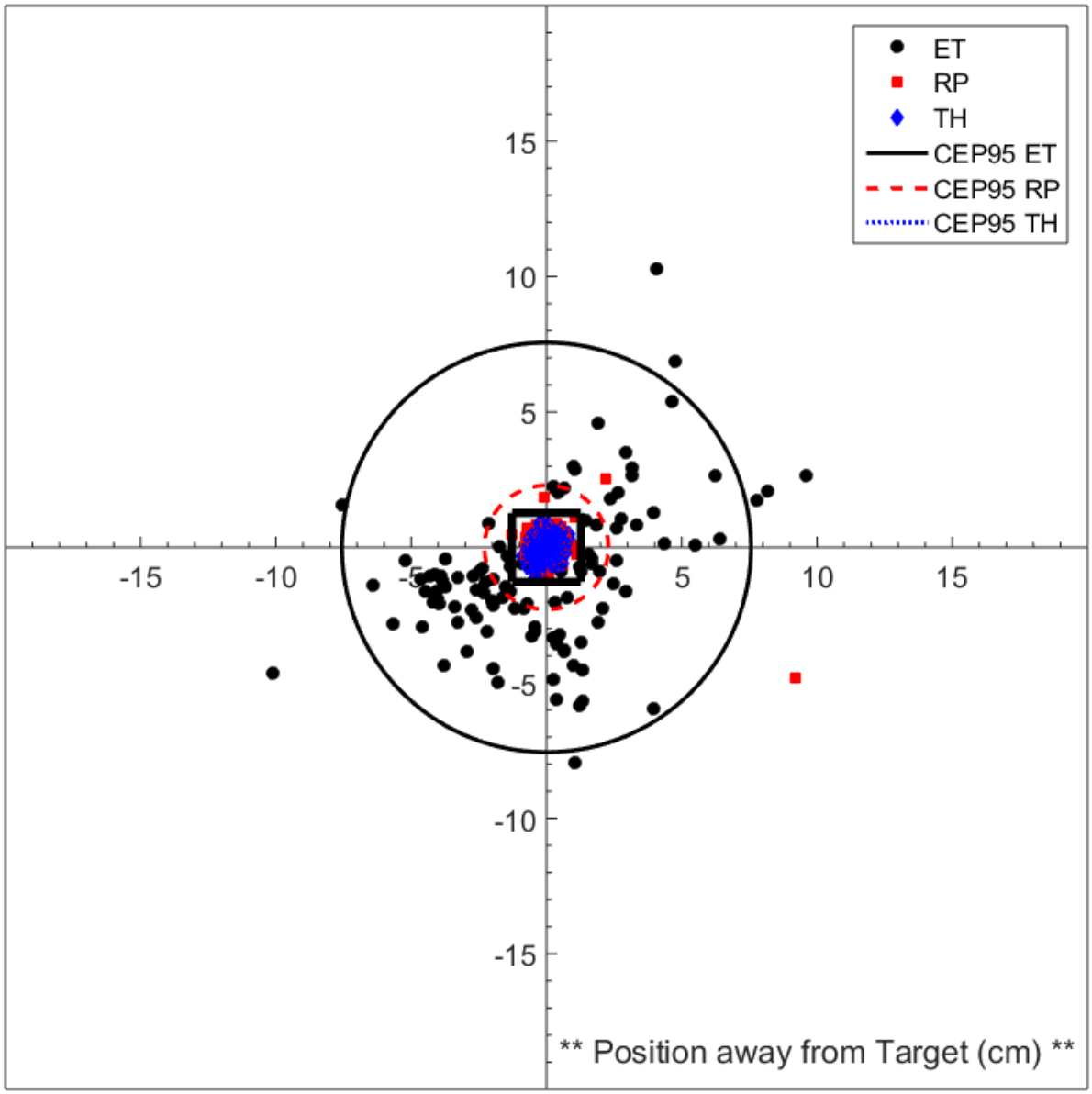

Normalized Workload & Maneuver Performance
Task: 2
Maneuver: HWL
HSI: All
Operator: 7
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

Task 3 – Accuracy



Accuracy at First Action

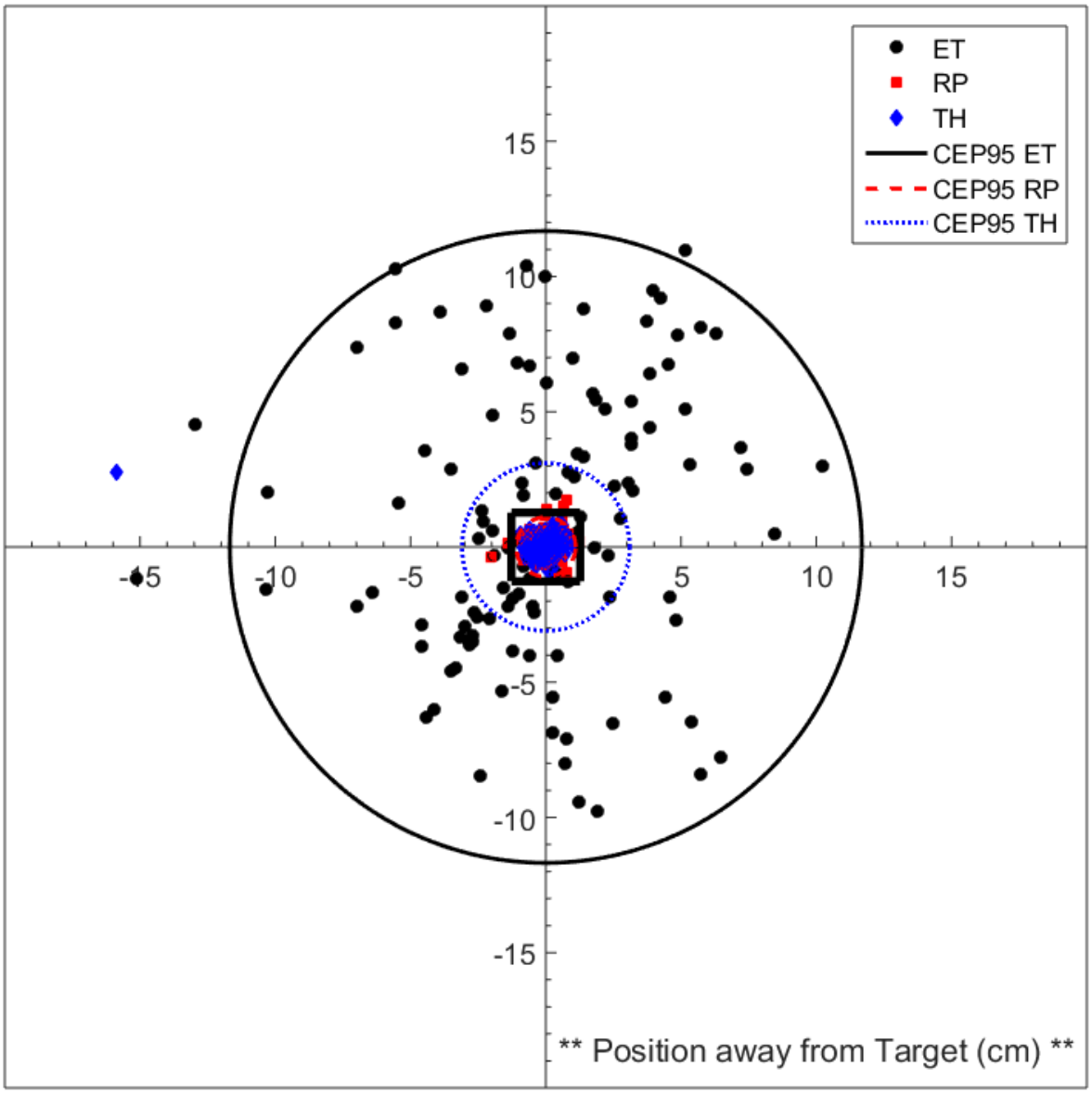
Task: 3 **HSI: All**
Maneuver: All **Operator: 2**



Accuracy at First Action

Task: 3
Maneuver: All

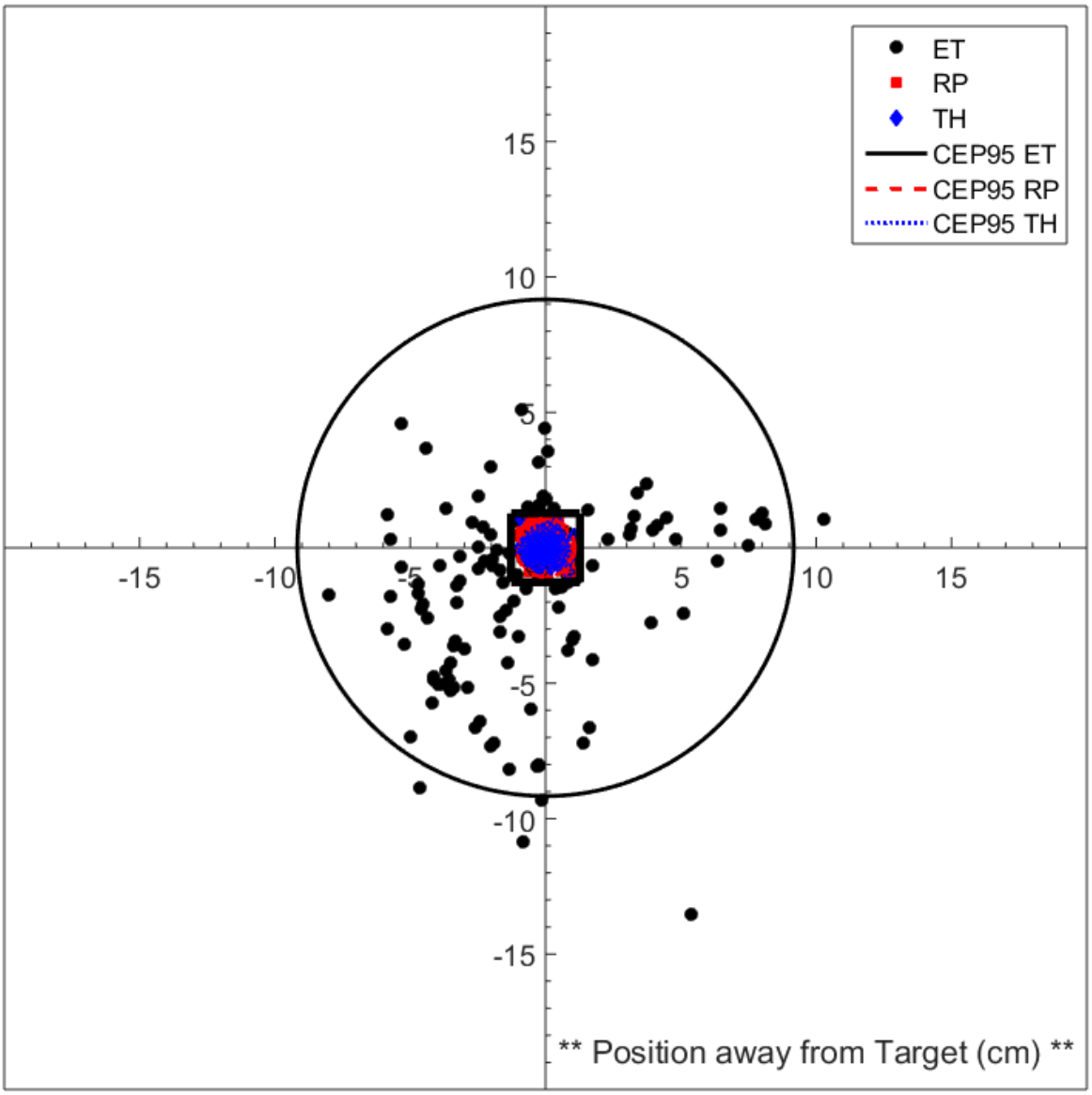
HSI: All
Operator: 3



Accuracy at First Action

Task: 3
Maneuver: All

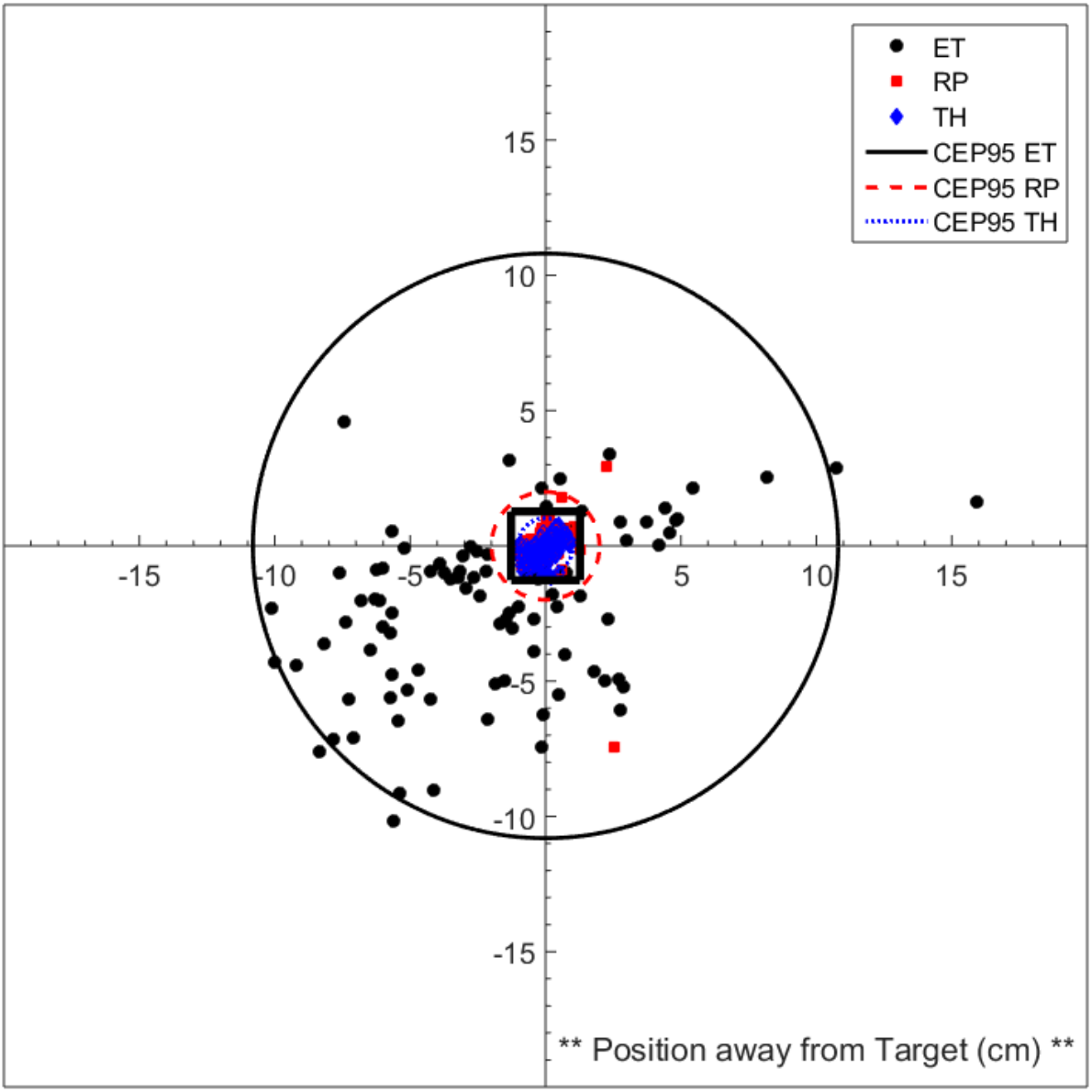
HSI: All
Operator: 4



Accuracy at First Action

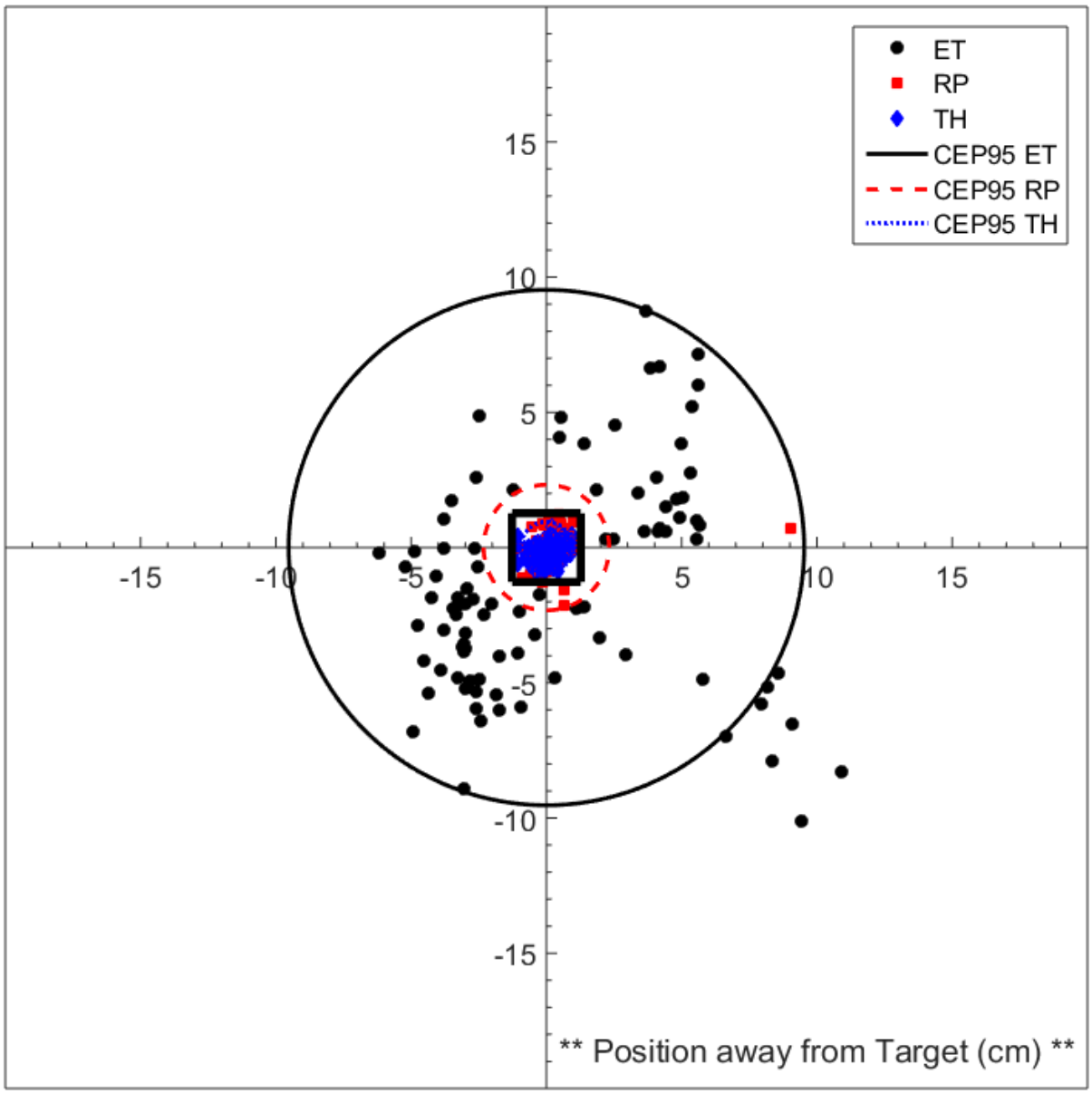
Task: 3
Maneuver: All

HSI: All
Operator: 5



Accuracy at First Action

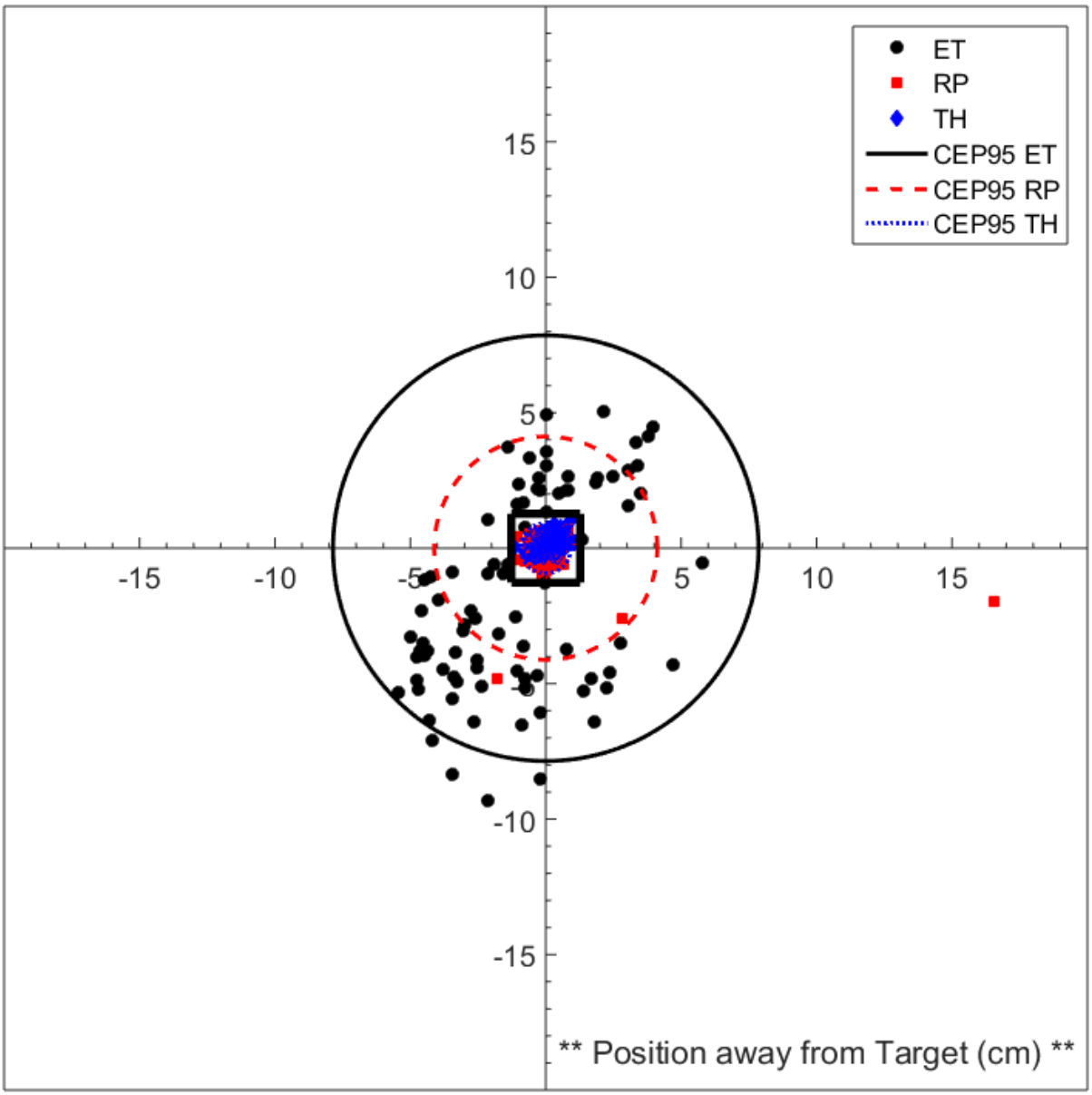
Task: 3 **HSI: All**
Maneuver: All **Operator: 6**



Accuracy at First Action

Task: 3
Maneuver: All

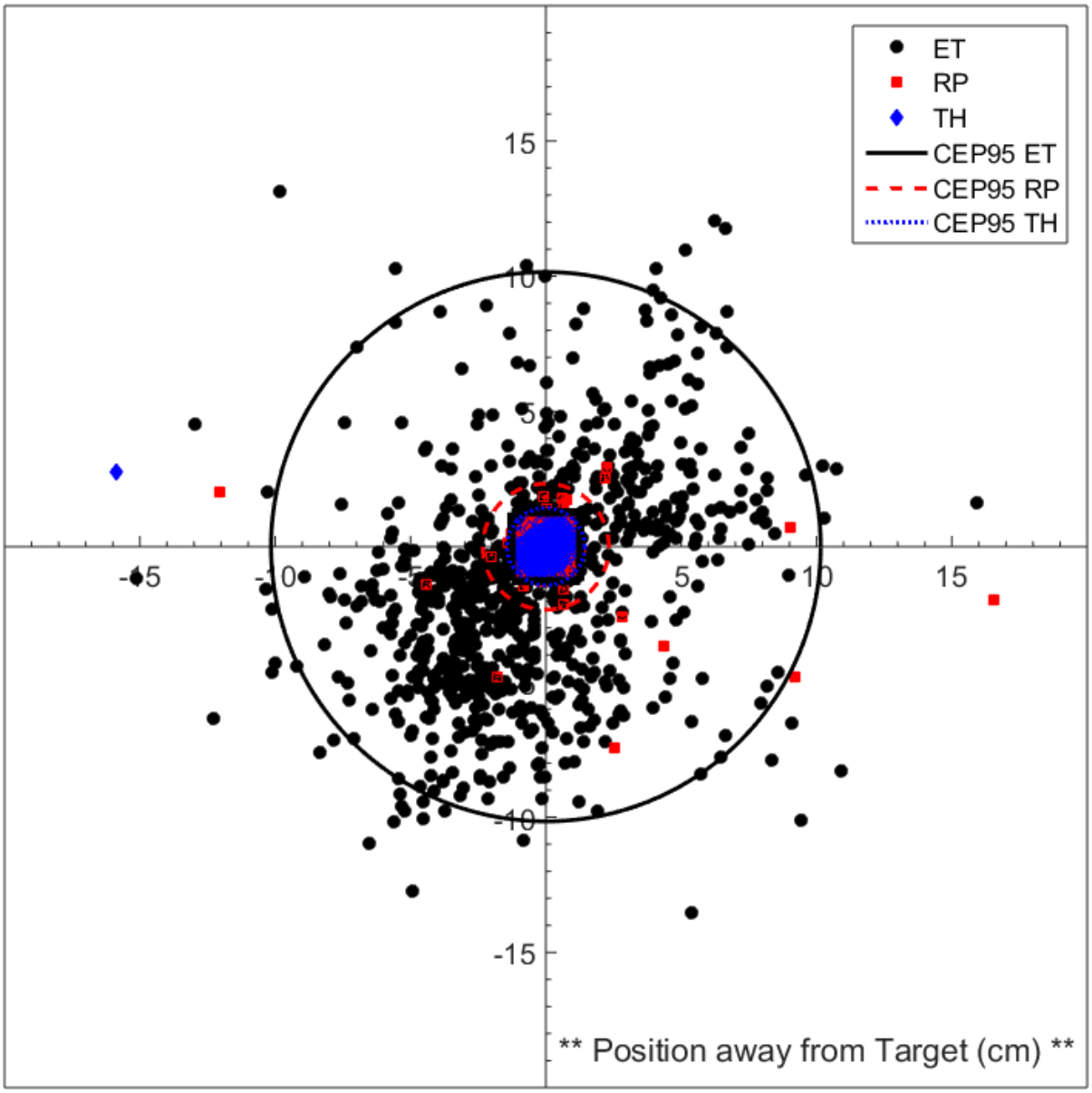
HSI: All
Operator: 7



Accuracy at First Action

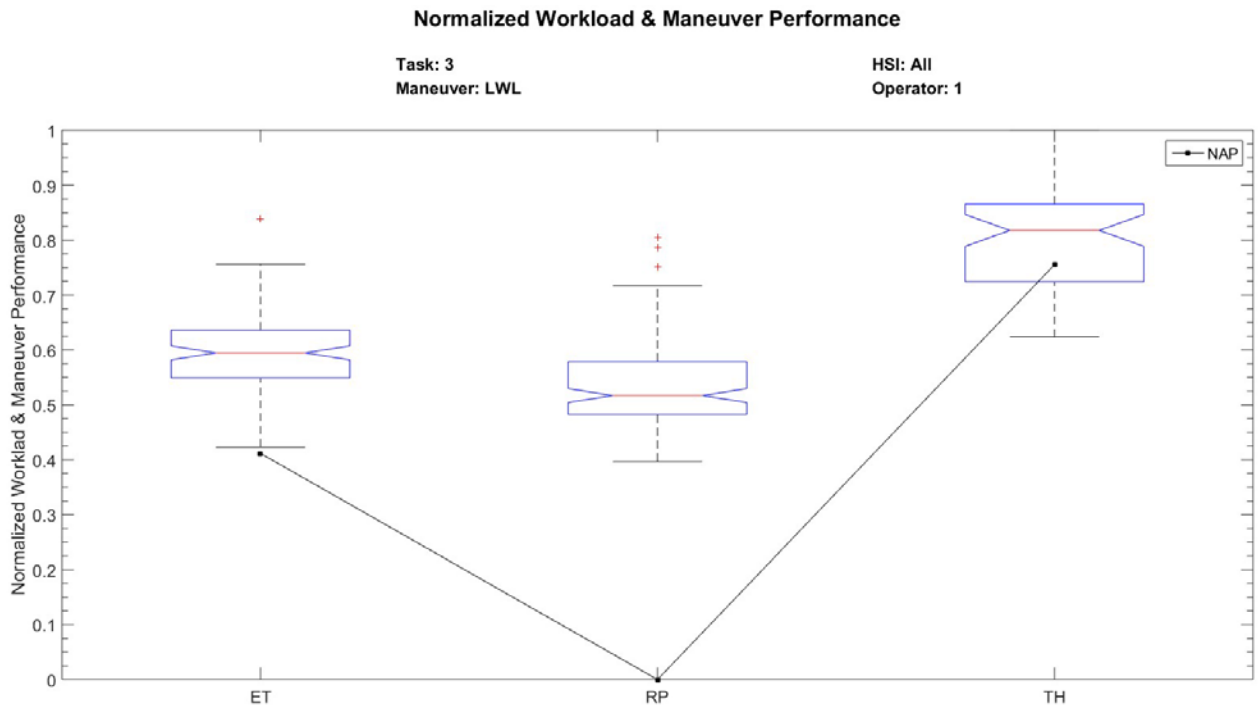
Task: 3
Maneuver: All

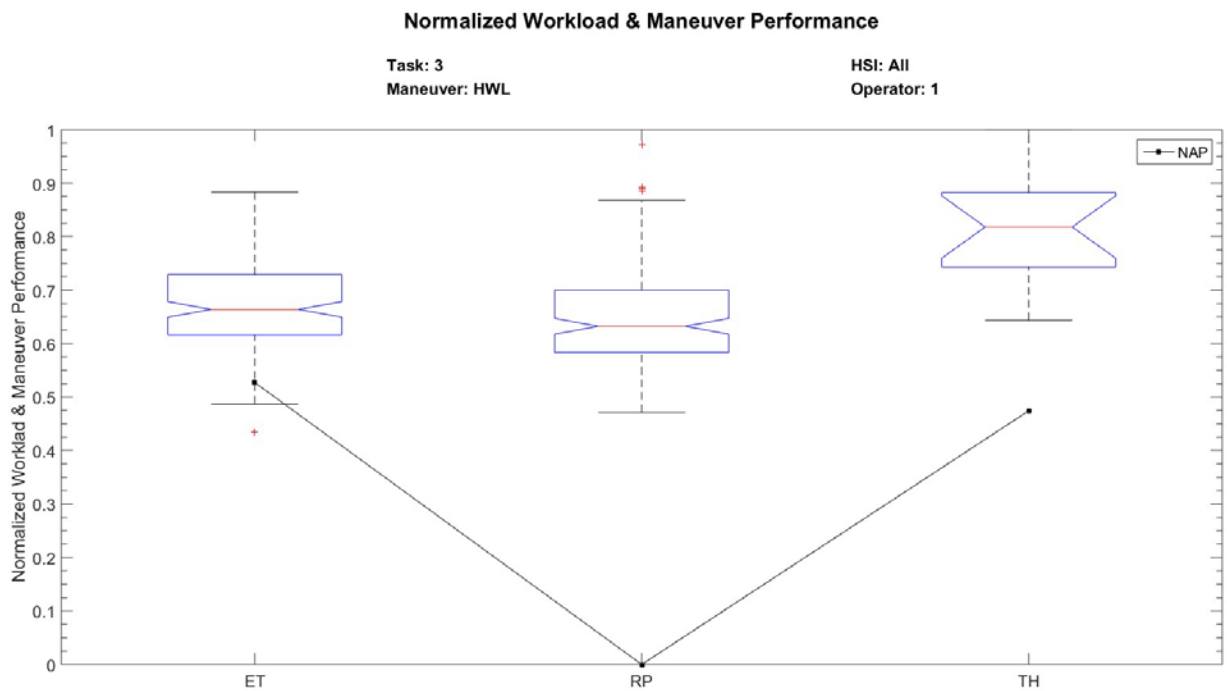
HSI: All
Operator: All

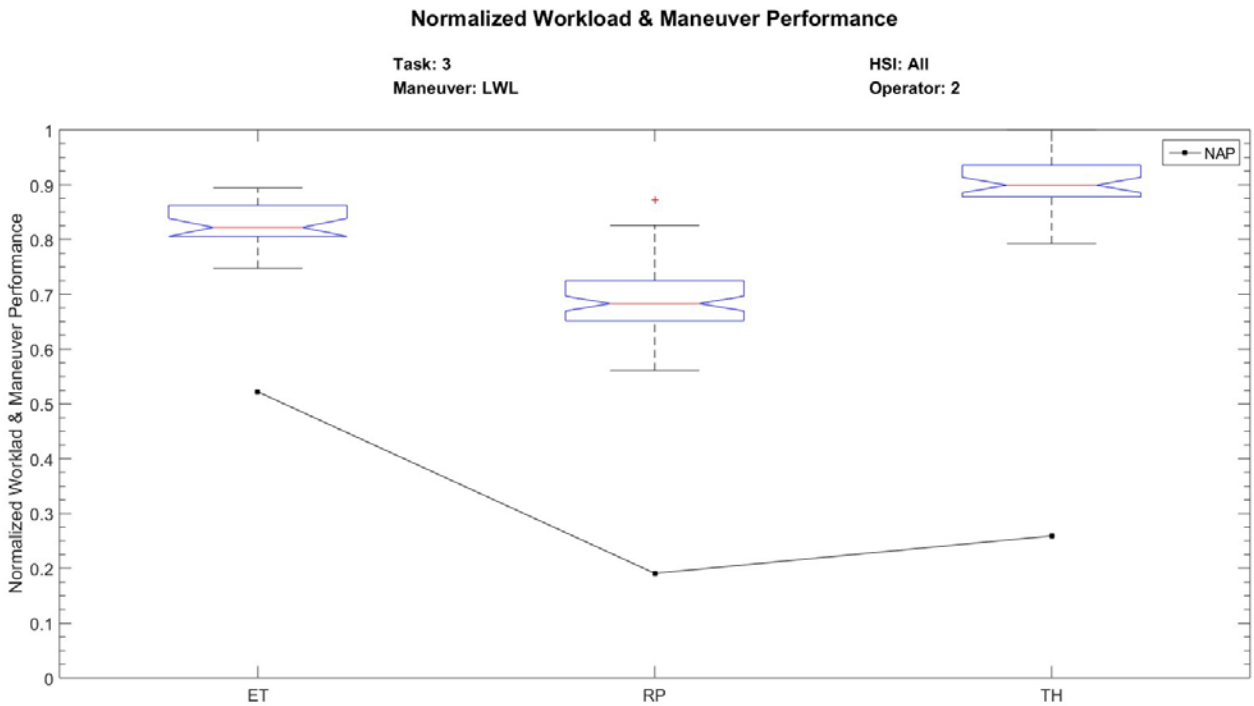


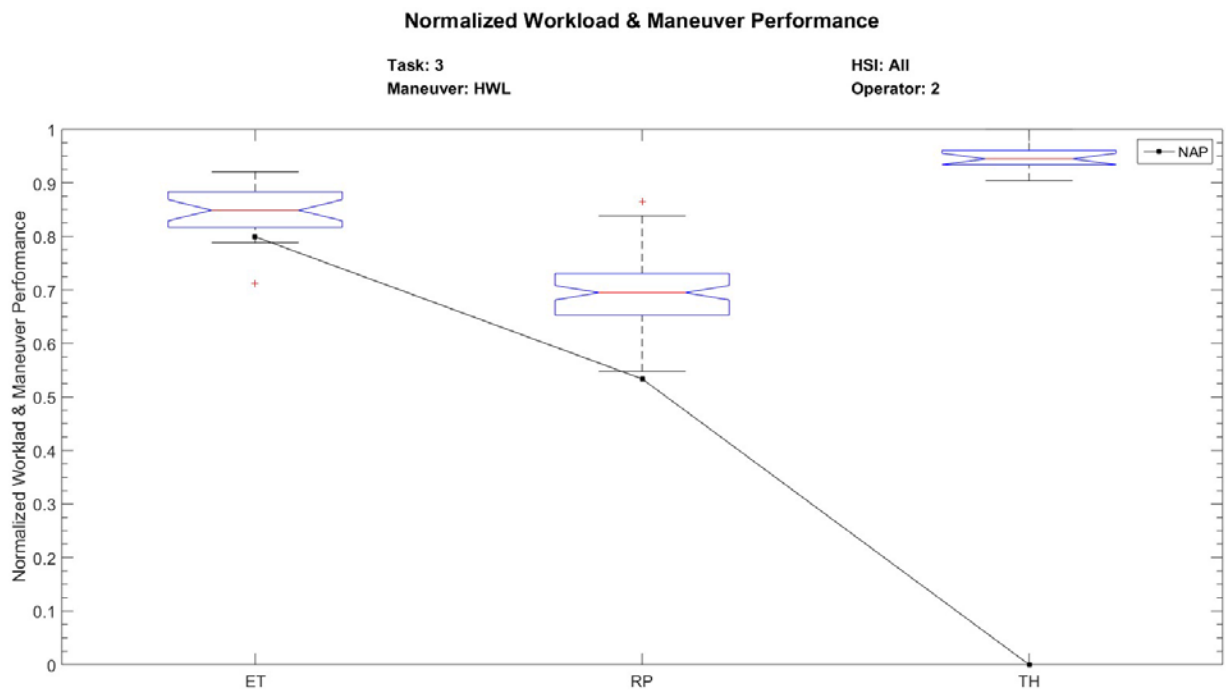
Op	Man	ET_R	ET_Y	ET_RMS	ET_CEP_95	RP_R	RP_Y	RP_RMS	RP_CEP_95	T_R	T_Y	T_RMS	T_CEP_95
1	All	0.83	1.601	6.89	11.031	0.4	2.069	1.307	2.704	0.803	1.643	0.474	0.779
2	All	0.732	1.748	4.325	7.560	0.404	2.065	1.109	2.290	0.731	1.749	0.481	0.841
3	All	0.685	1.796	6.504	11.681	0.625	1.858	0.592	1.100	0.293	2.162	1.433	3.098
4	All	0.662	1.82	5.037	9.167	0.745	1.735	0.623	1.081	0.712	1.769	0.485	0.858
5	All	0.739	1.741	6.207	10.806	0.434	2.039	0.978	1.994	0.802	1.645	0.623	1.025
6	All	0.838	1.588	6	9.528	0.416	2.055	1.129	2.320	0.75	1.727	0.552	0.953
7	All	0.831	1.599	4.913	7.856	0.243	2.206	1.867	4.119	0.806	1.639	0.565	0.926
8	All	0.733	1.747	5.812	10.154	0.399	2.07	1.125	2.329	0.548	1.937	0.738	1.430

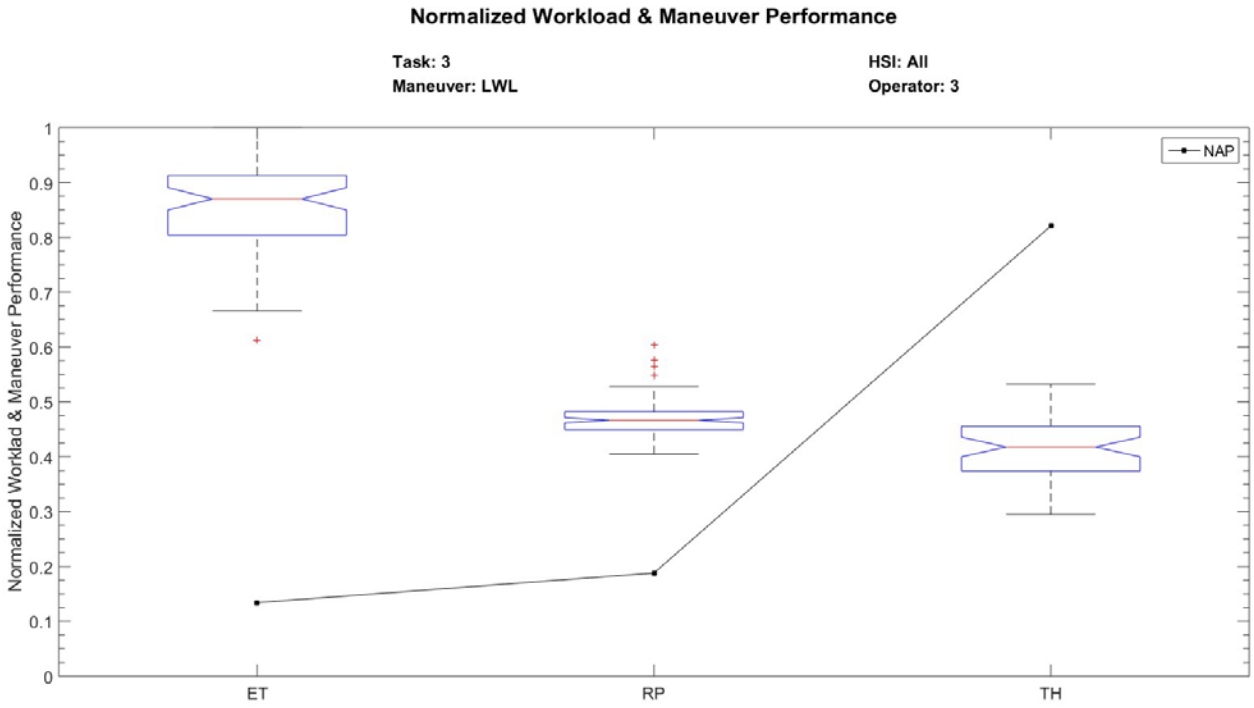
Task 3 – Workload

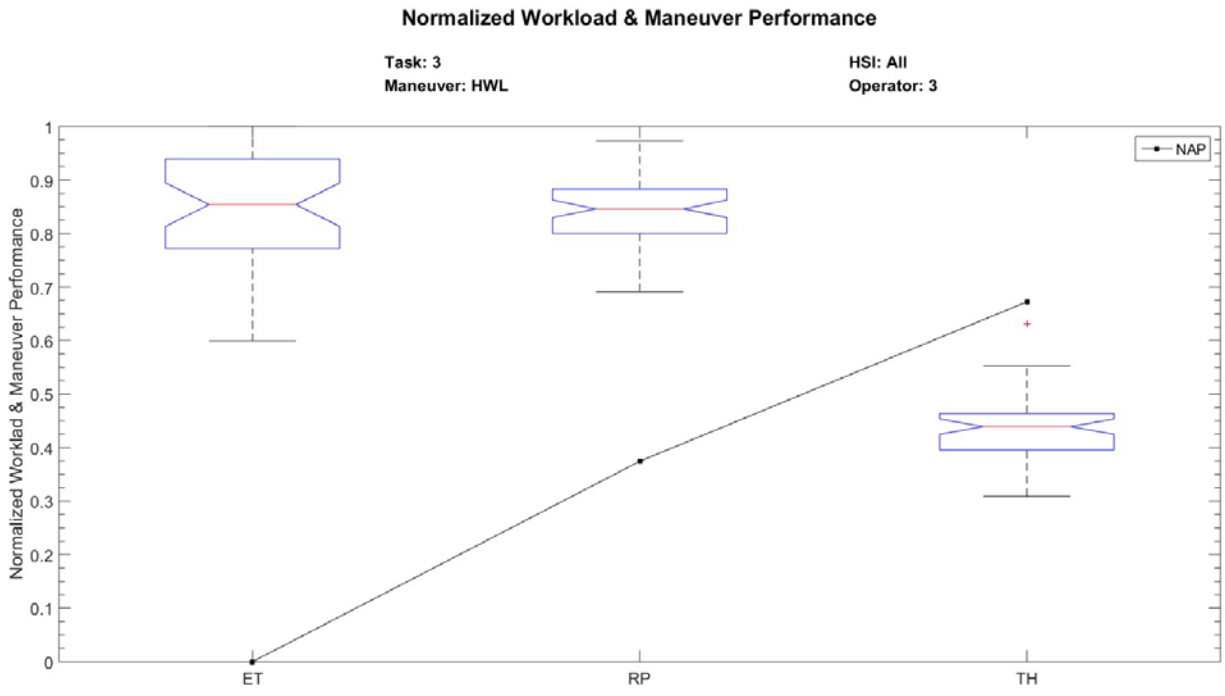


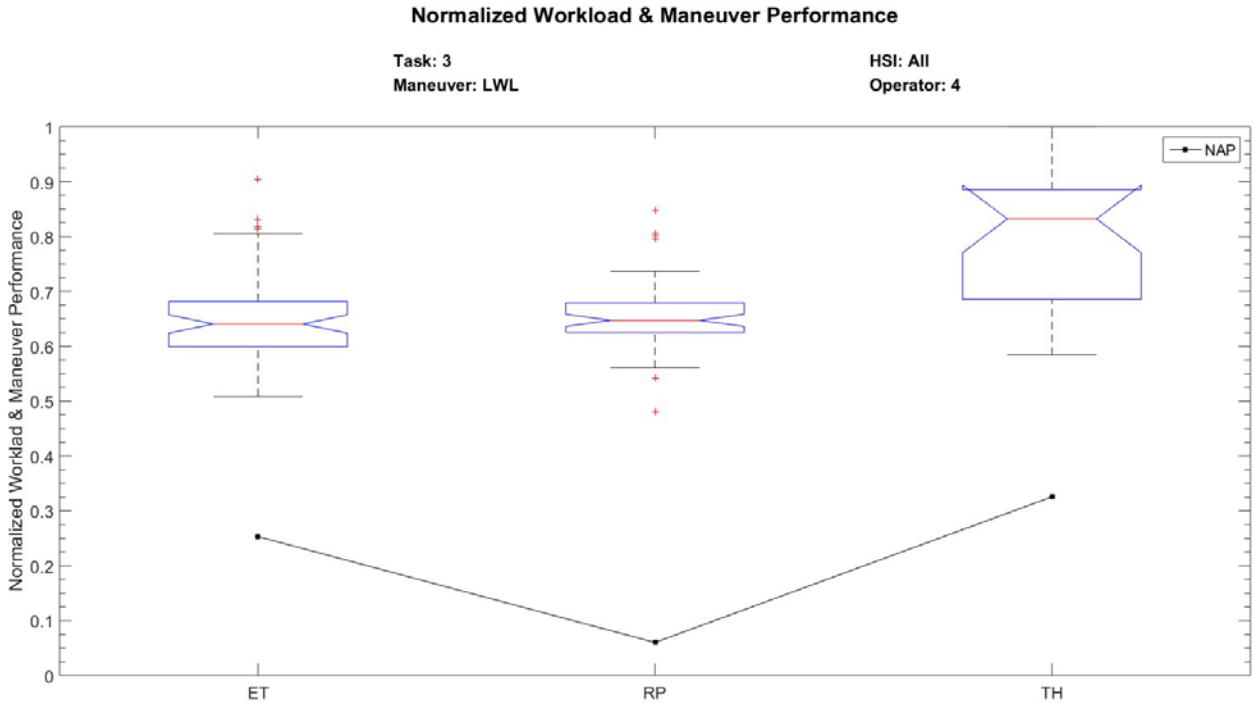


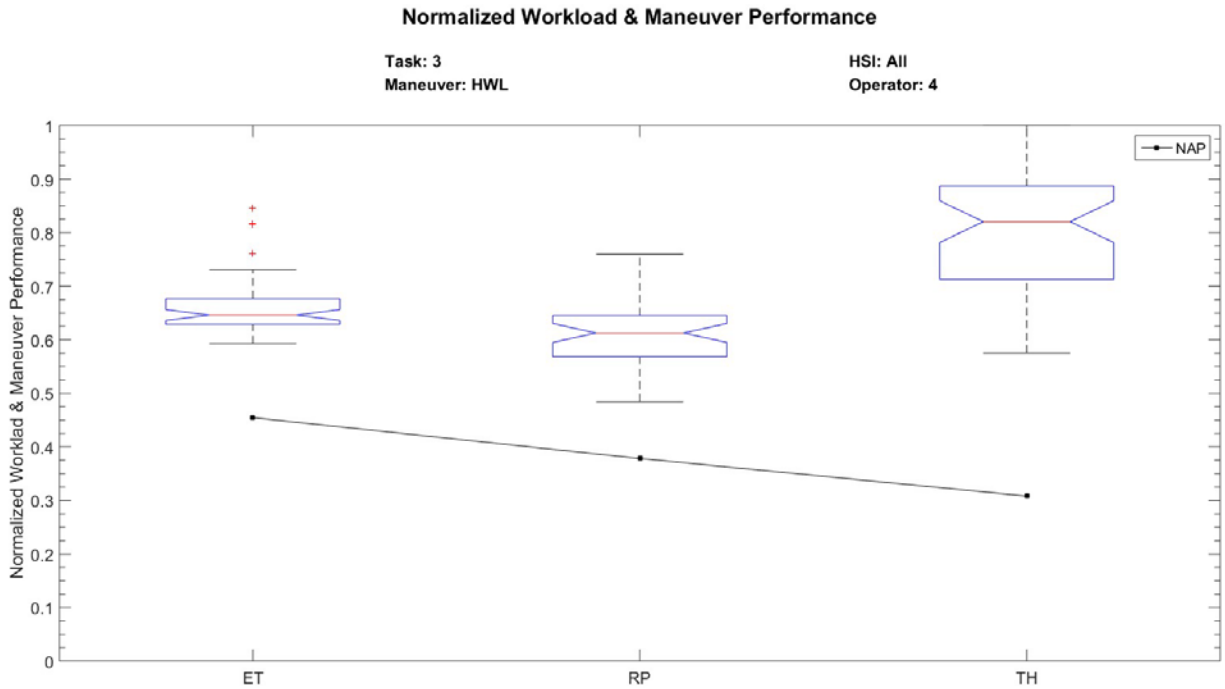

Normalized Workload & Maneuver Performance
Task: 3
Maneuver: LWL
HSI: All
Operator: 2
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

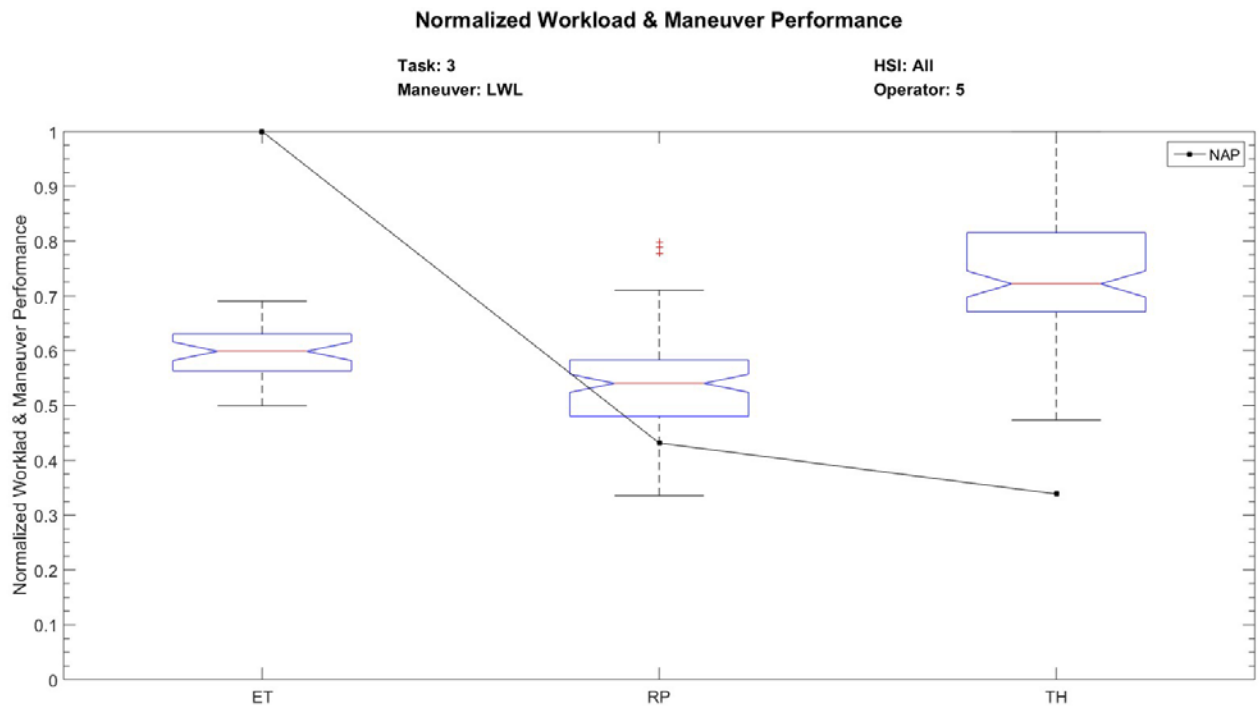

Normalized Workload & Maneuver Performance
Task: 3
Maneuver: HWL
HSI: All
Operator: 2
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

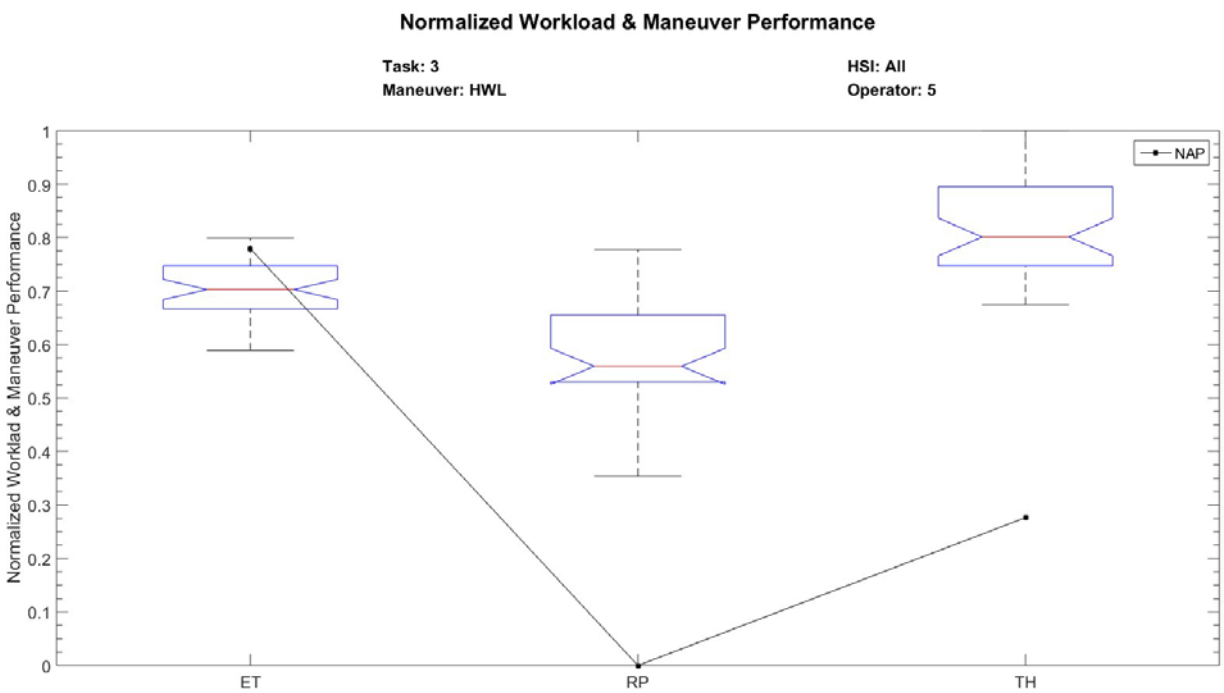

Normalized Workload & Maneuver Performance
Task: 3
Maneuver: LWL
HSI: All
Operator: 3
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

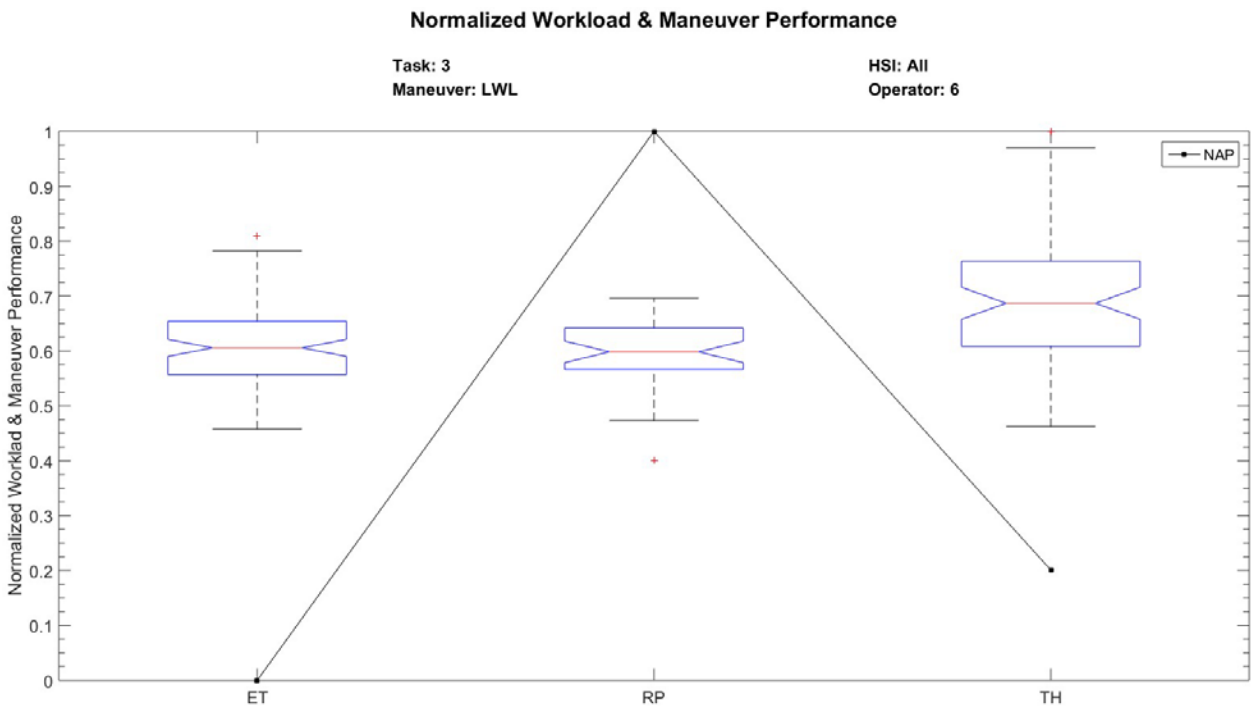

Normalized Workload & Maneuver Performance
Task: 3
Maneuver: HWL
HSI: All
Operator: 3
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

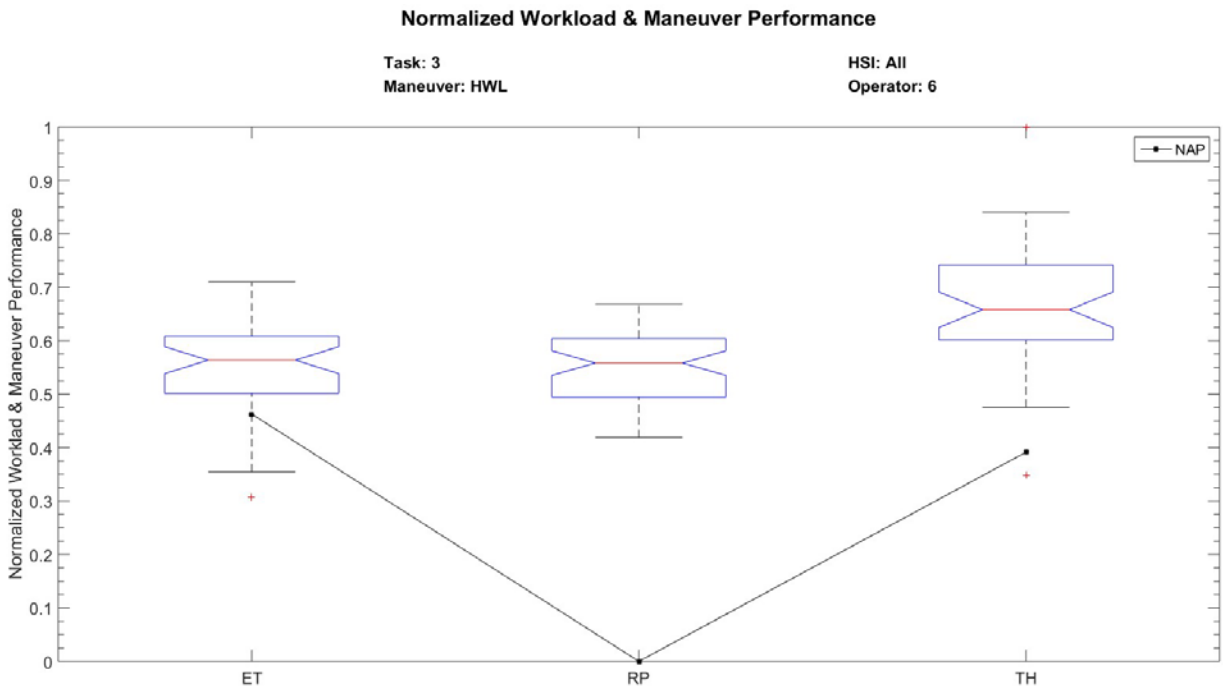

Normalized Workload & Maneuver Performance
Task: 3
Maneuver: LWL
HSI: All
Operator: 4
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

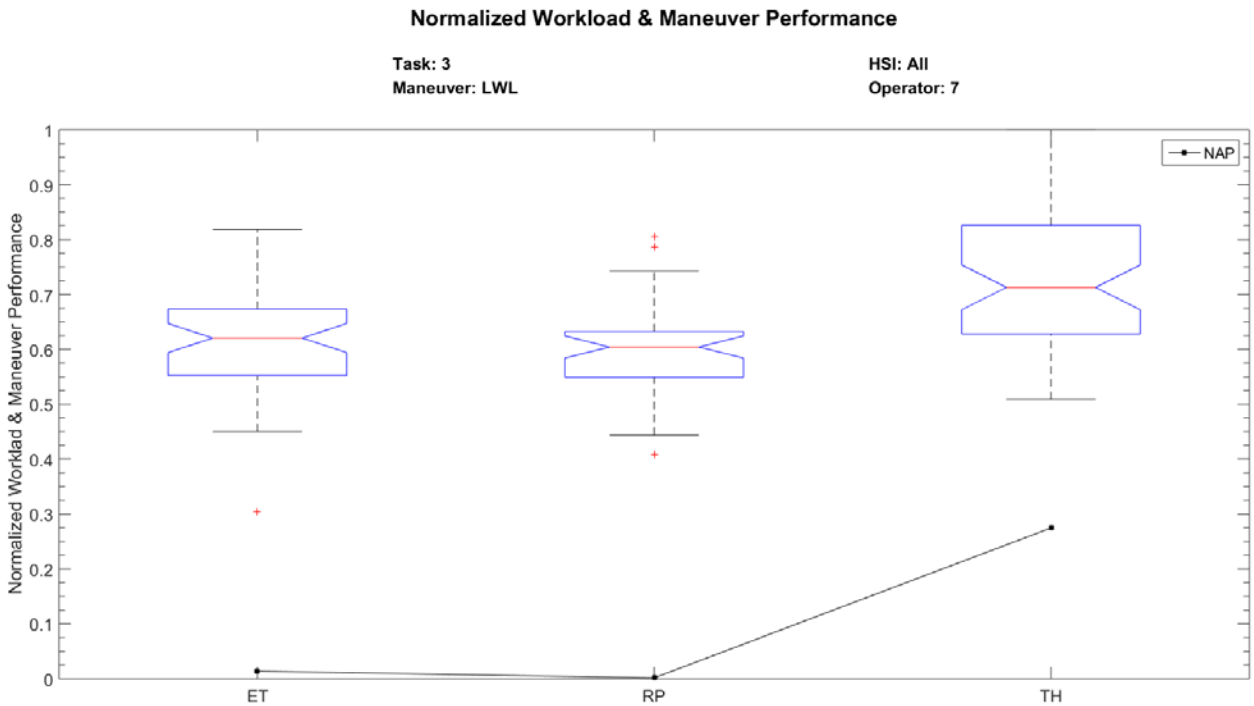

Normalized Workload & Maneuver Performance
Task: 3
Maneuver: HWL
HSI: All
Operator: 4
NAP
Normalized Workload & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

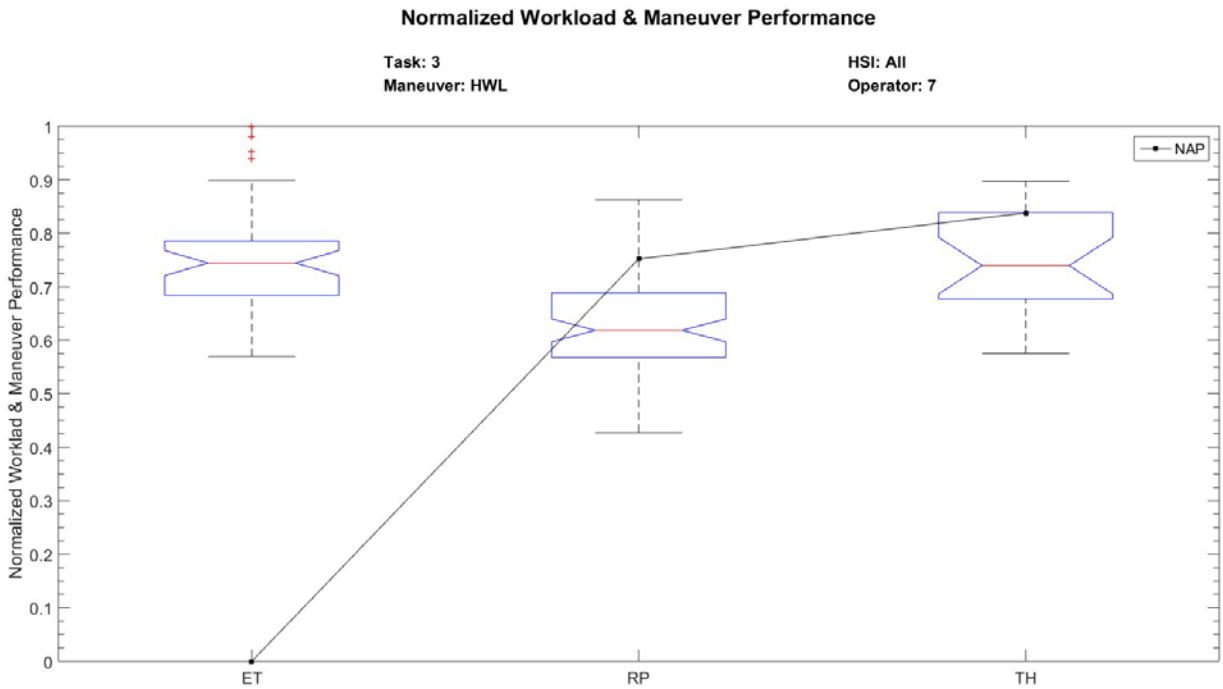

Normalized Workload & Maneuver Performance
Task: 3
Maneuver: LWL
HSI: All
Operator: 5
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH


Normalized Workload & Maneuver Performance
Task: 3
Maneuver: HWL
HSI: All
Operator: 5
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

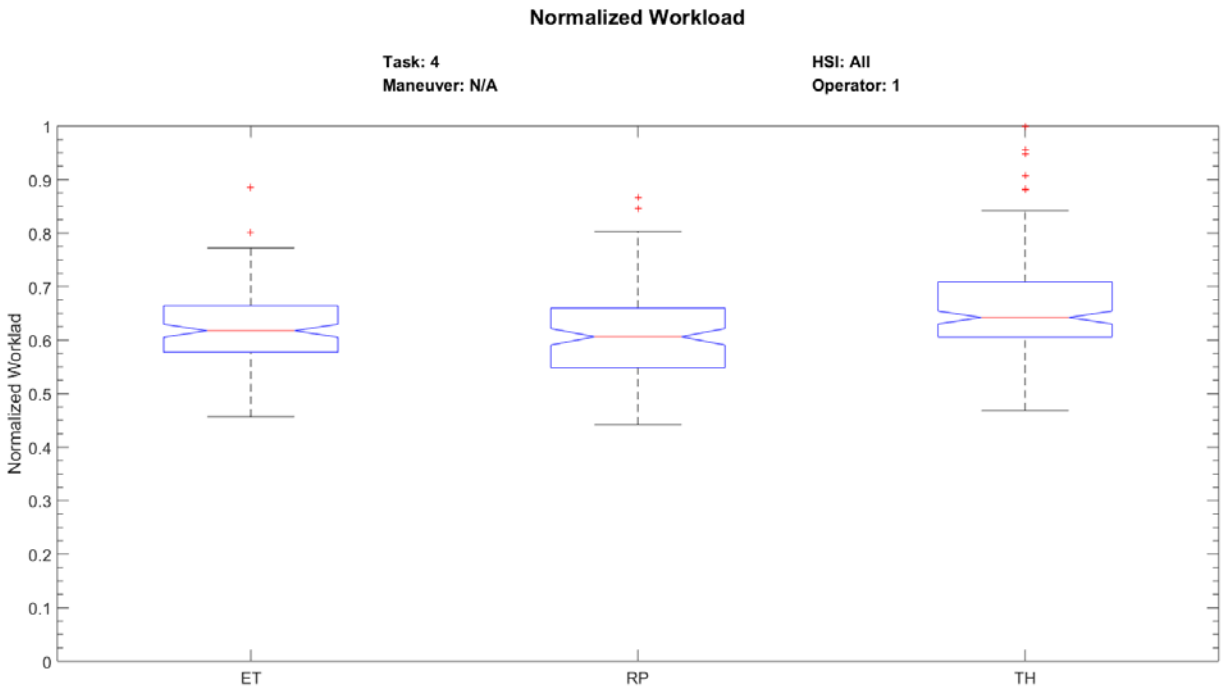

Normalized Workload & Maneuver Performance
Task: 3
Maneuver: LWL
HSI: All
Operator: 6
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

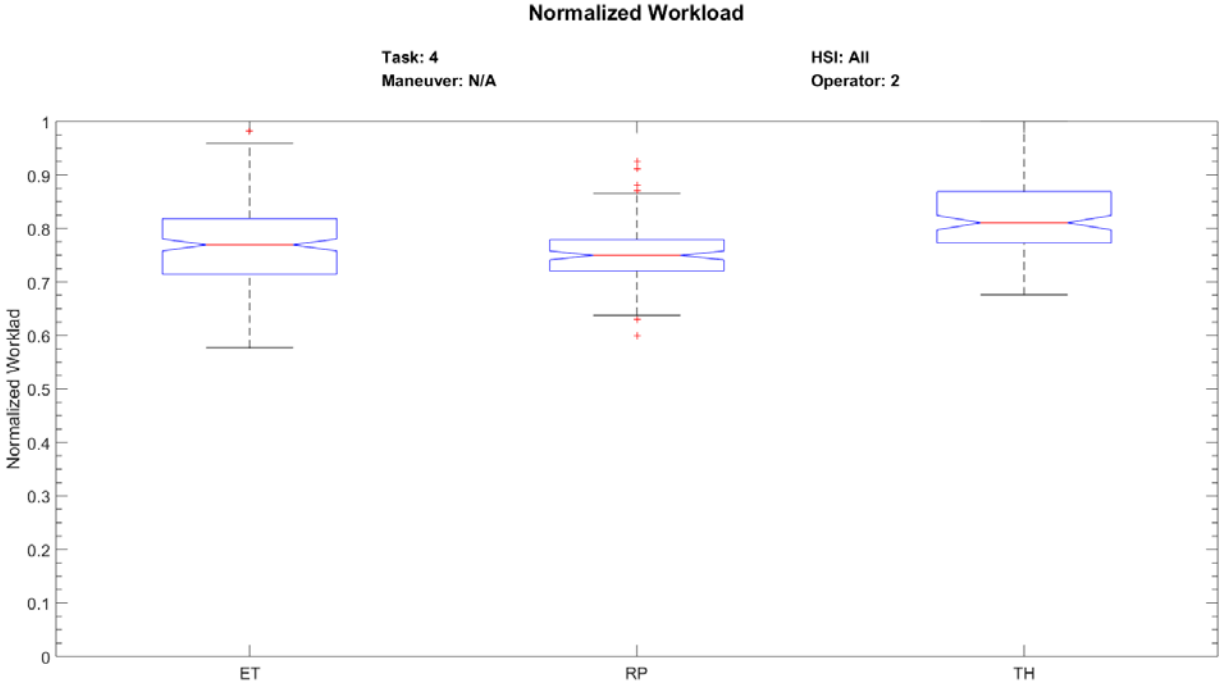

Normalized Workload & Maneuver Performance
Task: 3
Maneuver: HWL
HSI: All
Operator: 6
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

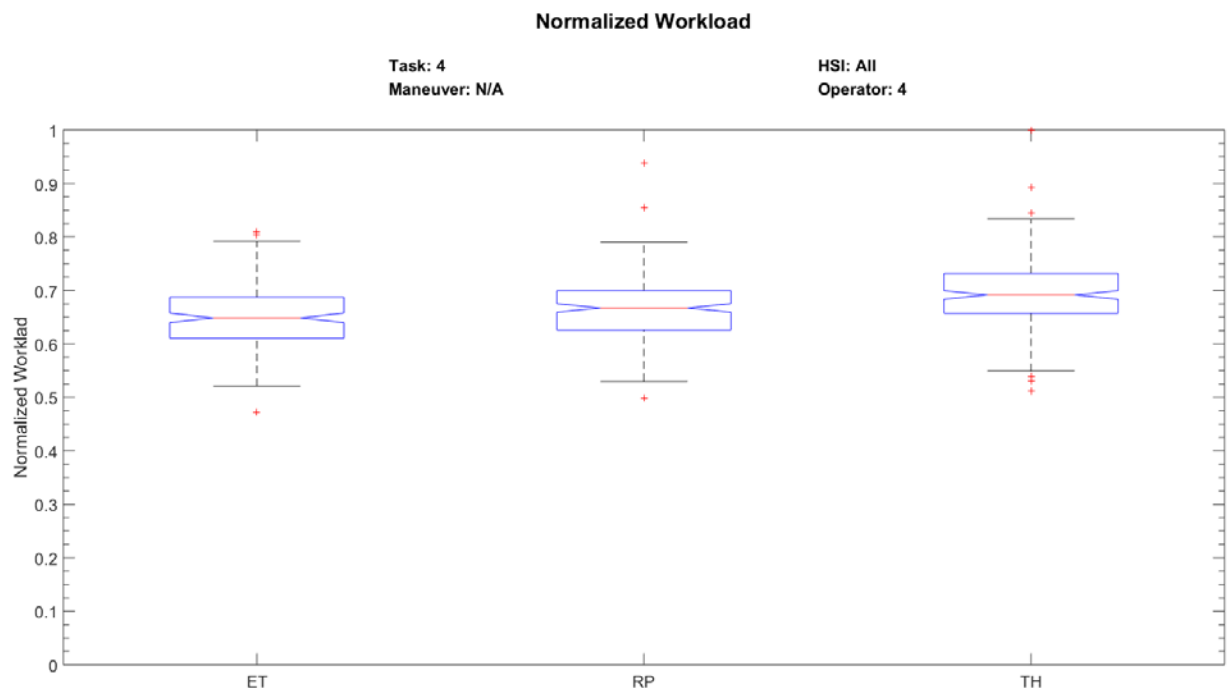

Normalized Workload & Maneuver Performance
Task: 3
Maneuver: LWL
HSI: All
Operator: 7
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

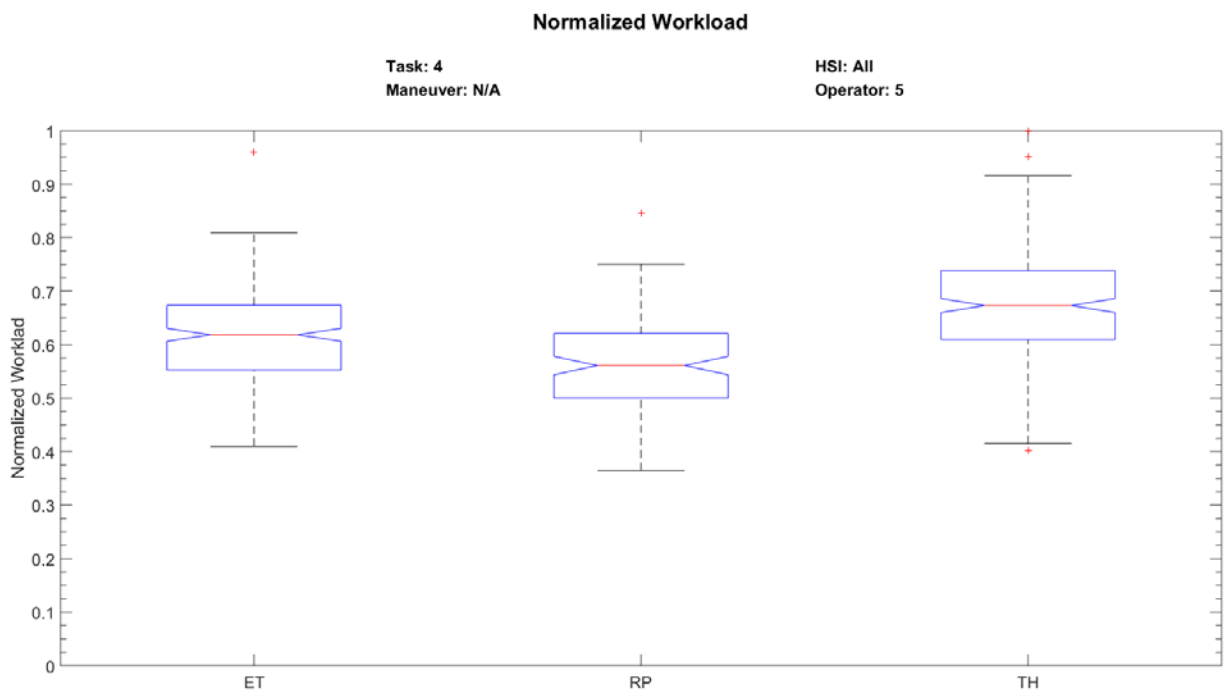

Normalized Workload & Maneuver Performance
Task: 3
Maneuver: HWL
HSI: All
Operator: 7
NAP
Normalized Worklad & Maneuver Performance
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ET
RP
TH

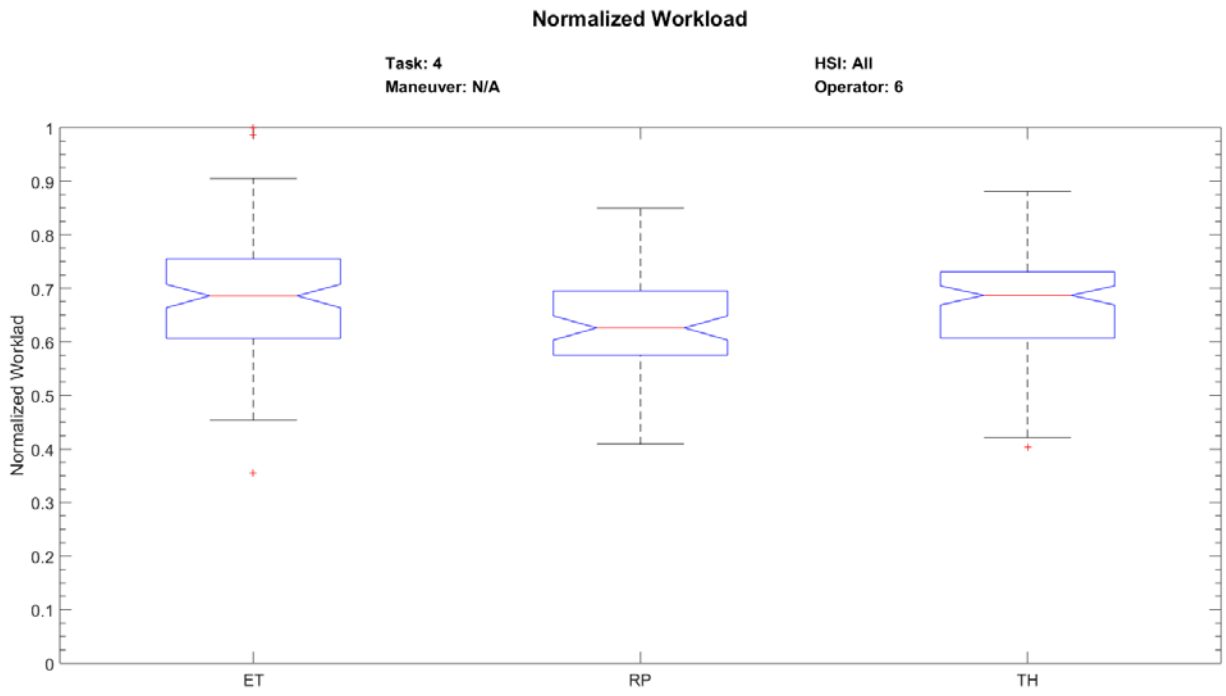
Task 4 – Workload






Normalized Workload
Task: 4
Maneuver: N/A
HSI: All
Operator: 4
Normalized Worklad
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
ET
RP
TH


Normalized Workload
Task: 4
Maneuver: N/A
HSI: All
Operator: 5
Normalized Worklad
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
ET
RP
TH


Normalized Workload
Task: 4
Maneuver: N/A
HSI: All
Operator: 6
Normalized Workload
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
ET
RP
TH

APPENDIX F – HSI USABILITY SURVEY RESULTS

An HSI usability survey was completed by each operator immediately after each flight and ground test for each HSI method. Within this survey, Military Utility will also be commented upon. An example survey is given below. Each additional HSI method survey is comprised of the same factors.

Very Unsatisfactory	Unsatisfactory	Marginally Unsatisfactory	Marginally Satisfactory	Satisfactory	Very Satisfactory
1	2	3	4	5	6
Task cannot be performed or the item is unusable or unsafe. Mission/task not accomplished due to equipment deficiencies or procedural limitations.	Major problems encountered. Task accomplished poorly or with great difficulty. Significant degradation of mission/task accuracy or accomplishment.	Minor problems. Task accomplished with some difficulty. Some degradation of mission/task accuracy or accomplishment .	The item or task meets its intended purpose with some reservations. Meets minimum requirements to accomplish mission/task.	The item or task meets its intended purpose; it could be improved to make it easier or more efficient.	The item or task is fine the way it is; no improvement required.

#	Factor	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.
1	Navigating across displays		
2	Slewing to portal of interest		
3	Slewing cursor to symbol and designating (single symbol displayed)		
4	Slewing cursor to symbol and designating in a dense target environment		
5	General accuracy throughout tasks		
6	Conducting a numerical data entry task		
7	Conducting a text data entry task		
8	Activating the appropriate data entry menu		
9	Usability for track designation during a tactical A-A engagement		
10	Comfort using Eye Tracking		
11	Overall usability using Eye Tracking		
12	Usability under G		
13	Usability under turbulence		

Additional usability comments:

Military Utility comments:

The following tables contain the full collection of HSI survey responses for all operators on all flights.

Table 1 – Operator 1, Flight 1, Eye Tracking Survey

#	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4	Actually pretty good, with some limitations under G that I have called out elsewhere. Generally speaking, eye tracking is great for gross navigation across displays. It allowed me to unconsciously plant the cursor. For precise navigation, it's poor; the cursor rarely snaps to the exact point I wanted. However, often that wasn't necessary.
2	4.5	For a gross slew action, eye tracking usually works. The only reservation is that the calibration seems to drift over time. If the calibration gets poor, it becomes more difficult to slew to the far left or right side of the display, as the eye tracker might think that you are actually looking off the MFD entirely.
3	4	The more calibration drifts, the more you have to slew after snapping the cursor.
4	4	Same.

5	3	Eye tracking accuracy is low.
6	3.5	Eye tracking works for data entry, but the snap location (due to generally poor accuracy) was usually pretty far from the intended location, leading to long slew times. For some reason the right half of the screen was less accurate than the left.
7	3.5	Same.
8	3.5	Same.
9	NA	Did not complete.
10	4	Other than the poor ergonomics of the helmet, eye tracking is actually pretty comfortable. The main problem is the two cameras that sit in the helmet and occlude your FOV. This makes cranium movement required to see the bottom of the display, which can be uncomfortable.
11	3	It's a 5 until calibration drifts, then eye tracking becomes marginal to unusable.
12	1	At 2 Gs eye tracking worked much like it did at 1 G. At 3.5 Gs eye tracking was completely unusable and I was unable to snap a cursor anywhere on the display despite multiple attempts.
13	NA	No effect of turbulence on eye tracking specifically observed.

Additional usability comments:

Military Utility comments:
This seems like a technology that I actually could find helpful, but it would need to overcome the calibration drift. If eye tracking's mean performance could be brought up to the level of the highest performance I witnessed, this technology could significantly improve my ability to maneuver a cursor across large displays.

Table 2 – Operator 1, Flight 1, Rhino Pointing Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4	Navigation works fine but the required cranium motion is annoying and causes significant neck strain by the end of the sortie.
2	5	RP certainly works for this, as you tend to look in the direction of your POI anyway.
3	3.5	See comment for line 1. I could pretty much always direct-designate a track, but this meant that I had to be extremely deliberate with my cranium motion. I found this requirement to be deliberate both physically annoying and mentally taxing. Also, if the symbol was towards the bottom of the display, I lost a peripheral view of the VHUD as I designated a track.
4	4	Somewhat better than #3, because on the expand operation I didn't experience the compulsion to designate as precisely.
5	4	Quite accurate if I took the time to refine it.
6	5	Quite good for this task. The main improvement would be to move the numerical keypad to the center of the display, as looking right to make the entries was physically cumbersome.
7	3	Text entry was way worse than numerical entry because of the size of the keyboard (extended across the entire display). This necessitated that I move my cranium many more degrees than during eye tracking and tactile data entry.
8	5	No different than item #6
9	NA	Did not perform
10	3	Caused pain.
11	4	Fairly usable, but the physical workload detracts from task performance
12	4	G didn't actually seem to affect RP very much. If anything, it actually stabilized the boresight somewhat. The left-and-right cranium movements felt more abnormal than the slight up-and-down deviations caused by G.
13	2.5	I experienced turbulence on 2 points: a data entry point and a Task 2 point (lines 21 and 22 on my execution sheet). Turbulence significantly degraded my rhino pointer. I still could point in the "general area" for RP, but the boresight cursor jumped significantly, causing me confusion about where it was pointing and the loss of a direct designation capability.

Additional usability comments:

Military Utility comments:
The idea of using your cranium to activate specific portals or areas of your display is intriguing and may have several applications. However, right now there are two serious problems that limit military utility. One is how deliberate and conscious your cranium movements have to be to precisely point on a large display that stretches from far left to far right; this is distracting and reduces my focus on other tasks. The other is the ergonomics; comfort and occlusion of the primary flight reference while rhino pointing must be improved

Table 3 – Operator 1, Flight 1, Traditional HSI Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	5	Slow to slew across large distances.
2	4	Need a better way to rapidly switch portals on different sides of the large display.
3	5	Highly accurate but lacks a magnetization or stickiness in the vicinity of a track.
4	4	The way this is programmed is way too slow. After expanding, the cursor does not maintain its relationship to tracks in the vicinity of the expand location; instead it remains in the same location on the display. This means that if you expand near a track on the edge of the display, that track now moves (zoomed in) to the center of the display, while the cursor remains far on the side of the display. This requires another slew across half of the entire display.
5	5.5	Cursor slew is extremely accurate though in general it needs magnetization or stickiness.
6	4	Accurate but slow
7	3	Accurate but extremely slow due to the larger keyboard
8	4	Accurate but slow
9	NA	Not accomplished
10	6	No problems with comfort
11	5	See previous comments.
12	4.5	It becomes somewhat harder to watch the cursor on the MFD under G; this is mainly a visual limitation and the cursor slew switch itself works fine under G.
13	5	Turbulence has minimal effect.

Additional usability comments:

Military Utility comments:
This is a known technology that has high military utility. It needs improvement based on the large display.

Table 4 – Operator 1, Flight 2, Eye Tracking Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4	See sortie 1.
2	4.5	See sortie 1. The same tendencies were noted in the operational task.
3	4	See sortie 1.
4	4	See sortie 1.
5	3	See sortie 1.
6	3.5	See sortie 1.

7	3	See sortie 1. I would definitely prefer not to use this method for text input; it's so many characters that it becomes quite irritating to snap and slew for every single character.
8	3.5	See sortie 1.
9	4	The mental burden of using eye tracking during the operational task was very low. I had quite a good calibration, and found that my ability to jump portals unconsciously and effortlessly was blazing fast. If the track I desired to designate lay near the edge of the screen, I often experienced delays because the system would not respond to a snap input (thinking my eyes were off the screen). Other than that it was honestly pretty awesome, just inaccurate and therefore wholly inadequate for precision designation. Luckily, the buttons for missile allocation were sufficiently large that the eye tracker usually enabled me to snap a cursor within their boundaries and actuate the button immediately.
10	4	Totally comfortable except when having to bias your eyes to get the cursor to "believe" you are looking at the screen. That gets awkward and mentally taxing.
11	3.5	Slightly increased score from sortie 1 due to the success had with the operational task. Keep in mind that this score depends strongly on the quality of the calibration -- I had a good one.
12	2	I was successful in getting the system to work under G (unlike in sortie 1). Unfortunately, significant pilot compensation had to be applied in one of two ways to get the cursor to even appear on the screen. I had to either (1) look about 6-9" BELOW where I wanted the track to appear (at 3.5 Gs), and then execute a cursor snap or (2) hold up the mask or helmet with one hand to undo the drooping effect caused by load factor. Since I could rarely compensate for the exact amount of G by lifting my mask upward, this further degraded the already low eye tracker accuracy.
13	3.5	Same as overall usability, as I didn't note any degradation.

Additional usability comments:

Military Utility comments:

Table 5 – Operator 1, Flight 2, Rhino Pointing Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4	See sortie 1.
2	5	See sortie 1.
3	3	The tendency with RP is to direct-designate a symbol even when it's not necessary. i.e., because you can see the cursor, you almost always try to point that cursor as accurately as you can. This requires more neck engagement and more mental focus and detracts from performance overall. It would actually be better for performance if you used RP more like ET and just "threw" the cursor in the direction of the track. However, the fact that the boresight is in the center of your nose, and the fact that you can see the boresight make this physically and mentally difficult to execute.
4	4	See sortie 1.
5	4	See sortie 1.
6	5	See sortie 1 and comment for line #3.
7	3	See sortie 1 and comment for line #3.
8	5	See sortie 1.
9	3	RP degraded performance in the operational environment. The tendency to try to direct designate (line #3) was a factor. However, it also became mentally taxing to track the two cursors. When rapidly switching between portals and between tracks, I would often try to slew my rhino boresight, or I would try to snap at an inappropriate time, forgetting that I had to move my rhino boresight deliberately first. Overall, tracking two cursors elevated mental workload significantly, and detracted from task performance.
10	3	See sortie 1.
11	3.5	Operational usability in question, due to direct-designation tendency and the mental workload from tracking two cursor locations.
12	4	See sortie 1.

13	2.5	See sortie 1.

Additional usability comments:

Military Utility comments:

Table 6 – Operator 1, Flight 2, Traditional HSI Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	5	See sortie 1. The task seems programmed to always force you to slew all the way from one side of the display to the other; this is not how the task should have been programmed, and I did not rate traditional HSI lower because of this.
2	4	See sortie 1.
3	5	See sortie 1.
4	4	See sortie 1.
5	4	Data entry accuracy using the touchscreen is questionable. Accuracy with the cursor is extremely high.
6	3.5	Touchscreen entry on a small pad was executable, but significantly detracted from task performance (or at least increased my subjective workload) during the LWL/HWL maneuvers. Under G, touchscreen is highly problematic. Luckily, the screen was low enough that I could brace my forearms on my legs while interacting with the screen; had I not been able to do this it would have been much harder to stabilize my hand while interacting with the screen.
7	2	Touchscreen was wholly inadequate for interacting with the whole screen, particularly under G. Bracing became awkward and cumbersome. It was a two-handed operation requiring me totally releasing the controls. Accuracy was terrible; I often hit a key two or more times while trying to hit it just once. I also hit adjacent keys due to turbulence or difficulty bracing.
8	4	Sort of fine.
9	5	Tactile input is a proven method for this and works great. Unfortunately it's extremely slow to get from one portal to another. Designating tracks on the right-hand portal and then on the main portal took long enough that I would become slightly task overloaded by the number of tracks that had appeared in the interim.
10	6	See sortie 1.
11	4	See sortie 1. Additional deficiencies as noted in #6 and #7, entirely related to using touchscreens while operating the aircraft precisely or while under G.
12	3	Cursor slew under G works great. Touchscreen under G spans the spectrum from degraded to unusable.
13	3	Cursor slew under turbulence works great. Touchscreen under turbulence spans the spectrum from degraded to unusable.

Additional usability comments:

Military Utility comments:

Table 7 – Operator 2, Flight 1, Eye Tracking Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	5	
2	5	
3	3	Fine tuning required with the CSDS but there were a few times when it was not required. Accuracy of the eye tracker degraded when the head was not looking straight forward and/or under 2 and 3.5 Gs. 3.5 Gs required compensation as the tracker thought I was looking about 6 inches higher than I actually was.

4	4	Gross tracking for the initial expand was good but see previous comment for fine tracking.
5	4	See previous
6	2	Not useful as a standalone data entry method with 1" square buttons. Required fine tuning with the CSDS.
7	2	See previous
8	2	See previous
9	N/A	Task 4 not executed
10	5	
11	3	The time consuming calibration process on the ground didn't even result in a perfect result. A quicker way to calibrate while flying is needed. Gross tracking is great, when not under G but a more rigid scene camera could help. Fine tracking is not quite there but there were times when it was spot on. The eye cameras also occluded the vision down toward the bottom of the display and the pilots kneeboard card.

Additional usability comments:

Military Utility comments:

Table 8 – Operator 2, Flight 1, Rhino Pointing Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	6	
2	6	
3	4	High frequency small movements, less than 1/8 inch (jitter) would happen with the boresight when trying to stabilize on one spot.
4	4	Good for initial expand but deliberate focus was required to get the cursor directly on the track to designate due to the jitter which took time away from cross checking flight parameters.
5	5	
6	4	Jitter caused some errors during input. May have been caused by requiring two inputs, TMS aft and then forward, to have a designation.
7	4	See previous
8	4	See previous
9	N/A	Task 4 not executed
10	5	
11	4	Adjustments to the fine tracking needed to make it much more useful (damping out the boresight).

Additional usability comments:

These comments are based on the baseline task only.

Military Utility comments:

I will complete the rest of the tasks prior to a final assessment but there are no show stoppers as of now.

Table 9 – Operator 2, Flight 1, Traditional HSI Survey

#	Rating	Comments

		Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4	Slew rate too slow for the large movements from side to side of the display.
2	4	Slew rate too slow for the large movements from side to side of the display.
3	3	Slew rate too fast for very fine tracking without a magnetized cursor.
4	3	Slew rate too slow for large movements and too fast for fine movements.
5	4	Good for other than fine tracking
6	3	Double taps of buttons occurred multiple times due to buffet in the jet even during basic maneuvering.
7	3	See previous
8	3	See previous
9	N/A	N/A
10	4	Good except for under G… removing hands and using the touchscreen at 2 and 3.5 G took my entire attention to get right.
11	4	Digital slew rate instead of gained based on deflection is the main issue with the CSDS. The touchscreen forced multiple double taps and errors when inputting data.
12	3	Using the touchscreen for data entry was challenging
13	4	Using the touchscreen for data entry was challenging but errors could be fixed before releasing the input

Additional usability comments:

Military Utility comments:

Traditional HSI has proven itself throughout its past and actual CSDS that are not digital are much more user friendly. A magnetic cursor would also help with the fine selection of tracks when needed. If data entry is planned at greater than 1 G, thought needs to be put in to figure out how to stabilize the users hand and to implement a way to avoid doubletaps.

Table 10 – Operator 2, Flight 2, Eye Tracking Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	5	
2	5	
3	3	Sometimes it's close and sometimes it's not but gets you in the ballpark. The accuracy was better on the right side of the screen versus the left which made it difficult to compensate for a consistent offset.
4	4	Great for gross acq but see previous comment.
5	2	Sometimes it's close and sometimes it's not but gets you in the ballpark. The accuracy was better on the right side of the screen versus the left which made it difficult to compensate for a consistent offset.
6	2	Not accurate enough with 1inch x 1inch buttons
7	2	Not accurate enough with 1inch x 1inch buttons
8	2	Not accurate enough with 1inch x 1inch buttons
9	3	Not accurate enough to direct designate but gross acquisition was often good enough to directly select the missile
10	4	Eye tracking cameras blocking view down toward kneeboard
11	3	Having to recalibrate throughout flight was a pain in addition to previous comments about lack of consistent accuracy

Additional usability comments:

Military Utility comments:

Table 11 – Operator 2, Flight 2, Rhino Pointing Survey

#	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.
1	5	Very quick and easy to use for the central part of the display… objectionable head movement required to get all the way to the edges
2	5	
3	4	Jitter throughout even with unnoticeable buffet. When the buffet increased it was impossible to direct designate with how much the boresight was bouncing around
4	4	See previous… no issues with gross acquisition
5	4	See previous
6	4	Objectionable amount of head movement required to get all around the keyboard.
7	4	See previous
8	4	See previous
9	4	Jitter throughout even with unnoticeable buffet. When the buffet increased it was impossible to direct designate with how much the boresight was bouncing around. For gross acquisition it is a very fast and with the large missile buttons easily selectable.
10	5	
11	4	Jitter in buffet, edges of screen require large head movements
12	4	The jitter seemed to stabilize but a large amount of effort is required to hit the corners of the display
13	4	Fine for gross tracking but fine tracking very difficult due to the boresight bouncing around.

Additional usability comments:

Military Utility comments:

Table 12 – Operator 2, Flight 2, Traditional HSI Survey

#	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4	Slow for large movements
2	4	See previous
3	4	Too fast for small movements without cursor magnetization
4	4	See previous
5	5	
6	5	
7	5	
8	5	
9	4	Too slow for large movements and too fast for small movements
10	5	

11	5	Touchscreen difficult under G without a place to anchor my hand.

Additional usability comments:

Military Utility comments:

Table 13 – Operator 3, Flight 1, Eye Tracking Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	3	When it works Eye tracking is very fast and easy to use. However eye tracking doesn't work often for me. Having to snap cursor multiple times and unsure where it will place often times.
2	3	Same as above.
3	2	Same comments for all when it works eye tracking feels like magic. However often doesn't work for me.
4	2	
5	2	
6	2	
7	2	
8	2	
9	N/A	
10	5	When it's working slewing and using eye tracking has small head movements and can be very comfortable.
11	2	
12	2	Under G seems to become less accurate and is uncomfortable trying keep head in correct placement for accurate eye tracking.
13	N/A	

Additional usability comments: Feels like magic when it works. For me the way the helmet and mask fit they interfere with each other, possible reason for issues experienced. Found opening my eyes wide and holding my head still and holding my breath in an exhale made eye tracking work better but during inhale or exhale and talking the eye tracking was moving even when staring at the same point.

Military Utility comments:

Table 14 – Operator 3, Flight 1, Rhino Pointing Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	5	Liked being able to quickly place cursor across large display quickly
2	5	Same above good for large gross movements quickly.
3	4	Good for large movements but fine pointing refinement with head movement can be uncomfortable or challenging.
4	3	Same as above. But the accuracy required for the designation is challenging. Good for the large movement of the cursor but Hotas slewing often needed at the end of the task.
5	4	Mostly good accuracy the Rhino cursor provides feed back to where it will place the cursor but fine tracking is hard.
6	3	Large head movements required to traverse entire display can be uncomfortable.
7	3	Same as above.
8	3	Same as above.

9	N/A	
10	4	For extremes of display more movement is required than would be desired. Causes longer time away from crosscheck.
11	4	Good for gross movements of the cursor across the display. Would be a good supplement to Traditional hotas.
12	3	Under 2G not to bad to move head for large movements under higher G becomes uncomfortable to move head and look down at display. The fine pointing with Rhino becomes very hard and but still good for gross movements if required.
13	3	In turbulence the bounce of the head takes away the fine pointing but for gross movements its good. Could be improved with some kind of filter for head movement.

Additional usability comments: Had the Rhino Boresight set fairly low made the head movements across the screen more comfortable. Was able to keep head in a upright position to move the cursor horizontally across the display easily and had small up or down head movements to place vertically on display.

Military Utility comments:

Table 15 – Operator 3, Flight 1, Traditional HSI Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4	Slewing across large display is slow and requires several seconds of constant input.
2	4	Same as above.
3	4	Can be slow due to distance. If cursor is close cuts the time down significantly. Can be very accurate.
4	4	Same as above.
5	5	Generally most accurate method.
6	4	Touch screen entry can be fast but prone to double entry and must take hand off controls or switch hands to reach opposite sides of the display.
7	4	Same as above.
8	4	Same as above.
9	N/A	
10	5	When hands on throttle no major head movements required generally as comfortable as flying in a normal aircraft.
11	4	Wish I could snap the cursor across the display quickly good for fine accurate placement of the cursor.
12	4	Slewing works under G but touchscreen entry takes hands off controls and becomes more difficult as G increases.
13	4	Same as above turbulence can lead to incorrect entry with touch screen.

Additional usability comments: Works how you would expect.

Military Utility comments:

Table 16 – Operator 3, Flight 2, Eye Tracking Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4	Today was working better than before but when it is working is almost like magic. Can be fast and fairly accurate to move across the display. Not having a second pointer that is slaved to where you are looking is less distracting.

2	4	Same as above. Love it when it works.
3	4	Can be very fast and put the cursor within an inch or so. Often not sure in which direction the cursor will be offset so there is a moment of reaction time to figure out what direction to slew after snappingthe cursor.
4	4	Same snapping near the symbol for expand its nice because it just has to be close not right on. Then snapping to symbol after expand is quick and just requires a small amount of refinement.
5	3	Not super accurate but good for moving the cursor quickly.
6	2	ET would have to be much more accurate to do direct entry with data but good for getting close. Without the feed back
7	2	Same as above.
8	3	Can place the cursor near the menu items better when the screen is less cluttered.
9	4	Eye tracking felt like magic when a track would pop up, look, and snap and the cursor was close and only required minor refinement. Designate, look at large Menu buttons, snap and select. Felt like the future when it works.
10	3	Overall Nice to not move your head around but seems to require holding your head still to have ET work.
11	3	Has a lot of potential, could supplement HOTAS greatly.

Additional usability comments: Needed to be calibrated often but worked better today. The operator had to pull the mask further down on his face as in the last flight it was touching the visor of the helmet where the eye tracking cameras are located. Had better results but was more uncomfortable.

Military Utility comments: Repeating from before when it is working feels like its magic or from the future. Had some hick ups and needed to be recalibrated frequently but was potential.

Table 17 – Operator 3, Flight 2, Rhino Pointing Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4	Fast to move across large display. But requires large head movements and takes away from cross check time when heads down and moving the cursor.
2	4	Same as above. Nice for quickly moving to another portal of interest.
3	4	Can move quickly across screen. The fine tuning of the cursor with only RP can be uncomfortable. Best technique is get the cursor close with RP and finish the fine turning with the slew.
4	3	Quick for large distances hard to place the cursor precisely in dense track. Can be confusing with multiple cursors in the same area as all of the tracks.
5	3	The cursor provides feedback and helps accuracy but fine tuning requires fine head movements that are uncomfortable and time consuming. Good for gross movement and fine tune with traditional slew.
6	3	Possible with rhino cursor to do direct entry of data but requires a decent amount of time. Better to move cursor large distance and slew or use touch screen.
7	3	Same as above.
8	3	With a large menu button can easily and quickly select a menu.
9	4	Made moving between portals quick. Best for gross movements across the screen or selecting the large menu buttons.
10	3	Rhino pointing in its current implementation requires significant head movement that can be uncomfortable. Under Higher G it is very uncomfortable especially for fine tracking like trying to do direct entry data or place the cursor right on the target with RP.
11	4	Has a lot of potential good for large movements.
12	3	Becomes harder to Use under G.

13		

Additional usability comments: Has potential, can be combined with HOTAS to improve the current interaction with Large displays. Possibly remove the pointer icon and use only for gross movement of the cursor or for selecting the portal of interest.

Military Utility comments:

Table 18 – Operator 3, Flight 2, Traditional HSI Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	3	Slow across large distance.
2	3	Slow across large distance. Using the touch screen to select portal of interest could be good way to quickly select potals
3	4	Slew rate slow if symbol is across the display but is predictable and fairly accurate.
4	3	Slow to slew long distances and having to slew from expand to the new location of the target can be slow.
5	5	Predictable and accurate just slow.
6	4	Can be fast and accurate, becomes harder under G and with ther current set up easy to make a double tap on selection.
7	4	Same as above. Must take hand of controls to input data.
8	4	With large enough menu buttons easy to select button with touch screen.
9	4	Predictable, slow slew makes moving between portals undesirable. If you could use touch screen to select portal could speed up task. However that requires you to take your hands off the controls.
10	5	
11	4	Only slow slew is biggest issue. Could be supplemented with RP or ET to have best results.
12	4	Touch screen becomes difficult under G HOTAS works.
13		

Additional usability comments: This is the existing method it works for what it is supposed to do but could be improved with RP or ET.

Military Utility comments:

Table 19 – Operator 4, Flight 1, Eye Tracking Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4	When the eye tracking was on and calibrated it was amazing for gross movements. Near the edges of the screen I had to compensate by looking to the center of the screen in order to get the ET to snap to my gaze. Again, this was good for moving across the screen but did not help in fine.
2	4	Same as 1
3	4	On its own, the ET would typically get close, but HSI slewing was required to designate. However, this was not found to be objectionable. Contrarily, I found it to be awesome. Just by looking I'm able to quickly slew my cursor "close enough" and quickly use tactile HSI to finely tune and designate. The two HSI methods complimented each other very well.
4	4	Same as 3
5	4	On its own, it gets "close enough" usually. BUT, with a good calibration and good conditions it was spot on and was like absolute magic.
6	3	Reference above comment. It requires repeating this process for each input.

7	3	Same
8	3	Same
9	NA	
10	6	Don't even have to move my head.
11	4	With a bit more maturity this would be amazing. This is opposite of my going in preconceptions. I am very interested in its future.
12	3	Currently, you have to compensate big time under G. I had to look at the bottom of the screen in order to get the tracker to recognize and snap. The cursor would then cue high on the screen. I assume this could be charted and corrected for. As it is now, it requires extensive compensation.
13	NA	Did not experience.

Additional usability comments: I think this is worthwhile to continue development. The ability to just look and throw my cursor is great.

Military Utility comments: This will easily speed up the kill chain once the tech matures a bit more. Continue to develop.

Table 20 – Operator 4, Flight 1, Rhino Pointing Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	5	Even with the additional jitter from actually flying, the gross movement was still very predictable and manageable.
2	5	Same comments as 1.
3	3	The additional jittering from actual flying became evident even not in buffet or g. With the cursor displayed, I found it took me longer to get a designation due to trying to get the cursor as close as I could before snapping. Also, moving your head under g is no bueno.
4	3	Same comments as 3.
5	4	Gross movement is good. It gets you close. But you have to move your head more than I thought you would. Fine movement is difficult especially under g.
6	3	Gross movement is good. The fine act of data entry gets tiresome (especially under g) and requires keeping your eyes on the screen for a significant amount of time and not monitoring flight conditions.
7	3	Same
8	3	Same
9	NA	
10	3	It works, but requires a lot of head down movement in the current configuration. This especially becomes an issue under g.
11	4	RP has great potential. As an idea it works. It still has a few inherent issues involved with head movement under g and maneuvering.
12	3	Uncomfortable to be heads down and moving under g.
13	3	Jitter becomes much more noticeable in turbulence/buffet.

Additional usability comments: I love the idea of RP. The execution of it has shown some of its potential weaknesses.

Military Utility comments: This is a potential solution for the gross movement side of large displays. With a few tweaks the fine tracking would be good as well as long as you're not maneuvering.

Table 21 – Operator 4, Flight 1, Traditional HSI Survey

#	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.
1	5	IAW the 5, it works fine but could be made more efficient.
2	5	
3	5	
4	5	
5	5	
6	4	The touchscreen is awesome unless you are maneuvering. Also, I still don't know how you would activate the buttons in the weather or in formation. I miss tactile buttons.
7	4	Same as 6
8	4	Same as 6
9	NA	
10	4	All good except touchscreen under g. That was not pleasant.
11	5	It works. Just subject to the same limits it always has been.
12	4	Touchscreen under g is not easy.
13	4	Touchscreen allows no place to anchor your hand to steady it.

Additional usability comments: NA

Military Utility comments: I think the combination of tactile HSI with another HSI such as the RP or ET would complement each other well. You could get the gross movement benefits of the ET/RP and the fine abilities of the tactile.

Table 22 – Operator 4, Flight 2, Eye Tracking

#	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.
9	4	Amazing for rapidly moving the cursor when it was in calibration. It greatly reduced my qualitative workload and was able to maintain my intercept geometry with much less effort.
10	5	
11	4	G and sun angle hurt it. With a little higher TRL this would be amazing.

Additional usability comments:

Military Utility comments:

Table 23 – Operator 4, Flight 2, Rhino Pointing Survey

#	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.
9	4	This system again worked very well for gross slewing across portals. With single tracks it was easy to get designations, but when the scope got cluttered it was very difficult get a precise designation and I would have to swap to tactile inputs.
10	4	Moving around and staying heads down during an intercept lead me to climb 1k ft. This shows that I was forced to not fly the jet as much in order to manipulate the system and designate.
11	4	It works. It just requires a lot of time heads down.

12	4	You have to fight it under G, which is not comfortable, but it works.
13	3	Even with slight buffet the cursor jumped very noticeable.

Additional usability comments:

Military Utility comments:

Table 24 – Operator 4, Flight 2, Traditional HSI Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
9	4	Really, the same comments as the first flight. It works, it is just limited by the slew rate. Which results in slow gross acquisition.
10	5	
11	5	
12	4	Touch screen under g still sucks.
13	4	Touch screen under buffet still sucks.

Additional usability comments: Fantastic for fine control.

Military Utility comments: Combine with another method to get faster gross movement and it would be awesome.

Table 25 – Operator 5, Flight 1, Rhino Pointing

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	5	Great at moving large distances quickly, intuitive direction
2	5	Easy to know which part of the screen the cursor is going to
3	5	Easy to move rhino-boresight to single target. Seeing cursor displayed on the screen gave complete assurance that the cursor would land at the intended location. Rhino-boresight moved to target location with a quick movement of the head.
4	5	Cursor can quickly be relocated to an area near the cluster to initiate the expand action. After expand, the rhino-boresight was easily relocated to designate.
5	4	Some jitter was present when attempting to stabilize on a single point. While attempting to lock onto the target, both the light vibrations in the aircraft and the natural oscillations of the head caused the rhino-boresight to wander within 5 mm of the target. Accuracy was dependent on when the snap and designate occurred.
6	4	Keeping the head down to fully complete the data task reduced flight condition SA.
7	3	Moving across the keyboard was easy, and seeing the motion of the rhino-boresight gave immediate feedback that the correct key would be selected when snapping and designating. Several times during the elevated-g maneuvers, the rhino-boresight disappeared from the MFD due to the drooping of the head. This led to a delay while the operator cycled his head in an attempt to reestablish the position of the rhino-boresight.
8	4	Moving across the menu tree was easy. Downgrade due to reason given in #6.
9	N/A	
10	3	The weight of the helmet and the necessary movement of the head caused discomfort throughout the flight, but especially during the elevated-g maneuvers.
11	4	Good robust system that always gave the expected result. Reduced rating for head discomfort and difficulty during elevated-g maneuvers.

Additional usability comments:

Military Utility comments:

Table 26 – Operator 5, Flight 1, Traditional HSI Survey

#	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4	Cursor slew rate limit requires waiting 2-3 seconds to cover full span of the display
2	5	Cursor slew direction is intuitive.
3	4	See comment on #1, and lack of variable cursor speed makes fine tracking difficult.
4	4	See comments #1 and #3.
5	3	Fine tuning was difficult since only maximum speed was available.
6	4	Touchscreen keypad was easy to use in a 1-g environment. I had to brace my right arm on my right knee for accurate numerical data entry during the elevated-g maneuvers.
7	4	See comment #6.
8	4	See comment #6.
9	N/A	
10	5	No issues using HOTAS. No large body movement required.
11	4	Slight rating reduction due to slew rate limiting and difficulty using touchscreen during elevated-g maneuvers.

Additional usability comments:

Military Utility comments:

Table 27 – Operator 5, Flight 2, Eye Tracking Survey

#	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.
1	5	ET was very quick to make large jumps across the screen. The lack of cursor gave no preview of where the cursor would land after snap, leading to some surprises.
2	4	If the portal of interest was near the center of the screen, the ET snap was often in the correct place. Slewing to points near the edges of the MFD made ET snaps more difficult, since the ET software often assumed that I was looking off the MFD.
3	3	The ET snap got the cursor close very quickly, but nearly every instance of the task required additional HOTAS slewing to get the cursor on the target, slowing the process and increasing workload.
4	4	The initial ET snap almost always got the cursor close enough for the expand command, and quickly snapped the
5	2	ET snaps were almost always over 5 cm away from the intended target. Additionally, the seemingly random direction of placement of the cursor after each snap increased workload and frustration.
6	2	The cursor nearly always snapped to a different key than desired. This required HOTAS slewing to redirect the cursor, at times taking longer than would have been necessary if only HOTAS was used. The ET accuracy was especially bad on the right side of the MFD where the number keypad is located, making the input of frequencies very difficult.
7	2	See #6
8	3	Since the menu selection buttons are close together, ET could get the cursor to the menu input area rapidly, but still required HOTAS slewing for terminal guidance.

9	3	The rapid ability to move between portals was helpful. Since the cursor was white and the snapping point couldn't always be predicted, the cursor would occasionally get lost in the clutter of the radar portals. Snapping back to the missile selection buttons was quick, but still often required additional HOTAS slewing to get the cursor in the correct location.
10	5	Minimal movement of the head increased comfort. The location of the ET cameras decreased FOV.
11	3	Task time and workload were increased due to the inaccuracy of ET. The lack of visible focus-boresight made it difficult to predict where the cursor would appear after snapping.
12	2	Under 2g the ET was degraded, but snapping was still possible with only slightly less accuracy. ET was unusable under 3.5g. To get the cursor to snap at all required pushing up the mask with the left hand and looking in the center of the screen where ET had the best reliability.
13	5	ET was relatively unaffected by large vibrations due to turbulence or buffet.

Additional usability comments:

Military Utility comments:

Table 28 – Operator 5, Flight 2, Rhino Pointing Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
9	5	RP allowed me to get the cursor back and forth between portals rapidly. After snapping the cursor close to the target, I used the HOTAS for fine tracking, then quickly jumped back to the large missile selection buttons to complete the task. I had no issue knowing exactly where the cursor would snap to since the RP boresight was constantly visible. Direct designate was possible for stand-alone targets, but not for targets in tight groups. The constant shake of the RP boresight made direct designation of a target in a dense group difficult.
10	3	Holding the head still for precise designations was tiring, and required breaks every 30 seconds or so. Designating the large missile selection buttons was easy and comfortable.
11	5	Improvement needed to filter out the jerky vibrations and possibly a head-movement gain increase to reduce the amount of head movement needed to drive the cursor across the screen.
12	4	Increased-g made large head movements more difficult, but not impossible since the cursor could still be found on the MFD. Additionally, by opposing the additional load of the helmet, the rhino-boresight actually became more stable on the MFD.
13	3	High levels of turbulence and buffet made RP less accurate requiring more HOTAS slewing, but not unusable.

Additional usability comments:

Military Utility comments:

Table 29 – Operator 5, Flight 2, Traditional HSI Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
9	4	Using the touchscreen to select missiles was very easy. Using quick touches with the right hand to quickly move the cursor between portals, then using the HOTAS to slew the cursor to its intended target was an efficient process. Removing a hand from one or both controls made maneuvering difficult, since engaging different targets required turns and trim/power changes.
10	4	The location of the cursor slew switch and the designate switch on different hands meant that one ability or the other wasn't possible when removing a hand to touch the touchscreen. Comfort on the controls was not an issue.
11	4	HOTAS was marginally effective, but the inability to quickly move across large screen distances without removing hands from the controls degraded its usability.
12	3	Moving a hand to the touchscreen during Task 3 under elevated-g was difficult.

Additional usability comments:

Military Utility comments:

Table 30 – Operator 6, Flight 1, Rhino Pointing Survey

#	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.
1	5.5	The RP cursor was quick and easy to make gross inputs.
2	5.5	Same as above.
3	4	Gross movements were quick. The amount of head movement was objectionable, especially when the track was on the far side of the screen and low.
4	4	Same as above. It was quick and easy to designate the track, but the head motion was objectionable. It was worse for the Task 2 because it required more head motion. With the weight of the helmet and the amount of connections to the helmet, the head motion was objectionable.
5	6	I was able to direct designation on 90% of the tasks without having to update the position with the slew switch
6	4	I would have preferred to enter the tasks in with the touchscreen. Data entry was slower and more cumbersome. Head motion was objectionable.
7	4	Same as above.
8	5	Selecting a menu without data entry was acceptable.
9	N/A	N/A
10	2.5	The head motion was objectionable and highly uncomfortable for the 1.3 hour sortie.
11	4	RP is definitely usable. The pointer was highly accurate and gross movements were quick. For baseline maneuvering, it was nice to use. However, as soon as I started maneuvering, RP was objectionable. For data entry, the tasks took longer and I was forced to look down for long periods of time, and I ended up being far off conditions.
12	2.5	The helmet was heavy to move around under G.
13	3.5	The cursor had jitter under G. The accuracy was lower and required updates from the slew switch.

Additional usability comments:

Military Utility comments:

Table 31 – Operator 6, Flight 1, Traditional HSI Survey

#	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4.5	Gross movements were pretty slow. The slew rate should be higher for gross movements. Fine adjustments were highly favorable.
2	4.5	Same as above.
3	4.5	Gross movements were pretty slow.
4	4.5	Same
5	6	Highly accurate and predictable.
6	4	Touch screen was a little sensitive, I would have preferred any actual key pad over the touch screen. No errors were discovered during the test
7	4	Same as above

8	5	
9	N/A	
10	5.5	I would have liked to keep my hands on the controls. However, data entry tasks were quick.
11	5.5	
12	4	
13	5	

Additional usability comments:

Military Utility comments:

Table 32 – Operator 6, Flight 2, Eye Tracking Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	5	When it worked, ET was phenomenal for making gross movements across the screen. Higher accuracy would have been preferred. However, ET was totally unreliable throughout the flight.
2	5	Same as above.
3	5	When it worked, it was great. It required minimal updates to the cursor with the slew switch. Higher accuracy would have been preferred.
4	N/A	Did not work.
5	2.5	The accuracy gets worse as time goes on and the calibration is changed. For Task 3, the accuracy was terrible. Often, I noticed that snapping made the cursor go farther away from the target than it already was.
6	2.5	Same as above. The accuracy was not good, making the task take longer.
7	2.5	Same as above.
8	4	The initial accuracy for activating menus seemed to work a lot better
9	4.5	The gross movement ability of ET was great for snapping the cursor from one end of the screen to the other.
10	5.5	When it worked, it was extremely comfortable because all that was required was for me to look at where I wanted the cursor to be.
11	3.5	This is a double edged sword. When it worked, it was extremely usable (5.5). Usually for the first 5 minutes after a calibration, the ET was fast and highly accurate, which was highly desirable. However, for the majority of the sortie, ET was not functional and the accuracy was dismal. It was completely unusable and made the tasks harder to complete than if I were to just slew the cursor with the cursor slew switch.
12	0	Completely unusable under G. ET did not function at either 2G or 3.5G
13	N/A	No noticeable turbulence was observed.

Additional usability comments:

Overall, ET did not work well for me. I have an eye condition that is not optimal for eye tracking. Furthermore, Task 2 was updated with a software change that changes the TMS functionality. The change in functionality caused the CS function to not work.

Military Utility comments:

Table 33 – Operator 6, Flight 2, Rhino Pointing Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.

1	5.5	The RP cursor was quick and easy to make gross inputs.
2	5.5	Same as above.
3	5	Quick gross movement with direct designation
4	5	Quick gross movement with direct designation
5	6	I was able to direct designation on 90% of the tasks without having to update the position with the slew switch
6	4	The amount of head movement for the data entry task was objectionable. Slewing the cursor back and forth on the screen was difficult. I would have preferred to use the touch screen
7	4	Same as above
8	5	Selecting a menu without data entry was acceptable. The menu selection didn't require the head motion that was observed in the actual data entry.
9	4	Similar to the data entry. Moving the cursor back and forth on the screen required substantial head motion. Gross movements were quick and it was easy to designate and then assign the missile.
10	2.5	The head motion was objectionable and highly uncomfortable for the long sortie
11	4	RP is definitely usable. The pointer was highly accurate and gross movements were quick. For baseline maneuvering, it was nice to use. However, as soon as I started maneuvering, RP was objectionable. For data entry, the tasks took longer and I was forced to look down for long periods of time, and I ended up being far off conditions.
12	3	The helmet was heavy to move around under G. However for Task 1 and 2, they were over quick. This became a larger issue for Task 3 due to the amount of head motion and how long it took to complete the task.
13	N/A	No noticeable turbulence was observed on the sortie

Additional usability comments:

Military Utility comments:

Table 34 – Operator 6, Flight 2, Traditional HSI Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4.5	Gross movements were pretty slow. The slew rate should be higher for gross movements. Fine adjustments were highly favorable.
2	4.5	Same as above.
3	4.5	Gross movements were pretty slow.
4	4.5	Same
5	6	Highly accurate and predictable.
6	5	The touch screen was still sensitive. However, for the second flight, I was able to anticipate the sensitivity better and did not have any error on the second flight.
7	5	Same as above
8	5	
9	3.5	It was predictable and accurate. However, slewing back and forth across the screen to designate and select a missile was objectionable. It was difficult to keep up with the targets that were popping up.
10	5.5	I would have liked to keep my hands on the controls. However, data entry tasks were quick.
11	5	It's a proven technology, but for a large area display, it was slow to slew the cursor across the screen

Additional usability comments:

Military Utility comments:

Table 35 – Operator 7, Flight 1, Eye Tracking Survey

#	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.
1	6	Look and TMS AFT gets your cursor across the display immediately.
2	6	Same as above
3	5	Direct designate is not an option. The initial slew is a great way to move the cursor across a large display, but refining with HOTAS slewing is almost always required regardless of the flight conditions because the cursor didn't go exactly to where I was looking.
4	5	Excellent for big movements, cursor refinement with HOTAS required each time prior to designation.
5	4	The accuracy was degraded under G based on the system setup. When the eye tracker though I was looking off the display, it would ignore the command to slave the cursor to the eye track which resulted in some miscues as the cursor was commanded to move but didn't.
6	4	It was nice to keep hands on the stick and throttle, but a slew was nearly always required to get the cursor to the right letter or number. It also took concentration to ensure that I was actually looking at the letter/number that I wanted to slave the cursor to because just looking in the general area would always result in the cursor snapping somewhere close, but not on the desired letter/number
7	4	No difference from above
8	5	
10	6	Not having to move my head, it almost seemed like mind control. I looked where I wanted the cursor to go, commanded it to slave, and it went there.
11	5	
12	2	Because the helmet and mask were not custom fitted or adjustable as much as it could be, the mask drooped under G and the camera required for eye tracking that looks at the display no longer looked in the right place. Even when the camera was adjusted, I needed to compensate by looking with an offset in order to get the cursor close when commanding to slave to the correct location on the display.

Additional usability comments:
When eye tracking was accurate, it was great. It was much nicer than having to move my entire head to control a cursor and allowed me to still look at VHUD data. However, it was the most sensitive to shifts in the mask and helmet. It was also the system that needed the most care and feeding. Without a feedback cursor (like the rhino pointing cursor) the pilot is unaware of any offsets in the system prior to slaving the cursor. Based on that, a second slew is always required.

Military Utility comments:
If technology allows eye tracking systems to be more robust and reliable, the accuracy it's capable of achieving when looking at a large area display would be extremely useful for making large cursor movement. As it was, the cameras that were tracking eye movement were mounted on the bottom of the visor and were within the pilots field of view. They also obscured the display.

Table 36 – Operator 7, Flight 1, Rhino Pointing Survey

#	Rating	Comments Provide comments for ratings that are marginally satisfactory (4) or lower.

1	5	
2	5	
3	5	All focus needs to go to the display to do this and the head has to be finely controlled to get the rhino pointer on the track. The pointer was jittery (could be filtered) but I could still place it close
4	5	
5	5	
6	3	It was workload intensive to keep moving the rhino pointer.
7	3	Same comment
8	5	
10	4	Takes a lot of head movement which becomes tiring while wearing a heavy helmet. Forces the pilot to move their head more than they normally would in order to get the cursor where they want it.
11	5	
12	5	
13	3	The jitter that exists under basic flight conditions is amplified in turbulence making it difficult to accurately point the rhino cursor, however large movements are still easy to make as long as precision is not required.

Additional usability comments:
RP worked well for moving the cursor quickly across a large area. I did not refine with the cursor slew control but instead simply direct designated with the Rhino pointer (TMS AFT/FWD). The Rhino pointer was jittery with aircraft motion and even while on the ground, but I felt like I was still able to designated close to the target.

Military Utility comments:
I would never use this for data entry or as a standalone way to manipulate a cursor, however it was effective at covering large distances on the display and mostly accurate enough to direct designate a track.

Table 37 – Operator 7, Flight 1, Traditional HSI Survey

#	Rating	**Comments** Provide comments for ratings that are marginally satisfactory (4) or lower.
1	4	Takes too long to cover large distances, but allows you to scan the VHUD and stay on parameters while the cursor is moving across the display.
2	5	
3	5	
4	5	
5	5	
6	5	
7	5	
8	5	
10	6	
11	5	
12	5	
13	5	

Additional usability comments:

The cursor was a fixed speed with no magnetic latching to the track so it was difficult to achieve fine control in close. Covering large distances across the display was extremely annoying

Military Utility comments:
Touchscreen data entry is still the best.

This page was intentionally left blank.

APPENDIX G – LIST OF ACRONYMS

Abbreviation	Definition	Units
AFB	Air Force Base	---
AFE	Aircrew Flight Equipment	---
AFRL	Air Force Research Laboratory	---
BAL	Basic Aircraft Limits	---
CATA	Collision Antenna Train Angle	---
CATS	Cognitive Assessment Tool Set	---
CEP95	Circular Error Probable of 95%	---
cm	Centimeters	---
CMS	Cursor Management Switch	---
CPC	Chaotic Physiological Classifier	---
CTF	Combined Test Force	---
DAS	Data Acquisition System	---
ECG	Electrocardiogram	---
EIAP	Environmental Impact Analysis Process	---
FAA	Federal Aviation Administration	---
FCP	Front Cockpit	---
FOV	Field of View	---
FTT	Flight Test Technique	---
G	Load Factor	---
GFE	Government Furnished Equipment	---
HIPAA	Health Insurance Portability and Accountability Act	---
HMD	Helmet Mounted Display	---
HOTAS	Hands On Stick and Throttle	---
HPW	Human Performance Wing	---
HSI	Human System Interface	---
HUD	Head-Up Display	---
HWL	High Workload	---
IMC	Instrument Meteorological Conditions	---
KCAS	Knots Calibrated Airspeed	---
KDE	Kernel Density Estimation	---
KIOW	Iowa City Municipal Airport	---
LCD	Liquid Crystal Display	---
LDTO	Lead Developmental Test Organization	---
LED	Light Emitting Diode	---
LWL	Low Workload	---
MFD	Multifunctional Display	---

NAP	Normalized Aviation Performance	---
OFRH	Operationally Flight Representative Helmet	---
OPL	Operator Performance Laboratory	---
PIC	Pilot In Command	---
RCP	Rear Cockpit	---
RMS	Root Mean Square	---
RP	Rhino-Pointing	---
SA	Situation Awareness	---
sec	Second	---
SP	Safety Pilot	---
STO	Specific Test Objective	---
TIM	Technical Information Memorandum	---
TMP	Test Management Project	---
TPS	Test Pilot School	---
TPV	Transition Probability Variance	---
USAF	United States Air Force	---
VHUD	Virtual Heads Up Display	---
VMC	Visual Meteorological Conditions	---

APPENDIX H – ABBREVIATED DATA ANALYSIS PLAN

<table>
<tr><th colspan="6">Task 1 - Cursor Designation Accuracy</th></tr>
<tr><th colspan="2"></th><th colspan="2">DAS Parameters</th><th colspan="2">Non-DAS Parameters</th></tr>
<tr><td colspan="2">Data Parameters</td><td>Operator
HSI Method
Airspeed (KIAS)
Altitude (ft)
G
Bank Angle (deg)
Turn Rate
Task Start/Completion Time (sec)
Cursor Dissociation
Cursor Position
Track Position
Designation Input
Cursor Snap</td><td>Available Real Time & Post Flight</td><td>Operator
Operator Comments
HSI Method
Maneuver</td><td>Captured Real Time by TC</td></tr>
<tr><td colspan="2">Qualitative Data</td><td colspan="4">Pilot comments were recorded by the TC via feed to the ground station.</td></tr>
<tr><td rowspan="3">Data Quality</td><td>Maneuver Quality</td><td colspan="4">Maneuver quality were recorded through the maneuver. For Maneuver 1 and 2, the aircraft must be within the data band. Any excursion outside of tolerance were corrected before continuing the maneuver.</td></tr>
<tr><td>Data Gathering Effectiveness</td><td colspan="4">Determined if effective real-time / post-flight. If the data was unusable, revert to ground test data and qualitative data.</td></tr>
<tr><td>Repeats</td><td colspan="4">None planned, but approved, fuel/time allowing. Will revert to ground test if data was unusable.</td></tr>
<tr><td colspan="2">Analysis Procedure</td><td colspan="4">The distance was plotted on an x-y plot showing the x-y distance between the track centroid and cursor at cursor dissociation, designate, or track timeout. The magnitude and direction away from the track centroid were calculated using the x-y coordinates of the cursor and track centroid.

CEP95 was calculated for each operator and HSI method by calculating the geometric mean and RMS. The R value was calculated by dividing the geometric mean by the RMS. The Y value was calculated using linear interpolation from a CEP95 table. The final CEP95 value was calculated by multiplying the Y value and the RMS.</td></tr>
<tr><td colspan="2">Data Products</td><td colspan="4">The x-y coordinates were plotted on an x-y cross plot, showing the distance and angle from the track centroid. CEP95 circles were plotted on the same plot to show the CEP95 for each HSI and operator. A cross-plot was provided for each operator with all three HSI methods.</td></tr>
</table>

Task 1 – Cursor Designation Time					
		DAS Parameters		**Non-DAS Parameters**	
Data Parameters		Operator HSI Method Airspeed (KIAS) Altitude (ft) G Bank Angle (deg) Turn Rate Task Start/Completion Time (sec) Designation Input	Available Real Time & Post Flight	Operator Operator Comments HSI Method Maneuver	Captured Real Time by TC
Qualitative Data		Pilot comments were recorded by the TC via feed to the ground station.			
Data Quality	**Maneuver Quality**	Maneuver quality were recorded through the maneuver. For Maneuver 1 and 2, the aircraft must be within the data band. Any excursion outside of tolerance were corrected before continuing the maneuver.			
	Data Gathering Effectiveness	Determined if effective real-time / post-flight. If the data was unusable, revert to ground test data and qualitative data.			
	Repeats	None planned, but approved, fuel/time allowing. Will revert to ground test if data was unusable.			
Analysis Procedure		The time between notification and designation were recorded for each task. The results from each PVI method were compared and presented in the final report.			
Data Products		A box and whisker plot of task completion times were presented containing the completion time for each user, HSI method, and maneuver type.			

<table>
<tr><th colspan="6">Task 1 – Operator Workload</th></tr>
<tr><th colspan="2"></th><th colspan="2">DAS Parameters</th><th colspan="2">Non-DAS Parameters</th></tr>
<tr><td colspan="2">Data Parameters</td><td>Operator
HSI Method
Airspeed (KIAS)
Altitude (ft)
G
Bank Angle (deg)
Turn Rate
Quantitative Workload Assessment</td><td>Available Real Time & Post Flight</td><td>Operator
Qualitative Workload Assessment
Operator Comments
HSI Method
Maneuver</td><td>Captured Real Time by TC</td></tr>
<tr><td colspan="2">Qualitative Data</td><td colspan="4">Pilot comments and qualitative workload assessments were recorded by the TC via feed to the ground station.</td></tr>
<tr><td rowspan="3">Data Quality</td><td>Maneuver Quality</td><td colspan="4">Maneuver quality were recorded through the maneuver. For Maneuver 1 and 2, the aircraft must be within the data band. Any excursion outside of tolerance were corrected before continuing the maneuver.</td></tr>
<tr><td>Data Gathering Effectiveness</td><td colspan="4">Determined if effective real-time / post-flight. If the data was unusable, revert to ground test data and qualitative data.</td></tr>
<tr><td>Repeats</td><td colspan="4">None planned, but approved, fuel/time allowing. Will revert to ground test if data was unusable.</td></tr>
<tr><td colspan="2">Analysis Procedure</td><td colspan="4">The TPV metric and normalized maneuver performance (for Maneuver 4 and 5) were plotted at a function of HSI method. The TPV metric was normalized within each operator by the maximum workload that was recorded during the maneuver.

The OPL provided a normalized maneuver performance score by taking the RMS of the normalized % deviations from the prescribed flight conditions. The RMS was normalized to a reference maneuver flown by each operator where no task was performed. In subsequent maneuvers, the NAP provided the deviation from the reference maneuver (either improved or degraded flying performance). The final NAP value was a number between 0 and 1, where 0 represented the maneuver with the highest deviation from desired parameters and 1 represented the least deviation from desired parameters.</td></tr>
<tr><td colspan="2">Data Products</td><td colspan="4">The TPV metric and normalized maneuver performance were plotted as a function of HSI method for each operator. Qualitative workload metrics (i.e., Bedford scale ratings and HSI survey) were not used.</td></tr>
</table>

Task 1 – HSI Usability					
		DAS Parameters		**Non-DAS Parameters**	
Data Parameters		N/A	N/A	Operator Operator Comments HSI Method	Captured Real Time by TC
Qualitative Data		Pilot comments were recorded by the TC via feed to the ground station.			
Data Quality	**Maneuver Quality**	N/A			
	Data Gathering Effectiveness	N/A			
	Repeats	N/A			
Analysis Procedure		A prescribed questionnaire was completed by each operator for each HSI method immediately following each sortie.			
Data Products		The questionnaires were included as an appendix in the final test report. The questionnaire results and operator comments were also summarized in the final report.			

Task 2 – Cursor Accuracy at Expand					
		DAS Parameters		**Non-DAS Parameters**	
Data Parameters		Operator HSI Method Airspeed (KIAS) Altitude (ft) G Cursor Dissociation Cursor Position Track Position Expand Input Cursor Snap	Available Real Time & Post Flight	Operator Operator Comments HSI Method Maneuver	Captured Real Time by TC
Qualitative Data		Pilot comments were recorded by the TC via feed to the ground station.			
Data Quality	**Maneuver Quality**	Maneuver quality were recorded through the maneuver. For Maneuver 1 and 2, the aircraft must be within the data band. Any excursion outside of tolerance were corrected before continuing the maneuver.			
	Data Gathering Effectiveness	Determined if effective real-time / post-flight. If the data was unusable, revert to ground test data and qualitative data.			
	Repeats	None planned, but approved, fuel/time allowing. Will revert to ground test if data was unusable.			
Analysis Procedure		The distance was plotted on an x-y plot showing the x-y distance between the track centroid and cursor at cursor dissociation, expand, or track timeout. The magnitude and direction away from the track centroid were calculated using the x-y coordinates of the cursor and track centroid. CEP95 was calculated for each operator and HSI method by calculating the geometric mean and RMS. The R value was calculated by dividing the geometric mean by the RMS. The Y value was calculated using linear interpolation from a CEP95 table. The final CEP95 value was calculated by multiplying the Y value and the RMS.			
Data Products		The x-y coordinates were plotted on an x-y cross plot, showing the distance and angle from the track centroid. For ET and RP, a line was plotted between the point at cursor dissociation and expand to indicate how far the operator had to slew to expand. CEP95 circles were plotted on the same plot to show the CEP95 for each HSI and operator. A cross-plot was provided for each operator with all three HSI methods.			

Task 2 – Cursor Designation Accuracy					
		DAS Parameters		**Non-DAS Parameters**	
Data Parameters		Operator HSI Method Airspeed (KIAS) Altitude (ft) G Cursor Dissociation Cursor Position Track Position Designation Input Cursor Snap	Available Real Time & Post Flight	Operator Operator Comments HSI Method Maneuver	Captured Real Time by TC
Qualitative Data		Pilot comments were recorded by the TC via feed to the ground station.			
Data Quality	**Maneuver Quality**	Maneuver quality were recorded through the maneuver. For Maneuver 1 and 2, the aircraft must be within the data band. Any excursion outside of tolerance were corrected before continuing the maneuver.			
	Data Gathering Effectiveness	Determined if effective real-time / post-flight. If the data was unusable, revert to ground test data and qualitative data.			
	Repeats	None planned, but approved, fuel/time allowing. Will revert to ground test if data was unusable.			
Analysis Procedure		The distance was plotted on an x-y plot showing the x-y distance between the track centroid and cursor at cursor dissociation, designation, or track timeout. The magnitude and direction away from the track centroid were calculated using the x-y coordinates of the cursor and track centroid. CEP95 was calculated for each operator and HSI method by calculating the geometric mean and RMS. The R value was calculated by dividing the geometric mean by the RMS. The Y value was calculated using linear interpolation from a CEP95 table. The final CEP95 value was calculated by multiplying the Y value and the RMS.			
Data Products		The x-y coordinates were plotted on an x-y cross plot, showing the distance and angle from the track centroid. CEP95 circles were plotted on the same plot to show the CEP95 for each HSI and operator. A cross-plot was provided for each operator with all three HSI methods.			

Task 2 – Cursor Designation Time					
		DAS Parameters		**Non-DAS Parameters**	
Data Parameters		Operator HSI Method Airspeed (KIAS) Altitude (ft) G Task Start/Completion Time (sec) Designation Input Cursor Snap	Available Real Time & Post Flight	Operator Operator Comments HSI Method Maneuver	Captured Real Time by TC
Qualitative Data		Pilot comments were recorded by the TC via feed to the ground station.			
Data Quality	**Maneuver Quality**	Maneuver quality were recorded through the maneuver. For Maneuver 1 and 2, the aircraft must be within the data band. Any excursion outside of tolerance were corrected before continuing the maneuver.			
	Data Gathering Effectiveness	Determined if effective real-time / post-flight. If the data was unusable, revert to ground test data and qualitative data.			
	Repeats	None planned, but approved, fuel/time allowing. Will revert to ground test if data was unusable.			
Analysis Procedure		The time between notification and designation were recorded for each task. The results from each PVI method were compared and presented in the final report.			
Data Products		A box and whisker plot of task completion times were presented containing the completion time for each user, HSI method, and maneuver type.			

<table>
<tr><th colspan="6">Task 2 – Operator Workload</th></tr>
<tr><th colspan="2"></th><th colspan="2">DAS Parameters</th><th colspan="2">Non-DAS Parameters</th></tr>
<tr><td colspan="2">Data Parameters</td><td>Operator
HSI Method
Airspeed (KIAS)
Altitude (ft)
G
Bank Angle (deg)
Turn Rate
Quantitative Workload Assessment</td><td>Available Real Time & Post Flight</td><td>Operator
Qualitative Workload Assessment
Operator Comments
HSI Method</td><td>Captured Real Time by TC</td></tr>
<tr><td colspan="2">Qualitative Data</td><td colspan="3">Pilot comments and qualitative workload assessments were recorded by the TC via feed to the ground station.</td></tr>
<tr><td rowspan="3">Data Quality</td><td>Maneuver Quality</td><td colspan="3">Maneuver quality were recorded through the maneuver. For Maneuver 1 and 2, the aircraft must be within the data band. Any excursion outside of tolerance were corrected before continuing the maneuver.</td></tr>
<tr><td>Data Gathering Effectiveness</td><td colspan="3">Determined if effective real-time / post-flight. If the data was unusable, revert to ground test data and qualitative data.</td></tr>
<tr><td>Repeats</td><td colspan="3">None planned, but approved, fuel/time allowing. Will revert to ground test if data was unusable.</td></tr>
<tr><td colspan="2">Analysis Procedure</td><td colspan="3">The TPV metric, normalized maneuver performance (for Maneuver 4 and 5), and normalized task performance were plotted at a function of HSI method. The TPV metric was normalized within each operator by the maximum workload that was recorded during the maneuver.

The OPL provided a normalized maneuver performance score by taking the RMS of the normalized % deviations from the prescribed flight conditions. The RMS was normalized to a reference maneuver flown by each operator where no task was performed. In subsequent maneuvers, the NAP provided the deviation from the reference maneuver (either improved or degraded flying performance). The final NAP value was a number between 0 and 1, where 0 represented the maneuver with the highest deviation from desired parameters and 1 represented the least deviation from desired parameters.</td></tr>
<tr><td colspan="2">Data Products</td><td colspan="3">The TPV metric and normalized maneuver performance were plotted as a function of HSI method for each operator. Qualitative workload metrics (i.e., Bedford scale ratings and HSI survey) were not used.</td></tr>
</table>

<table>
<tr><th colspan="6">Task 2 – HSI Usability</th></tr>
<tr><th colspan="2"></th><th colspan="2">DAS Parameters</th><th colspan="2">Non-DAS Parameters</th></tr>
<tr><td colspan="2">Data Parameters</td><td>N/A</td><td>N/A</td><td>Operator
Operator Comments
HSI Method</td><td>Captured Real Time by TC</td></tr>
<tr><td colspan="2">Qualitative Data</td><td colspan="4">Pilot comments were recorded by the TC via hot-mic feed to the ground station.</td></tr>
<tr><td rowspan="3">Data Quality</td><td>Maneuver Quality</td><td colspan="4">N/A</td></tr>
<tr><td>Data Gathering Effectiveness</td><td colspan="4">N/A</td></tr>
<tr><td>Repeats</td><td colspan="4">N/A</td></tr>
<tr><td colspan="2">Analysis Procedure</td><td colspan="4">A prescribed questionnaire was completed by each operator for each HSI method immediately following each sortie.</td></tr>
<tr><td colspan="2">Data Products</td><td colspan="4">The questionnaires were included as an appendix in the final test report. The questionnaire results and operator comments were also summarized in the final report.</td></tr>
</table>

Task 3 – Data Input Selection Miss Distance					
		DAS Parameters		**Non-DAS Parameters**	
Data Parameters		Operator HSI Method Airspeed (KIAS) Altitude (ft) G Cursor Position Data Item Centroid Position Designation Input Cursor Snap	Available Real Time & Post Flight	Operator Operator Comments HSI Method Maneuver	Captured Real Time by TC
Qualitative Data		Pilot comments were recorded by the TC via feed to the ground station.			
Data Quality	**Maneuver Quality**	Maneuver quality were recorded through the maneuver. For Maneuver 1 and 2, the aircraft must be within the data band. Any excursion outside of tolerance were corrected before continuing the maneuver.			
	Data Gathering Effectiveness	Determined if effective real-time / post-flight. If the data was unusable, revert to ground test data and qualitative data.			
	Repeats	None planned, but approved, fuel/time allowing. Will revert to ground test if data was unusable.			
Analysis Procedure		The distance was plotted on an x-y plot showing the x-y distance between the key centroid and cursor at cursor dissociation or designation. The magnitude and direction away from the key centroid were calculated using the x-y coordinates of the cursor and key centroid. CEP95 was calculated for each operator and HSI method by calculating the geometric mean and RMS. The R value was calculated by dividing the geometric mean by the RMS. The Y value was calculated using linear interpolation from a CEP95 table. The final CEP95 value was calculated by multiplying the Y value and the RMS.			
Data Products		The x-y coordinates were plotted on an x-y cross plot, showing the distance and angle from the key centroid. CEP95 circles were plotted on the same plot to show the CEP95 for each HSI and operator. Furthermore, the outline of the key size (i.e. a 1” by 1” square) was overlaid on the plot for reference. A cross-plot was provided for each operator with all three HSI methods.			

Task 3 – Data Input Error Rate					
		DAS Parameters		**Non-DAS Parameters**	
Data Parameters		Operator HSI Method Airspeed (KIAS) Altitude (ft) G Inputted Item Assigned Item	Available Real Time & Post Flight	Operator Operator Comments HSI Method Maneuver Assigned Item Inputted Item	Captured Real Time by TC
Qualitative Data		Pilot comments were recorded by the TC via feed to the ground station.			
Data Quality	**Maneuver Quality**	Maneuver quality were recorded through the maneuver. For Maneuver 1 and 2, the aircraft must be within the data band. Any excursion outside of tolerance were corrected before continuing the maneuver.			
	Data Gathering Effectiveness	Determined if effective real-time / post-flight. If the data was unusable, revert to ground test data and qualitative data.			
	Repeats	None planned, but approved, fuel/time allowing. Will revert to ground test if data was unusable.			
Analysis Procedure		The inputted line of data was compared to the assigned line of data. An error rate was calculated by taking the number of errors over total number of data selections (between menus and characters).			
Data Products		The error rate was presented in a table for each user, maneuver, and HSI method.			

Task 3 – Data Input Time					
		DAS Parameters		**Non-DAS Parameters**	
Data Parameters		Operator HSI Method Airspeed (KIAS) Altitude (ft) G Task Start/Completion Time (sec)	Available Real Time & Post Flight	Operator Operator Comments HSI Method Maneuver	Captured Real Time by TC
Qualitative Data		Pilot comments were recorded by the TC via feed to the ground station.			
Data Quality	**Maneuver Quality**	Maneuver quality were recorded through the maneuver. For Maneuver 1 and 2, the aircraft must be within the data band. Any excursion outside of tolerance were corrected before continuing the maneuver.			
	Data Gathering Effectiveness	Determined if effective real-time / post-flight. If the data was unusable, revert to ground test data and qualitative data.			
	Repeats	None planned, but approved, fuel/time allowing. Will revert to ground test if data was unusable.			
Analysis Procedure		The time between the first menu selected and the final key designated was recorded for each task. The results from each HSI method was be compared and presented in the final report.			
Data Products		A box and whisker plot of task completion times was presented containing the completion time for each user, HSI method, and maneuver type.			

<table>
<tr><th colspan="6">Task 3 – Operator Workload</th></tr>
<tr><th colspan="2"></th><th colspan="2">DAS Parameters</th><th colspan="2">Non-DAS Parameters</th></tr>
<tr><td colspan="2">Data Parameters</td><td>Operator
HSI Method
Airspeed (KIAS)
Altitude (ft)
G
Bank Angle (deg)
Turn Rate
Quantitative Workload Assessment</td><td>Available Real Time & Post Flight</td><td>Operator
Qualitative Workload Assessment
Operator Comments
HSI Method</td><td>Captured Real Time by TC</td></tr>
<tr><td colspan="2">Qualitative Data</td><td colspan="3">Pilot comments and qualitative workload assessments were recorded by the TC via feed to the ground station.</td></tr>
<tr><td rowspan="3">Data Quality</td><td>Maneuver Quality</td><td colspan="3">Maneuver quality were recorded through the maneuver. For Maneuver 1 and 2, the aircraft must be within the data band. Any excursion outside of tolerance were corrected before continuing the maneuver.</td></tr>
<tr><td>Data Gathering Effectiveness</td><td colspan="3">Determined if effective real-time / post-flight. If the data was unusable, revert to ground test data and qualitative data.</td></tr>
<tr><td>Repeats</td><td colspan="3">None planned, but approved, fuel/time allowing. Will revert to ground test if data was unusable.</td></tr>
<tr><td colspan="2">Analysis Procedure</td><td colspan="3">The TPV metric, normalized maneuver performance (for Maneuver 4 and 5), and normalized task performance was plotted at a function of HSI method. The TPV metric was normalized within each operator by the maximum workload that was recorded during the maneuver.

The OPL provided a normalized maneuver performance score by taking the RMS of the normalized % deviations from the prescribed flight conditions. The RMS was normalized to a reference maneuver flown by each operator where no task was performed. In subsequent maneuvers, the NAP provided the deviation from the reference maneuver (either improved or degraded flying performance). The final NAP value was a number between 0 and 1, where 0 represented the maneuver with the highest deviation from desired parameters and 1 represented the least deviation from desired parameters.</td></tr>
<tr><td colspan="2">Data Products</td><td colspan="3">The TPV metric and normalized maneuver performance were plotted as a function of HSI method for each operator. Qualitative workload metrics (i.e., Bedford scale ratings and HSI survey) were not used.</td></tr>
</table>

Task 3 – HSI Usability					
		DAS Parameters		**Non-DAS Parameters**	
Data Parameters		N/A	N/A	Operator Operator Comments HSI Method	Captured Real Time by TC
Qualitative Data		Pilot comments were recorded by the TC via feed to the ground station.			
Data Quality	**Maneuver Quality**	N/A			
	Data Gathering Effectiveness	N/A			
	Repeats	N/A			
Analysis Procedure		A prescribed questionnaire was completed by each operator for each HSI method immediately following each sortie.			
Data Products		The questionnaires were included as an appendix in the final test report. The questionnaire results and operator comments was also summarized in the final report.			

Task 4 – HSI Usability					
		DAS Parameters		**Non-DAS Parameters**	
Data Parameters		N/A	N/A	Operator Operator Comments HSI Method	Captured Real Time by TC
Qualitative Data		Pilot comments were recorded by the TC via feed to the ground station.			
Data Quality	**Maneuver Quality**	N/A			
	Data Gathering Effectiveness	N/A			
	Repeats	N/A			
Analysis Procedure		A prescribed questionnaire was completed by each operator for each HSI method immediately following each sortie.			
Data Products		The questionnaires were included as an appendix on the final test report. The questionnaire results and operator comments was also summarized in the final report.			

<table>
<tr><th colspan="6">Task 4 – Operator Workload</th></tr>
<tr><th colspan="2"></th><th colspan="2">DAS Parameters</th><th colspan="2">Non-DAS Parameters</th></tr>
<tr><td colspan="2">Data Parameters</td><td>Operator
HSI Method
Airspeed (KIAS)
Altitude (ft)
Quantitative Workload Assessment</td><td>Available Real Time & Post Flight</td><td>Operator
Qualitative Workload Assessment
Operator Comments
HSI Method</td><td>Captured Real Time by TC</td></tr>
<tr><td colspan="2">Qualitative Data</td><td colspan="4">Pilot comments and qualitative workload assessments were recorded by the TC via hot-mic feed to the ground station. At the completion of each operational task, the operator will provide a Bedford workload rating.</td></tr>
<tr><td rowspan="3">Data Quality</td><td>Maneuver Quality</td><td colspan="4">Maneuver quality were recorded through the maneuver. Any excursion outside of tolerance were corrected before continuing the maneuver.</td></tr>
<tr><td>Data Gathering Effectiveness</td><td colspan="4">Determined if effective real-time / post-flight. If the data was unusable, revert to ground test data and qualitative data.</td></tr>
<tr><td>Repeats</td><td colspan="4">None planned, but approved, fuel/time allowing. Will revert to ground test if data was unusable.</td></tr>
<tr><td colspan="2">Analysis Procedure</td><td colspan="4">The TPV metric was plotted at a function of HSI method. The TPV metric was normalized within each operator by the maximum workload that was recorded during the maneuver.</td></tr>
<tr><td colspan="2">Data Products</td><td colspan="4">TPV workload metrics were presented in a plot for each operator and HSI method.</td></tr>
</table>

Task 4 – Military Utility					
		DAS Parameters		**Non-DAS Parameters**	
Data Parameters		N/A	N/A	Operator Operator Comments HSI Method	Captured Real Time by TC
Qualitative Data		Pilot comments were recorded by the TC via feed to the ground station.			
Data Quality	**Maneuver Quality**	N/A			
	Data Gathering Effectiveness	N/A			
	Repeats	N/A			
Analysis Procedure		N/A			
Data Products		Questionnaires and operator comments regarding military utility were summarized in the final report.			

		Havoc	Knots	BMW	Clutch	Golden	Fuzzy
MOP 1.1	Baseline						
	Low G						
	High G						
	Bowtie 1						
	Bowtie 2						
	Bowtie 3						
MOP 1.2	Baseline (Each PVI type)						
	Low G (each PVI type)						
	High G (each PVI type)						
	Bowtie 1 (each PVI type)						
	Bowtie 2 (each PVI type)						
	Bowtie 3 (each PVI type)						
MOP 1.3							

APPENDIX I – DIGITAL DATA APPENDIX

Raw flight test data, additional figures, pictures/video, the data analysis plan, and the test matrix are available by contacting the USAF Test Pilot School.

The files are held in six folders. Their structure is listed below:

1. Raw Flight Test Data
2. Additional Figures
3. Media
4. Data Analysis Plan
5. Test Matrix
6. Test Hardware Manuals and Spec Sheets

This page was intentionally left blank.

APPENDIX J – FUTURE WORK

Have Rhino sought to test HSI methods roughly independent of the specifics of their implementations. Notwithstanding, the team gathered a significant quantity of incidental observations regarding the implementations as tested, and ideas for future implementation.

Overall, while RP and ET as implemented were representative of the current level of technology for head and eye tracking, the team determined that the current implementation is not ready for actual use in an aircraft. Following data collection, the team made small changes to the RP and ET implementations to try and improve usability in a direction that would make them closer to flight-ready. The team tested these changes both in flight and in the ground based simulator. The team also carried out additional ground test aimed at gathering information for future HSI implementations. Of course, the team had limited time available and more areas warrant investigation to improve the different HSIs.

A common comment throughout testing was how operators were using RP different than ET because there was a boresight displayed. ET was quick and required very little thought to make large gross acquisitions because you could not see the boresight. Operators were taking extra time to try and be very accurate with RP because the boresight was displayed. This led to more heads down time and a perceived higher workload when using RP over ET. The team programmed an alternate implementation called "Hidden RP" (HRP) in which the rhino-boresight was removed to make it more like ET. The following documents the results of the Hidden Rhino experiment.

HIDDEN RHINO

HRP was evaluated on the Flight 14. The concept was created based on pilot comments that the visual rhino boresight provided a tendency for the operator to achieve higher accuracy than what was required. The HRP concept made the rhino-boresight invisible; during task execution the operator moved his head to an estimated pointing solution and then snapped. Similar to the ET implementation (which was implemented due to perceived technology immaturity), the operator had no indication of to where on the screen the cursor would snap. After snap, the operator refined the cursor with TH in the usual way.

The accuracy results from the limited evaluation of HRP for Operator 6 are shown below. The results are also compared to Operator 6's RP results. The CEP95 were 1.145 cm (RP) and 12.832 cm (HRP). The HRP show similar results to ET, where the initial cursor snap function was quick, but inaccurate, forcing the operator to slew several cm before designating the track. Another observation to note here is the y-axis accuracy was significantly higher than the x-axis accuracy, meaning that the movement in elevation was more predictable than the head movement in azimuth. See the section within this appendix on charging rhino pointing (CRP) for another compelling result regarding the difference between azimuth and elevation for motion of the head.

HSI	R	Y	RMS	CEP95 (cm)
Hidden Rhino	0.557	1.927	6.659	12.832
Rhino Show	0.702	1.779	0.644	1.146

CEP 95 Results

HSI	Mean (sec)	Std Dev (sec)	T-Stat	DoF	P-Value	Result
Hidden Rhino	4.344	2.101	3.082	31.02	.00429	reject
Rhino Show	2.931	0.580				

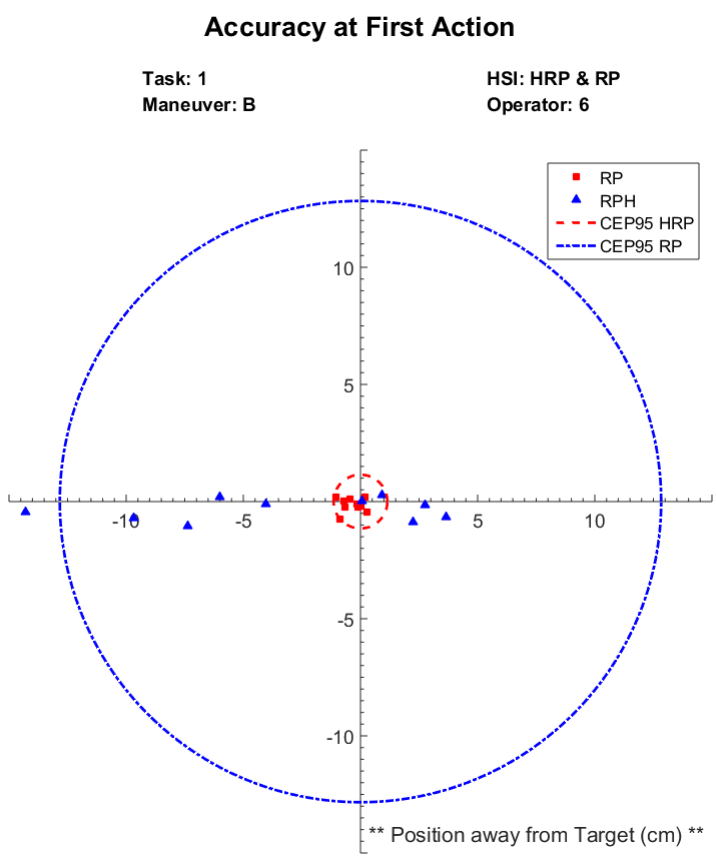
Time Results

HSI	Mean (cm/sec)	Std Dev (cm/sec)	T-Stat	DoF	P-Value	Result
Hidden Rhino	6.847	2.406	-0.221	17.97	0.828	Accept
Rhino Show	7.080	2.302				

Pseudo Velocity Results

When looking at the task completion time, RP again displayed a clear advantage, with a mean completion time of 2.931 sec compared to 4.344 sec for HRP. However, when assessing the pseudo velocity, both RP and HRP showed similar results with an average of 7.080 cm/sec for RP and 6.847 cm/sec for HRP. When including initial starting distance, HRP was as good, or nearly as good, as RP.

However, when assessing pilot comments, the HRP implementation significantly improved the comfort of using RP in general, strictly due to the fact that the hidden boresight prevented the operator from direct designating. The HRP implementation allowed the operator to perform the gross movement quickly (at the impact of accuracy), but then the operator spent the rest of the time slewing the cursor with the cursor slew switch, without added head movement. The HRP implementation also alleviated some of the operator complaints for Task 4, where the constant back-and-forth head motion and direct designation was highly objectionable. The constant back-and-forth gross acquisition was still present, however, less total head movement was required, increasing the level of comfort while performing the task as compared to the original RP implementation.



In the experiment detailed above, operators commented that they were quickly able to calibrate their brains for an approximate RP boresight and noted it was more predictable than ET.

Another issue that operators had while using ET was that if the system thought the operator was looking off of the display, it would not snap the cursor. This was mainly a problem on tracks close to the edge of the display and while at 3.5G. The test team implemented "clamping" which force the cursor to appear somewhere on the display even if the operator was looking off of the display. The cursor would appear at the closest location on the display to the location where the system though the operator was looking. Operators preferred this implementation over what was tested previously while using ET. This was also implemented with RP with no boresight displayed with preferable results.

Operators continuously commented on how good ET was when it was working, which did not happen consistently. There are relatively large technology improvements that need to be made prior to being useful. Accuracy was the biggest issue with the ET implementation. The Dikablis system was very sensitive to shift. That was because it only detected the pupil location and made an assumption of the normal vector without further correction. The system could not distinguish between a slip in the helmet from a shift of the pupil. This was partially – if not entirely – responsible for the errors in ET during the 3.5 G points. As much as the helmet shifted down on the operator's head, the system believed the operator was looking the same angular amount upward (above the screen). This explains the compensatory mechanism in which operators had to look 6-9 inches below their desired snap point to force the cursor to appear.

ET requires an implementation that is robust enough to remain accurate during aircraft operations. The operators moved their head around constantly and pulled increased G throughout the flights which caused the helmet to shift on the operator's head and induced the accuracy problems mentioned. The partial solution for this during testing was to make the helmet extremely tight; but an extremely tight helmet was very uncomfortable and not usable for normal operational missions. An absolutely perfect helmet fit to the pilot might mitigate the issue, but this might be cost prohibitive and would have to be tested. Alternatively, other ET systems available use the glint of the LED as a fixed point and express the gaze vector in terms of a vector difference of glint and pupil disc. Such an implementation would, in theory, correct for shifts of the helmet and thus would likely be a requirement for a fielded system.

Changes to the physical implementation of the eye tracking cameras is also needed to increase accuracy. The implementation during test had issues because it was looking for a bigger pupil. The eye cameras were not located the ideal distance from the operator's eye and were not adjustable. Ideally, the cameras could be mounted slightly closer to the eye. Yet this requirement needs to be balanced with the recommendation in this report to avoid obfuscating the operator's field of view.

By far the biggest issue that remained with RP after implementing HRP was the objectionable head movement required to utilize the edges of the display. This was most prevalent under G and led to many operator comments about the lack of comfort of RP. When ET was somewhat accurate, operators preferred it even over RP, due to apparent speed and comfort.

The team proposes to make RP more effective and user friendly by making it feel more like ET to the operator without an eye tracker. A system that doesn't require the operator to point their rhino-boresight all the way to the edges of the display would lead to increased comfort and usability. A potential solution is called "Charging Rhino Pointing" (CRP); the details of an experiment designed to provide the groundwork for implementation of CRP appears below.

CHARGING RHINO

In the Have Rhino experiment, it quickly became evident that RP head motions were excessive. With Hidden Rhino, RP behaved in a desirable way closer to ET; however, the operator remained conscious that the rhino-boresight was fixed normal to his face. Thus, the operator used large head motions to reach the edge of the display.

Consider an operator with a completely neutral neck position, gazing forward. If one asks this operator to rapidly adjust his gaze to a point directly in front of him, he will move neither his gaze nor his head. If one asks this operator to rapidly adjust his gaze to a point 1-2° to the right of his relaxed location, that operator will unconsciously move his head to almost completely center his gaze on that point. However, if one asks him to rapidly adjust his gaze to a point 45° off-center, the operator will move his head some proportion of the way, and use his eyes to move the remainder of the distance. At the extreme, if one asks the operator to rapidly adjust his gaze to a point 90° off his neutral gaze (along his shoulder), that operator will naturally move his head only about half the angular distance, and then transit the remaining distance with his eyes (try this at home). At around 90° and greater (behind his shoulder), the operator will actually start to roll his head on its side to enable his eyes to track farther.

The Have Rhino team thus developed the following hypothesis: there exists a function which maps desired gaze location to the head angle subtended in a comfortable movement to center the gaze on that location. More simply, the farther away from neutral an operator wants to look, the lower the proportion of that distance that his head moves. It should be possible to develop a function that determines that proportionality.

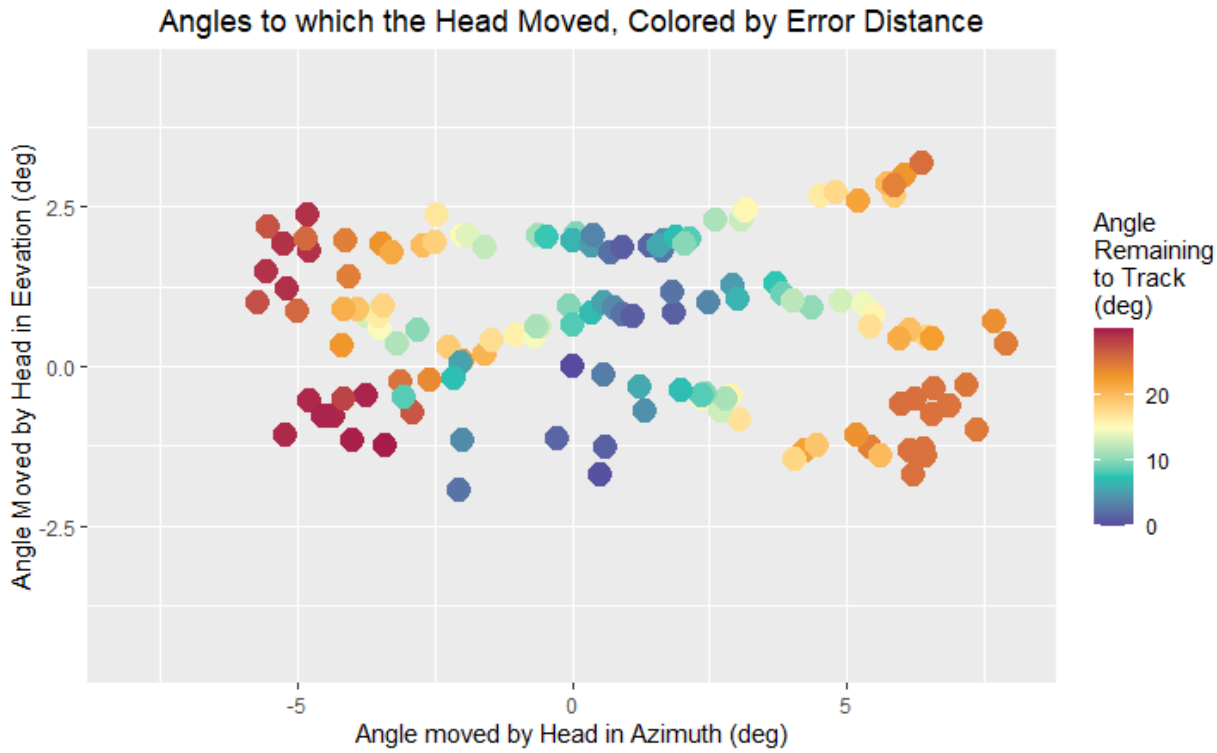
If such a function existed, RP could be dramatically improved. The head tracker would measure the angle moved by the head, and then actually displace the rhino-boresight farther outward using that function. From the operator's perspective, the rhino-boresight would be accelerated, or have a gain to its motion. This behavior is called charging rhino pointing (CRP).

To assess the truth of the hypothesis and potentially implement CRP, the team developed a fifth task, known as Task 5. This task simply displayed tracks one at a time in a lattice pattern across the display. At the first step, the MFD displayed a track directly in the center of the display. The operator comfortably centered his gaze on this track, and then pressed CMS designate. This set the zero point for the hypothesized function.

At each subsequent step, a track appeared at some given azimuth and elevation. The operator kept his gaze fixed on the center-point of the MFD until ready, and then rapidly shifted his gaze in a comfortable manner to the displayed track. With his gaze focused on the track, the operator pressed CMS designate, which caused the system to record: (1) the azimuth and angle of the track and (2) the azimuth and angle of his head from the zero point location. The operator returned his gaze comfortably to the neutral point and these steps were repeated for azimuths and elevations across the entire display and for statistical power.

This task was actually programmed and tested on the ground using the Have Rhino systems. The following figure shows a result from that test. Each dot represents an azimuth and elevation to which the head actually moved. The dots are colored according to how far away (in total angle) the track still was when the designate actuation was performed. The result affirming the hypothesis is immediate. The farther away the track, the lower the proportion of the total angle that the operator moved his head. Interestingly, it's not obvious that the function (or gain parameter) has the same value in azimuth and elevation. Rather, it appears that this operator might have almost always moved the full distance in elevation, while conserving

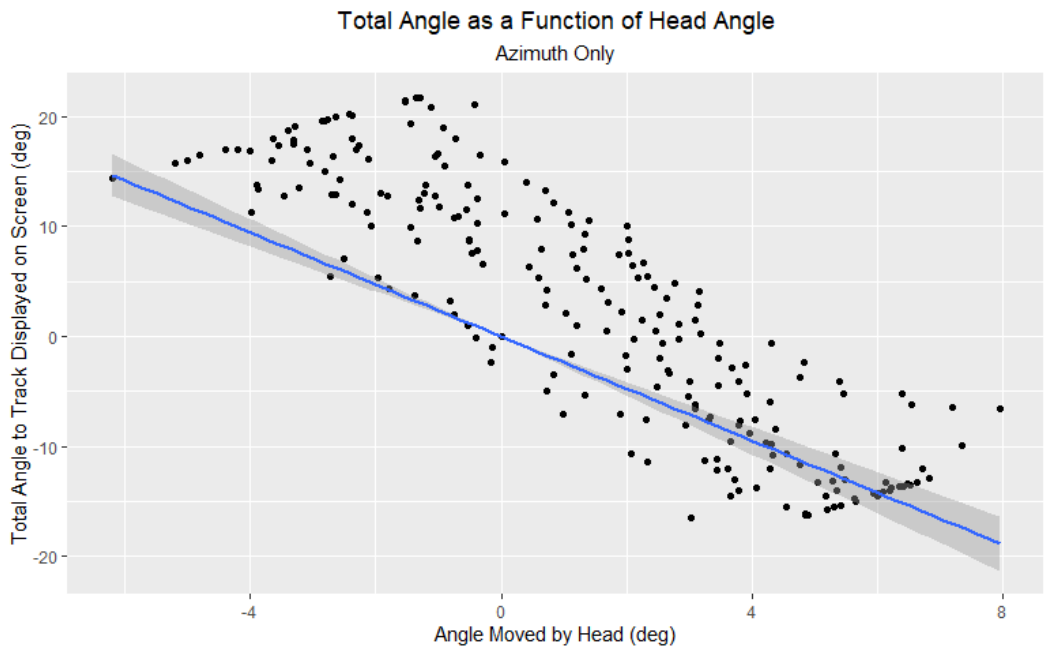
azimuthal motion. The distribution also appears asymmetric. Whether the asymmetry results from operator preference or experimental setup merits future research.

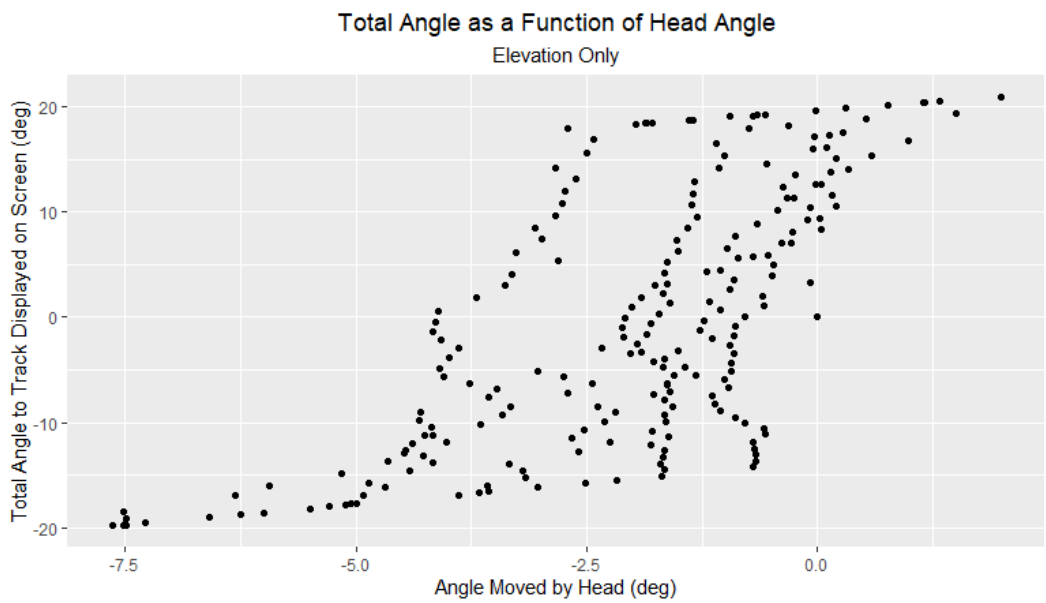


Breaking the results down by azimuth and elevation reinforces the questions remain unanswered. The plot on the next page shows a scatter of the azimuth angle transited by the head (on the x-axis) against the total azimuth angle to the track (on the y-axis). A best fit line (using linear regression) is displayed; this line has a clear (and curious) bias only because it was forced to pass through the origin, under the intuition that an operator will always center his gaze at the most comfortable location.

Despite the small bias, the slope of the data fits quite well, and the linearity is compelling. According to this data set, for every 5° of track displacement from center in azimuth, a comfortable head movement to look at it is about 2° of azimuth. To program CRP using this implementation, whenever an operator moved his head 2° in azimuth, the rhino-boresight would swing out to 5° in azimuth; when he moved his head 4° in azimuth, the rhino-boresight would accelerate to 10° in azimuth. If the mapping turned out not to be linear, it would follow whatever functional relationship existed.

Below the azimuth plot appears the results in elevation. It's not clear that a linear relationship exists at all. The operator might have attempted to fully move his head in the elevation each time. Alternatively, some sort of exponential relationship might be patterned, in which the operator barely moves in elevation until the elevation rises above a certain level, at which point the head rapidly moves in elevation to close the distance. The data are not clear and future research is warranted. In particular, the elevation data is complicated by the fact that the comfortable zero point in the center of the Have Rhino MFD was approximately 35° depressed from a neutral, forward-resting gaze. The Charging Rhino mapping might change dramatically and non-linearly 35° away from the neutral point. Again, further research is warranted.





In short, a method of RP with the speed and comfort of ET, but with increased accuracy and predictability would fulfill almost all of the requirements of this report. CRP offers an intriguing avenue towards just such an implementation; however, so does a dramatic improvement to the TRL of ET.

Overall ET and RP have some changes that can be made to their implementation methods to increase their usability. If ET is able to overcome the accuracy issues it would be able to supplement current TH methods to increase operator effectiveness with a lower workload. RP may be the quicker solution to the large area display problem but the large head movements are prohibitive in its use especially for extended periods and while under G. RP also may be a great backup to ET should the system fail and leave the operator with just head tracking.

TOUCHSCREEN BRACING

Commercial off the shelf (COTS) solutions to touchscreen bracing already exist. One such solution is shown in the photographs below:

Such COTS solutions may or may not resolve the combination of ergonomic and accuracy issues seen in this report. In particular, the method illustrated requires using the thumb for braced action; in Have Rhino, operators did not uniformly use the thumb to input touchscreen data. Various such options should be evaluated for operator comfort, speed and ease of use, and overall accuracy improvement.

This page was intentionally left blank.

APPENDIX K – LESSONS LEARNED

The preceding appendix includes the technical lessons learned during Have Rhino that pointed toward future work. However, several programmatic and execution-phase lessons were also learned during the conduct of test. To improve the planning, execution, and reporting processes for future TMPs, those lessons are summarized below.

Ground simulations prior to flight

Extensive modeling and simulation prior to airborne test execution was massively helpful for understanding the system and increasing the effectiveness of later airborne test sorties. AFRL provided a virtual reality headset that simulated RP functionality several weeks prior to test execution. The test team was able to learn the best techniques for rapid selection with RP and the optimal offset angle. Upon arrival at OPL's facility, the test team dedicated the first two days to ground familiarization sorties without data requirements. The OPL programmers finalized the HOTAS functionality and software configuration based on operator feedback during these ground simulations.

Running ground sorties with the full control room also increased airborne test effectiveness. All test team members had the opportunity to practice radio calls and computer operation prior to attempting airborne. Standardized card briefing techniques and radio brevity were also established during these simulated missions.

Full test team verification of major software configuration changes

During the early ground familiarization sorties, large functionality changes were necessary to optimize the system. Following a ground sortie, the operator provided feedback to the OPL programmers on how best to improve the system. Most changes were discussed with the full group, but several major changed were implemented without team buy-in to expedite the process. This went unnoticed until later data review that revealed some bias in specific data. Full test team approval should always be sought prior to making system modifications that could affect the data.

Abridged airborne card briefing

The airborne card briefings were substantially abbreviated due to the repetitive nature of many of the tasks. Since all operators were well-versed on the content of each test card and the steps or each procedure, card briefs were shortened to the HSI method, task, and maneuver, followed immediately by a countdown to task initiation. Repeated instances of a task required only a new initiation countdown with no additional information. This efficient procedure optimized test conduct and use of airborne time.

Contractor data analysis assistance

The test team was presented with my data analysis challenges that could not be easily solved with the level of coding experience they possessed. Rather than spend additional time learning complex data analysis processes, the test team turned to the OPL engineers for support. The OPL engineers were able to find simple solutions that would not have been considered by the test team. A considerable amount of time was saved by requesting data analysis guidance from the contractor rather than reinventing the same process.

Ability to recalibrate inflight

Both RP and ET relied on precise alignment to perform accurately. During ground familiarization sorties, it became apparent that both RP and ET calibrations might drift during flight, thereby invalidating any data collected. By standardizing a recalibration process for both HSI methods, and building the recalibration step into the test cards, the test team was able to protect data integrity and avoid early RTBs.

This page was intentionally left blank.