# Enhancing HMD-Based F-35 Training through Integration of Eye Tracking and Electroencephalography Technology

Meredith Carroll[1], Glenn Surpris[1], Shayna Strally[1], Matthew Archer[1],
Frank Hannigan[1], Kelly Hale[1], and Wink Bennett[2]

[1] Design Interactive, Oviedo, Florida
{meredith,glenn.surpris,shayna.strally,matt.archer,
frank.hannigan,kelly}@designinteractive.net
[2] Air Force Research Laboratory, Wright- Patterson Air Force Base, Ohio
Winston.Bennett@wpafb.af.mil

**Abstract.** The ever increasing complexity of knowledge, skills and abilities (KSAs) demanded of Department of Defense (DoD) personnel has created the need to develop tools to increase the efficiency and effectiveness of training. This is especially true for the F-35, the first 5th-generation aircraft to use an HMD as the primary instrument display. Additionally, the F-35 can perform operations previously performed by multiple operators, which potentially places incredible strain on the pilot's cognitive resources by exposing him to large amounts of data from disparate sources. It is critical to ensure training results in pilots learning optimal strategies for operating in this information rich environment. This paper discusses current efforts to develop and evaluate a performance monitoring and assessment system which integrates eye tracking and Electroencephalography (EEG) technology into an HMD enabled F-35 training environment to extend traditional behavioral metrics and better understand how a pilot interacts with data presented in the HMD.

**Keywords:** Training, Performance Assessment, Eye tracking, EEG, Helmet-mounted display, Heads-up display, F-35.

## 1 Introduction

According to the Air Force Transformation 2010, "the ultimate source of air and space combat capability resides in the men and women of the Air Force…first priority is ensuring they receive the precise education, training, and professional development necessary to provide a quality edge second to none"[1]. As technology progresses, the extensive knowledge, skills and abilities (KSAs) required of Department of Defense (DOD) personnel increases, and the demand for efficient yet effective training intensifies. This training need is particularly evident with 5th-generation tactical aircraft such as the F-35 Lightning II (formerly referred to as the Joint Strike Fighter).

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC/HCII 2013, LNAI 8027, pp. 21–30, 2013.

In the 1970's heads-up displays (HUDs) were introduced to tactical combat aircraft, which projected essential flight information onto the cockpit glass. This allowed pilots to continue to keep their eyes directed outside of the aircraft without being required to look down at important gauges. Several decades later, the development of the helmet-mounted display (HMD) allowed the HUD to be placed inside the pilot's helmet. The F-35 Lightning is the first 5th generation aircraft to use an HMD as the primary instrument and sensor display. Additionally, the F-35 is capable of performing air-to-air combat and air-to-ground strikes while flown by a single operator. Tactical information has been added to the HMD to aid the F-35 pilot in performing the additional tasks. This translates to an increase in cognitive demands for an F-35 pilot, with the amount of data from different sources potentially exceeding an individual's natural cognitive processing limits. As noted by Endsley (2001), data does not equal information, and may not be useful "unless it is successfully transmitted, absorbed and assimilated in a timely manner by the human" [2]. To date, the amount of information to be displayed often already exceeds available display space. Although essential information is provided on the HMD, the pilot must periodically transfer attention to other areas of the cockpit, such as the Multi-function Displays (MFDs), a paper copy of the Checklist, or cockpit control panels, in order to view more detailed information throughout a flight. Attentional demands in the cockpit shift frequently and rapidly if an emergency such as engine failure occurs, adding to the cognitive stress already amplified in an emergency situation. Given this, it is critical to ensure training results in pilots learning optimal strategies for operating in this information rich environment, including appropriate attention allocation between the different displays and pieces of information displayed within.

## 2 Training Needs / Opportunity

The application of HMD systems in tactical aircraft and simulation environments has substantial implications for performance assessment, proficiency tracking, and training. Much of the interaction that occurs with an HMD is unobservable, including gaze location/durations and cognitive processing of various information inputs that may not have an overt behavioral response. In order to effectively diagnose deficiencies/inefficiencies in performance and provide targeted feedback, it is necessary to obtain process level measures of performance that include capture of unobservable perceptual and cognitive tasks. To achieve this, there is a need for practical tools and instrumentation to better capture important data that can be assimilated in real-time to more accurately assess pilot performance, including data presented in the HMD, the interaction of the pilot with the data, and reactions and actions taken based on the data. With this enhanced data capture and performance monitoring capability, improved After Action Reviews (AARs) and debriefings will be possible that may substantially enhance training effectiveness and efficiency.

The current training practices for the F-35 lightning were investigated to ensure that the research effort to develop a precision performance assessment system, referred to as the Helmet-Mounted Display ASsessment System for the Evaluation of eSsential Skills (HMD ASSESS), is designed to address the training needs of the F-35. The current training program for F-35 transition pilots is 8 weeks long. The transition pilots are comprised of legacy aircraft experts such as experienced F-16 or F-22 pilots. These pilots will become F-35 instructors upon the completion of the program. Training begins with a week of military lectures, followed by 3 weeks of lectures and academic courses specific to the F-35. A pilot training aid (PTA) laptop simulator is flown by transition pilots during these early phases of the course. The last phase of the training program is a mixture of 8-10 F-35 Full Mission Simulator (FMS) sessions and 4-5 actual flights in the F-35. The PTA and the FMS are the two main simulators used in the transition curriculum. The PTA has a large touchscreen monitor that displays both the out-the-window view of the aircraft as well as the touchscreen instrumentation (i.e. Main Forward Display). In addition to the touchscreen monitor, the PTA also has a full replication of the F-35 Hands-On Throttle and Stick (HOTAS). The PTA is mainly used during academic lectures to familiarize the pilot with the controls and procedures for the F-35. An HMD is not used in conjunction with the PTA.

The FMS is a high fidelity flight simulator which contains a full 1-to-1 replication of the F-35 cockpit surrounded by a dome with almost 360 degrees of visual coverage. The pilot trainee is outfitted with an HMD visor that reveals a HUD fixed on the center windscreen. Additionally, a de-cluttered, un-fixed version of the main HUD with a reduced selection of essential symbols (e.g., airspeed, altitude) appears on the HMD when the pilot turns his/her head off bore-sight (i.e., left, right, up, or down). The simulator sessions in the FMS are 1.5 hours in duration and are preceded by a 1 hour pre-brief and followed by a 1 hour debrief. Each trainee in the FMS has the individualized, one-on-one attention of an instructor. The instructor has an operator station where he can launch scenarios and insert abnormal aircraft conditions. During the training session, the instructor can also view the pilot's performance unfolding from a series of view, including the field of view (FOV) in the cockpit due to a head-tracker associated with the HMD.

The debrief then provides the opportunity for the instructor to playback any flight segment during the simulator session and review notes, exceptional performance, and trainee performance errors. Control inputs, the pilot's FOV, and other simulator information can be accessed by the instructor to facilitate this debrief. Instructors depend on overt behavioral actions and communications to identify performance errors. One limitation of this approach is the inability of the instructor to determine the specific instruments the pilot is monitoring, both within the HMD and on the MFD. Heads up/heads down status can typically be inferred based on the FOV presented by the HMD, however, the specific information that the pilot is visually integrating is not accessible. Given that a large portion of the task is monitoring information presented by a range of instruments; this limits the instructors understanding of how pilot performance is unfolding.

Without sufficient data collection and diagnosis of performance data, evaluations and feedback provided by instructors may not address the underlying sources of poor performance. There are multiple reasons for this, including: 1) instructors may not be able to detect all errors due to the high workload associated with monitoring a complex scenario; and 2) instructors are unable to monitor subtle physical behaviors such as scanning patterns or attention allocation. As a result, instructors may not drill down far enough to expose the root cause of training deficiencies. For example, during irregular flight training such as warning or error procedures, a trainee may fail to take appropriate action to correct the aircraft parameters during a warning indicator. This could be due to several reasons including 1) he/she is not monitoring/scanning the relevant content in the cockpit, 2) he/she is monitoring the relevant content in the cockpit, but does not detect that they are out of tolerance, or 3) he/she detects they are out of tolerance but does not understand appropriate actions to take to mitigate. Additionally, there may be overarching error patterns undetected by the instructor, such as tendency to allocate unnecessary attention heads-down/ within the main forward display (MFD) or to symbols not relevant to the task at hand. Having data that can help instructors to determine the root cause of errors could provide key information regarding the general nature of the failures, which could potentially facilitate development of more effective training interventions.

Given the increased responsibilities and cognitive workload of the F-35 pilot, pilot's cognitive interactions with the HMD and other instrumentation are ever more important to ensuring that training feedback is as accurate and helpful as possible. To this end, objective measures of pilot information processing efficiency and effectiveness are required. Since information interaction within a head-mounted display (HMD) is limited almost entirely to perceptual and cognitive processes such as visual scan and information processing, there is a need for innovative solutions that can accurately and reliably capture this 'unobservable' behavior in order to 1) understand how an HMD is impacting pilot performance and 2) design training to effectively maximize performance. With the advancement in physiological monitoring technology such as eye tracking and EEG there is an opportunity to make these unobservable processes accessible to instructors to increase the accuracy and effectiveness of training feedback.

## 2.1 Eye Tracking and EEG

Visual attention can provide important insights to the information used in task performance, such as the importance of various features or cues [3]. Several studies [4; 5; 3; 6] have used eye tracking to extract information about scan strategies. These studies have demonstrated that eye tracking can aid in the assessment of perception through measurement of visual attention during observation via gaze, scan path, and fixation data. These measures can provide a means for increasing the granularity of performance feedback and hence the effectiveness of debriefs based on these measures. Additionally, mobile eye tracking technology has been successfully implemented in both the commercial flight deck [7] and military fighter jet [8] simulation environments to measure scan path sequence, visual attention allocation, overall situational awareness, and fixation times. These measures are particularly useful for assessing HMD interactions, as the HMD is an area of the cockpit where the pilot is solely monitoring information visually and is not performing observable direct control inputs. EEG has been successfully used in previous studies [9] along with electrocardiogram (ECG) sensors [10] to measure trainee workload in the aircraft simulation environment. Cognitive workload is of particularly interest in the F-35 environment due to the previously-stated consolidation of duties. Such a measure could allow the identification of times when cognitive overload led to performance failures as opposed to skill decrements, allowing for feedback to more accurately target the root cause of errors.

Eye tracking and EEG measures have been successfully implemented together in a number of desktop-based environments to provide this deep diagnostic evaluation of performance [11, 12, 13, 14, 15]. Eye tracking and electroencephalography (EEG) can be used in combination to access such “unobservable” perceptual cognitive processes as scan strategies [11], attention allocation [12, 13, 14] and cognitive workload [15, 16]. Thus, eye tracking and EEG emerged as the most suitable combination of physiological measures to incorporate into HMD ASSESS to address the training needs of the F-35 Lightning II. The initial version of HMD ASSESS intended for use in the F-35 training environment will be limited to utilizing eye tracking measurements, with EEG measurements reserved for the version of HMD ASSESS used in conducting research. However, incorporating EEG in the training assessment version of HMD ASSESS is the end goal when EEG technology becomes more deployment friendly.

These measures can provide a means for increasing the granularity of performance feedback and hence the effectiveness of debriefs based on these measures. Specific advancements required to realize the benefit of such metrics in FMS include 1) integration of hardware into an HMD; 2) analysis techniques that can reliably identify visual focus, such as when focus is on the Heads-Up Display (HUD) versus out the window, on which instrument the pilot is fixating, and cognitive state (e.g., cognitive overload); and 3) display techniques for visualizing the data in a format usable by pilot instructors during assessment and debrief.

## 3 HMD ASSESS Approach

The HMD ASSESS development effort aims to create a precision performance assessment system which integrates advanced sensor technologies including eye tracking and EEG to measure “unobservable” perceptual and cognitive processes such as visual scan, attention allocation and cognitive workload during HMD-based performance. Based on these granular-level process measures, HMD ASSESS will diagnose performance deficiencies (e.g., failures in monitoring and detection) and inefficiencies (e.g., times of cognitive overload, distraction or inefficient scan strategies) and provide Real-time and After Action Review (AAR) summaries of individualized performance issues. These summaries can be used to 1) support training instructors in identifying skill decrements which need to be effectively remediated to achieve criterion performance and 2) assist system designers in gaining an understanding of how a pilot is interacting with the system and 3) identify specific problem areas within the display. HMD ASSESS will thus provide a comprehensive understanding of pilot performance within an HMD enabled environment, a task previously unachievable due to the unobservable nature of these processes. The resultant precision performance assessment system is intended to improve training effectiveness by providing instructors with access to previously unobservable perceptual and cognitive processes, allowing them to pinpoint the root cause of performance deficiencies (e.g., issues with attention allocation) and effectively tailor the debrief to address the problem.

This effort commenced with the development of a taxonomy which delineated the data presented in the F-35 HMD and the expected pilot interactions with this information. F-35 instructor pilots and other domain experts were interviewed throughout the design process, with interviews conducted in an iterative manner. Utilizing the taxonomy as a foundation for what needs to be measured in order to understand pilot interactions with the HMD, a conceptual design of HMD ASSESS was developed and evaluated by F-35 Subject Matter Experts (SMEs) who provided input leading to the redesign of several HMD ASSESS metrics, diagnostic methods and displays. The resulting HMD ASSESS conceptual model consists of four main components, including 1) Measurement component, 2) Diagnosis component, and 3) Instructor Displays component.

These components are discussed in the following sections and a use case is presented to illustrate the tool concept of operations.

## 3.1 HMD ASSESS Conceptual Design

**HMD ASSESS Measurement.** The HMD ASSESS measurement and data capturing component will log the occurrence of relevant events during the training session. The measurement component will receive events from a variety of available data sources, including the simulation system or another instructor learning station as appropriate (e.g., when warnings are provided or HOTAS inputs received from the pilot), eye tracking (e.g., ocular fixations relative to pre-defined high or low priority areas of the cockpit for a specific segment of flight or emergency scenario), EEG hardware (e.g., cognitive workload levels ) and input devices available to the user. The measurement component will assess events received for inclusion in the diagnostics to facilitate system flexibility required for integration into multiple training simulations. This component will be the hub for integrating the available data sources and calculating metrics to support the diagnostics.

*Taxonomy Development.* Based on an analysis of F-16 and F-35 operations, and advanced HMD systems (including the Helmet Mounted Display System, the Joint Helmet Mounted Cueing System, and the Helmet Mounted Integrated Targeting system), an HMD-ASSESS Taxonomy was developed to serve two purposes. First, it provides a preliminary understanding of unique and common data displayed across HMD systems as well as when and how pilots interact with HMD presented data. Second, it provides a foundation for identifying metrics to assess this interaction as it identifies when pilots should be monitoring different pieces of information and potential errors in doing so. The taxonomy provides a breakdown of the following information for 40 symbols provided in the F-35 HMD interface and 8 MFD displays, including a description and location of the symbol, other locations in the cockpit the same information can be found, the associated tasks performed when interacting with the symbol, and common errors associated with monitoring the symbol.

**HMD ASSESS Diagnosis.** The diagnostic component will utilize the raw metrics output from the measurement component and run a series of algorithms utilizing constraint-based modeling approaches to identify key performance decrements and the underlying causes of these decrements such as insufficient attention allocation, cognitive state and occurrence of tunnel vision. This will be used to identify critical performance issues on which instructors should focus their training interventions such as AAR debrief and future training scenario selection and manipulation. Output of the performance diagnosis will provide instructors with pilot generated errors, root-cause error analysis, and consolidated error pattern analysis.

**HMD ASSESS Real-Time Display.** The HMD ASSESS will include a real-time presentation of pilot trainee eye scan data displayed over a video feed displaying the area of the cockpit where the trainee is currently looking. The real-time display will be viewable in the instructor station, so the instructor can monitor where the pilot is looking and flag errors if desired, in addition to the errors identified automatically by the system.

**HMD ASSESS After Action Review Displays.** An AAR screen generator will be implemented in HMD ASSESS that displays a variety of data to assist instructor in pilot performance assessment and debrief, including:

- Graphical representation of pilot eye scan performance
- Diagnostic information regarding pilot performance decrements
- Performance summaries of pilot behavioral and eye scan performance

Diagnostic outcomes will be fed forward to the display component which will present a single AAR screen containing 1) a playback mode showing real-time trainee eye scan data relative to pre-defined high or low priority areas of the cockpit for a specific training segment, 2) an overview mode showing a summary of all eye scan data relative to pre-defined high or low priority areas of the cockpit for a specific training segment, 3) a multi-level timeline which contains performance feedback and allows the instructor to zoom into specific segments of flight (i.e., taxi, take-off, approach, etc.), emergency scenarios (e.g., engine flame-out or rudder failure), instructor flags or system identified errors (e.g., when the pilot misses a required task for a specific segment of flight or emergency scenario), 4) a summary list of all instructor flags and system-identified errors, and 5) summaries of the total time the trainee fixated on different areas of the cockpit (e.g., the total time the pilot was heads up or heads down, the total time spent looking at HMD symbols or MFD pages). In summary, the AAR screens provide the instructor with the ability to select a specific segment of flight in order to review scan patterns, errors, and visual allocation timing information with the pilot trainee.

## 4 HMD ASSESS Use Case

HMD-ASSESS is designed to be utilized during the actual training session and debrief. A use-case was developed to demonstrate the HMD ASSESS concept of operations for F-35 FMS training sessions and is presented in summary in this section.

A typical training session in the FMS may include several abnormal malfunctions from which a pilot must attempt to recover. During this particular training session, the instructor has inserted an Integrated Power Package (IPP) failure into the scenario. As the pilot trainee attempts to recover from the IPP failure, he performs three key errors: 1) the pilot misses a critical checklist item (i.e., arming the backup oxygen system); 2) the pilot spends too much heads down time looking at his checklist and fails to scan his primary flight instruments (altitude, attitude, airspeed) at the necessary intervals; 3) the pilot develops tunnel vision on an area of the cockpit irrelevant to the appropriate task, e.g., determining the best place to land, resulting in a delay in conducting a critical checklist item (i.e., open RAM door).

After the training session in the simulator has ended, the instructor uses the HMD ASSESS after action review displays to facilitate his debrief to the pilot trainee as follows. The instructor is interested in assessing the students handling of the IPP failure, so the instructor clicks on this segment of the timeline and the timeline automatically zooms into the IPP failure event. The instructor points out overall timing summary for that segment to the pilot, including total time heads up vs. heads down and total time in high priority areas. The instructor can illustrate to the pilot trainee that he spent a large amount of time heads down while handling the IPP Failure.

The instructor then clicks on the first system identified error, which automatically zooms the timeline down to a system default of 30 seconds on either side of the error. The instructor plays back the error and points out that, based on the eye tracking data, the pilot was distracted from reading the checklist by focusing on blinking lights on the IPP Panel.

The instructor then moves on to the next error (i.e., breakdown in a periodic eye scan of flight instruments), by selecting the error from the error summary list. The instructor wants to show the pilot how he failed to scan his primary flight instruments frequently enough. By using the Overview mode containing a summary of all eye tracking data for 30 seconds on either side of the error, the instructor illustrates to the trainee that a scan of these three primary flight instruments did not occur during this time period. The instructor confirms this by pointing out the timing summary which shows that the pilot spent very few seconds looking at the altitude, attitude, and airspeed instruments for the specified window of time.

The instructor then points to the section of the timing summary that shows the total time spent on each MFD page for the segment of flight in focus. He uses this data to illustrate that the pilot spent only 30 seconds looking at the navigation page and flight instruments because he started to look for the nearest airport to land too early, instead of following the checklist steps. This caused the pilot to delay in opening the RAM (i.e., air intake) door, which resulted in systems overheating more quickly.

As illustrated in the se case, HMD ASSESS will allow an instructor to more accurately and efficiently diagnose a performance issue. Instructors will be better able to direct a pilot's attention during overwhelming flight scenarios and prevent pilots from making common mistakes with regard to visual attention allocation.

## 5 Future Research

Development of the HMD ASSESS prototype is currently underway. HMD ASSESS will be integrated with the PTA initially, with the ultimate goal of implementing the system in the F-35 FMS. As HMD ASSESS has an iterative lifecycle and development process, the initial HMD ASSESS prototype will be verified and validated, and then revised as needed following implementation. The effort will culminate with a training effectiveness evaluation to assess the impact HMD ASSESS has on performance assessment and training effectiveness.

**Acknowledgements.** This research was sponsored by the U.S. Air Force Research Laboratory under contract FA8650-12-C-6303.

## References

1. United States Air Force. The Edge Air Force Transformation 2010 (2008), `http://permanent.access.gpo.gov/lps40477/edgeweb.pdf` (retrieved June 1, 2011)
2. Endsley, M.R.: Designing for Situation Awareness in Complex Systems. In: Proceedings of the 2nd International Workshop on Symbiosis of Humans, Artifacts and Environment, Kyoto, Japan (2001)
3. Raab, M., Johnson, J.G.: Expertise-based differences in search and option-generation strategies. Journal of Experimental Psychology: Applied 13(3), 158–170 (2007)
4. Jarodzka, H., Scheiter, K., Gerjets, P., van Gog, T.: In the eyes of the beholder: How experts and novices interpret dynamic stimuli. Learning and Instruction 20(2), 146–154 (2009)
5. Mello-Thoms, C., Ganott, M., Sumkin, J., Hakim, C., Britton, C., Wallace, L., Hardesty, L.: Different search patterns and similar decision outcomes: How can experts agree in the decisions they make when reading digital mammograms? In: Krupinski, E.A. (ed.) IWDM 2008. LNCS, vol. 5116, pp. 212–219. Springer, Heidelberg (2008)
6. White Jr., K.P., Hutson, T.L., Hutchinson, T.E.: Modeling human eye behavior during mammographic scanning: Preliminary results. IEEE Transactions on Systems, Man, & Cybernetics Part A: Systems & Humans 27(4), 494–505 (1997)
7. Weibel, N., Fouse, A., Emmenegger, C., Kimmich, S., Hutchins, E.: Let's look at the cockpit: exploring mobile eye-tracking for observational research on the flight deck. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp. 107–114. ACM (March 2012)
8. Wetzel, P.A., Anderson, G.M., Barelka, B.A.: Instructor use of eye position based feedback for pilot training. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 42(20), pp. 1388–1392. Sage Publications (October 1998)

9. Carroll, M., Fuchs, S., Hale, K., Dargue, B., Buck, B.: Advanced Training Evaluation System: Leveraging Neuro-physiological Measurement to Individualize Training. In: The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC), vol. 2010(1), National Training Systems Association (January 2010)
10. Schnell, T., Hamel, N., Postnikov, A., Hoke, J., McLean III, A.L.: Physiological Based Simulator Fidelity Design Guidance (2012)
11. Carroll, M.B., Kokini, C., Moss, J.: Multi-Axis Performance Interpretation Tool (MAPIT) Training Effectiveness Evaluation Report (Program Interim Rep., Contract No. N00014-10-C-0091). Office of Naval Research, Arlington (2010a)
12. Hale, K.S., Fuchs, S., Axelsson, P., Baskin, A., Jones, D.: Determining gaze parameters to guide EEG/ERP evaluation of imagery analysis. In: Schmorrow, D.D., Nicholson, D.M., Drexler, J.M., Reeves, L.M. (eds.) Foundations of Augmented Cognition, 4th edn., pp. 33–40. Strategic Analysis Inc., Arlington (2007)
13. Hale, K.S., Fuchs, S., Axelsson, P., Berka, C., Cowell, A.J.: Using physiological measures to discriminate signal detection outcome during imagery analysis. Human Factors and Ergonomics Society Annual Meeting Proceedings 52(3), 182–186 (2008a)
14. Hale, K.S., Fuchs, S., Berka, C.: Driving EEG cognitive assessment using eye fixations. In: Applied Human Factors and Ergonomics 2nd International Conference, Las Vegas, NV, July 14-17 (2008b)
15. Carpenter, A., Fuchs, S., Whitlow, S., Fiore, S., Hale, K.: Auto-Diagnostic Adaptive Precision Training for Baggage Screeners (Screen-ADAPT). Phase I Final Report. Department of Homeland Security Contract No. N10PC20028 (2010); Carroll, M., Fuchs, S., Hale, K., Dargue, B., Buck, B.: Advanced training evaluation system: Leveraging Neuro-physiological measurement to individualize training. In: Proceedings of the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) Annual Meeting, Orlando, FL (2010b)
16. Vision Systems International - JHMCS, `http://Vsi-hmcs.com` (retrieved September 15, 2010)