Analysis of Eye-Tracking Data During Conditions Conducive to Loss of Airplane State Awareness

Evan T. Dill[1], Steven D. Young[2], Taumi S. Daniels[3], and Emory Evans[4]
NASA Langley Research Center, Hampton, VA, 23665

In the constant drive to further the safety and efficiency of air travel, the complexity of avionics-related systems and of the procedures for interacting with them appear to be on an ever-increasing trend. While this growing complexity often yields productive results with respect to system capabilities and flight efficiency, it typically places a larger burden on pilots to manage increasing amounts of information and to understand intricate system designs. This can be problematic as too much information and/or ineffective provisions of information can potentially overwhelm and/or confuse pilots, and as a result, increase the likelihood of loss of airplane state awareness (ASA). One way to gain more insight into this issue is through experimentation using more objective measures. This study summarizes an analysis of eye-tracking data obtained during a high-fidelity flight simulation study that included most of the complexities of current flight decks, as well as several planned for the next generation air transportation system. Multiple analyses were performed to understand how the 22 participating airline pilots were observing ASA-related information provided during different stages of flights and in response to specific events within these stages. Also, study findings are compared to data presented in similar previous studies to assess trends or common themes regarding how airline crews apply visual attention in complex flight deck and operational environments.

Nomenclature

ADI	=	Attitude Direction Indicator
AIME	=	Automation and Information Management Experiment
AOA	=	Angle of Attack
AOI	=	Area of Interest
ASA	=	Airplane State Awareness
ATC	=	Air Traffic Control
CAST	=	Commercial Aviation Safety Team
CMF	=	Cockpit Motion Facility
EFB	=	Electronic Flight Bag
EICAS	=	Engine Indicating and Crew Alerting System
FMA	=	Flight Mode Annunciator
HSI	=	Horizontal Situation Indicator
HUD	=	Head-Up Display
IB/CB	=	Inattentional Blindness / Change Blindness
ND	=	Navigation Display
OTW	=	Out-the-Window
PF/PM	=	Pilot Flying / Pilot Monitoring
PFD	=	Primary Flight Display
RFD	=	Research Flight Deck
VSD	=	Vertical Situation Display

---

[1] Aerospace Research Engineer, Safety-Critical Avionics System Branch, 1 South Wright Street, Member.
[2] Aerospace Research Engineer, Safety-Critical Avionics System Branch, 1 South Wright Street, Associate Fellow.
[3] Aerospace Research Engineer, Electromagnetics and Sensors Branch, 8 North Dryden Street.
[4] Aerospace Research Engineer, Flight Systems Software Branch, 5 North Dryden Street.

I. Introduction

THE basis of this study is a flight simulation experiment completed in 2016 called the Automation and Information Management Experiment (AIME)[1]. AIME was conducted using the Research Flight Deck (RFD) within the Cockpit Motion Facility (CMF) at NASA's Langley Research Center (Fig. 1). The RFD mimics most of the interfaces provided on the B-787 aircraft while providing a flexible environment for emulating scenarios of interest. Eleven two-pilot airline crews participated in the study, all of which had no less than 3000 hours of commercial airline flight experience. The crews completed more than 220 simulated flight scenarios using the Memphis and Denver International Airports as test sites. Each crew executed a mix of 17 different scenarios, some that were based on one or more reference events from past accidents/incidents or studies where loss of ASA was a contributing factor[2,3]. Data were collected using an experimental design that allowed for the manipulation of information, operational complexity, system performance and uncertainty across these scenarios. Flight crews were immersed in high density traffic and adverse weather environments that included many concepts either currently emerging in the industry or planned for the near future (e.g. expanded data link services, synthetic and enhanced vision systems, and interval management automation). In addition, the study emulated off-nominal (and complex) situations such as unexpected weather events, traffic deviations, equipment failures, poor data quality, communication errors, unexpected clearances, and changes to flight plans.

Figure 1. CMF/RFD simulator.

During each simulated flight, many types of data were collected, including eye-tracking data for both crewmembers. This eye-tracking data were collected to allow for objective evaluation of the crews' attention to the various displays and areas of interest (AOI), and their respective awareness of airplane state at critical junctures of the flights. The AOIs defined, analyzed, and reported here are shown in Fig. 2. They include two head-up displays (HUD), two primary flight displays (PFD), two navigation displays (ND), two electronic flight bags (EFB), and a Mode Control Panel (MCP). A final AOI (not shown in Fig. 2) is Out-the-Window (OTW).

Figure 2. AOI's overlaid on the RFD.

Furthermore, attention to important sub-AOIs on the PFD and ND are also defined, analyzed, and reported. For the PFD, sub-AOIs include the flight mode annunciator (FMA), airspeed indicator, altitude indicator, angle-of-attack (AoA) indicator, horizontal situation indicator (HSI), and air traffic control (ATC) message list area. The ND was divided into four sub-AOIs: the vertical situation display (VSD), the engine-indicating and crew-alerting system (EICAS) display, the EICAS message list area, and the navigation display (ND). These sub-AOIs are overlaid on their respective displays in Figure 3.

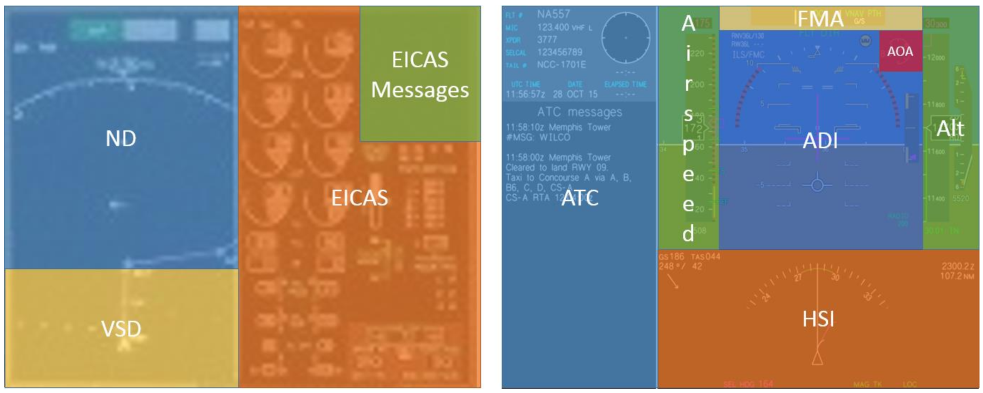

Figure 3. Left: Sub-AOIs for ND Muli-Function Display AOI; Right: Sub-AOIs for PFD AOI.

Using the derived point cloud measurements combined with their associated AOIs and sub-AOIs, multiple analyses are performed to determine where crews were dedicating their attention, and how this varied across the different scenarios. These results help us to understand how pilots observe the information provided during different stages of the flights and in response to specific events. The following sections describe the method used to perform the data processing and analysis and examples of some of the more interesting results.

II. Methodology

To collect information regarding which AOIs pilots were referring to throughout the course of each flight, the RFD platform was equipped with a Smart Eye Pro[4] head- and eye-tracking system. Through the use of multiple cameras, this system produces, for each pilot, estimates at 50 Hz for eye position and gaze direction relative to a predetermined coordinate frame. The system provides other parameters used here, including an associated confidence metric (scaled 0:1) and an estimate of the AOI being observed for each measurement.

Raw data collected through the Smart Eye Pro system were first pre-processed to eliminate outliers, reduce noise, better define the boundaries of each AOI, and yield the capability of identifying new AOIs and sub-AOIs. First, each of the eye location and gaze direction measurements was projected onto a unit surface or plane. The results of this projection can be seen for the entirety of a single flight in the upper right quadrant of Fig. 4. For further visualization and evaluation of the eye-tracking system's outputs, the lower left quadrant of Fig. 4 shows the point cloud of data with corresponding AOIs overlaid. Upon comparing these results to the RFD cockpit shown on the upper left quadrant of Fig. 4, it can be seen that the projected points and corresponding AOIs obtained through the eye-tracking system produce a reasonable representation of the RFD's true layout.

While the raw data from the eye-tracking system produced a realistic depiction of the RFD's layout, a number of points can be observed that appear to have incorrectly identified AOIs based on their projected spatial location. To minimize the points that are either incorrectly identified or have erroneous locations, the eye-tracking data were filtered using a set of defined heuristics. First, points were eliminated when provided with a low confidence metric as reported by the eye-tracking system. All points with an associated confidence less than 60% were removed. Next, points were eliminated based on their proximity to other points with common AOI identification. For each AOI, an ellipse was created around the center (i.e. spatial average) of the subset of points to encompass 98% of the points associated to that AOI. All points not contained within this ellipse were eliminated. Lastly, moving-average filters were employed to account for the motion of the pilot's gaze direction and eye location. Given the 50-Hz update rate combined with the expected actions of the pilot, it was assumed that large excursions in motion should not be observed between samples. Therefore, a sliding window average was applied to each channel (i.e., x, y, and z) of the aforementioned parameters to help smooth out any such noise present on the measurements. This was accomplished through an implementation of equation (1) where a new measurement is formed, $p_{i,smoothed}(t_k)$, by averaging the current measurement, $p_i(t_k)$, with the previous two measurements, $p_i(t_{k-1})$ and $p_i(t_{k-2})$. The effects of this filtering technique is shown in the lower right quadrant of Fig. 4. These results appear to be less cluttered, less noisy, and more indicative of eye-movement behavior than the raw data provided by the eye-tracking system.

$$p_{i,smoothed}(t_k) = \frac{p_i(t_k) + p_i(t_{k-1}) + p_i(t_{k-2})}{3} \tag{1}$$

Using the filtered point cloud, sub-AOIs were assigned to points contained on the PFDs and NDs. Figure 5 depicts these sub-AOIs overlaid onto the cumulative point cloud for an example flight.

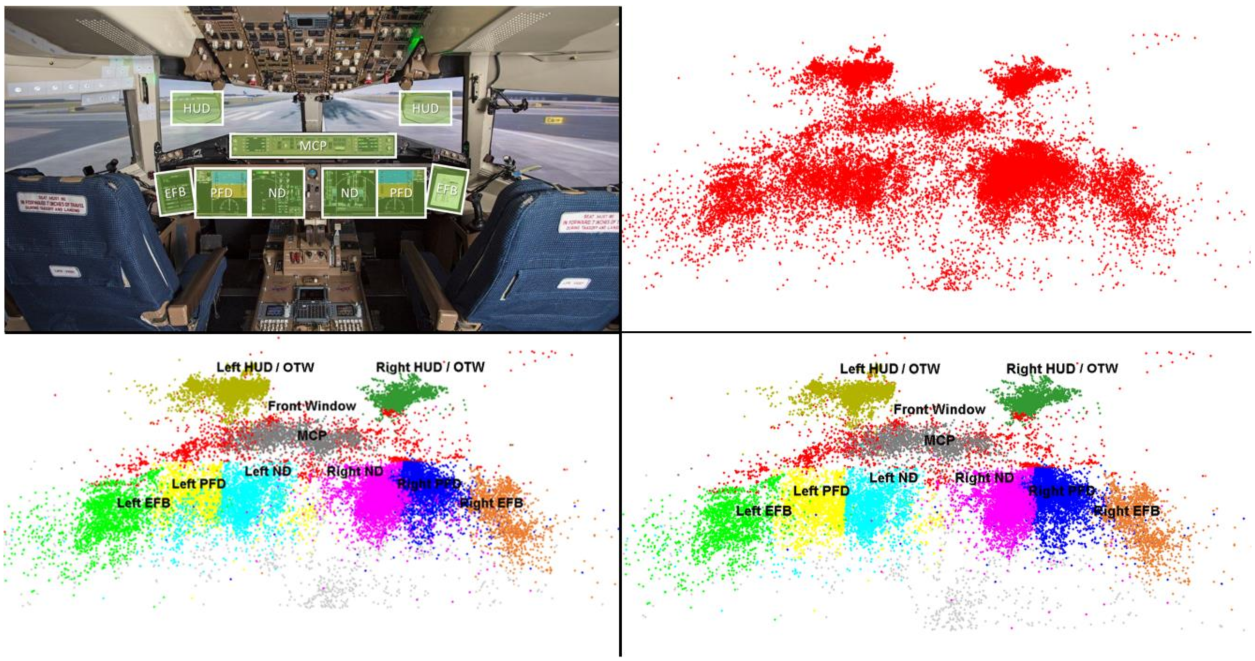

Figure 4. Upper Left: AOI's overlaid on RFD; Upper Right: Raw compilation of gaze samples; Lower Left: Unfiltered gaze samples with system-identified AOIs overlaid; Lower Right: Filtered gaze samples.

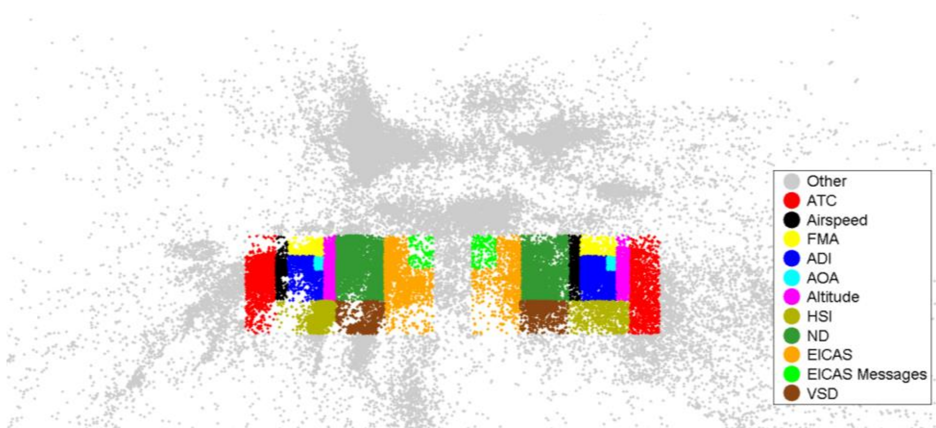

Figure 5. Filtered gaze samples with overlaid sub-AOIs for the PFD and ND MFD displays.

III. Results and Analysis

By means of the method described above, each pilot's engagement with the different AOIs was tracked and analyzed across the all flights (~220 flights). For this paper we present three types of findings: aggregate pilot gaze distributions, FMA engagement during scenarios involving radar altimeter failures affecting auto-flight system state, and airspeed/altitude indicator engagement during scenarios involving unexpected auto-throttle disconnects. Other analyses and findings will be published separately.

For the data presented in the following sub-sections, the two pilots are distinguished based on their assigned role and using the terms pilot flying (PF) and pilot monitoring (PM). Here, we consider the PM term equivalent to the pilot not flying term, which is also often used in the industry. These roles and terms should not be confused with seat position (left/right). In this study, a given crew took a seat position and stayed in that seat for the duration of their participation (2 days). They were, however, asked to change roles (PF/PM) occasionally, but to always complete at least two flights consecutively before switching. Scenarios were randomly sequenced both within and across the crews with the pilot role for each seat position briefed and set prior to each flight.

A. Aggregate pilot gaze distributions

As previously described, the projection data from the head- and eye-tracking system can be used to identify where each pilot is looking at any given time. By aggregating these data over all flights, distributions can be established regarding the ratio of time per flight that pilots engage each AOI. The average distribution for PFs and PMs is displayed in Fig. 6. PFs spent more time eyes-out due to the fact that most flights were in IMC, while PMs spent more time eyes-in (monitoring the instruments). PFs transitioned to eyes-out as visibility allowed during the latter part of the approaches and while landing. These may have been more equal had there not been added demands on the PMs to attend to EFB-related tasks (e.g. charts, checklists, ATC datacomm). It is unclear why statistically PMs spent more time engaging the HUD; however, this is believed to be an artifact of the data collection since looking at the HUD could be misinterpreted as looking OTW (and vice-versa). It is probably more appropriate to sum OTW looks with HUD looks (i.e. 28% for PFs; 27% for PMs).

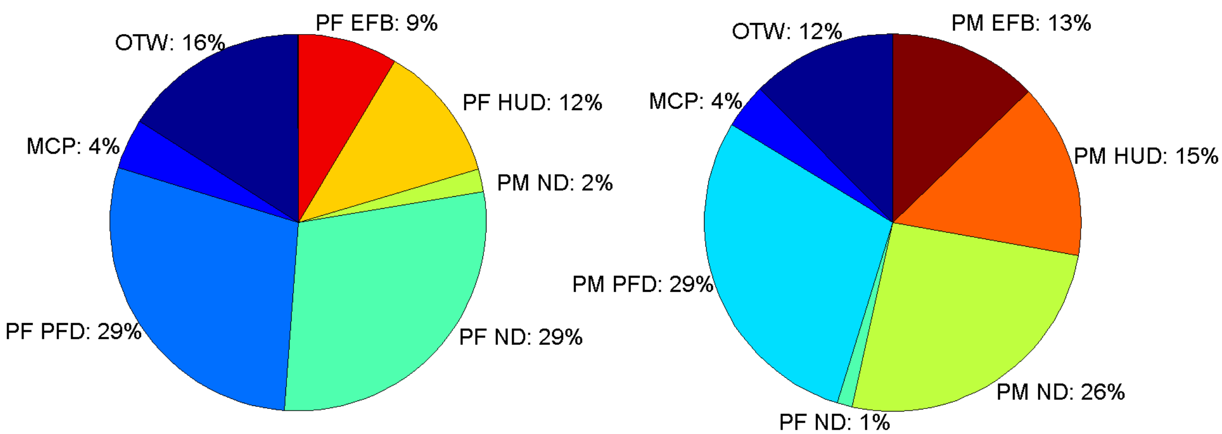

Figure 6. Average amount of time each pilot spent looking at the specified AOIs. (Left: PF; Right: PM)

At a coarse level, we see that there are four primary consumers of visual attention, accounting for 95% of the available time. For PFs this breaks down to 29% PFD, 29% ND, 28% OTW/HUD, and 9% EFB. For PMs the decomposition is 29% PFD, 26% ND, 27% OTW/HUD, and 13% EFB. These results can be compared to a previous study in the same facility but with different airline crews and different scenarios[5]. Between the two studies, AOI engagement distributions for the PFs changed significantly, while the PMs remained somewhat similar. For the PFs, ND usage increased from 16% to 29%, PFD usage went from 20% to 29%, and EFB usage went from 6% to 9%.

The increased usage of these AOIs came at the expense of time spent looking out the front window (or at the HUD) which decreased from 54% to 28%. These changes can be attributed to the different types of scenarios used in the studies. While the previous study also consisted of approaches to the KMEM airport, they did not involve the complex off-nominal situations encountered here. Dealing with onboard system failures and unexpected ATC clearances tended to draw the PFs attention inside much more in the current study. With respect to the PMs, PFD and ND usage remained within 1% of each other, while EFB usage increased from 7% to 13%. Increased EFB usage by the PMs is primarily due to working through the electronic checklists during the system failures and other off-nominal conditions.

These findings can be compared to the results of older studies such as by Mumaw[6] and Huettig[7]. However, when comparing these results, caution must be exercised due to different platforms, scenarios, system functions, and training/proficiency levels. In the present study, the PFD average of 29%, is less than the findings of (Ref. 6) and (Ref. 8) who reported 35% and 40%, respectively. The percent engagement with the ND found here (27.5% for the PF and PM combined) was slightly higher than the 25% and 20% results found in these same studies, respectively. Together, these results suggest how the growing complexity of systems and operations is tangibly increasing demands for visual attention inside the flight deck, with an unknown and difficult to quantify effect on safety margins.

Along with the total distributions, AOI engagements can be evaluated with respect to their use over time during the flights. For approaches that reached the runway threshold, Figures 7 and 8 depict the average engagement with each AOI by the PFs and PMs, respectively. These data indicate, as expected, a substantial increase in the amount of time each pilot spends looking OTW after reaching the final approach fix (typically 3-4 min prior to the runway). Earlier in the flights, both PFs and PMs engage the PFDs slightly more than the NDs. However, over the course of the flights, there is a similar trend of decreasing attention to the EFB by the PF, and to the ND by the PM.

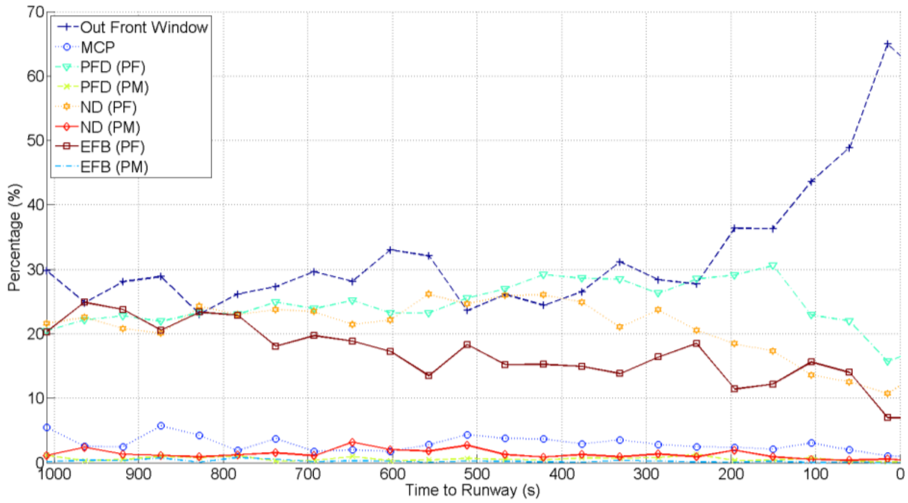

Figure 7. Average engagement of PFs with different AOIs over time.

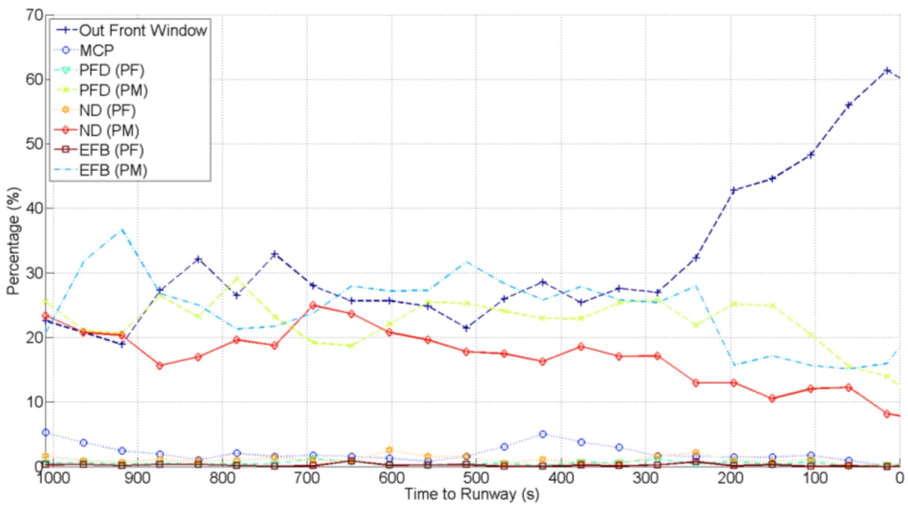

Figure 8. Average engagement of PMs with different AOIs over time.

B. FMA engagement during scenarios involving radar altimeter failures affecting auto-flight system state

The eye-tracker also provided interesting insight into behavior following the radar altimeter failure condition utilized in one of the scenarios. Similar to the reference CAST accident event (Turkish Air, 2009), a radar altimeter failure was triggered to occur late in the approach at ~1600 ft altitude. In the RFD, the only indicators to the crews of this condition were (1) the small numeric indicator of radar altitude near the center of the ADI, and (2) the auto-flight system mode change indicated as “IDLE” and “ROLLOUT FLARE” on the FMA (at the top center of the PFD) (Fig. 9). If neither pilot notices one of these indicators, it is very unlikely they will diagnose the problem and its effect (e.g., on the auto-throttles). In the reference accident, this ultimately led to a stall condition and impact with the ground short of the runway.

During the simulated flights and likely because none of the 22 pilots had ever experienced such a failure, reactions varied widely[1]. For example, Fig. 10 illustrates the significant variations in pitch following the failure as the pilots “fought” the auto-flight system, not realizing the auto-flight system mode changes to track runway centerline and hold the throttles at idle thrust. Although decoupling/disconnecting the auto-flight system was the expected decision, several crews delayed 30-60 seconds before taking this action. In other words, manual disconnect of the auto-throttles and/or the auto-pilot was the most effective way to mitigate the problem; this did occur in all of the flights, but with large variability in the time to take this action.

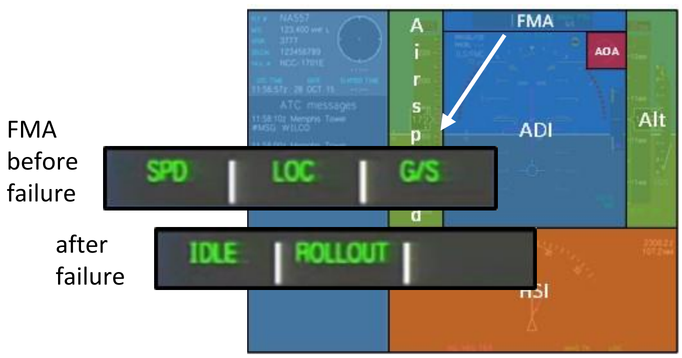

Figure 9. Change of FMA on PFD following radar altimeter failure and subsequent mode changes.

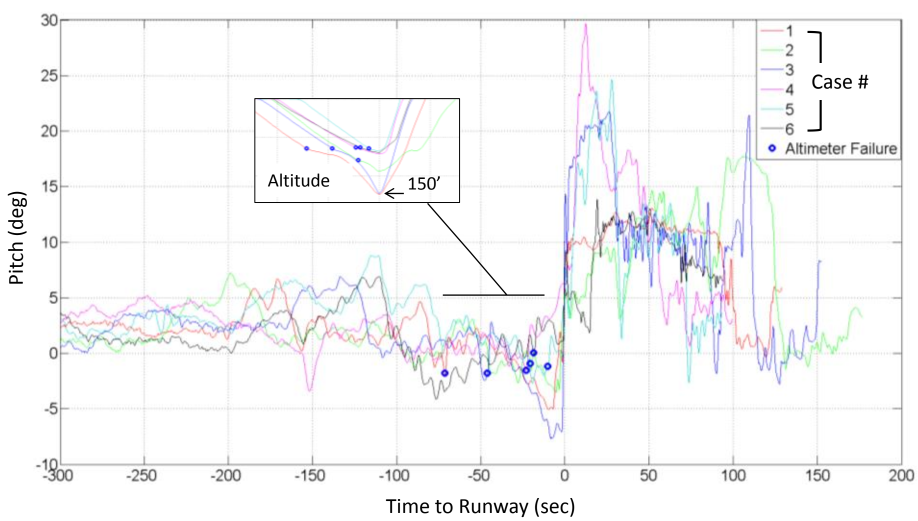

Figure 10. Pitch attitude and minimum altitude for radar altimeter failure scenario flights.

Figures 11 and 12 provide insight from the eye-tracker as to why this occurred. Figure 11 shows the variability in time-to-notice and dwell time when crews first looked at the FMA after the failure and resultant mode changes. For five of the flights, the PMs did not look at the FMA for more than 18 seconds after the failure (and mode changes) occurred. In general, PFs were more attentive to the FMA, with eight of 11 looking at it in less than five seconds.

Referring to the corresponding cases in Fig. 10, the PF in case #1 was more stable in pitch most likely due to good recognition of the FMA state change by the PM (two sec to first look, then 11 sec dwell) and then by the PF (20 sec to first look and 10 sec dwell). In contrast, the large pitch variability for the PF in case #3 is likely due to 35 sec time to first look at the FMA by the PM. The glance to the FMA in this case by the PF must not have been sufficient for recognition as there was no discussion on the audio recordings of noticing the FMA state change by the crew (and the PM did not look at the FMA for another 30 sec).

These data as well as previous research[8] suggest that some pilots may have experienced Inattentional Blindness (IB) and/or Change Blindness (CB) with respect to the radar altitude indication and/or the FMA mode change indication. For the flights shown here, the crews in some cases did not mention the mode change or the radar altimeter failure indications, although the eye-tracker suggests they were looking at or near these indicators. Crew conversations following the failure were more along the lines of "what just happened?" and/or "why is it doing that?" Figure 12 illustrates how the pilots generally applied their visual attention to the PFD (aggregated over all flights and all pilots) versus how they applied their attention just after the failure and resultant mode changes. The latter is a composite of data collected within 30 sec after the failure. In Fig. 12 (and later in Fig. 15 and 16), the brighter the color of a pixel, the more that location was measured as a look over the total number of samples. In Fig. 12, notice on the left how generally pilots spread their attention across the display space, but with more attention paid to ATC, airspeed, altitude, and pitch. For the off-nominal failure cases (on the right side), most of the attention goes to pitch as one might expect, but there is also still attention applied across the airspeed, altitude and FMA areas, suggesting the FMA and radar altitude indicators were in field-of-regard (i.e. looked at, but not necessarily seen).

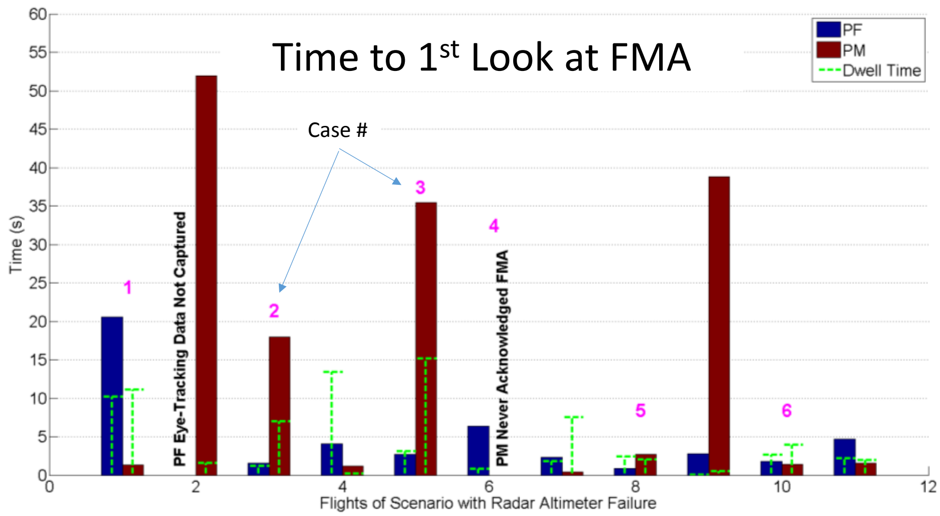

Figure 11. FMA looks for radar altimeter failure scenario flights.

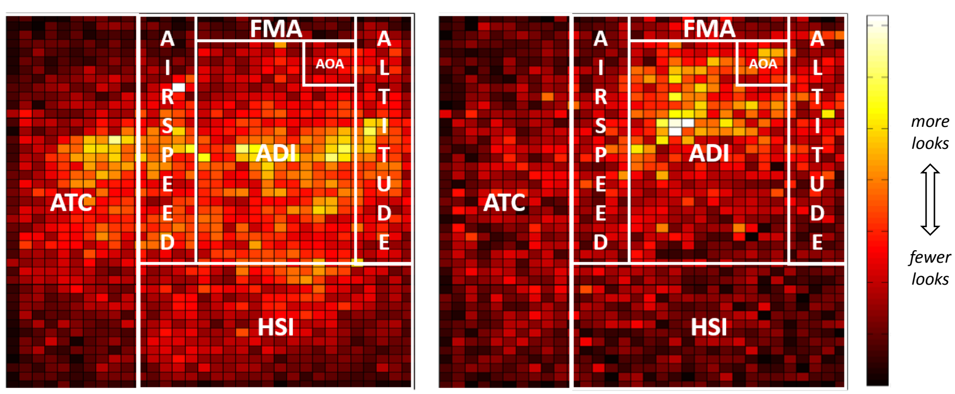

Figure 12. Distribution of visual attention to PFD for radar altimeter failure scenario flights. (Left: all pilots, all flights; Right: 0-30 sec following failure and mode changes)

C. Airspeed indicator engagement during scenarios involving unexpected auto-throttle disconnect

Several scenarios were designed in an attempt to induce low energy states for the aircraft. Two were triggered by manipulation of the auto-throttles (A/T) – disconnects or disengagements at points during the flight where the crews may be distracted or working other tasks. If not noticed, airspeed will decay until either the crew "manually" speeds up, or they re-engage the automation. In the first scenario, the A/T disconnects about three minutes after an ATC clearance is received during the descent to hold present position at ~4900 ft and with an airspeed of ~230 knots. To comply with the clearance, crews build a holding pattern in the FMS and begin flying this pattern per standard procedures. At about mid-way through the first turn, the A/T disconnects. This is indicated on the displays, but the audible tone is suppressed. The aircraft continues to track lateral and vertical path, but airspeed will begin to gradually decrease until the crew intervenes. In the second scenario, A/T disconnect occurs later in the descent when airspeed is ~180 knots, and just after ATC has requested a runway change to a parallel runway whose threshold is displaced (i.e. it is farther away than the original runway). If the crew accepts the runway change, energy will remain low for a longer period of time. The A/T disconnect is timed for this scenario in such a way as to potentially go unnoticed (or unattended to) while working the runway change procedure.

Figures 13 and 14 illustrate the main effect observed for these two scenarios. Without A/T support (the time after the black circle symbol in the figures), airspeed generally decreases until at least one pilot notices, or a stall warning is issued. As shown in Fig. 13, for two of the flights airspeed dropped more than 50 knots before either pilot reacted. In the most dramatic event (case #1), airspeed drops over 80 knots before the PM said "watch your airspeed"; then there is a brief stickshaker warning as the PF begins to recover. Ironically, part of the recovery of airspeed procedure includes lowering the flaps to five and then 10 deg, which in this case led to a brief flap overspeed warning as airspeed was quickly increased. In contrast, for the flights shown in Fig. 14, the effect on airspeed varied significantly across the crews; but loss of airspeed due to not noticing the A/T disconnect is less significant. This is because the aircraft is very late in the approach and at a relatively slow speed. The larger effect here is on the decision whether to accept the runway change, continue to the original runway, or go-around. Crews here are already concerned about being too slow, so airspeed dropping further is more carefully monitored. However, there is more spread in airspeed fluctuations as the crews make the decision how to proceed and then control airspeed accordingly.

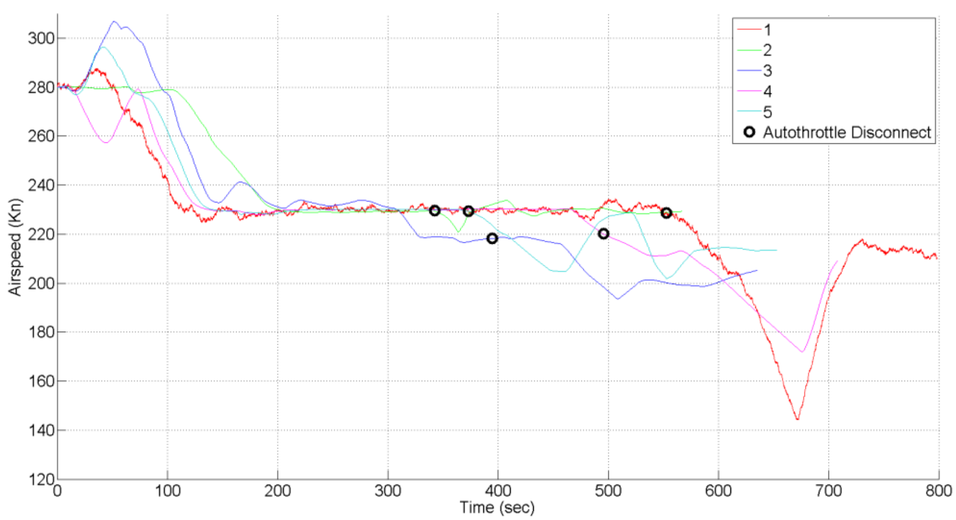

Figure 13. Airspeed during A/T disconnect scenarios while in holding pattern.

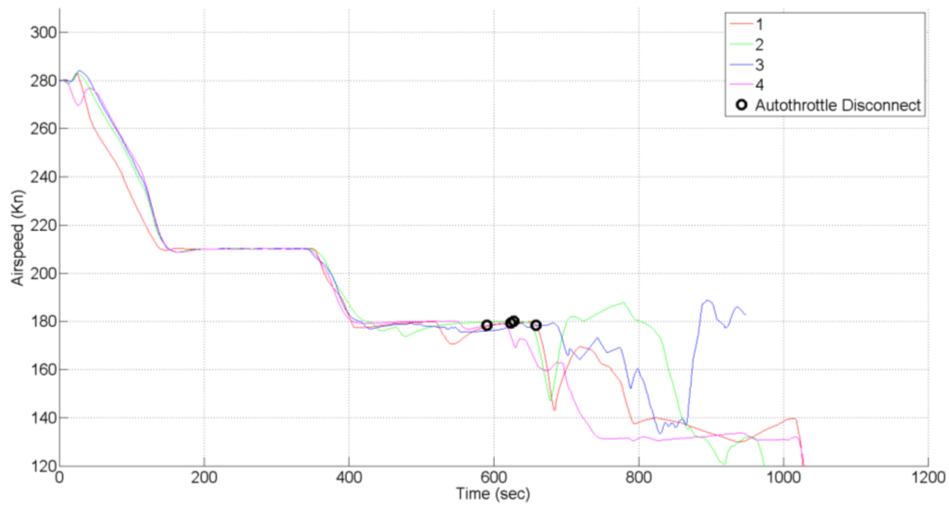

Figure 14. Airspeed during A/T disconnect scenarios following runway change clearance.

To illustrate more generally how the eye-tracking system can help to understand how crews apply attention in these sorts of unexpected loss of energy situations, refer to Table 1. Here we see that during the initial loss of airspeed, the number of looks at the PFD airspeed indicator and the magnitude of the airspeed loss varied considerably across the crews and the two scenario types. But more looks/sec did not necessarily correspond to faster recognition and intervention to return to appropriate airspeed. For the data in this table, looks are only counted if the dwell time was greater than one second; with the majority of dwell times between 1-3 sec, and only a few larger than five seconds.

Table 1. Applying visual attention to the PFD airspeed indicator following A/T disconnects.

	Duration of initial airspeed loss (sec)	*Initial airspeed loss (kts)*	*Looks during loss (PF)*	*Looks during loss (PM)*
Holding pattern case (Fig. 13)				
1	117.9	83.8	9	7
2	22.0	8.5	2	3
3	50.3	21.8	3	5
4	107.7	40.4	9	2
5	79.1	24.5	6	5
Runway change case (Fig. 14)				
1	29.7	38.5	4	3
2	21.9	30.2	3	2
3	19.4	13.5	2	2
4	17.5	9.9	3	1

Figures 15 and 16 reinforce these observations by showing how the pilots generally applied their visual attention to the PFD (aggregated over all flights and all pilots) versus how they applied their attention just after the A/T disconnects. The right side of these figures is a composite of data collected within 30 sec after the disconnects. Notice, for example, in Fig. 16 that much more attention goes to the airspeed and altitude indicators for the runway change scenario flights. These data provide a clear and objective measure of the pilots' attentional priority regarding the low energy state during this situation. As for the related decision-making performance, there was a fairly equal split by the crews as to which option to take (i.e., continue to original runway, change runways, or go-around).

A final point regarding Fig. 15 and 16 – the right side of both also provide additional evidence of IB/CB (i.e. looked at, but not seen). After the A/T disconnects, crews are clearly including in their visual scans looks at the indicators of this automation mode change on the PFD. Yet their actions, as well as the audio recordings, suggest the state change was not always recognized as quickly as might be expected given these look distributions.

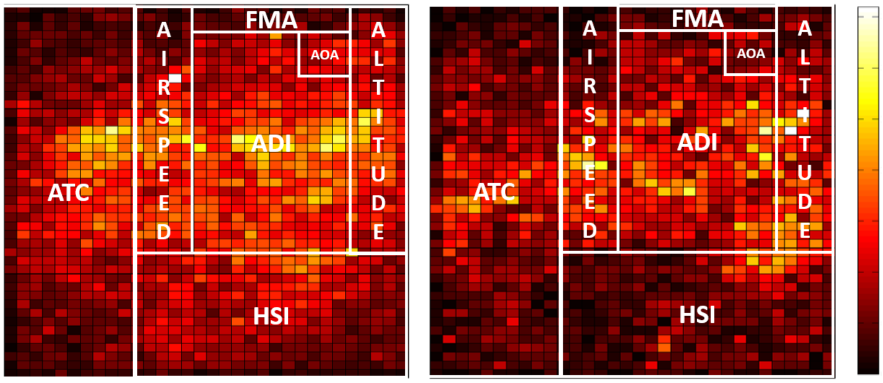

Figure 15. Distribution of visual attention to PFD for A/T disconnect in holding pattern. (Left: all pilots, all flights; Right: 0-30 sec following A/T disconnect)

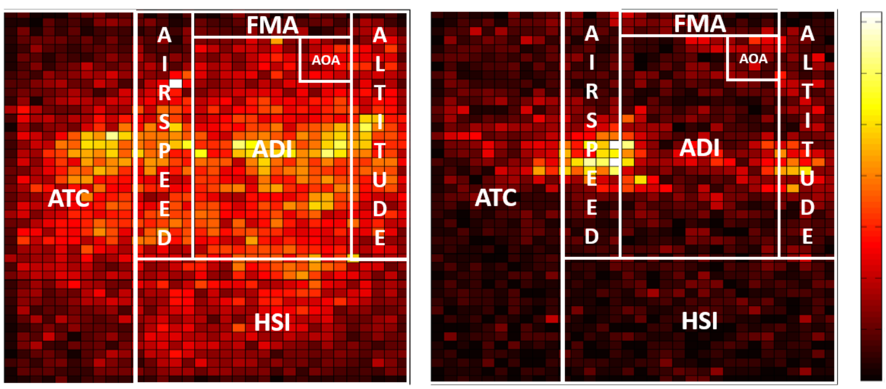

Figure 16. Distribution of visual attention to PFD for A/T disconnect following runway change clearance. (Left: all pilots, all flights; Right: 0-30 sec following A/T disconnect)

IV. Conclusion

The analysis of eye-tracking data presented in this paper is intended to provide insight into how pilots apply visual attention during complex and often off-nominal situations where loss of airplane state awareness can manifest. While eye-tracking data alone cannot tell the complete story, it can provide important and objective clues. Several of these are presented in this paper. Subsequent AIME studies are planned to build on the lessons-learned in this study, while also evaluating new system designs that can mitigate potential vulnerabilities (e.g., IB/CB) in an age of increasingly autonomous and complex systems.

Acknowledgments

The authors would like to thank NASA Langley's CMF/RFD simulation development team, led by Victoria Chung and James Barnes, for their dedication to creating a realistic and complex environment for the crews to experience. Thanks also to Lon Kelly and Kyle Ellis for their help in capturing and understanding the eye-tracking data. The work presented here was funded by the Airspace Technology Demonstration project within NASA's Airspace Operations and Safety Program.

References

[1]Young, S.D., Daniels, T.S., Evans, E.T., Dill, E.T., Uijt de Haag, M., Etherington, T. (2016). *Flight Simulation Study of Airplane State Awareness and Prediction Technologies,* Proceedings of the IEEE Digital Avionics Systems Conference. Sacramento, CA.

[2]Commercial Aviation Safety Team, (2014). *Airplane State Awareness Joint Safety Analysis Team Final Report Analysis and Results.* http://skybrary.aero/bookshelf/books/2999.pdf.

[3]Commercial Aviation Safety Team, (2014). *Airplane State Awareness Joint Safety Implementation Team Final Report Analysis and Recommendations.* http://skybrary.aero/bookshelf/books/3000.pdf.

[4]Smart Eye Pro, (2009). *API Manual,* Revision 477, Gothenburg Sweden.

[5]Dill, E.T., Young, S.D (2015). *Analysis of Eye-Tracking Data with Regards to the Complexity of Flight Deck Information Automation and Management – Inattentional Blindness, System State Awareness and EFB Usage,* Proceedings of the AIAA Aviation Conference. Dallas, Tx.

[6]Mumaw, R.J., Sarter, N.B., Wickens, C.D., (2001). *Analysis of Pilots' Monitoring and Performance on an Automated Flight Deck,* Transactions of the 11th International Symposium on Aviation Psychology. Columbus, OH.

[7]Huettig, G., Anders, G., & Tautz, A., (1999). *Mode awareness in a modern glass cockpit attention allocation to mode information*, Proceedings of the 1999 Ohio State University Aviation Psychology Conference. Dayton, OH.

[8]Mack, A., Rock, I., (1998). *Inattentional Blindness*, MIT Press, Cambridge, MA.