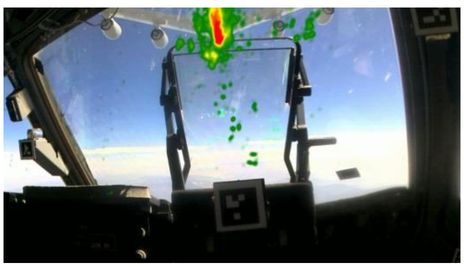
412TW-PA-19537

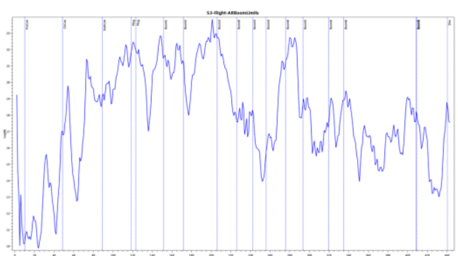
4
1
2
T
W

Objective Measures of Pilot Workload

Patrick Martin,
Lt Col Paul Calhoun,
Air Force Test Center

Dr. Tom Schnell, Colton Thompson
Operator Performance Laboratory
University of Iowa





JUNE 2019

STUDY RESULTS

Conducted with Funding from Ball Aerospace and USAF

Contract Number: 18S0234C

412TH TEST WING
EDWARDS AIR FORCE BASE, CALIFORNIA
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

Table of Contents

Table of Figures

1. Executive Summary

This project was an engineering feasibility study to determine the value of using physiological based workload assessment technology in the USAF test and evaluation process at the Air Force Test Center, Edwards AFB. The feasibility study was requested by the 412 Test Wing, Air Force Test Center, Edwards AFB, California. The responsible test organization was the 418th Flight Test Squadron at Edwards AFB. Testing was conducted by the 418th Flight Test Squadron with the Operator Performance Laboratory (OPL) from the University of Iowa. The OPL provided, operated and collected data with test equipment called the Cognitive Assessment Tool Set (CATS) kit. Testing was conducted from 28 August 2018 to 12 February 2019 and consisted of four sorties totaling 18.5 flight test hours. Testing was conducted by the Combined Test Force IAW test plan, 412TW-TP-18-47, Objective Measures of Workload Feasibility Study [1].

A total of seven evaluator pilots performed takeoff, AR, normal flight operations, and landing tasks to determine the utility of the CATS system in a flight test environment. The CATS was set up inside an instrumented C-17 aircraft acting as a receiving aircraft against a legacy KC-135 tanker. Alongside CATS, OPL used the Dikablis Professional eye-tracking glasses in conjunction with D-Lab software from Ergoneers.Each pilot was fitted with electrocardiogram (ECG or EKG) electrodes prior to flight. The eye-tracker was used exclusively in the right seat, and only on the August flights. During the course of these flights, ECG data was successfully obtained with little-to-no complications. Flight One produced useful eye-tracking data, but Flight Two did not result in successful eye tracking from the Dikablis glasses due to inproper fit. For the final two flights in February, due to technical infeasibility of the current eye tracking setup, OPL made the decision to forego additional eye tracking data and instead focus on the CATS workload data collection.

In addition to live flights, workload data was gathered from simulator tasks in Simulator B at the Test Pilot School. Pilots flew through 4 scenarios of varying difficulty in which they were tasked with following an aircraft while keeping it within a set of horizontal bars in the HUD, all while performing a secondary auditory response task. OPL gathered data from seven pilots total, but only five of these seven participated in the live flights, with the two additional pilots only instrumented with OPL's ECG amplifier.

Results indicated the pilots subjectively rated the ECG as comfortable and non-evasive. The utility of the ECG was reliable for the most part, there were a few instances of leads disconnecting resulting momentary data drop-outs, but were quickly detected and corrected by the pilot. The Dikablis glasses eye-tracker was generally uncomfortable and compatable with head-sets for only short durations. The eye-tracker could be better implemented if integrated into a helmet. Using the ECG and workload data, the difficulty level of each task (and even levels for distinct aspects of each task) was able to be determined. ECG data was sensitive to both low and high workload flight conditions and consistent across users. When combined with eye-tracker and aircraft data, the measure showed good diagnostisity. ECG data were unique to each pilot, but of the various tasks performed, station keeping on-boom during a turn generally produced the highest workload, followed by boom limits operations. Comparatively, takeoff and landing tasks seemed fairly straightforward and low-workload. Post-task pilot questionnaires (TLX and Bedford) backed up the ECG workload scores.

Eye-tracking data showed that during aerial refueling tasks, the singular point that attracted the attention of the pilot was the KC-135 pilot director lights (PDL's), and could be delineated down to the "captain's bars", which were used as a visual reference point to maintain formation. Regardless of the presence of the HUD, the captain's bars held each pilot's attention exclusively during refueling, with only occasional glances to their instruments or the tanker wing. Unsurprisingly, pilots maintained good situational awareness with a robust scan-pattern as they shifted their gaze much more during Landing and Takeoff tasks, alternating between the HUD, instruments, and forward and right-side windows.

This study has shown that OPL’s methods and instrumentation can reliably provide physiological data including workload and eye-tracking that can help better evaluate the effort and attention required by the flight crew to successfully complete aerial refueling operations. The ECG based workload assessment system was deployed in minutes and required no training or special skills from the pilot. The system provides a relative workload number in real-time from the second the ECG amplifier was turned on and no further, calibration, modification, or refinement was needed to generate the figures shown in this report. Other than commercial power for the laptops, there was zero integration with the aircraft and was acceptably nonintruisive as the pilots were not tethered in any way. Unique flight events were conviently tagged by the test team to eliminate the need for aircraft systems integration. The ECG system could easily be deployed simultaneously on the boom operator and on a pilot in a single seat cockpit at the same time and get the total team workload picture established with relative ease.

2. Introduction

2.1 Background

Evaluating the level of cognitive demand that aircraft systems place on operators is an important part of flight test. Subjective mental workload procedures are based on an operators personal jugment of the workload associated with task performance. Although subjective scales are well understood by the test pilot community they can lack in sensitivity, diagnosticity and are prone to bias. Developmental flight test could be improved with more sensitive and diagnostic testing to both pinpoint when and how operator demands are affected. In the past, use of biometric sensors during developmental flight test has been intrusive to the operators, frought with data precision errors, and cumbersome in a test environment. However, current technology has led to smaller less intrusive, higher fidelity, wearable sensors suites,that have been approved for in aircraft use.This an opportune time to develop biometric-based objective workload measures. There has also been considerable work on validating the human performance models that correlate physiological affects with workload. These objective data can lend themselves to more robust data analysis that can influence system development of human-machine interfaces in the future.

Aerial Refueling (AR) represents a cornerstone of the USAF air warfare readiness capability. Aircraft must be proven ready to be a tanker or receiver and new AR technologies must be assessed by the 418th FLTS (412 TW) to be certified for operational use. Certification includes test procedures of both the tanking and receiving aircraft separately and then again as certified pairs. Many factors such as gross weight, flight speed, maneuvering (banking, climbing), structural dynamics, bow waves, etc. influence the operational behavior of a tanker and receiver pair. Some of these factors interact only when the pair is formed up into a coupled system during receiving. Many components of performance make up the approval battery of certification. These include (but are not limited to) handling qualities, adequacy of controls and displays, adequacy of communications and procedures. Additionally, crew workload and attention requirements are an important factor in consideration of the pairing certification. If it takes an inordinate amount of concentration and off-scale flying skills to keep station during receiving, both the receiving pilot and the boom operator will “wear out” quickly or in a worst case, be unable to refuel. This can have dramatic consequences. Added circumstances of actual air warfare such as minimal fuel at join-up, bad weather, battle stress, etc. may compound the problem and cause the AR task to be unsuccessful.

Traditional AR installations such as in the KC-135 tanker involve an intrusive modification to the tanker airframe. The boom operator is situated in a prone (belly down) position looking out through a window in the tail section of the tanker. This installation is very costly because the pressurized airframe must be cut open and a structural window must be installed. Newer tanker configurations such as KC-46, KC-767, and KC-30 use a system of remote, tail-mounted cameras and a boom operator who is situated in a crew compartment within the aircraft. A structural window modification is no longer needed. The cameras are intended to provide a stereoscopic image of the area that was formerly seen through the window in the tail in the legacy airframes.

The boom operator in the tanker and the pilot in the receiving aircraft are a coupled human-machine system. If the boom operator is having trouble due to high workload, this is going to affect the receiving pilot and make his job harder. Therefore, assessment of both the boom operator and pilot workload is of great interest to the 418th FLTS as part of AR certification.

2.2 Physiological Assessment using the Cognitive Assessment Tool Set (CATS)

In well over a decade of physiological based assessment work, the Operator Performance Laboratory (OPL) has investigated many physiological sensors (e.g. EEG, EMG, etc.) and arrived at the conclusion that the electrocardiogram (ECG) waveform is the best signal for workload assessment. The research community has known for a number of years that human physiological signals in general, and ECG specifically, are deterministically nonlinear (also known as chaotic) systems [2-4]. Chaotic systems are often not well represented via the normal scalar time series. The dynamics of the system are obscured in the single dimension, but become apparent when transformed into a multi-dimensional embedded phase space [6].

Electrical conductive systems of the heart respond strongly to changes in a person's attention and cognitive effort. The OPL developed a unique, deterministically nonlinear dynamical classifier to assess cognitive workload [7, 8]. We refer to this approach as the Chaotic Physiological Classifier (CPC) method. In this study, we used the CPC method to transform raw ECG data in a cognitive workload metric. We call this metric the Transition Probability Variance (TPV)-Based Workload, or OPL Workload. Electrocardiogram (ECG) data was obtained from the pilot participants throughout the flight. The ECG time series data was then transformed into phase space using the CATS software tool [9]. This step established the Ergodicity Transition Matrices (ETMs) that represented the dynamics of the ECG signal in phase space for the different conditions in the flight. The transitions within the ETM were summarized in a single metric, termed the transition probability variance (TPV). TPV calculates the variance of the probabilities of transition from one cell to another different cell of the course-grained ETM. The TPV therefore captured the variability in the dynamics of the ECG signal as the participant underwent different levels of cognitive loading.

OPL workload was calculated continuously for the EP throughout each flight and a scale was established to put ITPV in a range that is similar to the Bedford Workload Scale [10]. However, we want to stress that we are making no attempts to exactly match the rubrics of the Bedford scale. Our procedure simply scales the minimum and the maximum observed values of an EP's experiment into a 1..10 range. The experimenter thus must ensure that during the experiment, there were some very high workload tasks and some low ones. Sometimes, we use surrogate tasks such as the Multi-Attribute Task Battery (MATB) [11] to administer tasks that are distributed along a range of difficulty levels. In this flight test, we used the organic evolution of the flight to select a range of tasks, including some that are relatively easy (e.g. takeoff), some that somewhere in the middle (e.g. on boom, position 0), and some that are very hard (e.g. on boom with turn and roll rate). TPV varies inversely to the degree of workload with higher TPV numbers seen under low workload conditions and low TPV numbers seen under high workload conditions. Thus, the raw data of relative workload is 1/TPV, which we call ITPV. ITPV itself is a very small number that is not practical for use by researchers. To scale the CATS ITPV into the range of 1 to 10 we start by taking the log of ITPV. Then, we determine the maximum and the minimum values of the dataset of the flight and linearly map the log(ITPV) numbers onto the chosen scale (e.g. 1…10). Any other scale could be chosen, such as for example perhaps going only to 7 or 8.

Understanding and monitoring the changes in the cognitive workload in pilots can offer critical quantitative information about their progression and performance while undergoing training. In this study, flight technical and physiological data was collected using the Cognitive Assessment Tool Set (CATS). CATS provides a real-time assessment of cognitive workload to allow performance assessment and training adaptations. CATS is a solution to collecting physiological data and quantifying cognitive workload, while being minimally invasive to the objectives of the training.

The CATS system is a multi-sensor, body-worn, fully self-contained, untethered physiological based data collection and analysis system that can assess aircrew cognitive performance in real world crew stations under a large range of mission parameters such as long endurance missions, high altitude, fatigue, stress, or thermal comfort. CATS has been used in numerous airborne flight environments including fast jets, large transport aircraft, and rotorcraft [12-17]. Figure 1 shows the CATS sensor kit and Dikablis eye tracker being worn by a pilot. In this figure, the pilot is wearing a 3-electrode electrocardiogram (ECG) montage under the flight suit. The ECG amplifier is located in a flight suit pocket and transmits the sensor data onto the CATS computer. Based on avionics integration needs, the integrator can chose to run CATS on a small form factor computer that is also worn on the pilot or on an aircraft mounted processor. The signals can be transmitted off the pilot onto the aircraft avionics using either a wireless connection or through a quick disconnect cable. The Dikablis binocular eye tracker comprised of two eye cameras and one scene camera. The signals for those cameras were delivered to the eye tracking laptop by means of an Instrumentation Vehicle Interface (IVI) as shown in Figure 1.

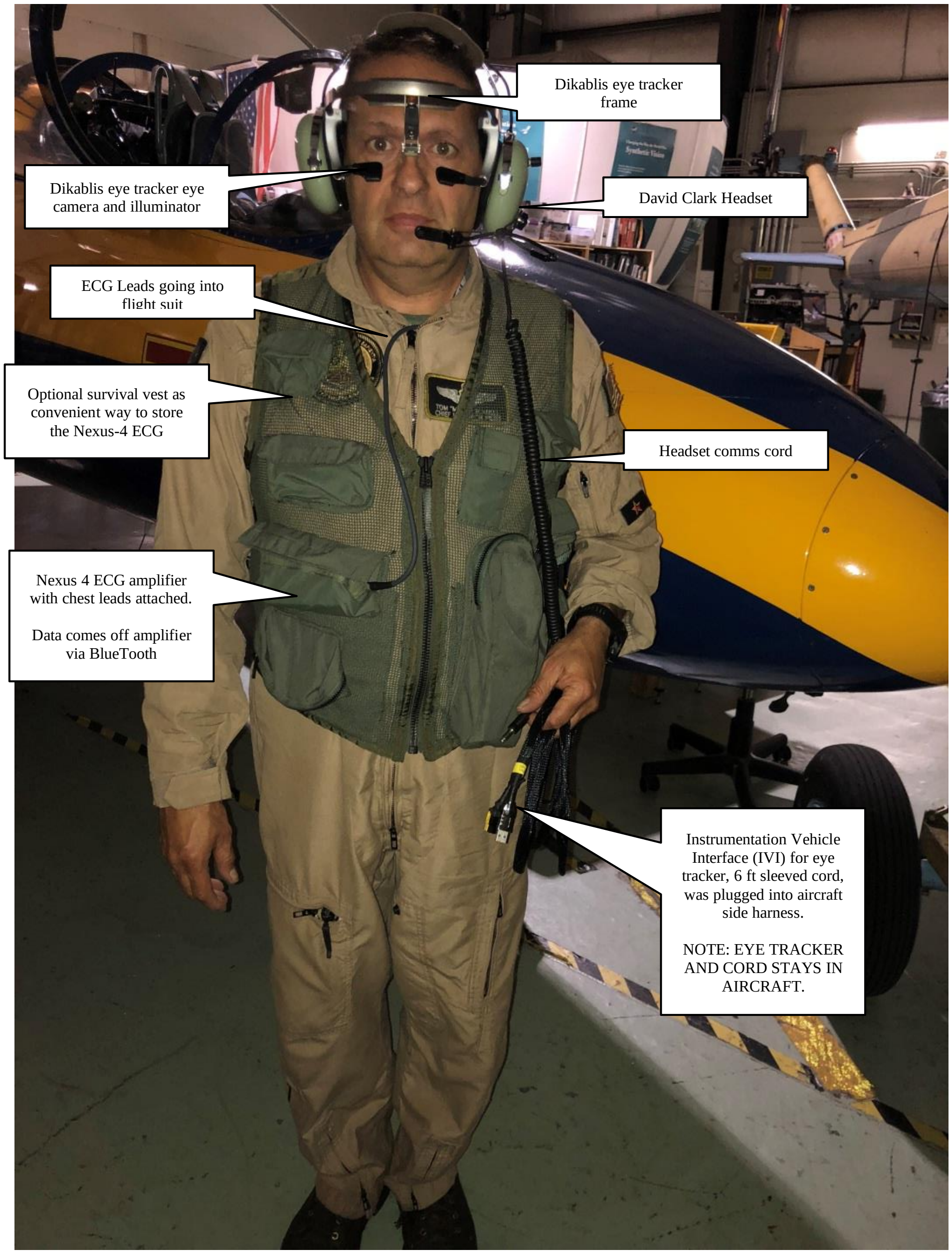


Figure 1. CATS Sensor Kit and Dikablis Eye Tracker

Figure 2 shows a high level process flow of the CATS system. The entire processing chain can operate on an aircrew-worn, fully self-contained physiological data acquisition and processing package with no physical tethers to the platform. The CATS architecture is net-centric so that it is possible to separate the acquiring sensor systems from the processing computing components, should that be desirable from an aircraft systems integration point of view.

The Cognitive Assessment Tool Set (CATS) [9] was developed by OPL. Body-worn sensors acquire physiological signals from the person. The CATS system is sensor agnostic as we are using specially developed providers (they are like device drivers) to translate sensor manufacturer proprietary protocols into the CATS provider protocol. This gives the customer great freedom in the choice of sensors and eliminates the cost for expensive, manufacturer specific, data collection systems licenses. The CATS provider system is based on a network protocol, where data can come through any data network from anywhere on or off the aircraft, to be stored within CATS. This is a very powerful capability that facilitates simple integration of the CATS framework in complex systems. Aircraft state, for example is acquired on the platform in various avionics systems such as the Time-Space-Position-System (TSPI) which comprises of Global Positioning System and an Inertial Navigation System (GPS/INS), Air data computers, Flight Management System, mission computers, etc. An aircraft-side CATS provider streams that platform specific data to the data intake on the CATS computer. Using the provider data protocol, the data now flows to the CATS data management system (Data Synch and Storage box in Figure 2) where the data is sorted by time and stored in a relational database with the time stamp as primary key. This automatically synchronizes all incoming data, no matter when it arrived at the data manager.

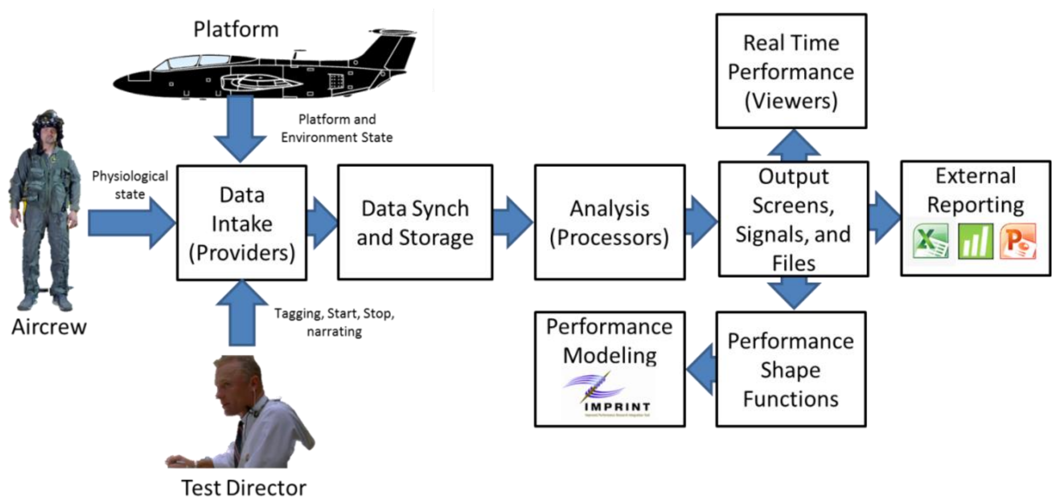


Figure 2. High Level Process Flow of CATS

All data analysis in CATS is done by what we refer to as processors. Processors accomplish simple mathematical functions like sorting, truncating, taking the Root Mean Square (RMS). Processors also accomplish more complex functions such as, but nowhere near limited to, calculating Shannon Entropy, calculating flight technical errors (FTEs) and weapons impact performance, calculating Ergodicity Transition Matrices (ETMs, for workload assessment), calculating workload (from the ETMs) and many other physiological based functions. Some processors run in real time and act upon the incoming data stream for real-time processing such as for example the real-time workload assessment. Other processors run when requested by a user for off-line data analysis or data output such as for use in external analysis and reporting tools such as Minitab, Excel, Powerpoint, etc. Users can enter data queries to extract combinations of data from the storage system and view it on graphs in CATS or export that data as Tab Delimited Text File. Graphs in CATS can be directly copied and pasted into external applications such as Powerpoint or Word.

Based on over a decade of experience in airborne physiological assessment, OPL derived a set of Most Important Requirements that govern the design of an aircrew physiological monitoring system. At the top of the MIR list are the requirements for the system to be unobtrusive and rugged, for real-world use, and capable of supporting unimpeded emergency egress (bailout or ejection). In our laboratory, OPL has used the Cognitive Assessment Tool Set (CATS), for many years in airborne flight test applications in fixed and rotary wing applications.

To give the avionics integrators a wide range of choices as to where to place the sensor and processing hardware, OPL gave the CATS system the capability of being fully self-contained on the aircrew without any physical tether to the aircraft. This drive of integration allowed for miniaturization of the system for research purposes. Given the distributed, net-centric, concept of CATS, the avionics designer can thus choose to locate some of the hardware on the pilot and some on the aircraft, but in any case, the form factor is small.

Our research system and capabilities demonstrator is strictly based on Commercial-Off-The-Shelf (COTS) parts. The current CATS system can run on processors as small as an UP-Board, a very small (85.6mm x 56.5mm x 15.0mm) PC that consumes only 5 watts of power. Sensors generally require additional hardware such as amplifiers or processors. Our COTS system comprised of the Up-Board and Nexus amplifier are about the size of a pack of cigarettes each. The amplifier supports a relatively wide range of body-worn physiological appliances such as EXG (electrocardiogram ECG, electromyogram EMG, electroencephalogram EEG, electrooculogram EOG), respiration belt, galvanic skin response (GSR), and body temperature sensors. The processor and amplifier are integrated in the pouches of the aircrew vest, although other locations can be easily designed for. The body-worn physiological sensor wires are routed through a hole in the flight suit to the amplifier in the vest. For a production implementation, these sensor pads would likely be integrated in a stretchy t-shirt or similar garment.

The aircrew wearing the CATS sensor kit are completely free to move around and ingress or egress the aircraft as needed. The CATS system has a startup script that enables data collection on all sensors as soon as power is applied to the processor. This one-touch startup was instantiated from another most important requirement (MIR) that states that data collection must be very simple. A tablet interface is incorporated through a secure (encrypted) Wi-Fi connection to the processor for additional system intervention such as starting and stopping data collection, tagging events in real-time, or activating higher level functions of CATS. OPL uses this capability as a research and maintenance function and use this capability in research flight tests with great success. Through aircraft data links, the CATS system can be accessed from a remote ground station for additional off-board interventions. This capability is useful in single seat aircraft or two-seat aircraft like OPL’s L-29 jets, where the front seat pilot must focus on flight related activities more so than experimental data collection activities.

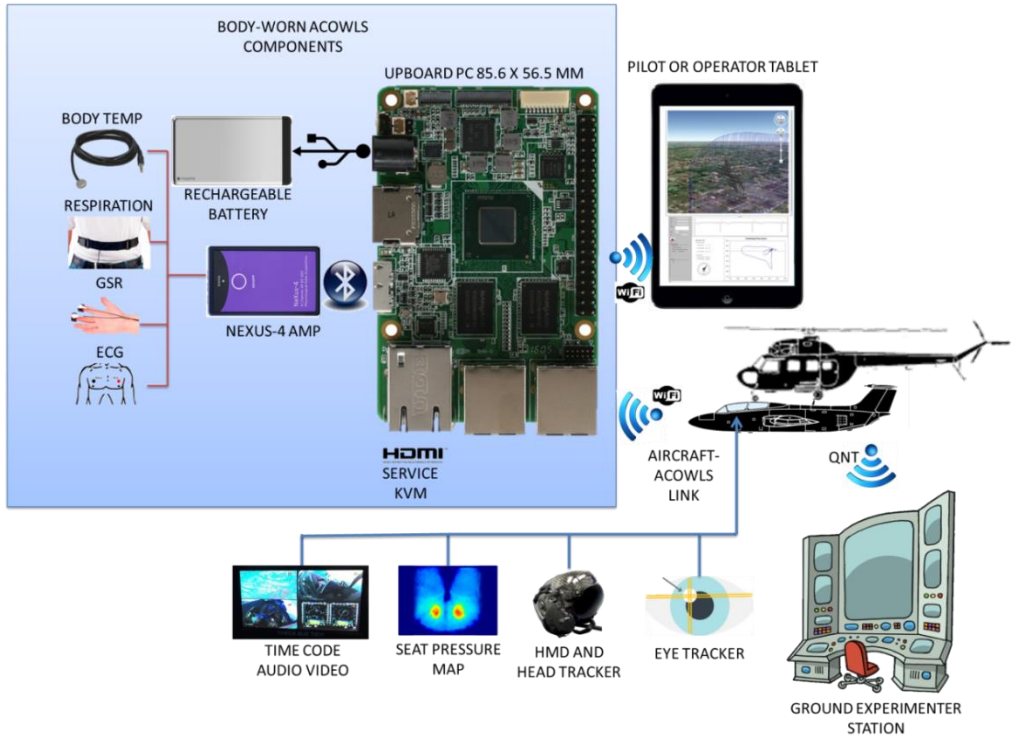


Figure 3. CATS Architecture Diagram

Over the years, OPL has developed providers for many COTS sensor systems to connect to CATS. The provider depends on an Application Program Interface (API) offered by the sensor manufacturer. Each provider acquires data from its COTS sensor, time stamps the data packet, and routes it to the CATS processor. For generic sensors such as ECG, GSR, respiration, eye tracking, etc. CATS already has data structures and processing capabilities built in. Switching to another brand of eye tracker, for example, is simply a matter of reconfiguring the provider to the API of the new eye tracker, leaving the data packets that are outbound to CATS unchanged. Thus, no changes are needed in CATS to operate on the new eye tracker.

CATS also supports multi-channel recording of audio/video streams. Video recording synchronization is accomplished with SMPTE time code audio that is generated by the data collection system and sent to the DVRs. Additional data that flows from the aircraft data concentrator to CATS may include button pushes, aircraft state, control input movement, etc. These variables provide very important contextual data that supports data and effects analysis. For example, in CATS, engineers can quantify control input (elevator, aileron) entropy, which is a measure of smoothness of control. This measure relates strongly to stress and overload conditions.

1. Technical Approach

1.1 Data Collection

OPL's Cognitive Assessment Tool Set (CATS) is a software application that collects data from various sensor and simulator systems (e.g. aircraft state), creates timestamps, synchronizes, and records the data into a MySQL database. Sensor and simulator data devices produce data at different rates and utilize proprietary data formats. CATS utilizes custom-defined providers that collect and collate the data from these variant sources. CATS allows the experimenter to effectively record and annotate data from human-in-the-loop (HITL) studies in a single, simple-to-operate user interface. Figure 4 shows a high-level view of the CATS architecture.

The CATS user interface allows the experimenter to select an experiment to run and then enter relevant information concerning the run under consideration, such as the participant identification number and the flight test number. Once the experimenter has entered all applicable experiment information, the recording of the run is started by clicking the record button. Throughout the run, the experimenter places tags in the data to indicate when something of note has taken place, such as the start or end of a use case. CATS features a number of worksheet tabs that can be accessed to view information concerning sensor data while recording. This ensures the ability to collect high-quality data throughout the entire experiment. One such tab screen is the Workload screen, which displays the real-time workload data related to the ECG sensor. Additional digital eye gaze data such as fixation durations, distance between subsequent fixations, pupil diameter, and many more, are synchronized and stored in the CATS relational database.

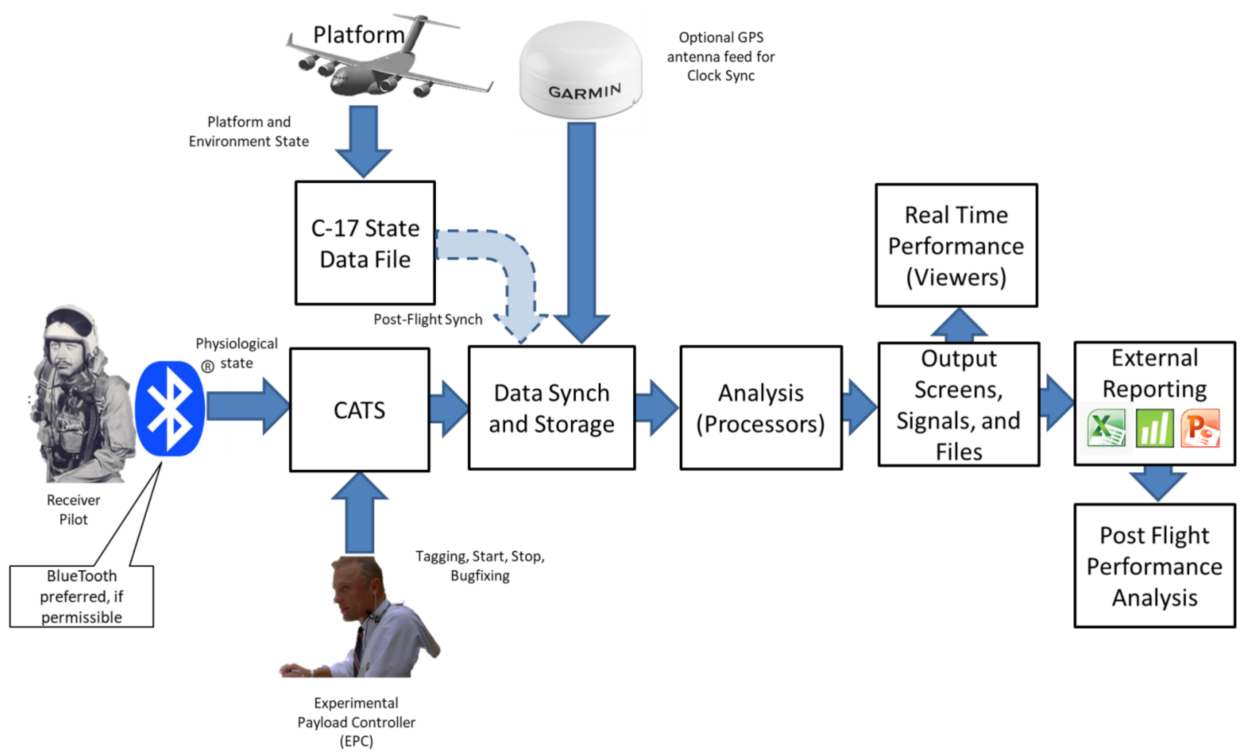


Figure 4: High Level Data Flow Diagram of CATS System on C-17

1.2 Instrumentation

For this study, CATS was provided with two main data sources: electrocardiogram (ECG) and eye-tracking.

For the purposes of acquiring the desired data for OPL, the following equipment was used:

1. 2 Dell Latitude laptop computers
    a. Laptops connected via Ethernet crossover cable
2. Nexus ECG device (wireless, other than electrode leads)
    a. With sets of electrodes and leads for each of the pilots
3. Dikablis Professional eye-tracking glasses
    a. With glasses connected to small Dikablis video processing box
    b. Dikablis box connected to Dell Laptop and power via cable harness
    c. 8 fiducial QR markers placed around right side of cockpit
4. One small lapel microphone placed in headset ear cup to pick up in-flight audio

The ECG data was collected using a Nexus-4 ECG amplifier. This system is a battery-powered, portable, wireless Bluetooth biofeedback system about the size of a pack of cigarettes. The pilot whose turn it was to fly was connected to the Nexus-4 system. The ECG leads connect on the Hiroshe connectors at the face plate and once connected, the amplifier was carried in a zipped up flight suit pocket. Data was transmitted via Bluetooth from the amplifier to the CATS laptop. The advantage of the Bluetooth connection is that other than the electrode leads, this configuration was completely unobtrusive to the crewmember wearing the system. This was beneficial in the case of seat swaps or emergency egress, when compared to a wired connection.

Figure 5: Nexus-4 ECG Amplifier

Eye-tracking data was somewhat secondary to the workload data from CATS. The goal was to identify and mark key points of interest in the cockpit and track what areas the pilots were most focused on during flight operations. OPL used the Dikablis Professional Glasses from Ergoneers in conjunction with Dikablis eye-tracking software known as D-Lab. Alongside CATS, D-Lab records the eye-tracking video and tracks eye movements such as saccades and fixations. Once recording is complete, D-Lab post-processing of the eye-tracking data can produce gaze maps, heat maps, and other visualizations to help analyze the results, as well as output eye-tracking statistics related to all of the above.

Figure 6: Dikablis Professional Glasses (photo from Ergoneers, Inc.)

The Dikablis glasses were only operated from the right-hand seat, as that is where the fiducial markers were placed. In future flights, it would be possible to operate from both seats if markers were placed on both sides. The glasses were worn by the right-seat pilot, with a cable running from the glasses to the Dikablis box placed in a pouch behind the seat. A cable was then run from this box directly behind the cockpit seats and along the bulkhead to the crew rest area where the OPL laptops were located. Outside of the actual glasses being worn by the pilots, this setup resulted in next to zero footprint in terms of cables or markers interfering with flight operations.

Two OPL members sat in the crew rest area operating the laptops, one for the eye-tracking and another for CATS. These laptops were secured with Velcro to the bench/bunk directly in front of the jump seats. One OPL member had to occasionally leave the jump seat to position and calibrate the Dikablis glasses on the pilot before data could be collected.

2. Design of Experiment

A total of seven evaluator pilots performed takeoff, aerial refueling, landing, and simulator tasks to determine the utility of the CATS and eye-tracking systems in the context of an engineering feasibility study. There was a total of seven separate task cards for flights and four for simulation, designed to put the pilots in high workload scenarios in different phases of aerial refueling and other flight operations. These tasks included:

Flight:

- Normal Takeoff
- Normal Flight Operations
- AR Boom Limits
- Contact Station Keeping
- Normal Approach
- Offset Approach
- Normal Landing

Simulator:

- Low Difficulty Boundary Avoidance
- Medium Difficulty Boundary Avoidance
- High Difficulty Boundary Avoidance
- "Impossible" Difficulty Boundary Avoidance

Pilots performed all aerial refueling tasks using the legacy control system. A full listing of the test cards can be found in Appendix A.

Before stepping to the aircraft, each pilot was fitted with the majority of the required instrumentation, both from OPL and the other groups involved. This included OPL's ECG electrodes and the 711th's ECG belt and EEG cap. Once onboard the C-17, the pilot flying would connect to the Nexus system and (if in the right-hand seat) put on the Dikablis glasses just before beginning flight operations. During debrief pilots completed subjective questionnaires to evaluate the level of comfort and intrusiveness.

Each task was performed with one pilot acting as pilot flying, one as pilot monitoring, and a third in the crew rest area. Tasks were performed for a given seat configuration with each pilot performing as many tasks as possible for each piece of instrumentation without requiring a seat swap. Once a seat swap was executed, the tasks were repeated to allow different instrumentation configurations. This process was repeated for the duration of the flight, completing as many tasks as possible before having to return to base.

For tasks in the simulators, pilots were outfitted with the same instrumentation as during the flights (but not the Dikablis eye tracker). The task consisted of flying closely behind a fighter aircraft and keeping them within two horizontal bars in the HUD, as well as keeping the wing-tips of the aircraft inside the HUD crosshair. The pilot had stick control but not throttle or rudder, and the fighter aircraft continuously changed pitch, intending to cause a Pilot-Induced Oscillation (PIO) for the test pilot. This task was made more difficult by varying the width of the horizontal bars, and by introducing delays in the control inputs Additionally, the pilots were given an auditory task in which a monotone beep was played every three seconds, occasionally changing pitch. The pilot had to keep track of how many times the tone changed, and pressed a button held in their left hand every time this change occurred.

3. Flight Summaries

3.1 Flight 1

Date: 28 August 2018
Flight Duration: 5hr 21min (18:41 – 24:01 UTC)

For the first flight, there were 3 instrumented pilots (2 in the cockpit, 1 in the crew rest area), with seat swaps allowing pilots to perform multiple tasks from different seats. Two seat swaps were performed. Pilots 1 and 2 performed each task, while Pilot 3 only completed Boom Limits and Station Keeping.

ECG data was collected in CATS from each pilot for every task. In several instances, the ECG data may be unusable due to muscle EMG artifacts obscuring the ECG readings; however, the rest of the data is legitimate and provided the desired results.

Eye-tracking data was only obtained from the right seat on roughly half the tasks, and the 711th Tobii glasses were used for the other tasks. The glasses were calibrated adequately for 2 of the 3 pilots, but outside factors made calibration difficult for the third pilot. One of the factors leading to this difficulty in calibration was the fact that Pilot 3 glasses, which also prevented the use of the Tobii eye-tracking glasses.

3.2 Flight 2

Date: 30 August 2018
Flight Duration: 3hr 1min (15:46 – 18:47 UTC)

The second flight was shorter, and only had two instrumented pilots, with a third pilot (not instrumented) to assist with seat swaps. Ultimately, due to time constraints, no seat swaps were performed.

ECG data was collected for each pilot on every task that pilot performed. However, due to shorter flight time, the operation hit joker time and was forced to return to base, so not all task cards were completed for each pilot.

Eye-tracking data was not obtained for this flight. Complications with the glasses' calibration process made any data gathered for the first pilot invalid. An attempt to recalibrate was made, but to no avail. Had the flight been long enough to involve a seat swap, the glasses could have been used with the other pilot, which may have yielded better results.

3.3 Flight 3

Date: 5 February 2019
Flight Duration: 6 hr 24 min (18:06 – 24:30 UTC)

Flight three was the longest of the four PWS flights. Three pilots were aboard and instrumented. Pilots B and E flew on flight three after previously flying on a previous PWS flight, but only Pilot E performed AR tasks.

Due to airspace restrictions and weather, this flight operated at a higher altitude block than the other three flights, which led to some interesting workload results, discussed further in the Results section.

ECG Data was obtained for each pilot, but OPL did not use eye-tracking. All other instrumentation was the same as previous flights.

3.4 Flight 4

Date 7 February 2019
Flight Duration: 3 hr 51 min (17:53 – 21:44 UTC)

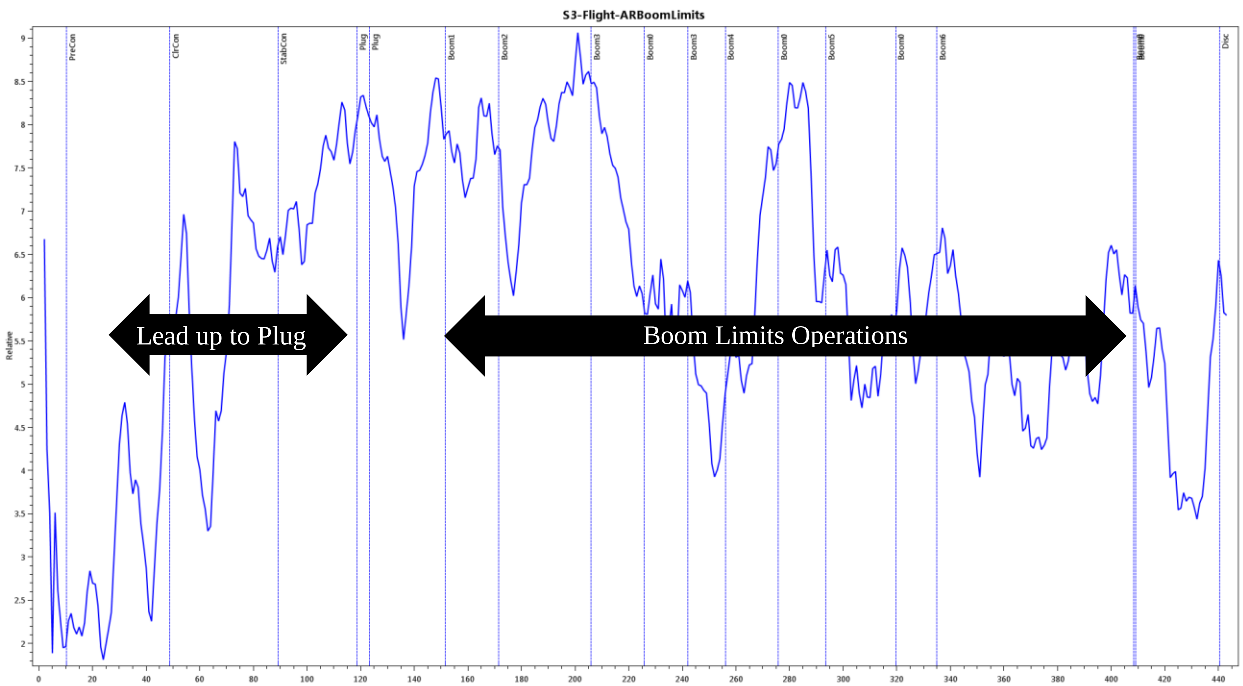
The final flight had three pilots aboard and instrumented. Pilot A and Pilot C flew on flight four after having flown on flight one as well.

Flight four went very smoothly, with successful ECG data collection for all pilots and few operational issues regarding the OPL crew.

4. Results

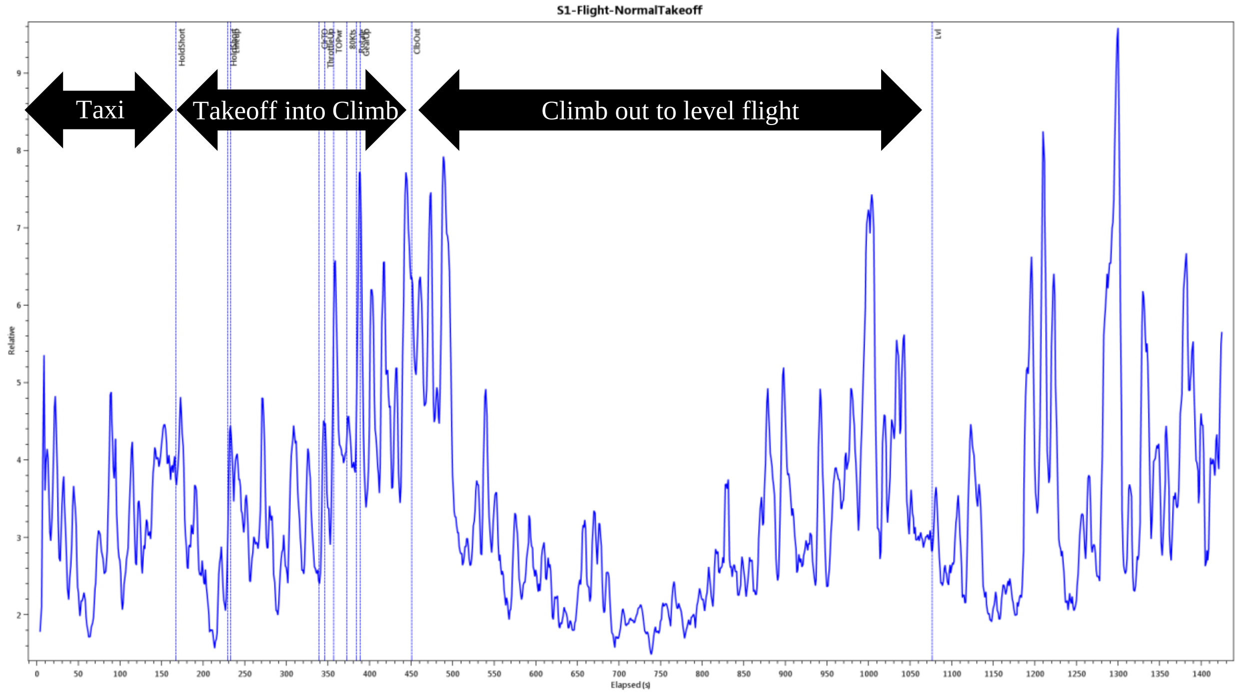
4.1 Workload Results

Workload scores obtained from CATS have allowed for direct comparisons of workload and task. This report shows summary data in graph form and the actual workload data files are available upon request. Every task for each pilot was recorded in a separate dataset to keep the results organized and uncluttered. For example, Figure 7 below shows the scaled (relative) workload output for a boom limits task, and how workload fluctuates as the task progresses. The trace shows that at the beginning, workload is relatively low, but as the gap between the C-17 and KC-135 closes and the aircraft is cleared for and then stabilizes for contact, workload increases. As the pilot progresses through the boom positions, workload is ever-changing, but remains at a heightened state. In many high workload tasks such as this one, we routinely see the highest workload at the beginning of the task with a gradual reduction as the pilot becomes accustomed to the high workload task. Towards the end of the task, workload goes down as the end of the task is approaching. With one final spike just before disconnection, the task is complete.



NOTE: Y Axis shows Log(ITPV) which was scaled to 1…10 Scale, then the data was smoothed to plot the graph

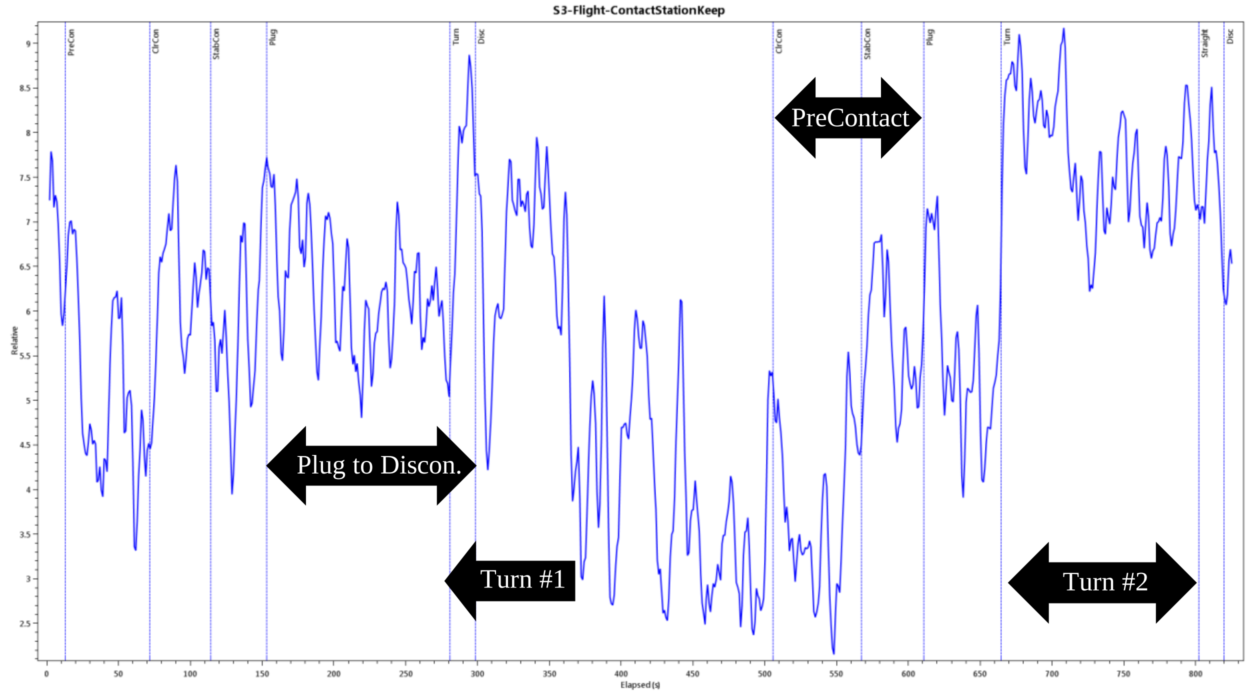
Figure 7: Workload Data for Boom Limits Task



NOTE: Y Axis shows Log(ITPV) which was scaled to 1…10 Scale, then the data was smoothed to plot the graph

Figure 8: OPL WL data for takeoff and climb

On takeoffs, (Figure 8), a gradual increase of workload as the aircraft taxis and starts off down the runway can be seen during the first half of the task. There is a peak just after rotation as the aircraft begins climb out, but workload drops swiftly shortly thereafter as altitude increases and the relatively straightforward climb continues.

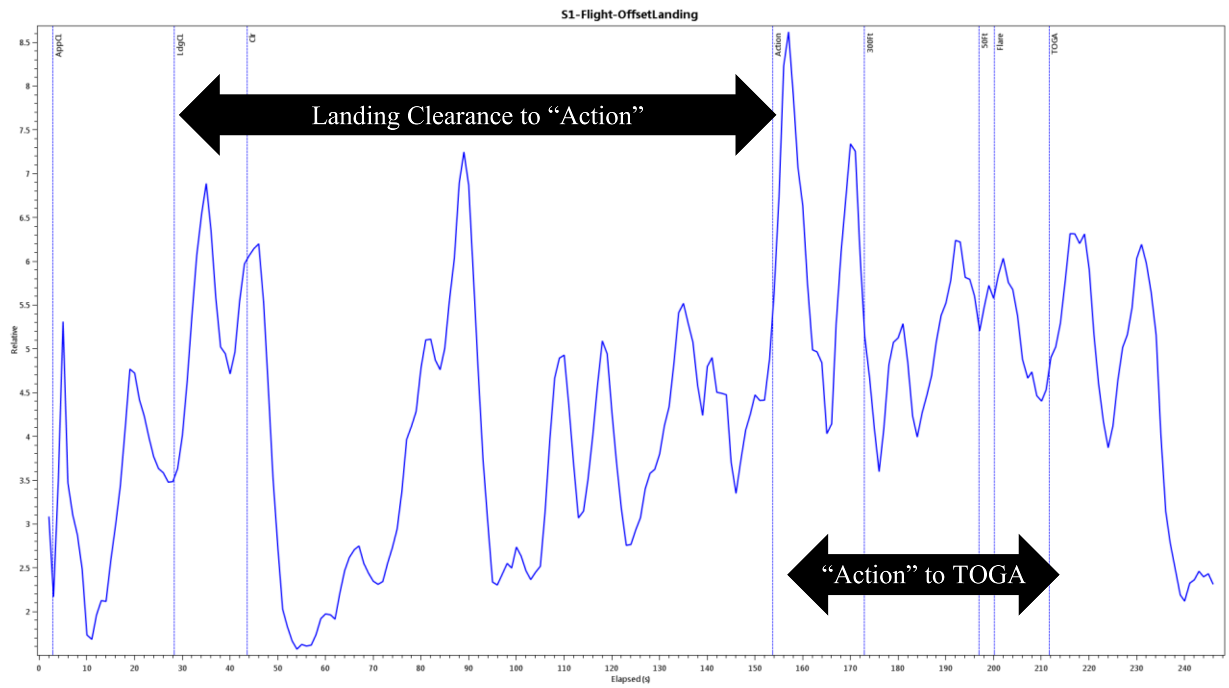


NOTE: Y Axis shows Log(ITPV) which was scaled to 1…10 Scale, then the data was smoothed to plot the graph

Figure 9: OPL WL data for Station Keeping task

From Figure 9 above, the workload trace during boom station keeping is shown. Workload is highest when the turn begins, and when the aircraft rolls out back to level flight. This is consistent with the turn rate compensation needed to roll in and out of the turns. During the duration of the turn, workload remains at a heightened level. Additionally, the moment the C-17 disconnects from the boom, workload immediately drops, as the pilot is given a brief moment of respite to regroup himself and regain formation.

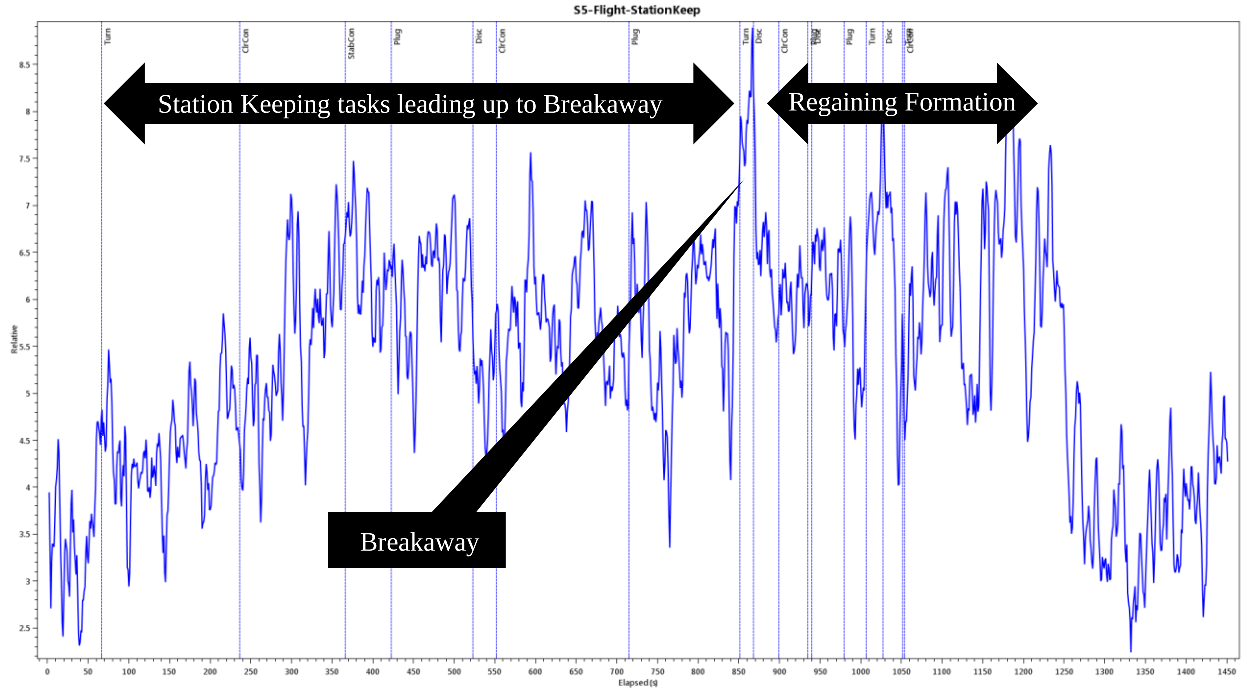
During the offset landing task (Figure 10), workload was shown to gradually increase leading up to touchdown. At the moment the pilot monitoring called "ACTION," there was a pronounced jump in workload as the pilot flying sought to correct the flight path and center the aircraft with the runway.



NOTE: Y Axis shows Log(ITPV) which was scaled to 1…10 Scale, then the data was smoothed to plot the graph

Figure 10: OPL WL data for an Offset Landing task

Unexpected occurrences were also captured in the workload data. For example, during Flight #3 there was a “breakaway” situation during boom limits. The breakaway call is reserved for situations that are judged to be unsafe. In this instance, the breakaway call came from the boom operator, who said that contact was missed, and that the C-17 was already approaching the forward limit of the envelope while the KC-135 had the autopilot disengaged. As can be seen in FIGURE, this led to a distinct spike in the workload of the pilot flying (and certainly everyone else aboard either aircraft as well).

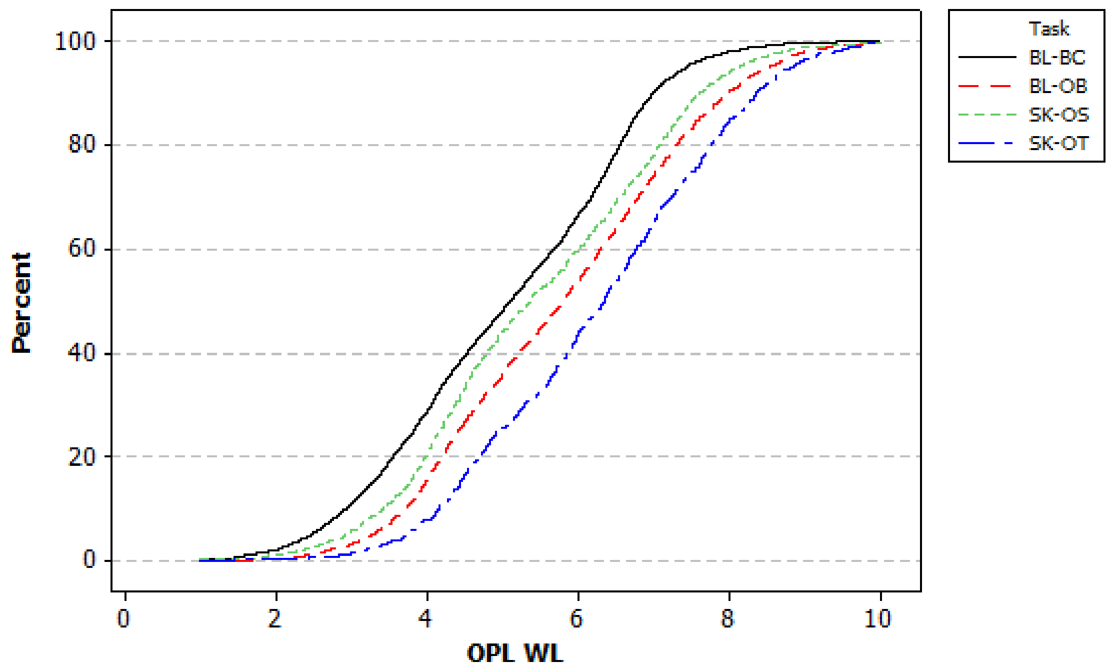


NOTE: Y Axis shows Log(ITPV) which was scaled to 1…10 Scale, then the data was smoothed to plot the graph

Figure 11: OPL Workload data for Station Keeping task including breakaway

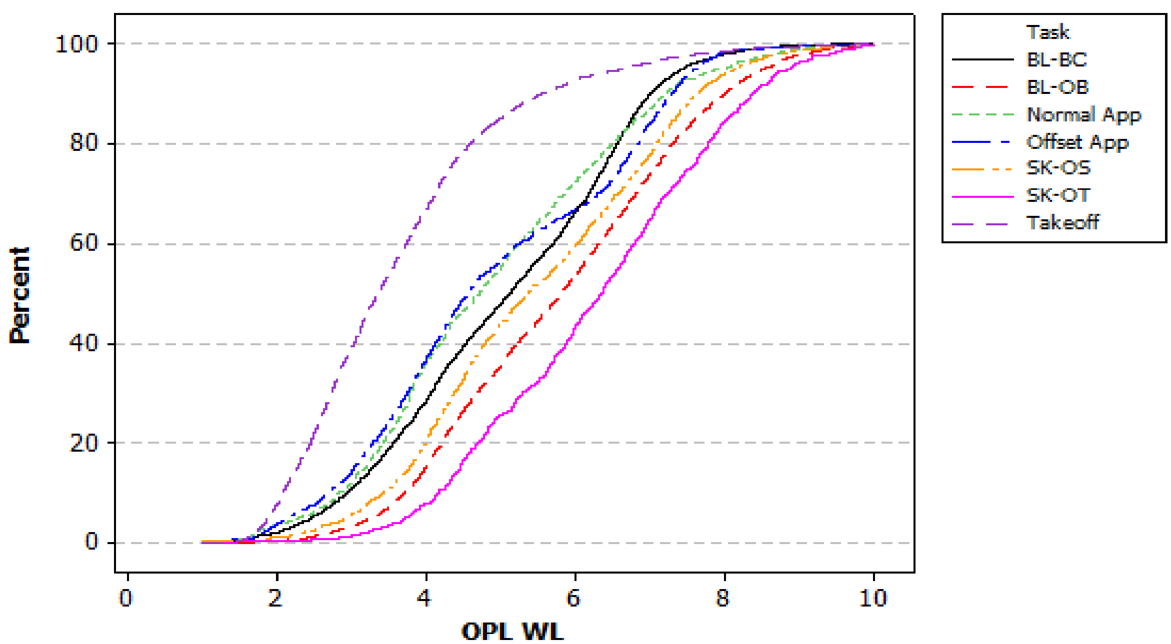
The workload data was then separated even further, not only by task but by certain points of the task. Boom limits (BL) was separated into two groups: before contact (BL-BC) and on-boom (BL-OB). Pilots also noted during the test flights that during station keeping tasks, the portions during the turn were far more difficult than flying straight ahead on the boom. The station keeping (SK) data was then separated into two groups as well: on-boom, straight (SK-OS) and on-boom, turn (SK-OT). Figure 12 shows that the pilots were correct that the turning portion of station keeping was consistently higher workload than the straight sections, and that on-contact boom limits was higher workload than pre-contact.

When the workload scores of all tasks are overlaid in the same plot (Figure 13), it is possible to see the workload difference in each task. Station keeping and boom limits are the most difficult tasks, while takeoff is comparatively easier. One interesting finding is that normal approach appears to be higher workload than offset approach. This may be true, but a likely cause is the order in which the tasks were performed. Since the normal approach came before the offset approach, the pilot may have been more relaxed for the second approach when he was more familiar with the landing conditions and had already performed a touch-and-go just moments before. Had the task order been flipped, it is probable the offset landing would show a higher workload.



NOTE: X Axis shows Log(ITPV) which was scaled to 1…10 Scale

Figure 12: Empirical CDF comparing station keeping tasks while flying On-boom, Straight (SK-OS), On-boom, Turning (SK-OT), boom limits on boom (BL-OB), and boom limits before contact (BL-BC)



NOTE: X Axis shows Log(ITPV) which was scaled to 1…10 Scale

Figure 13: Empirical CDF comparing workload results of all tasks

Digging deeper into the data, workload scores were separated based on the boom position during the boom limits task (Figure 15). As shown in Figure 14 below, there were 6 boom positions. The workload data was split into three groups: Vertical movement (position 1 and 2), horizontal movement (3 and 4), and axial movement (5 and 6).

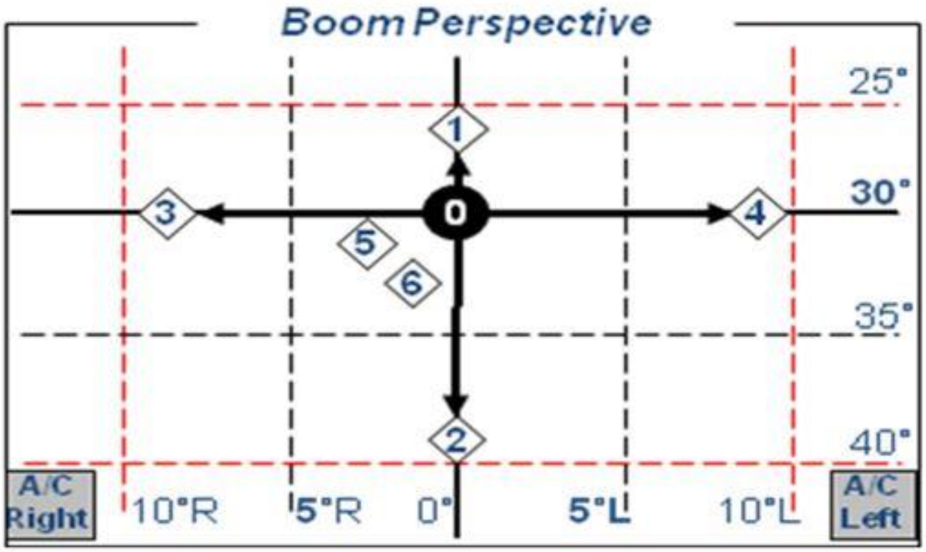
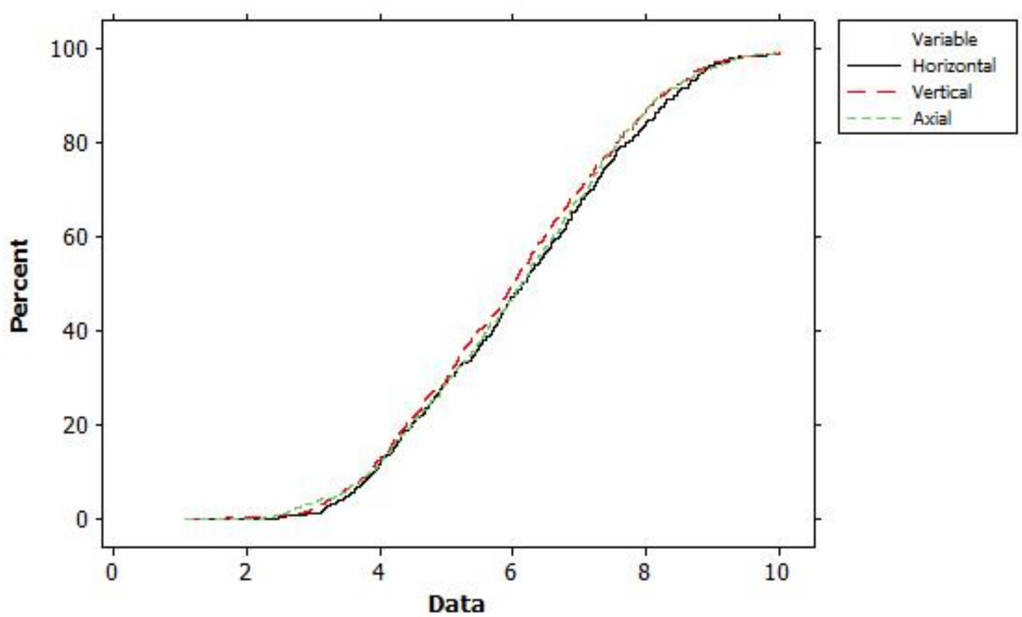


Figure 14: Boom positions shown from perspective of tanker



NOTE: X Axis shows Log(ITPV) which was scaled to 1…10 Scale

Figure 15: CDF comparing OPL WL scores for boom limits maneuvers

A hypothesis was made that some movements during boom limits may be more difficult than others, while the workload data shows that there was no discernible difference between the three movement modes. Horizontal movement workload scores tended to be insignificantly higher than the other movement modes. There may not have been a significant comparison to be made in this case; however, the data has shown the capability to make such a comparison. Given a broader set of tasks and more opportunities to gather similar data, interesting comparisons and trends may become apparent in time.

One such comparison that presented itself following the February flights focuses on Flight #3. Due to weather conditions, airspace restrictions and traffic in the area, the aerial refueling tasks had to be performed at a higher altitude than the other three flights. In addition to the turbulent conditions that day, this higher altitude resulted in less responsive throttle controls leading to a higher degree of difficulty in maintaining refueling formation. Figure 16 below compares the workload scores observed during AR maneuvers (Boom Limits and Station Keeping) for all four flights.

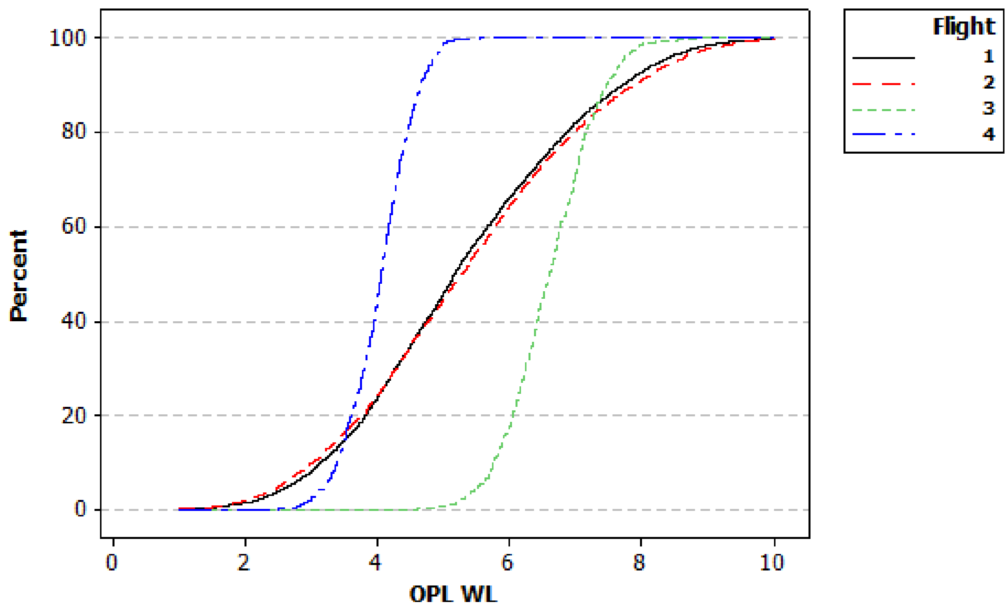


Figure 16: CDF showing OPL Workload scores for aerial refueling tasks only, separated by flight

Looking more closely, there were two pilots (Pilots B and E) on Flight #3 that had previously flown on PWS Flight #2 in August, but only Pilot E performed refueling tasks on both flights. Looking directly at the workload scores from Pilot E's AR tasks, a direct comparison can be made between the two flights in Figure 17. The only difference between the flights other than the weather and altitude was the absence of OPL's Dikablis eye tracker. If anything, the absence of the Dikablis glass could have made the task *easier*, based on pilot feedback from the August flights (Section 5), but the added difficulty of the maneuvers eliminated any benefit from the reduced instrumentation.

During the refueling tasks, Pilot E commented that the throttle "felt loose" and that the most difficult part was closing the gap between aircraft during pre-contact. Because of the higher altitude and decreased air density, the throttle input response became more sluggish, making fine inputs (such as those required for formation flying) more difficult. To counteract this, Pilot E used only the inner two throttle levers for finer control for forward and aft movement in the envelope.

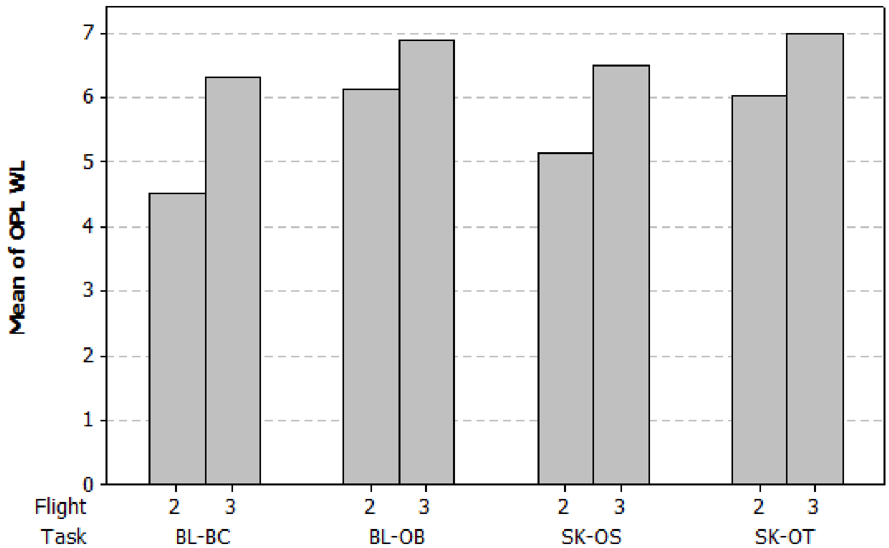


Figure 17: Mean OPL Workload Scores for Pilot E (L-R: Boom Limits–Before Contact, Boom Limits–On Boom, Station Keeping–Straight, Station Keeping–Turn)

In Figure 18 through Figure 20, a direct comparison can be made between OPL WL scores and pilot-provided Bedford & TLX scores. The results show a definite correlation between the two. Pilots rated station keeping and boom limits tasks as highest workload, and the OPL WL scores reflect this. Takeoff and landing emerged as the easiest tasks, a sentiment which is also reflected in the OPL WL data. Bedford and TLX scores were only provided for boom limits as a whole, and not separated by on-boom and pre-contact, so no direct comparison can be made.

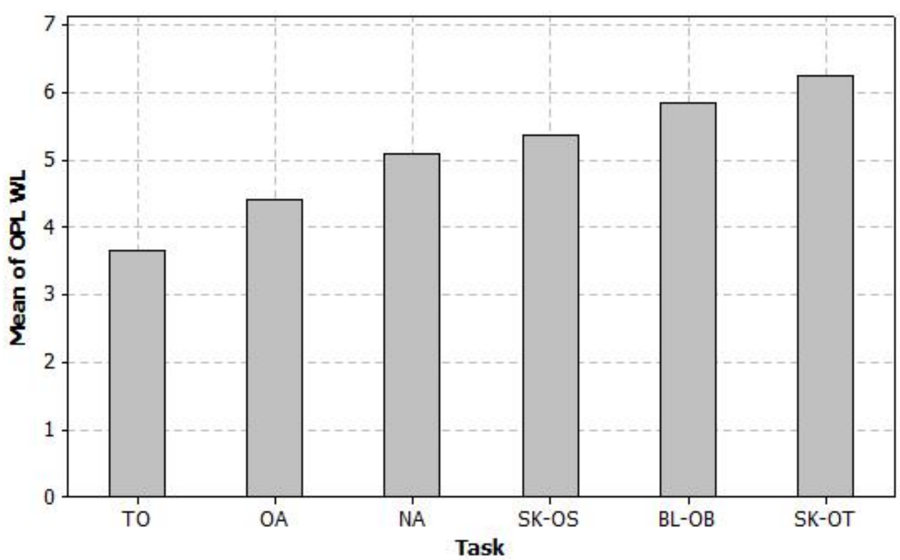


Figure 18: Mean Relative Workload separated by Task (L-R: Takeoff, Offset Approach, Normal Approach, Station Keeping – Straight, Boom Limits – On Boom, Station Keeping – Turn)

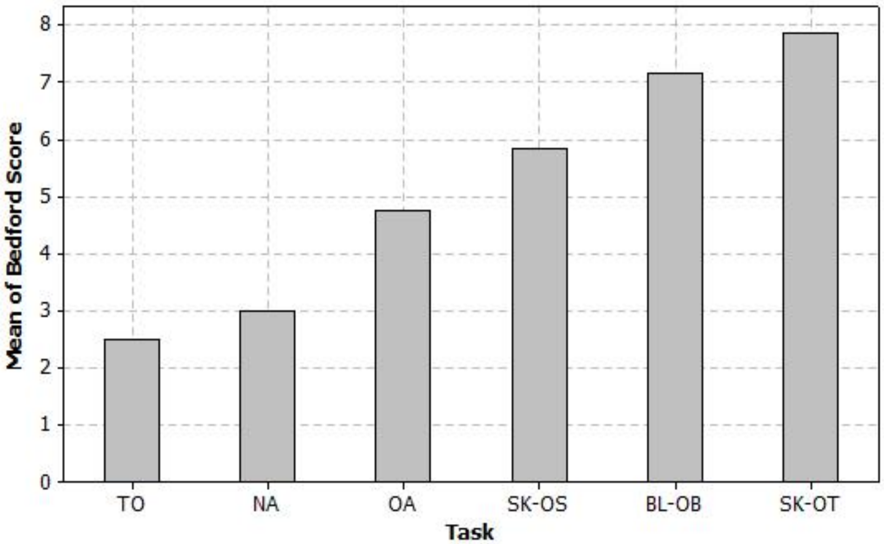


Figure 19: Mean Bedford Workload Scores separated by task (scale of 0 - 10)

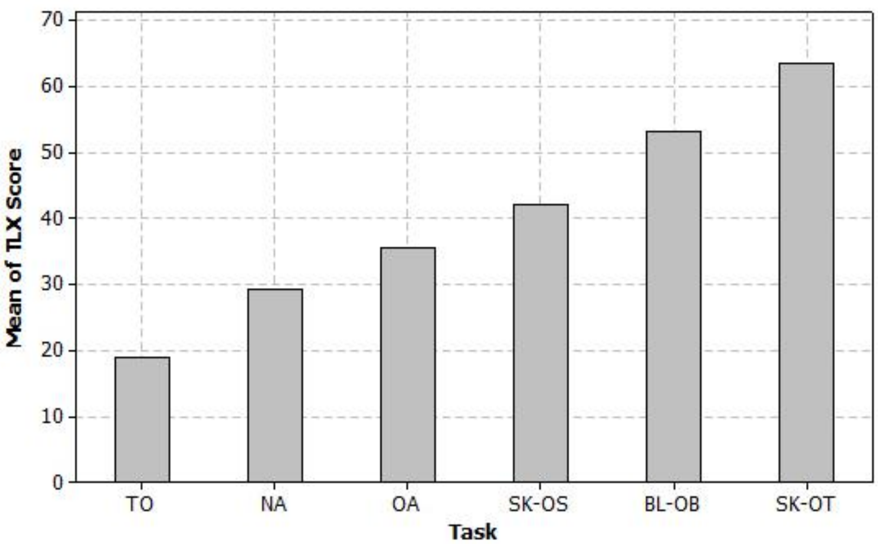


Figure 20: Mean TLX score separated by task (scale of 0 - 100)

An interesting note about these comparisons: while there is a good correlation between OPL Workload and the subjective scores, the subjective scores were occasionally shown to be unreliable. For instance, as mentioned above, the conditions of Flight #3 resulted in a higher-than-normal workload environment. It could be expected that the Bedford scores would reflect that, but looking again at Pilot E (who was on both Flight #2 and #3), this is not the case. Figure 21 shows that Pilot E actually reported *lower* Bedford scores for two-thirds of the AR tasks during Flight #3, while the OPL Workload scores reflect the assumption that Flight #3 was more difficult.

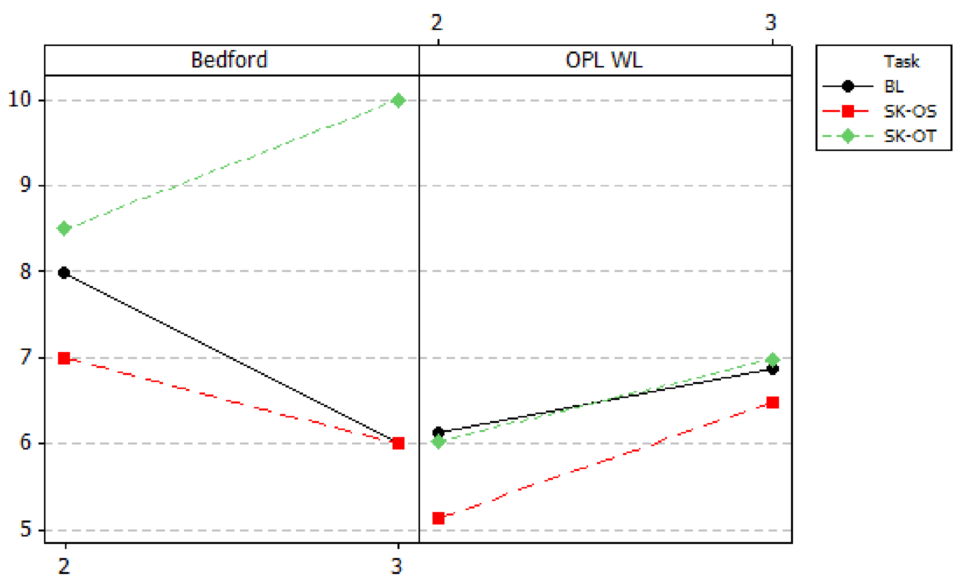


Figure 21: Difference in Bedford and OPL WL scores between Flights #2 and #3

We generated Table 1 to allow a basic comparison of OPL WL with corresponding Bedford ratings. It should be noted that the OPL WL was a model-less ITPV based method because we did not have the ability to perform a battery of surrogate tasks that would have allowed for development of a meaningful discrete classifier. Also, no baseline was made for simple straight and level and simple turning maneuvers. The ITPV method used herein is very simple and provides reasonable results. For better calibration, we would have had to perform a better baseline battery of tasks, particularly one that included surrogate tasks to delineate the extreme ends of the spectrum at very easy and very hard. Figure 22 shows a simple linear regression model that can be used to estimate the Bedford rating (BF) when entering OPL WL scores into the equation BF=-0.550 + 1.114 OPL WL The adjusted R2 is relatively low (38.1%). This should not be a surprise as the average OPL WL measurement is aggregated across multiple minutes and any subjective ratings such as Bedford (or TLX) are notoriously undiagnostic of the real workload [16, 17]. As such, any correlation between a dynamic measured variable and a subjective lump sum, post-run variable should be taken with a grain of salt. It is our position that the objective measurement method we show herein is a simple tool to obtain a relative comparison of workload between test conditions. Our method can also be used to generate absolute measures of workload but for that, the user needs to invest time in collecting a serious battery of surrogate task measures so that a classification model can be developed.

Table 1. Pilot's Average OPL Workload

	TO		BL-BC		BL-OB		SK-OS		SK-OT		NA		OA	
Pilot	OPL WL	BF	OPL WL	BF	OPL WL	BF	OPL WL	BF	OPL WL	BF	OPL WL	BF	OPL WL	BF
A	3.4	3.0	4.5		5.2	6.5	4.3	5.0	5.4	8.0	4.6	3.0	4.3	6.0
B	3.8	2.0	4.6		6.4	8.0				7.5	5.9	3.0		3.0
C			5.4		6.2	7.0	6.1	6.0	7.7	8.0	3.6	3.0	3.7	4.0
D			4.9		5.8	6.0	6.0	5.0	6.1	7.0				
E	4.5	2.0	4.5		6.1	8.0	5.1	7.0	6.0	8.5	4.9	3.0	4.6	4.0
F	3.2	2.0	3.5		4.1	6.0	4.2	4.0	4.6	7.0	3.7	4.0	4.1	4.0
G			6.5		7.7	10.0	7.2	8.0	8.2	10.0	6.8	2.0	6.9	3.0
AVG	3.8	2.3	4.9		5.9	7.4	5.5	5.8	6.3	8.0	4.9	3.0	4.7	4.0

Note: BL=Boom Limits, SK=Station Keeping, BC=Before Contact, OB=On Boom, OS=On Boom-Straight, OT=On Boom-Turn, NA=Normal Approach, OA=Offset Approach

NOTE: OPL WL=OPL Workload based on Log(ITPV) which was scaled to 1…10 Scale, BF=Bedford Workload Scale. Gray shaded area denotes Pilot did not perform task or submit Bedford score for that portion of the task.

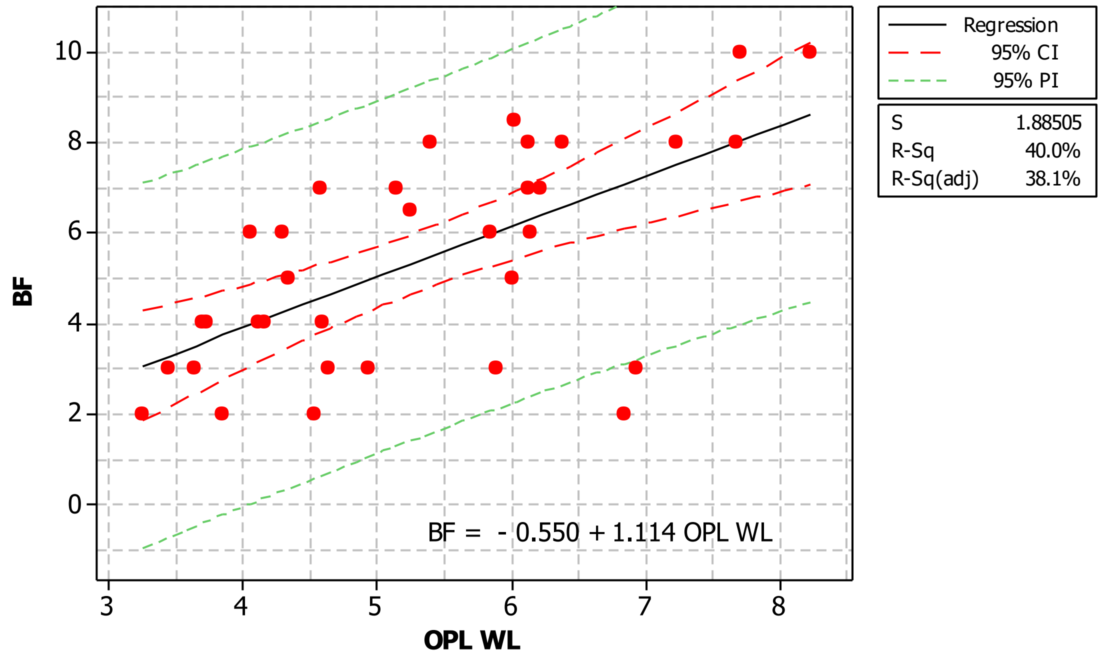


Figure 22. Simple Linear Regression Line for BF as a Function of OPL WL

4.2 TPS Simulator Workload Results

Five of the seven pilots from the PWS flights also participated in simulator tasks. On 11 and 12 February, the pilots were outfitted with the same instrumentation used in the C-17 flights. Two pilots performed the simulator runs on the 11^{th}, and three more went on the 12^{th}.

Each pilot had a quick familiarization run followed by four separate test cards in the simulator. Each test card consisted of the same task: following closely behind an F-16 while keeping the F-16 in between two horizontal bars in the HUD (Figure 23), as well as keeping the wingtips of the F-16 from crossing the opposite point of the HUD crosshair. As the F-16 flew, it would make slight changes in its pitch and altitude, but its heading was constant. The task was designed to cause a Pilot-Induced Oscillation in the simulator aircraft. Each test card followed the same task of following the F-16, but the difficulty of each was modulated by changing the flight controls of the simulator and the width of the HUD bars. The pilot was allowed up to 2 vertical and 1 lateral violations for “desired” performance, and up to 3 vertical and 2 lateral violations for “adequate” performance. Further violations would result in a “not complete” rating for the test card. Figure 23 shows how violations were determined.

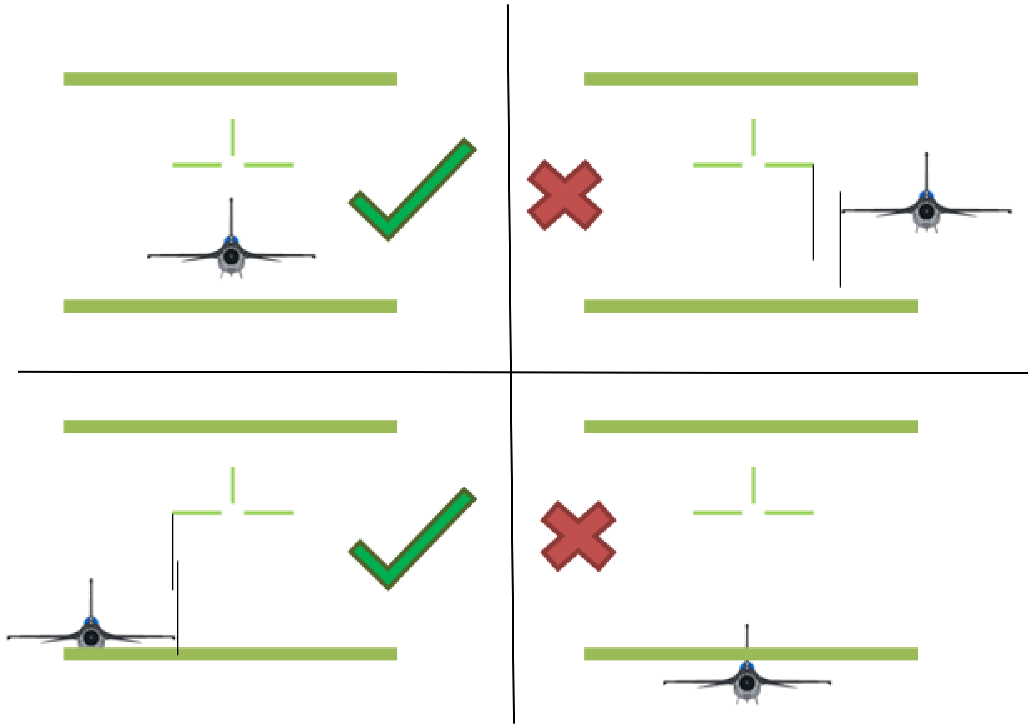


Figure 23: Representations of simulator HUD and boundary constraints for chase aircraft

Due to limitations of the simulator, each run would only last 1 minute 15 seconds before the chase aircraft overtook the F-16. At this point, the TC would quickly reset the simulator and the test card would continue. Each run was repeated four times per test card, for a total of five minutes of data collection. This led to an issue with the flight controls, however; as each time the simulator was reset, the stick was centered. For example, if the pilot was performing a right roll when the simulator was reset, the new "zero point" of the stick would be to the right of its physical center position. This essentially caused the aircraft to become increasingly out of trim. It was not noticed and corrected until the final pilot of the study, so every other pilot had the additional control challenge.

While performing the primary flight task, the pilots had to perform a secondary task of keeping track of audio cues. Throughout each test card, a brief tone would be played every 3 seconds. Roughly 15% of these tones would be at a different pitch than the others. The pilot was instructed to keep track of these pitch changes, both mentally and by pressing a button held in the pilots' left hand.

Each of the four tasks was designed to roughly correspond to the four branches of the Bedford Workload Scale and were performed in a random order. For the "impossible" task, the goal was for the pilots to fail the task, but in such a way that would hold the pilots' engagement by seeming somewhat attainable.

The order in which the pilots performed each difficulty level was carefully considered. We did not want the order of scenarios to influence the results in any way. It was finally decided to run the order as follows:

1. Familiarization Run
2. Medium Difficulty
3. Low Difficulty
4. Impossible Difficulty

5. Hard Difficulty

It was hoped that by running the scenarios in a non-linear manner it could reveal how adept the pilots were at accurately judging their own workload, rather than just giving a higher rating each successive attempt if they knew it was getting linearly more difficult. It can be seen in Figure 24 below that the Bedford ratings given by the pilots were very much in line with the intended difficulty of each task.

The workload results seen in Figure 25 show that the average workload score overall did not vary widely by task, but it does line up with the intended difficulty level of each task.

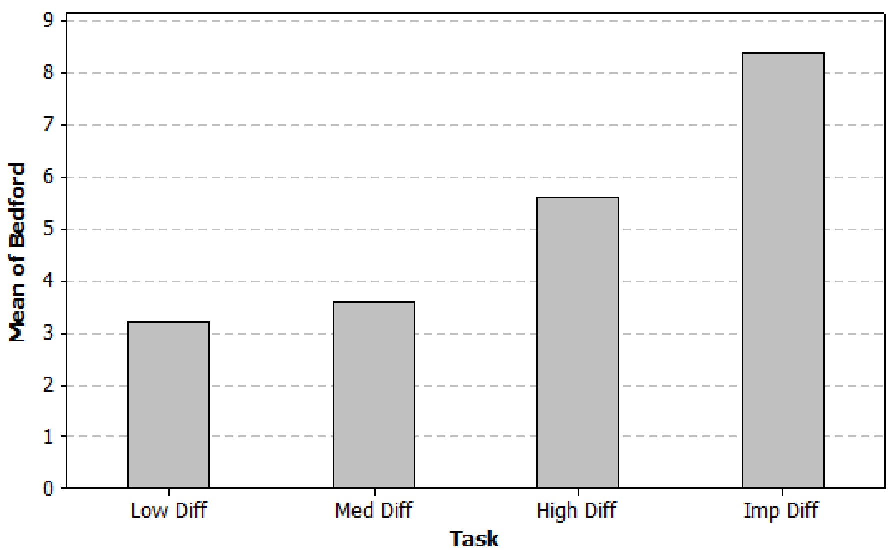


Figure 24: Mean Bedford Score by simulator scenario

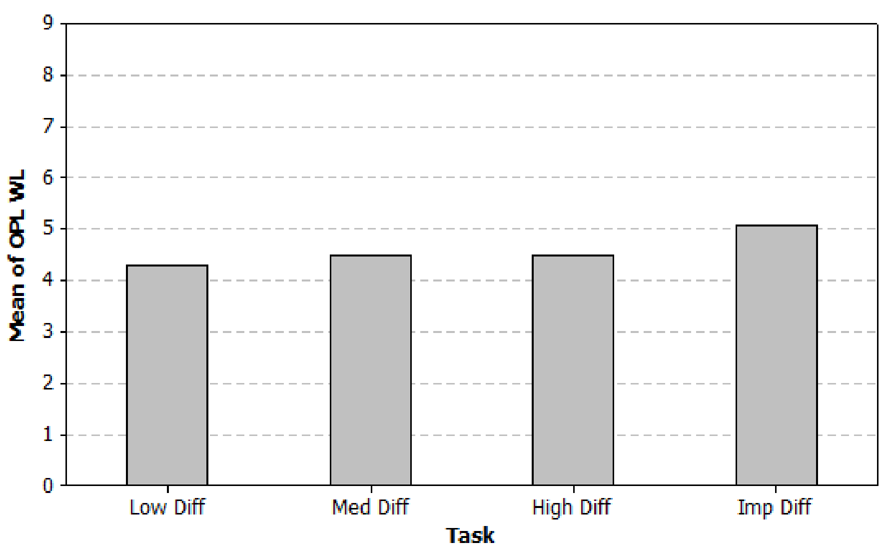


Figure 25: Overall mean OPL Workload by simulator task

When the workload of all pilots is averaged together, the results become slightly flattened. Due to many factors with workload measurement (physiological differences, instrumentation issues, etc.) each subject may have different workload scores for the same level of workload. When a single pilot is examined (Figure 26) the variance in workload becomes clearer. There is still not a stark difference between each, but it is certainly more noticeable.

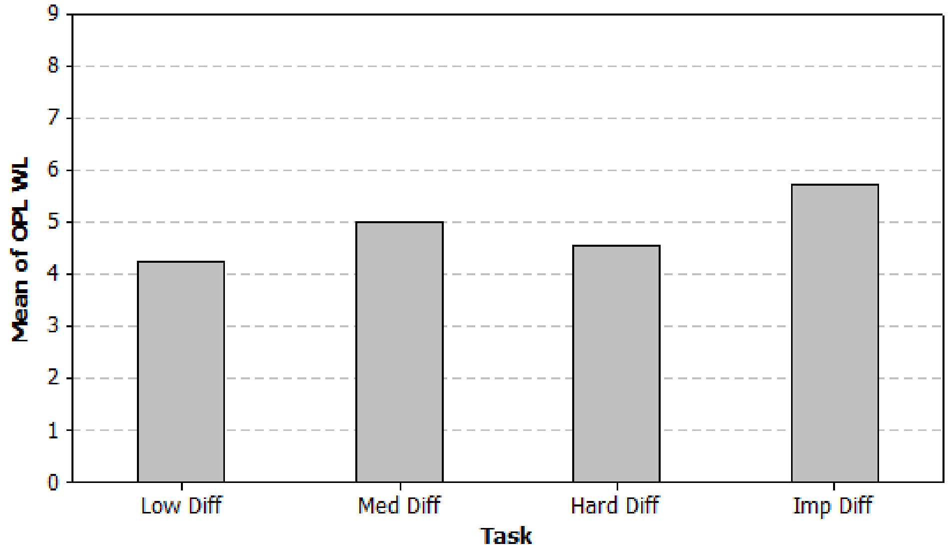


Figure 26: Mean OPL Workload by simulator task for only one pilot

From the time series plots of OPL Workload scores (examples shown in Figure 27 through Figure 29), workload increases and decreases can be linked to specific events. In Figure 27 below, a high difficulty task is shown. It is possible to see exactly when the pilot was struggling to stay in bounds, and how each time the simulator was reset, the pilot appeared to compose themself a bit, and workload dipped.

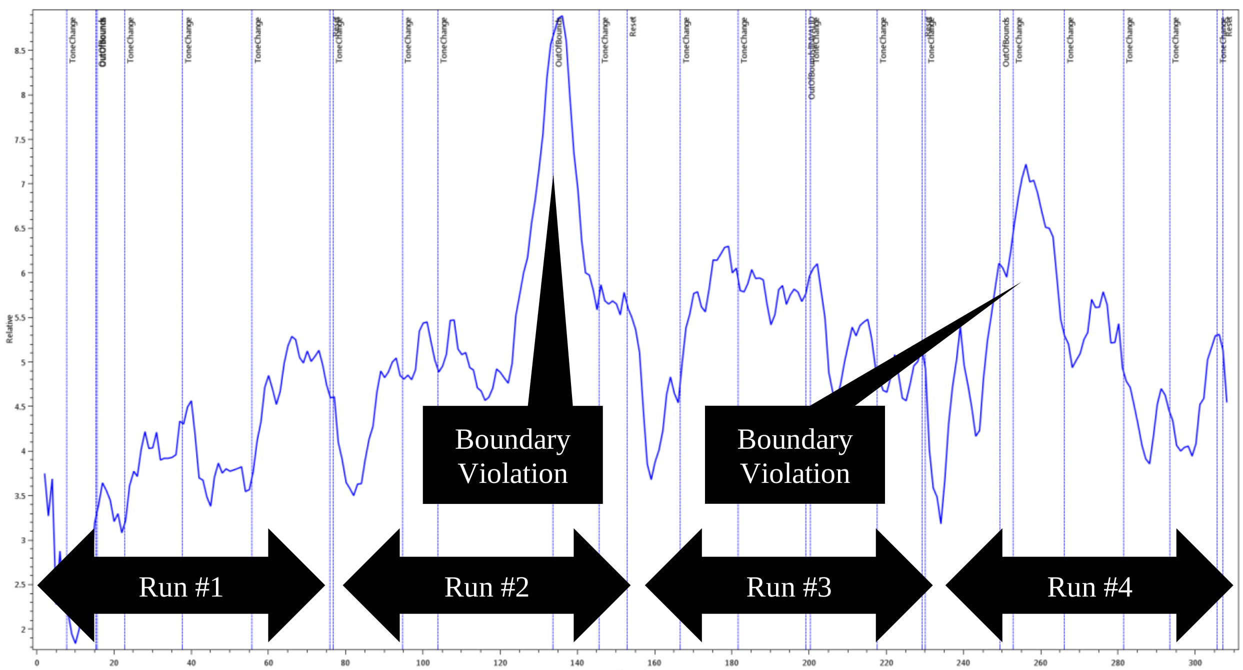


Figure 27: Elapsed OPL Workload for single pilot, high difficulty task

The two figures below show two examples of the "Impossible" task, and two different pilots workload responses. Each vertical line in the plot is either a tone change in the auditory task, or an instance of a boundary violation. In the first, the pilot immediately violates the boundary and enters a state of increased workload. Control is briefly regained before the first reset. Run #2 is very similar. The final two runs have a much lower workload, so the overall trend is that workload decreases from beginning to end.

In the second case, the opposite trend is observed. Workload bounces up and down quite a bit as the pilot violates the boundaries occasionally but stays pretty low overall. As the task continues, workload goes up by quite a bit as the pilot works harder and harder to keep the aircraft within the boundaries, and by the fourth run nearly succeeds.

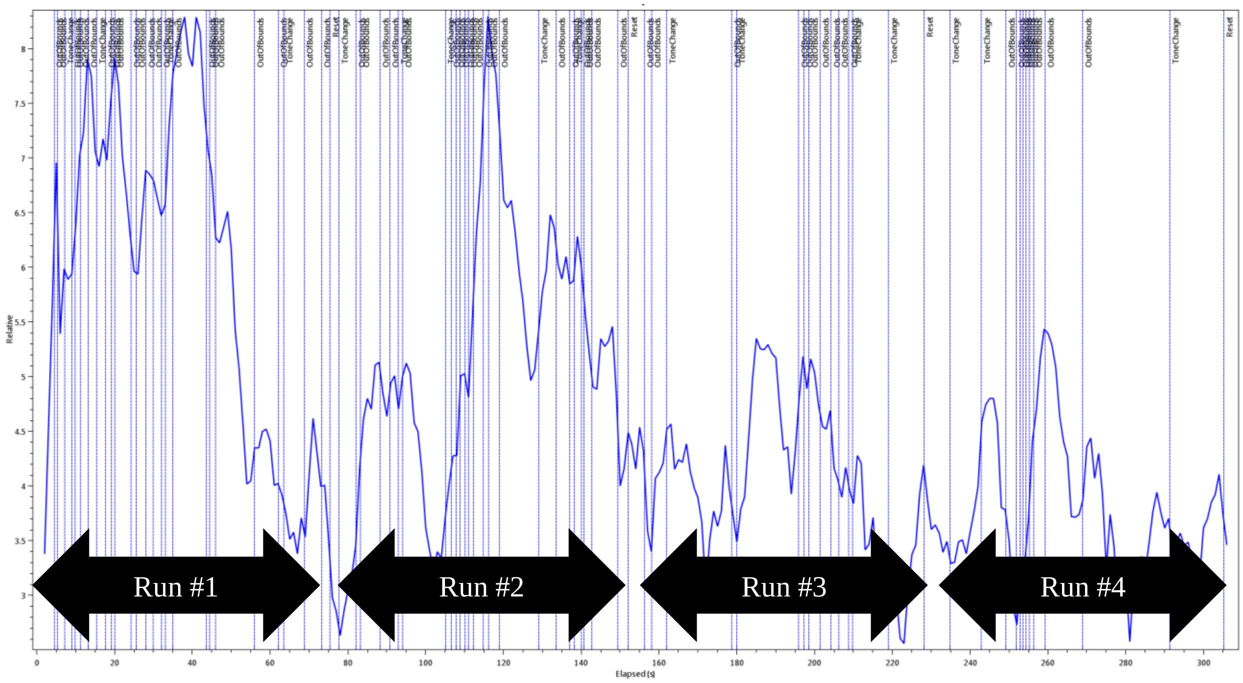


Figure 28: Elapsed OPL Workload for single pilot, impossible difficulty task

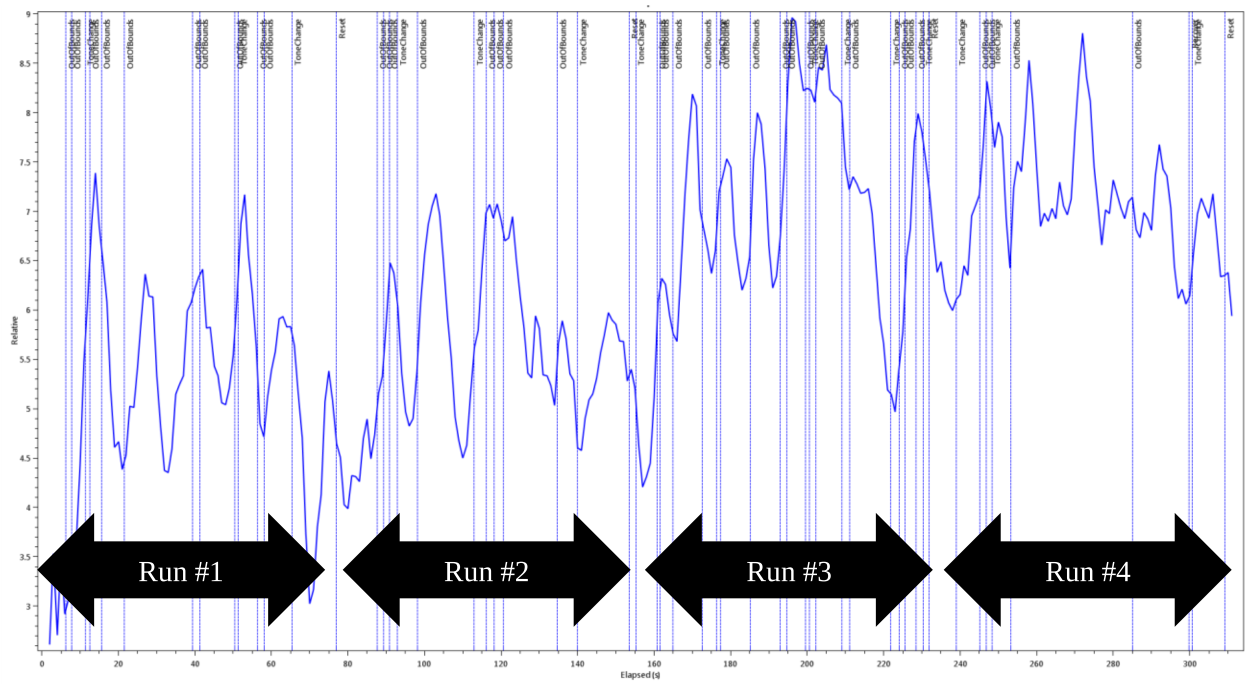


Figure 29: Elapsed OPL Workload for single pilot, impossible difficulty task

Two additional pilots did the simulator tasks on February 13th but were only outfitted with the OPL ECG leads and did not perform the auditory oddball task. These pilots were TPS instructors who were willing to help us out. We wanted to gather a bit more data, but since the conditions were different than the previous five pilots, these data sets will not be included with the previous five pilots. These two pilots were labeled as Pilot 8 and Pilot 9. They came into the simulator room, were instrumented with the CATS ECG leads, ran through the scenarios, and were out the door in roughly 30 minutes each.

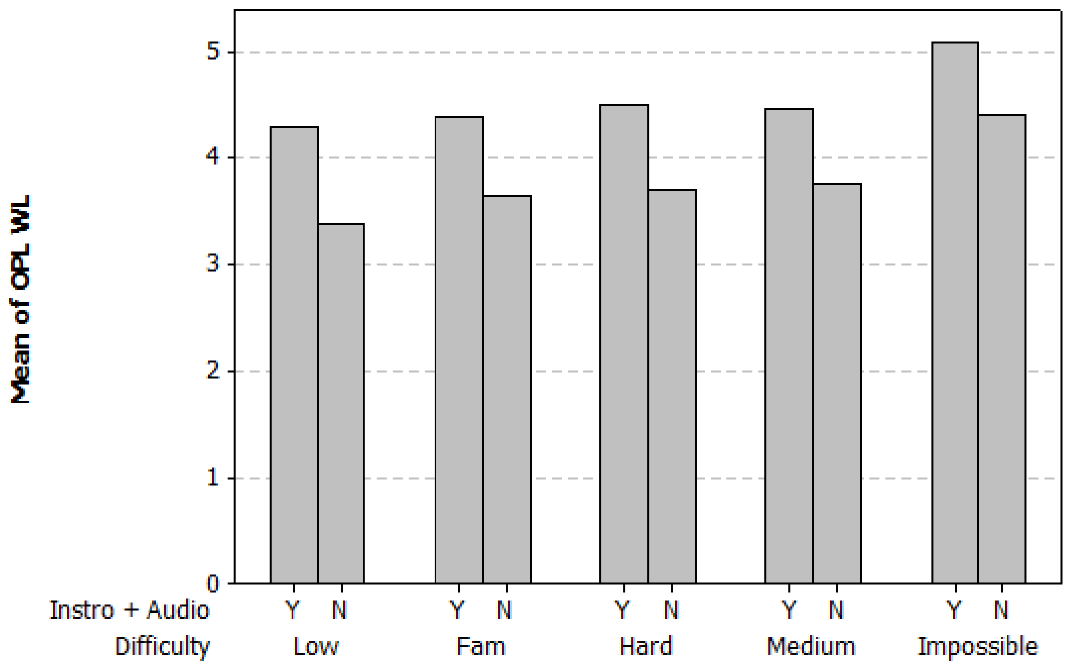
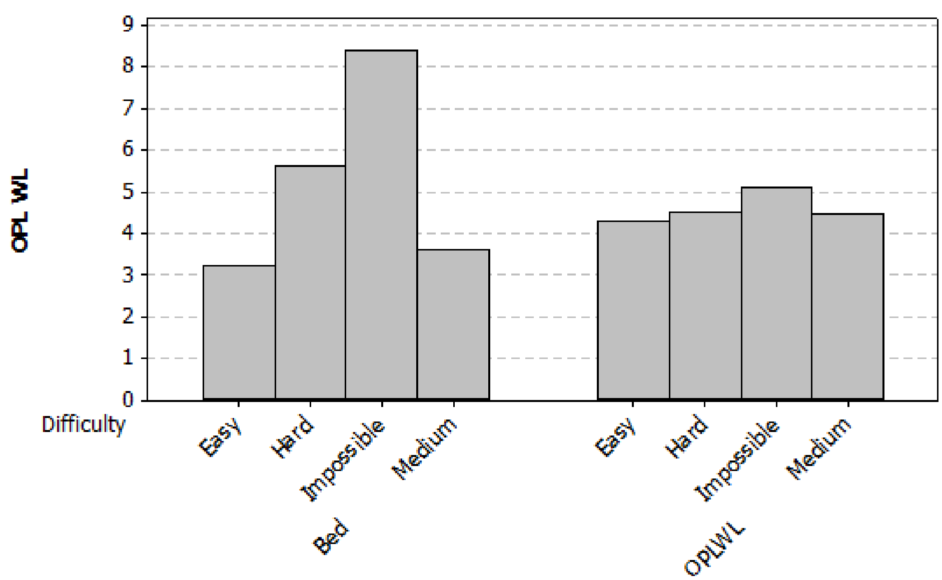


Figure 30: Mean OPL Workload during simulator runs both with and without full instrumentation

The workload results of Pilots 8 and 9 show the same trend in difficulty level but overall a lower workload across all tasks. This was expected, as these pilots were likely more recently active on the simulator, were minimally instrumented, and had no secondary task to perform simultaneously.



`

Figure 31: Mean OPL WL vs Bedford scores in simulator tasks

Figure 31 further illustrates the correlation between objective workload measurement and subjective self-reporting of perceived workload. Both the Bedford scores and OPL's workload measurement accurately identify the relative difficulty of each task; however, the OPL workload metric does not give such pronounced differences between levels of workload between tasks. There could be a number of causes for this. Subjects may be over- or underrating their workload based on perceived difficulty of the task. For example, in a very hard task followed by an easier task, the pilot may rank the workload of the easier task as lower than it objectively was because it was noticeably easier than the previous task. With OPL's objective workload, this bias is completely out of the equation. Additionally, workload may not have been as high simply due to the environment of the simulator. Pilots would conceivably work harder during an actual flight versus simply sitting in a simulator on the ground.

4.3 Eye-Tracking Results

Due to the nature of the aerial refueling tasks, the eye-tracking data was rather narrow in scope. During refueling, the pilots would use the parallel stripes ("captain's bars") on the underside of the tanker as a point of reference. This resulted in an essentially static gaze. The eye-tracker was collecting data, but since the pilots only looked at one point with occasional glances to check their instruments, there is only so much to glean from that data. Additionally, there were several complications with the eye tracking results. Even when the eye-tracker was calibrated and operational, unforeseen issues occasionally prevented good data collection. For example, when the sun was in certain positions relative to the aircraft, the fiducial markers didn't register due to shadow or glare. For the second flight, eye-tracking yielded no usable results due to difficulties calibrating the glasses with one of the pilots. These factors together resulted in less data than was desired, but the data that *was* obtained yields some interesting results.

Figure 32: Eye-tracker scene camera view during aerial refueling task

One aspect that was tracked was how much the pilots used their HUD versus simply looking out the window. Using the fiducial markers placed around the cockpit, so-called Areas of Interest (AOI) were created in D-Lab. These areas can be seen above in Figure 32. One such area is the box outlining the HUD, while the window area around the HUD is a separate AOI. Three fiducial markers can be seen outlined in red. Using these AOIs, the software then calculated exactly how much the pilot was looking at each area.

The HUD was used only part of the time, so the data sets are limited, but it is interesting to see how the pilots tend to use the HUD during refueling operations.

Figure 33 shows the attention ratio for both the HUD and the window area surrounding the HUD. The data set was taken from a single pilot performing the boom limits task. The HUD dominated the pilot's attention, most likely because of how he used the HUD as his reference point when flying behind the tanker. He only looked out the window around the HUD about 22% of the time, mostly when checking the tanker. The "other" category is for times when the pilot was not looking at any specified area of interest, such as glancing around the cockpit, or simply because the pupils were not detected for brief amounts of time.

The concept of the fiducial markers worked exceptionally well and greatly facilitates automatic world-model based analysis of glance occupancy in areas of interest (AOIs). The ergonomics of the headset left much to be desired and the flexible eye-camera mounts are all but impossible to adjust accurately. At OPL, we use the Dikablis eye tracker integrated in flight helmets. Using that concept, the eye tracker is highly reliable and very accurate.

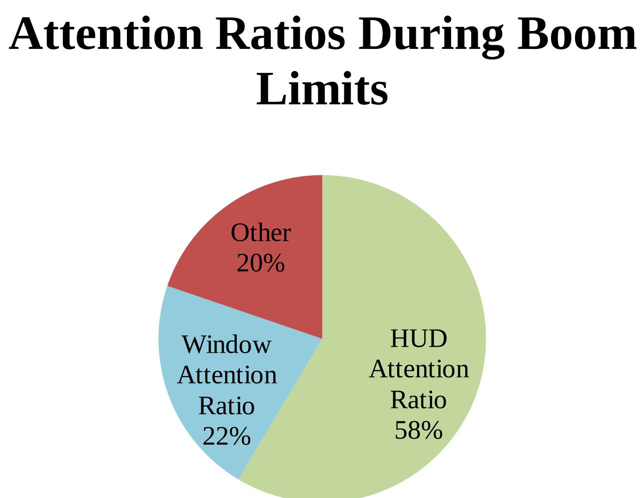


Figure 33: Attention Ratios for Boom Limits operations

The heat map of the HUD usage for a single pilot is shown in Figure 34. As the pilot's gaze lingers at any particular point, the color shifts from green to yellow to red. This pilot tended to place the captain's bars right at the top of the HUD, which he used as a reference to keep in formation with the tanker. He also occasionally glanced at the wing and engine of the tanker, presumably to keep an eye on the position and attitude of the aircraft. It is clear that with key flight data displayed in the HUD that would otherwise only be found in the MFD, the pilot could remain focused on the tanker and not have to constantly move his head down to look at his instrument panel. Instead he would only have to briefly glance at the HUD to get the information required.

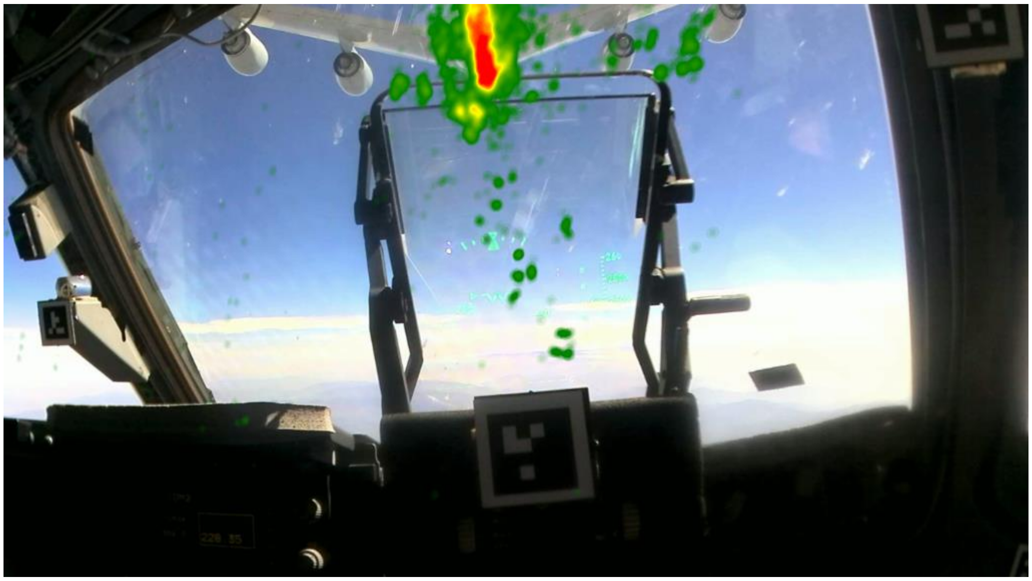


Figure 34: Heat map showing which areas the current pilot most frequently focused on during Boom Limits

The heat map also further illustrates just how singly focused the C-17 pilots were during refueling operations. Once the aircraft enters AR mode, the pilot can shift their focus almost entirely on the tanker they are flying behind. However, that the tanker is moving relative to the QR markers in the frame. Because the gaze coordinates are based entirely on the QR markers, as the C-17 approaches the tanker, the tanker moves in the scene camera frame. Since the heat map is cumulative based on the gaze coordinates up to that point, this results in a slight skewing of the gaze markers and heat map relative to the tanker's current position. We printed a 4 ft x 4ft fiducial marker that we initially wanted to stick to the underside of the tanker. This would have provided a tanker AOI that could be used for automated data analysis. The idea of the large fiducial was dropped due to the short time limit during the TRB/SRB.

Figure 35: Scene camera view during landing task

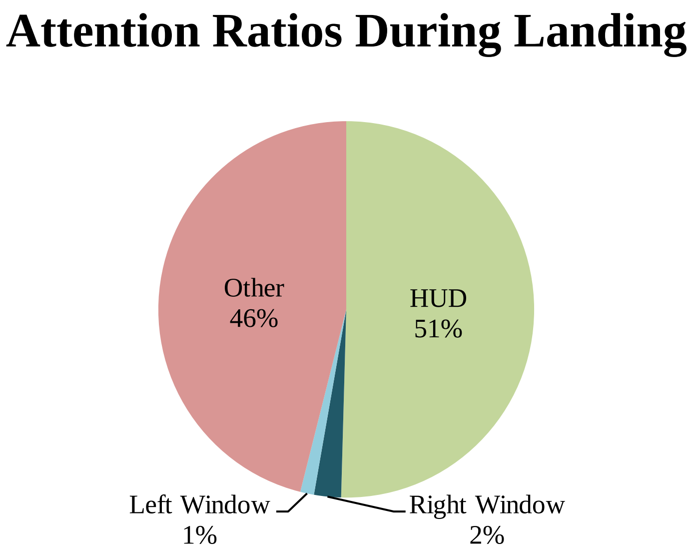


Figure 36: Attention Ratios for landing operations

During landing tasks, the HUD was the focal point roughly the same amount of time, but the gaze of the pilots shifted away from the forward windscreen much more (Figure 36). The total percentage of time spent looking through the forward window was just over 50%. The rest of the time, 46%, the pilot was looking either at his instruments or out the side window. Due to the difficulties with the software recognizing the fiducial markers on the side and around the MFD, any gaze data other than the HUD and forward two windows aren't 100% reliable. Another cause of data variance is that with the pilots moving their head and scanning, a lot of the eye movement is registered as "transition time" moving from one AOI to another. Due to these factors, the category of "other" was used for any gaze time not involving the HUD or forward windows.

Figure 37: Heat map showing which areas the current pilot most frequently focused on during Landing

The heat map for landing (Figure 37) shows a much more typical distribution typically seen for flight using a HUD. Obviously, the key focus of the pilot was the runway ahead of him, but he also spent a good majority of his time focused on the right side of the HUD (presumably the altitude indicator). The area to the bottom left most likely represents glances to the airspeed indicator, but due to pupil detection inconsistencies or misaligned calibration, the gaze may have appeared lower than it actually was.

4.4 HSI Comfort, Intrusiveness, and Utility Results

Seven pilots across four flights completed the questionnaires and assessed comfort, intrusiveness and operational utility in a flight test environment using the 412 TW Six-Point General Purpose Scale (Table B1). The scale ranges from “1-Very Unsatisfactory” to “6- Very Satisfactory”. The mean scores were compared to the HSI 412 TW Six-Point General Purpose Scale Evaluation Criteria (Table B2). See Table 2 for the mean scores and overall ratings.

Table 2. Pilots Subjective Comfort, Intrusion, and Utility Means and Ratings

Factor	Mean	Overall Rating
ECG		
Comfort	5.0	*Satisfactory*
Intrussiveness	4.6	*Satisfactory*
Utility	4.4	*Marginally Satisfactory*
DIKABILIS		
Comfort	3.0	*Marginally Satisfactory*
Intrussiveness	3.8	*Marginally Satisfactory*
Utility	4.0	*Marginally Satisfactory*

The ECG was easy to place on pilots with an installation time of 5-10 minutes, and the systems small footprint made it comfortable and non-intrussive for the most part. Future improvements could be made to improve the utility of the ECG leads connectivity to the pads during flight, such as reinforcing with more adhesive. During flight there were some complaints the leads attachments were uncomfortable under the seat harinesses. Two pilots had minor redness and irritation at the sight of the ECG leads pads, but resolved by the next day. The Dikablis eye-tracker was uncomfortable when worn under the David-Clarke Headsets, however this could have been confounded with the pilots also wearing a multi-lead EEG scalp cap that was also interfering with the headsets. The eye-tracker was also bothersome due to the frequency of having to recalibrate the glasses to the cockpit markings. Although it was relatively quick to calibrate it disrupted flight efficiency.The eye-tracker was intrusive to the visual field when conducting AR tasks for some pilots.

5. Discussion

In conducting this pilot workload study of aerial refueling, the objective was to examine the feasibility of using physiological instrumentation to monitor workload metrics. Using CATS software, OPL was able to monitor and record live workload data while pilots were performing a range of tasks, all while being minimally invasive towards the pilot flying. The CATS workload system worked completely flawlessly and provided workload readings that were consistent with the self-reported workload and known task difficulty levels. We want to make a point that the eye tracker was not a part of our workload assessment method. It was an extra appliance that we did not depend on for assessment of workload. While our ECG system was minimally invasive, we do not believe that the remaining instruments such as the eye tracker or EEG cap were noninvasive.

The key takeaway for us was the successful gathering of ECG-based workload data during aerial refueling operations. The data has shown that OPL Workload, in the limited data sets due to the limited number of flights, has proven effective at conveying the relative difficulty of a variety of aerial refueling, takeoff, and landing tasks. During the course of the flights, OPL researchers were able to monitor in real time the effort the pilots were exerting as they performed each task. Following the flight, the OPL WL data was able to be immediately exported and put into graphs that could easily be compared on a task-by-task or pilot-by-pilot basis.

The tasks performed in the simulator have shown that OPL workload scores derived from ECG monitoring can be directly correlated to the difficulty of the task. The tasks were designed with four levels of difficulty in mind, and the workload data matches the presumed difficutly of each task. Due to the physiological nature of OPL's workload scores, the difficulty cannot be assigned to any specific workload number, but only compared relative to other workload scores obtained from that individual pilot.

As was discussed in February, the next phases of this study would include a more "off-the-shelf" implementation of the instrumentation. We believe that we are very close to this point with our ECG and CATS systems. The system is already simple enough to set up that only a few minutes of instruction is required to begin collecting workload data in the CATS software. With a few additions to help streamline the application and more effectively stow the wiring from the electrodes, the pilots would quickly forget they were instrumented at all during flight. Of course it is unlikely the instrumentation would be forgotten in cases such as in the PWS study where the pilots had to deal with no less than five separate forms of collection equipment.

In future studies utilizing OPL's workload analysis, it may be beneficial to isolate each of the data collection systems (OPL ECG, EEG caps, eye-tracking glasses, etc.) and gather data separately, as OPL did with Pilots

8 and 9 from TPS. These additional pilots were able to quickly don the sensors and perform the task, then quickly remove them and continue with their day. When the pilots are inundated with so many sensors and systems, it may be hard to truly evaluate each system independently, as each system gets lost in the suffle a bit and the combined encumbrance of all the instruments amplifies the negative aspects of each.

The pilots in this study were fitted with multiple sensors to the point where the sensors may have caused discomfort. We believe that we could get a very good initial screening of pilot workload with a fully non-integrated ATS system deployed on both the boom operator and the pilot.

As for the eye-tracking, further data acquisition would be required to draw any meaningful conclusions. Eye tracking did not reveal any surprising patterns. The pilots themselves were not overly enthusiastic about the Dikablis system. The eye-cameras tended to encroach into the pilots' peripheral vision, hindering their ability to use established points of reference when flying in formation. When coupled with the EEG caps, the strap of the glasses also caused some discomfort by pressing on the eleetrodes on the back of the pilots' head.

Another problem faced with eye-tracking was the issue of calibration. In order to properly match eye movement with gaze path, the eye cameras must have a clear, centered image of the eye. OPL attempted to facilitate the pilots by positioning these eye cameras as low as possible so as to minimally interfere with pilots' vision. This worked for the first flight, though there were still some complaints about the eye tracker getting in the way of visual reference points. For the second flight, however, calibration failed altogether. In post-flight testing of the glasses, simply moving each eye camera up an inch or two resulted in nearly flawless calibration. Obviously calibrating in an office indoors is massively different than performing the same calibration in flight, but a lot of the adjustment and re-adjustment that delayed tasks during the flight could be avoided by simply repositioning the eye cameras ahead of time. In our own flight testing at the Iowa City home base, we use the same eye tracker but integrated on a flight helmet. This provides for a rugged, stable platform with less need for constant recalibration and readjustment.

It would also be beneficial to have each pilot sit down with an OPL member before the day of the flight and find what camera positions work best for each. This way, rather than having to guess and check mid-flight, OPL would have a reference for each pilot and how best to position each eye camera. All of the above factors contributed to discontinuing OPL's eye-tracking for the latter two flights of the Pilot Workload Study.

Appendix A: Test Cards

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	TWR:	Risk:
C-17					Low

Event:	T/S/TestPoint. No:	Event No:
Normal Takeoff PWS	XXX-104-XX	

Initial Condition:	*Sys Configuration:*	*Other Items:*
Alt: Field **A/S:** Static **Thrust:** As Required **Weight:** A/R **CG:** A/R **Flaps/Slats:** 1/2 EXT ____ **Gear:** Down	N/A	N/A

Limits:

TEST PROCEDURES

1. **(TC)** Verify Go-Pro Cameras and Eye Glasses remain OFF until before taking runway or as required per flight operations
2. **(TC)** Record pilot flying and instrumentation status
3. **(PF)** Perform normal takeoff per Flight Manual procedure
4. **(TC)** After critical phase of flight, call "*Point Complete*"
5. **(PF)** Provide pilot workload comments

PILOT COMMENTS:

UNCLASSIFIED

1/1

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	TWR:	Risk:
C-17					Low

Event:	T/S/TestPoint. No:	Event No:
Normal Takeoff PWS	XXX-104-XX	

Instrumentation/Data:	Prof Card Approval
Veridian, 711th, UI, SMU Instrumentation *Go / NoGo Criteria* Veridian, 711th, UI, SMU Instrumentation	Eng Review: OPS Eng Review: OPS Review: DOO Coord: DO Approval: Approved:

Aircrew Notes:

Time			**Pilot: A / B / C / D**
HR	**MIN**	**SEC**	
			INSTRUMENTATION ON
			711 HPW: ECG, EEG, BIOHARNESS, SMART EYE (Left Seat), EYE GLASSES (Right Seat)
			UI OPL: ECG, CATS, EYE GLASSES (Right Seat)
			SMU: WRISTBAND
			VERIDIAN
			GO-PRO CAMERAS (Wi-Fi OFF)
			Start Time:
			Stop Time:

Before Takeoff Data: ATIS				
IDENT	TIME	WIND	VIS	CLOUDS
TEMP/DEW	ALT SETTING	PA	RWY	NOTES

TOLD				
T/O Thrust		VGO		
TOGW/CG		VROT		
Flap	IDX/POS	STAB	VMCO	
VR	CFL	T/O RUN	VMFR	
VBMAX	XWND / HWND		VMSR	VMCG

UNCLASSIFIED

1/1

Figure 38: Normal Takeoff

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	TWR:	Risk:
C-17					Low

Event:	TIS/Test Point No:	Event No:
Normal Flight Operations PWS	XXX-104-XX	

Initial Condition:	*Sys Configuration:*	*Other Items:*
Alt: A/R **A/S**: A/R **Thrust**: A/R **Weight**: A/R **CG**: A/R **Flaps/Slats** A/R **Gear**: A/R	N/A	N/A

Limits:

TEST PROCEDURES

1. **(TC)** Record pilot flying and instrumentation status
2. **(TC)** Record **START** and **END** times for each phase of flight

UNCLASSIFIED

1/1

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	TWR:	Risk:
C-17					Low

Event:	TIS/Test Point No:	Event No:
Normal Flight Operations PWS	XXX-104-XX	

1/1

Instrumentation/Data:	Prof Card Approval
Veridian, 711th, UI, SMU Instrumentation *Go / NoGo Criteria* Veridian, 711th, UI, SMU Instrumentation	Eng Review: OPS Eng Review: OPS Review: DOO Coord: DO Approval: Approved:

Aircrew Notes:

Time			**INSTRUMENTATION ON**
HR	**MIN**	**SEC**	
			711 HPW: ECG, EEG, BIOHARNESS, SMART EYE (Left Seat), EYE GLASSES (Right Seat)
			UI OPL: ECG, CATS, EYE GLASSES (Right Seat)
			SMU: WRISTBAND
			VERIDIAN
			GO-PRO CAMERAS (Wi-Fi OFF)

START Time			**Flight Phase**	**Pilot**	**END Time**		
HR	**MIN**	**SEC**			**HR**	**MIN**	**SEC**
			TAKEOFF				
			CLIMB				
			CRUISE				
			DESCENT				
			APPROACH				

UNCLASSIFIED

Figure 39: Normal Flight Operations

UNCLASSIFIED

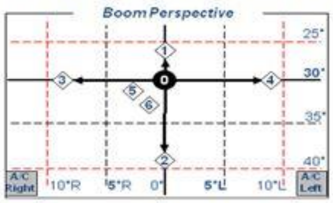
Aircraft:	Serial No:	Flight No:	Revision:	T/WR:	Risk:
C-17A					Low
Event: AR Boom Limits – C-17 PWS KC-135			TIS/TestPoint No: XXX-101-XX		Event No:

Initial Condition:	Sys Configuration:	Other Items:
Alt: 25,000 +/- 5000 PA **A/S**: 265 KCAS +/- 10 kts **Thrust**: TLF **Wt**: A/R **CG**: A/R **Flap Slt Ind**: UP/RET **Gear**: UP	Alt: A/R A/P & ATS: Off MFD: PFD & Confg Pg Pre-contact CL Completed	KC-135 tanker A/P: ON

Limits:

TEST PROCEDURES

1. **TC:** Record pilot flying and instrumentation status.
2. **BO:** Clear Receiver to **CONTACT** position.
3. **RP:** Stabilize in the **CONTACT** position.
4. **BO:** When stabilized, initiate contact.
5. **BO:** Verbally direct receiver to boom locations (cross pattern), hold for 5 second at each point. Stabilize within 3° of limit.



6. **BO:** Direct receiver to inner limit (POINT 5) and outer limit (POINT 6). Stabilize within 2 feet of the limit.
7. **RP:** Follow BO commands to boom locations.
8. **TC:** Record times at each point.
9. **BO:** Clear Receiver to **PRE-CONTACT** position.
10. **RP:** Stabilize in **PRE-CONTACT** position. Give comments on task.
11. **TP/BO:** Give comments on maneuvers.
12. **RP:** Reconfigure aircraft as required.
13. **RP:** Complete Bedford Workload Scale.

UNCLASSIFIED

1/1

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	T/WR:	Risk:
C-17A					Low
Event: AR Boom Limits– C-17 PWS KC-135			TIS/TestPoint No: XXX-101-XX		Event No:

1/1

Instrumentation/Data Req'd:	Author:
Veridian, 711th, UI, SMU Instrumentation *Go / NoGo Criteria:* Veridian, 711th, UI, SMU Instrumentation Tanker A/P ON	Phone: Eng Review: OPS Eng Review OPS Review: Date:

Aircrew Notes:

Tanker GW__________ Tanker CG ________
C-17 GW: __________ C-17 CG: ________

Time			Pilot: A / B / C / D
HR	**MIN**	**SEC**	
			INSTRUMENTATION ON
			711 HPW: ECG, EEG, BIOHARNESS, SMART EYE (Left Seat), EYE GLASSES (Right Seat)
			UI OPL: ECG, CATS, EYE GLASSES (Right Seat)
			SMU: WRISTBAND
			VERIDIAN
			GO-PRO CAMERAS (Wi-Fi OFF)
			Start Time:
			Position 1
			Position 2
			Position 3
			Position 4
			Position 5
			Position 6
			Point Complete:

Bedford Scale Rating: __________ **COMMENTS:**

UNCLASSIFIED

Figure 40: AR Boom Limits

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	T/WR:	Risk:
C-17A					Low

Event:	T/S/Test Point No:	Event No:
Contact Station Keep – KC-135	XXX-102-XX	

Initial Condition:	*Sys Configuration:*	*Other Items:*
Alt: 25,000 +/- 5000 PA **A/S:** 265 KCAS +/- 10 kts **Thrust:** TLF **Wt:** A/R **CG:** A/R **Flap Slt Ind:** UP/RET **Gear:** UP	Alt: A/R A/P & ATS: Off MFD: PFD & Confg Pg Pre-contact CL Completed	KC-135 tanker A/P: ON

Limits:

Note:

TEST PROCEDURES

1. TC: Record pilot flying and instrumentation status.
2. **BO:** Clear Receiver to **CONTACT** position.
3. **RP:** Stabilize in the **CONTACT** position.
4. **BO:** When stabilized, initiate contact.
5. **RP:** Maintain the aircraft within the center of boom. Maintain contact for 5 minutes.
6. **TP/RP:** When ready within the 5 minutes, perform a turn (min 90° of heading change) with 30° angle of bank, entering and exiting the turn at a roll rate of 3 degree/second or less.
7. **TP/RP:** Return to straight and level to complete 5 minutes in contact.
8. **BO/RP**: When ready, perform normal disconnect.
9. **BO:** Clear receiver to **PRE-CONTACT** position.
10. **RP:** Stabilize in **PRE-CONTACT** position.
11. **RP:** Reconfigure aircraft as required.
12. **RP:** Complete Bedford Workload Scale.

UNCLASSIFIED

1/1

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	T/WR:	Risk:
C-17A					Low

Event:	T/S/Test Point No:	Event No:
Contact Station Keep – KC-135	XXX-102-XX	

Instrumentation/Data Req'd:	*Author:*
Veridian, 711^{th}, UI, SMU Instrumentation	*Phone:*
	Eng Review:
Go / NoGo Criteria:	*OPS Eng Review:*
Veridian, 711^{th}, UI, SMU Instrumentation	*OPS Review:*
Tanker A/P ON	Date:

Aircrew Notes

Tanker GW__________ Tanker CG _______

C-17 GW: __________ C-17 CG: ________

Time			**Pilot:** A / B / C / D
HR	**MIN**	**SEC**	
			INSTRUMENTATION ON
			711 HPW: ECG, EEG, BIOHARNESS, SMART EYE (Left Seat), EYE GLASSES (Right Seat)
			UI OPL: ECG, CATS, EYE GLASSES (Right Seat)
			SMU: WRISTBAND
			VERIDIAN
			GO-PRO CAMERAS (Wi-Fi OFF)
			Start Time:
			Point Complete:

Bedford Scale Rating: __________ **COMMENTS:**

UNCLASSIFIED

1/1

Figure 41: Contact Station Keeping

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	T/WR:	Risk:
C-17					Low

Event:	T/S/Test Point No:	Event No:
Offset Landing PWS	XXX-103-XX	

Initial Condition:	*Sys Configuration:*	*Other Items:*
Alt: Approach **A/S**: Approach **Thrust**: As Required **Weight**: A/R **CG**: A/R **Flaps/Slats**: As Required **Gear**: Down	Alt : Set to LOCAL MFD 2/3 – PFD	N/A

Limits: Stable on centerline by 300 feet AGL or execute Go-Around

TEST PROCEDURES

1. **(TC)** Record pilot flying and instrumentation status
2. **(PF)** Line up approach on edge of runway
3. **(TC)** Verify:
   a. Gear: DOWN
   b. Flap/Slat Position: IAW MC Landing Data
   c. Spoilers: ARMED
4. **(PM)** At 500 ft AGL call *"ACTION"*
5. **(PF)** Capture runway centerline and landing aimpoint
6. **(PM)** At 300 ft AGL call *"300 STABLE"* or *"GO-AROUND"*
7. **(PF)** Land, touch-and-go, or go-around. Execute a go-around if aircraft not stabilized by 300 feet AGL. Note touchdown point.
8. **(PF)** Complete Bedford Workload Scale.
9. **(TC)** Call *"Point Complete"*

LANDING DATA 3: ACTUALS

GA EPR ______	XWND/HWND ____/____	VAPP ______
VBO ______	VBO LDG DIST ____/____	VMCO ______
FLAP ______	IDX/POS ____/____	VMFR ______
APPR PATH ______°	VMCA2 ______	VMSR ______
LDG DIST ______		REF GS ______

UNCLASSIFIED 1/2

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	T/WR:	Risk:
C-17					Low

Event:	T/S/Test Point No:	Event No:
Offset Landing PWS	XXX-103-XX	

Instrumentation/Data:	Prof Card Approval
Veridian, 711th, UI, SMU Instrumentation *Go / NoGo Criteria:* Veridian, 711th, UI, SMU Instrumentation	Eng Review: OPS Eng Review: OPS Review: DOO Coord: DO Approval: Approved:

Aircrew Notes: Recording will be stopped and glasses-based eye trackers removed as soon as practical after exiting active runway for full stop landing.

Time			**Pilot: A / B / C / D**
HR	**MIN**	**SEC**	
			INSTRUMENTATION ON
			711 HPW: ECG, EEG, BIOHARNESS, SMART EYE (Left Seat), EYE GLASSES (Right Seat)
			UI OPL: ECG, CATS, EYE GLASSES (Right Seat)
			SMU: WRISTBAND
			VERIDIAN
			GO-PRO CAMERAS (Wi-Fi OFF)
			Start Time:
			Stop Time:
			In Chocks

ATIS				
IDENT	TIME	WIND	VIS	CLOUDS
TEMP/DEW	ALT SETTING	PA	RWY	NOTES

Bedford Scale Rating:__________ **COMMENTS:**

UNCLASSIFIED

Figure 42: Offset Landing

UNCLASSIFIED

Aircraft: C-17	Serial No:	Flight No:	Revision:	T/WR:	Risk: Low
Event: Normal Approach PWS			TIS/Test Point No: XXX-104-XX		Event No:

Initial Condition:	*Sys Configuration:*	*Other Items:*
Alt: Approach **A/S:** Approach **Thrust:** As Required **Weight:** A/R **CG:** A/R **Flaps/Slats:** As Required **Gear:** Down	Alt : Set to LOCAL MFD 2/3 – PFD	N/A

Limits:

TEST PROCEDURES

1. **(TC)** Record pilot flying and instrumentation status
2. **(PF)** NORMAL or STEEP approach.
3. **(PF)** Approach Type: Full Stop / Touch & Go / Low App __AGL
4. **(TC)** Verify:
   a. Gear: DOWN
   b. Flap/Slat Position: IAW MC Landing Data
   c. Spoilers: ARMED
5. **(PF)** Perform approach
6. **(TC)** Call ***"Point Complete"***
7. **(TC)** Collect pilot comments after critical phase of flight

LANDING DATA 3: ACTUALS

GA EPR	XWND/HWND	VAPP
________	____/____	________
VBO	VBO LDG DIST	VMCO
________	____/____	________
FLAP	IDX/POS	VMFR
________	____/____	________
APPR PATH	VMCA2	VMSR
________°	________	________
LDG DIST		REF GS
________		________

UNCLASSIFIED

1/1

UNCLASSIFIED

Aircraft: C-17	Serial No:	Flight No:	Revision:	T/WR:	Risk: Low
Event: Normal Approach PWS			TIS/Test Point No: XXX-104-XX		Event No:

Instrumentation/Data:	*Eng Review:*
Veridian, 711th, UI, SMU Instrumentation	*OPS Eng Review:*
	OPS Review:
	DOO Coord:
Go / NoGo Criteria:	*DO Approval:*
Veridian, 711th, UI, SMU Instrumentation	Approved:

Aircrew Notes: Recording will be stopped and glasses-based eye trackers removed as soon as practical after exiting active runway for full stop landing.

Time			Pilot: A / B / C / D
HR	**MIN**	**SEC**	
			INSTRUMENTATION ON
			711 HPW: ECG, EEG, BIOHARNESS, SMART EYE (Left Seat), EYE GLASSES (Right Seat)
			UI OPL: ECG, CATS, EYE GLASSES (Right Seat)
			SMU: WRISTBAND
			VERIDIAN
			GO-PRO CAMERAS (Wi-Fi OFF)
			Start Time:
			Stop Time:
			In Chocks

ATIS

IDENT	TIME	WIND	VIS	CLOUDS
TEMP/DEW	**ALT SETTING**	**PA**	**RWY**	**NOTES**

PILOT COMMENTS:

UNCLASSIFIED

Figure 43: Normal Approach

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	T/WR:	Risk:
C-17					Low
Event: Normal Landing PWS			TIS/Test Point No: XXX-104-XX		Event No:

Initial Condition:	*Sys Configuration:*	*Other Items:*
Alt: Approach **A/S**: Approach **Thrust**: As Required **Weight**: OPT **CG**: OPT **Flaps/Slats**: As Required **Gear**: Down	Alt : Set to LOCAL MFD 2/3 – PFD	N/A

Limits:

TEST PROCEDURES

1. **(PF)** NORMAL, STEEP, or ENGINE OUT (#____Eng) approach.
2. **(PF)** Approach Type: Full Stop / Touch & Go / Low App __AGL
3. **(TC)** Verify:
   a. Gear Position: _________
   b. Flap/Slat Position: ________
4. **(TC)** At taxi of 40 knots, call ***"Point Complete"***
5. **(TC)** Collect pilot comments after completing "after landing" checklist

LANDING DATA 3: ACTUALS

GA EPR	XWND/HWND	VAPP
________	____/____	______
VBO	VBO LDG DIST	VMCO
________	____/____	______
FLAP	IDX/POS	VMFR
________	____/____	______
APPR PATH	VMCA2	VMSR
________°	________	______
LDG DIST		REF GS
________		______

UNCLASSIFIED

1/1

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	T/WR:	Risk:
C-17					Low
Event: Normal Landing PWS			TIS/Test Point No: XXX-104-XX		Event No:

Instrumentation/Data: Veridian, 711th, UI, SMU Instrumentation *Go / NoGo Criteria:* Veridian, 711th, UI, SMU Instrumentation	Eng Review: OPS Eng Review: OPS Review: DOO Coord: DO Approval: Approved:

Aircrew Notes: Recording will be stopped and glasses-based eye trackers removed as soon as practical after exiting active runway for full stop landing.

Time			Pilot: A / B / C / D
HR	MIN	SEC	
			INSTRUMENTATION ON
			711 HPW: ECG, EEG, BIOHARNESS, SMART EYE (Left Seat), EYE GLASSES (Right Seat)
			UI OPL: ECG, CATS, EYE GLASSES (Right Seat)
			SMU: WRISTBAND
			VERIDIAN
			GO-PRO CAMERAS (Wi-Fi OFF)
			Start Time:
			Stop Time:
			In Chocks

ATIS				
IDENT	TIME	WIND	VIS	CLOUDS
TEMP/DEW	ALT SETTING	PA	RWY	NOTES

PILOT COMMENTS:

UNCLASSIFIED

Figure 44: Normal Landing

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	TWR:	Risk:
TPS SIM					Low
Event: SIM – Boundary Avoidance Task PWS			TIS/TestPoint No: XXX-201-XX		Event No:

Initial Condition:	*Sys Configuration:*	*Other Items:*
Alt: 15K MSL **A/S:** 300 KTS **Thrust:** As Required **Weight:** A/R **CG:** A/R **Flaps/Slats:** UP / RET **Gear:** UP	N/A	N/A
Limits:		

TEST PROCEDURES

1. **(PF)** Perform Primary and Secondary tasks for 5 minutes

PRIMARY TASK: Keep target aircraft's fuselage inside horizontal bars. Bottom of fuselage crossing top bar is a violation. Top of fuselage crossing bottom bar is a violation. Keep crosshairs inside wingtips. Left tip of cross hairs outside of right wing is a violation. Right tip of cross hairs outside of left wing tip is a violation.

DESIRED: ≤2 longitudinal axis violations, ≤1 lateral axis violation

ADEQUATE: ≤3 longitudinal axis violations, ≤2 lateral axis violations

SECONDARY TASK: Pilot will hear a periodic tone repeating. Identify higher pitch tones, click and provide total number of high pitch tones at the end of task. Keep counting tones during simulator task resets.

2. **(TC)** Record the following during maneuver:

	#	Performance
Long. Violations		Desired / Adequate / NC
Lateral Violations		Desired / Adequate / NC
Tone Count Reported		**Actual:**

3. **(PF)** Complete Bedford and NASA TLX Workload Scales
4. **(PF)** Provide any comments on maneuver

UNCLASSIFIED

1/1

UNCLASSIFIED

Aircraft:	Serial No:	Flight No:	Revision:	TWR:	Risk:
TPS SIM					Low
Event: SIM – Boundary Avoidance Task PWS			TIS/TestPoint No: XXX-201-XX		Event No:

Instrumentation/Data:
711th, UI-OPL

Go / NoGo Criteria

Aircrew Notes:

Time			Pilot: A / B / C / D / E / F / G
HR	**MIN**	**SEC**	
			INSTRUMENTATION ON
			711 HPW: BIOSEMI, BIOHARNESS, EYE GLASSES
			UI OPL: ECG, CATS
			SIM TASK: __________

Bedford Scale Rating: __________ **NASA TLX:**

COMMENTS:

UNCLASSIFIED

1/1

Figure 45: Simulator Task

Appendix B: Human Systems Integration Scales

Table B1. 412 TW Six-Point General Purpose Scale

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

Table B2. HSI 412 TW Six-Point General Purpose Scale Evaluation Criteria with Descriptors

Mean	Equivalent Descriptor*	Equivalent Rating*
Mean equal to 6.0	Excellent	Satisfactory
Mean equal to or between 5.1 and 5.9	Good	Satisfactory
Mean equal to or between 4.5 and 5.0	Adequate	Satisfactory
Mean equal to or between 2.5 and 4.4	Borderline	Marginal
Mean equal to or between 2.0 and 2.4	Deficient	Unsatisfactory
Mean equal to or between 1.1 and 1.9	Unacceptable	Unsatisfactory
Mean equal to 1.0	Unsafe	Failed

*These descriptors and ratings are not reflective of the final, overall HSI rating. They are designed to merely show the correlation of the questionnaire data with the 412 TW Rating Criteria.

References

1. Martin, Patrick, USAF, *Objective Measures of Workload Feasibility Study*, 412TW-TP-18-47, Air Force Test Center, Edwards AFB, California, August 2018.
2. Govindan, R., K. Narayanan, and M. Gopinathan, *On the evidence of deterministic chaos in ECG: Surrogate and predictability analysis.* Chaos, 1998. **8**(2): p. 495-502.
3. Kozma, R.a.F., W., *Classification of EEG Patterns using Nonlinear Dynamics and Identifying Chaotic Phase Transitions.* Neurocomputing, 2002. **44-46**: p. 1107-1112.
4. Owis, M., et al., *Study of features based on nonlinear dynamical modeling in ECG arrhythmia detection and classification* IEEE Transactions on Biomedical Engineering, 2002. **49**(7): p. 733-736.
5. Govindan, R., K. Narayanan, and M. Gopinathan, *On the evidence of deterministic chaos in ECG: Surrogate and predictability analysis.* Chaos: An Interdisciplinary Journal of Nonlinear Science, 1998. **8**(2): p. 495-502.
6. Richter, M. and T. Schreiber, *Phase Space Embedding of Electrocardiograms.* Physical Review E, 1998. **58**(5): p. 6392-6398.
7. Engler, J. and T. Schnell, *Deterministically Nonlinear Dynamical Classification of Cognitive Workload*, in *I/ITSEC 2013*. 2013: Orlando, FL.
8. Schnell, T. and J. Engler, *Entropic skill assessment of unmanned aerial systems (UAS) operators.* Journal of Unmanned Vehicle Systems, 2014. **2**(02): p. 53-68.
9. OPL, *Cognitive Assessment Tool Set (CATS) User Manual* O.P.L. (OPL), Editor. 2014, University of Iowa: Iowa City, Iowa.
10. Roscoe, A. and G. Ellis, *A subjective rating scale assessing pilot workload in flight. A decade of practical use*. 1990, Royal Aerospace Establishment: Farnborough, UK.
11. NASA. *Multi Attribute Task Battry Official Web Site*. 2013; Available from: https://matb.larc.nasa.gov/.
12. Dideriksen A., et al. *Define Expert: Characterizing Proficiency for Physiological Measures of Cognitive Workload*. in *Submitted to Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. 2018. Orlando, Florida.
13. Schnell, T. and C. Thompson, *Objective Measures of Pilot Workload* 2018, United States Air Force, 412th TW, 418th Flight Tesrt Squadron: Iowa City, Iowa.
14. Schnell T, et al. *A Comparison of Helmet-Mounted Display Symbologies During Live Flight Operational Tasks*. in *19th International Symposium on Aviation Psychology*. 2017. Dayton, Ohio, USA, May 8 - May 11, 2017.
15. Geiselman, E.E., H.P. Williams, and T. Schnell. *Use of a live, virtual, constructive simulation approach to evaluate visual symbology on a helmet-mounted display for spatial disorientation prevention.* in *Proceedings of IMAGE Society 2017 Conference*. 2017. Dayton, OH.
16. Schnell T., Reichlen C., and R. C., *Spatial Disorientation Threat Characterization for F-35 Representative Helmet-Mounted Display Use in the Flight Environment* 2017, US Air Force Research Laboratories: Dayton, OH.
16. Schnell T., Reuter C., and Cover M., *Flight Deck Visual and Auditory Display Counter-Measures to Spatial Disorientation and Loss of Energy State Awareness*, N.A.a.S.A. (NASA), Editor. 2017, Langley Research Center Hampton, VA 23681.
18. Dexter H. Kalt, R.T., Farhad H. Choudhury, Gomer C. Custer, Thomas J. Pitsor, David A. Benson, *Aerial Refueling Test Methods*. 2015, Aerial Refueling Systems Advisory Group. p. 74.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 07-11-2019	**2. REPORT TYPE** Technical Paper	**3. DATES COVERED** *(From - To)* 4 Nov 2019
4. TITLE AND SUBTITLE Public Release for 'Objective Measures of Pilot Workload' Study Results		**5a. CONTRACT NUMBER** 12100 **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Lt Col Paul Calhoun, Mr. Patrick Martin, Dr. Thomas Schnell		**5d. PROJECT NUMBER** **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES) 418 FLTS, 35 N Flightline Rd., Edwards AFB, CA 93524 Operator Performance Laboratory, 1801 S Riverside Dr., Iowa City, IA 52246		**8. PERFORMING ORGANIZATION REPORT NUMBER** 412TW-PA-19537
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) 412th Test Wing 195 E Popson Ave Edwards AFB CA 93524		**10. SPONSOR/MONITOR'S ACRONYM(S)** N/A **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release A: distribution is unlimited.

13. SUPPLEMENTARY NOTES
Results presented under approved public release for Society of Experimental Test Pilots Annual Symposium, Anaheim, CA, 26-28 Sep 2019; this is a follow up request to publically release the technical paper corresponding to the presentations.

14. ABSTRACT
A presentation was given sharing the results of a test endeavor to determine if biometric sensors are a feasible means of objectively measuring operator workload in flight. The presentation exhibited the test effort by the 418 FLTS and associated partners to accomplish initial testing, and now we are seeking public release of the technical paper containing the study results. No proprietary, FOUO, or otherwise protected material is used herein.

15. SUBJECT TERMS
Objective Measures, Pilot Workload, Biometric, Physiological, Sensors

16. SECURITY CLASSIFICATION OF: Unclassified			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** 412 TENG/EN (Tech Pubs)
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	None	53	**19b. TELEPHONE NUMBER** *(include area code)* 661-277-8615