Eye Tracking Metrics for Workload Estimation in flight Deck Operation

Kyle Ellis, Graduate Student; Thomas Schnell, Associate Professor
Operator Performance Laboratory, University of Iowa
kkellis@engineering.uiowa.edu; tschnell@engineering.uiowa.edu

Abstract. Flight decks of the future are being enhanced through improved avionics that adapt to both aircraft and operator state. Eye tracking allows for non-invasive analysis of pilot eye movements, from which a set of metrics can be derived to effectively and reliably characterize workload. This research identifies eye tracking metrics that correlate to aircraft automation conditions, and identifies the correlation of pilot workload to the same automation conditions. Saccade length was used as an indirect index of pilot workload: Pilots in the fully automated condition were observed to have on average, larger saccadic movements in contrast to the guidance and manual flight conditions. The data set itself also provides a general model of human eye movement behavior and so ostensibly visual attention distribution in the cockpit for approach to land tasks with various levels of automation, by means of the same metrics used for workload algorithm development.

1. INTRODUCTION

Operators in today's aircraft flight decks find themselves in various situations that change their cognitive workload. Research to improve the interaction between the operator and the aircraft interface is benefited by being able to analyze operator state quantitatively as opposed to the historical standard of subjective feedback. This eliminates the subjective bias across subjects and standardizes feedback to provide more accurate analysis of operator state in different testing scenarios in flight deck operations. The empirical data found within the data set is useful in creating human eye movement simulation models. The eye movement metrics, the experimental procedure, and findings are described in this paper. Together, they are part of ongoing research at the University of Iowa's Operator Performance Laboratory (OPL) initiatives to both model human attention and pilot workload in flight deck environments, funded through NASA research grants.

2. BACKGROUND

There are several opportunities to advance flight decks of the future through utilization of real-time pilot workload assessment. Current avionics are not aware of pilot real-time capabilities and limitations resulting from varying workload levels. In flight deck operations there exists the potential for information overload in various phases of flight and various circumstances. Several systems within the flight deck itself, such as the flight management system and autopilot, are very effective at making easy procedures easier and hard procedures harder in situations with dynamic changes; such as unexpected occurrences in flight. If the avionics could be aware of pilot state, they could provide dynamic displays with situationally appropriate information.

The concept of the intelligent flight deck is currently being defined by a NASA project within the Aviation Safety program. The OPL at the University of Iowa is working a project entitled Operator State Sensor Investigations and Operator Feedback

Algorithms. One aspect of this project is to interpret operator workload and overall cognitive state effectively to optimize the flight deck interface.

There are several ways to characterize operator state, including electroencephalogram (EEG), electrocardiogram (ECG, heart rate), galvanic skin response (GSR), respiration rate, flight technical performance, and eye tracking to name a few. Eye tracking is appealing in flight deck operations due to its technical readiness level being higher than that of other sensors or measures of operator state. Since flying on instruments is a visually prescribed activity that is likely to be influenced by over/under-loading the pilot, eye movement behavior provides a rich data set to investigate its ability to characterize operator state.

3. EYE TRACKING METRICS

Generally, we speak of two types of eye movements: Fixations and Saccades with respect to attention allocation. Cf. Jacob and Karn [1], define a fixation as a single point of gaze vector within a threshold of two degrees for a minimum duration of 200ms. However, definitions do vary from user to user. Saccadic movement is simply derived based upon the definition of the fixation, by counting a saccade as the movement from one fixation to the next. Saccadic movements are measured by saccadic distance (deg) and velocity (deg/sec). Their Euclidian distance can be derived by determining the plane on which the fixation is occurring and identifying the distance between that specified location and the eye gaze origin.

Fixations are the time in which an individual processing the visual data within the foveal field (<2 deg). Since the foveal field is so limited, saccadic movements are necessary to bring to focus and process information across a person's overall field of view. For example, reading a book consists of several fixations and saccades that trace in spatial segments across the page. A pilot's eye scan behavior is similar, making fixations at specific instruments to obtain information, performing a saccade to fixate upon the next instrument.

Further metrics may be derived from these two general eye movements. Statistical analysis of each general metric, such as average fixation duration, fixation frequency, fixation duration max and standard deviation of the fixation duration are generally conducted. Scan-path/link analysis is used to quantify saccadic movements and fixation location patterns. These statistics are then used to observe if a correlation exists between the eye tracking metrics and the independent test variables.

Eye tracking entropy is the level of randomness observed in eye movement behavior by evaluating the gaze vector X and Y coordinates. For real-time evaluation, the standard deviation of the X and Y gaze vector components are calculated over a moving time window of 30 seconds [2]. A moving window of 30 seconds was chosen in an attempt to collect enough fixations to evaluate with statistical significance when calculating the standard deviation. With this calculation, changes in entropy values indicate a change from the current scan-path and fixation trends, presumably induced by a change in the flight deck and/or operator.

Equation 1. Entropy Equation

$$Entropy = H = \sum p_i \log_2(1/p_i)$$

Each area of interest or fixation point is associated with a state-space probability of subject focus (p_i). By assuming when situations are in high entropy, or high levels of randomness, the probability of looking at everything an equal number of times will transition between all areas of interest and stimuli at near equal frequencies. The state-space probability changes over time as scan-path trends change, therefore, changing the entropy value [4]. In theory, as workload increases the observed scan-path becomes less random [3, 4].

Since each display on the flight deck provides specific information utilized by the pilot in different ways, it is important to characterize the data in two ways: General metric analysis across the flight deck, and specific analysis of the metrics within pre-defined areas of interest (AOIs). Areas of interest used in this experiment were broken down specifically to interpret data on the simulator flight deck interface. They included the Multifunction Display (MFD), Mode Control Panel (MCP), Out the Window (OTW), Computer Display Unit (CDU), and a higher resolution of the Primary Flight Display (PFD) with the standard electronic flight information system (EFIS) display configuration, including the airspeed indicator (ASI), altimeter (ALT), heading indicator (HDG), flight mode annunciator (FMA), and the attitude indicator (AI).

4. APPARATUS

A pilot-in-the-loop study was conducted in the Operator Performance Laboratory's flight deck simulator that is based on the Boeing 737-800 form factor. The simulator is comprised of a flight deck with complete glass cockpit displays, five outside visual projectors, functioning mode control panel (MCP) with autopilot and auto throttle, and standard Boeing 737 flight controls. The head down display (HDD) panel was configured to represent the standard Boeing EFIS display on the PFD. The MFD displayed a moving map depicting the current flight plan and corresponding waypoints, as well as other useful information as would be found standard on a typical 737 glass cockpit.

Figure 1. OPL Flight Deck Simulator

A three camera Smarteye eye tracking system was installed in the simulator cockpit and the camera angles were optimized for the left seat only as shown in Figure 1. OPL Flight Deck Simulator The eye tracking camera configuration was optimized to obtain pilot eye gaze vectors with quality spatial resolution down to one degree and no greater than two degrees of gaze point variation for critical areas of interest as specified earlier. Cameras were mounted directly beneath the glare shield of the flight deck to minimize display obstruction.

5. DESIGN OF EXPERIMENT

The experiment was designed with the intention of yielding a wide range of induced workload across pilots. A total of 12 pilots with at least a private pilot license and IFR rating were asked to fly a simulated instrument approach to runway 9R at KORD (Figure 2) with two design factors, 1). Level of automation (none, flight director, coupled), and 2). Visibility at decision height (DH) (sufficient or insufficient to land)

Level of automation varied among a fully coupled autopilot mode with automated flight controls and auto-throttle, flight director mode with auto-throttle and a flight director overlaid on the PFD, but flight control were manipulated manually by the pilot and manual approach with no automated controls requiring the pilot to control both the yoke and the throttle. Pilots were allowed to utilize only IFR low-altitude en-route charts, approach plates, the localizer glide slope and course deviation indicators as well as a moving map on the MFD. Pilots were also given an approach checklist broken down by waypoint and were required to make standard radio calls and frequency changes.

The "land or go-around" visibility at DH changed between three tenths and a single tenth of a nautical mile. This required pilots to make a decision 200 feet above touchdown height to either land or go-around upon visualization of the runway end identifier lights. This condition only changed the approach scenario from the last waypoint to decision height.

Figure 2. KORD 9R Approach Flight Plan

In flight assessment was administered at each waypoint to increase the resolution of pilot workload using the Bedford workload scale. The Bedford scale is a 1-10 workload rating assessing the current workload perceived by the pilot. Pilots were trained in pre-flight briefing on the scale's decision tree to aid in assessing their workload appropriately according to the scale definition.

The Bedford scale (figure 4) breaks down into subcategories of satisfactory workload (1-3), tolerable workload (4-6), possible to accomplish task workload (7-9), and impossible to accomplish task due to high workload (10).

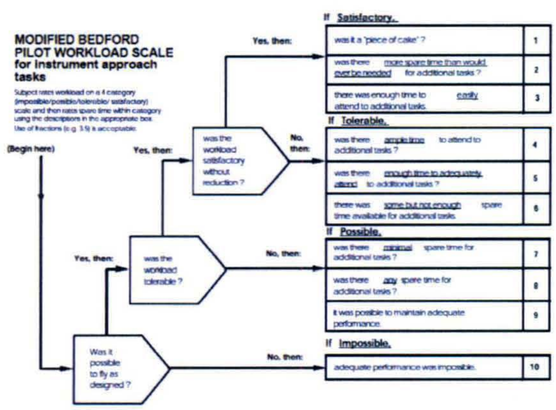

Figure 3. Bedford Workload Scale

6. HYPOTHESIS

Two hypotheses were generated in this experimental procedure: 1.) Workload will have a monotonically increasing relationship with increasingly manual flight conditions. 2.) Saccade length will have a monotonically decreasing relationship with increasingly manual flight conditions. It is reasonable to consider other eye tracking metrics to be in place of saccade length for hypothesis two. This analysis simply addresses saccade length as a simple indicator of fixation dispersion, similar to that of visual entropy or nearest neighbor index analyzed across the flight deck.

The hypotheses are developed by observing typical pilot eye scan behavior in varying levels of flight automation and their respective visual demands. Pilot demand is increased by reducing the level of automation provided. The level of automation provided changes the pilots scan behavior, indicated in this analysis by a reduction in average saccade length with increasing manual flight control. This follows the logic that a pilot is required to more closely monitor the aircraft state gauges on the PFD looking elsewhere less of the time when not flying on autopilot (Full Auto), thereby decreasing the average saccade length.

7. DATA SET AND ANALYSIS

Data were processed and analyzed by Minitab version 14. Analysis of the subjective results crossed with the testing conditions indicated that test conditions yielded significant variance in induced workload.

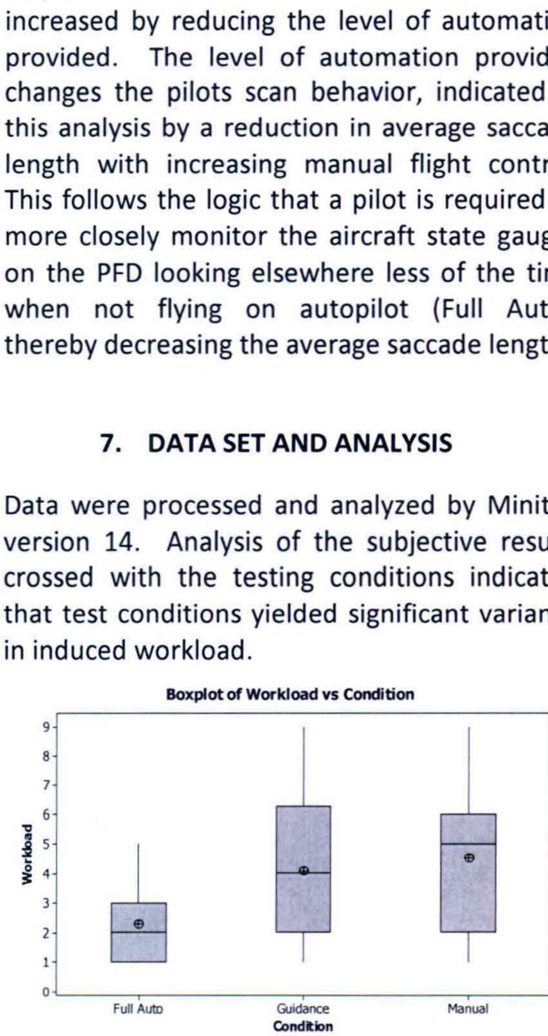

Figure 4. Workload vs. Condition

ANOVA analysis of the effect of condition on workload indicates significant variance among the automation conditions ($F(2,136)=16.35$, p<0.001). A Tukey pair wise comparison test indicated significance between the full automation condition and the guidance condition (t=4.237, p<0.0001), and significant variance between full automation condition and manual condition (t=5.404, p<0.0001). No significant variance was found between the guidance condition and the manual condition (t=1.006, p>0.05).

Analysis of variance of the saccade length observed against automation condition yielded significant results (F(2,136) = 11.50, p<0.001). Tukey pair wise comparison tests indicate significance between the full automation condition and the guidance condition (t=-4.732, p<0.001) and the full automation condition and the manual condition (t=-2.962, p<0.05) There was no significant variance between the Guidance and Manual conditions (t = 1.006, p>0.05).

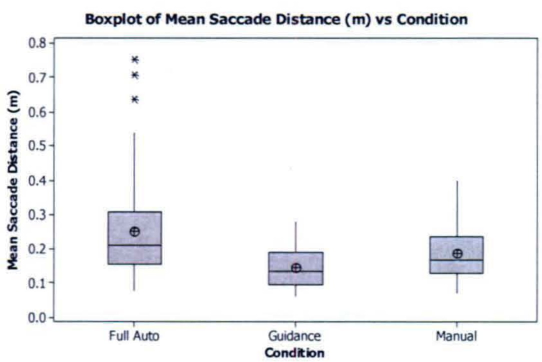

Figure 4. Saccade Length vs. Condition

8. DISCUSSION

The lack of variance between the Guidance and Manual conditions is explained by the layout of the standard EFIS display. The source of Guidance information is the flight director displayed on top of the AI on the PFD. A pilot operating in the Manual condition would use the AI combined with other instruments on the PFD to fly the correct flight path. This results in insignificant differences between these two conditions when average saccade length is analyzed across the entire flight deck and not limited to the PFD area of interest.

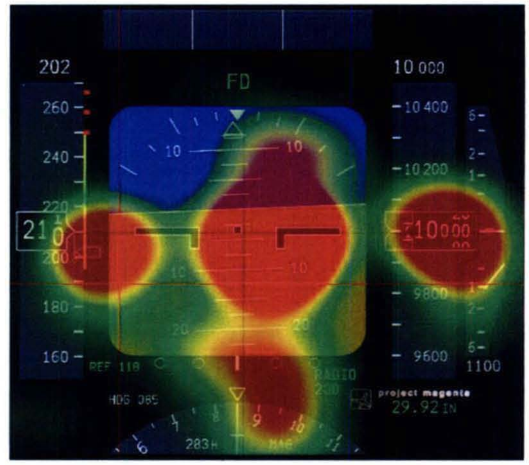

Figure 5. 737 EFIS Manual Condition Heat Map

Figure 5. 737 EFIS Manual Condition Heat Map shows a pilot's fixation heat map with no guidance (flight director) provided.

9. FUTURE WORK

The data set itself is a plentiful amount of eye movement behavior of pilots performing an instrument approach. Using this data as an empirical data set, the metrics can be averaged over the pilot population and used to feed into a human computer model for eye movement behavior. Research done at the OPL is performing such research that will utilize this data set. A project done in collaboration with NASA entitled Integrated Alert and Notification (IAN) (Grant: NNX08BA01A) will develop a digital human model capable of interacting in a digital simulation to identify optimal configuration within the flight deck.

This work is presented as a separate paper in this conference (Cover & Schnell [5]

Stemming from this research is the development of a software tool capable of receiving data and characterizing pilot workload in real time and through use of multi-channel inputs, such as EEG, EKG, Respiration rate sensors or any other types of sensors research proves useful in characterizing human workload. Cognitive Avionics Tool Set (CATS) [6] software is currently in development by OPL. Combination analysis tool and real time classifier is a useful graphical user interface for post processing and analysis.

Future analysis of the raw data will include other metrics such as visual entropy [4], nearest neighbor indexing [3], scan path indexing [7] and fixation mapping [8] to further look into what derivations of saccadic eye movement behavior yield trends that substantially correlate to pilot workload.

Further use of the data set as an empirical data source continues on various projects with the OPL and NASA. One intention is to use the data to drive an eye movement behavior model for future flight deck human performance simulations.

10. ACKNOWLEDGEMENTS

We would like to acknowledge the National Aeronautics and Space Administration under project NNL07AA00A and their team at NASA Langley for their insight and support. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NASA, the U.S. Government, or any other organization.

11. REFERENCES

[1] Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *The Mind's eye: Cognitive The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, 573 - 603.

[2]Ellis, K., & Schnell, T. (2009). *Eye Tracking Metrics For Workload Estimation In Flight Deck Operations.* Iowa City, IA: University of Iowa.

[3] Di Nocera, F., Terenzi, M., & Camilli, M. (2006). Another look at scanpath: distance to nearest neighbor as a measure of mental workload. In D. de Waard, K. Brookhuis, & A. Toffetti, *Developments in Human Factors in Transportation, Design, and Evaluation* (pp. 1-9). Maastricht, the Netherlands: Shaker Publishing.

[4] Gilland, J. (2008). *Driving, Eye Tracking and Visual Entropy: Exploration of Age and Task Effects.* University of South Dakota.

[5] Cover, M., & Schnell, T. (2009). *Modeling Pilot Behavior for Assessing Integrated Alerting and Notification Systems on Flight Decks.* Iowa City, IA: University of Iowa

[6] Schnell, T., Keller, M., & Macuda, T. (2007). Pilot State Classification and Mitigation in a Fixed and Rotary Wing Platform. *Aerospace Medical Association annual conference* (p. 377). New Orleans, LA: Aviation Space and Environmental Medicine.

[7] Simon, P., Rousseau, F., & Angue, J.-C. (1993). *Quantitative Analysis of Mental Workload Influence on Eye Scanning Movements.* Le Touquet, France: Systems, Man and Cybernetics.

[8] Wooding, D. S. (2002). *Fixation Maps: Quantifying Eye-Movement Traces.* New Orleans: ETRA.