Eye-Tracking Analysis from a Flight-Director-Use and Pilot-Monitoring Study

Peter M. T. Zaal*
Metis Technology Solutions, Inc.
NASA Ames Research Center
Moffett Field, CA, USA

Thomas J. J. Lombaerts†
KBR Wyle Services, LLC
NASA Ames Research Center
Moffett Field, CA, USA

Randall J. Mumaw‡
San José State University
NASA Ames Research Center
Moffett Field, CA, USA

Dorrit O. Billman§
NASA Ames Research Center
Moffett Field, CA, USA

Isabel C. Torron¶
San José State University
NASA Ames Research Center
Moffett Field, CA, USA

Saad Jamal‖
San José State University
NASA Ames Research Center
Moffett Field, CA, USA

*Megan C. Shyr***
NASA Ames Research Center
Moffett Field, CA, USA

Michael S. Feary††
NASA Ames Research Center
Moffett Field, CA, USA

Eye tracking may be a useful tool to investigate pilot monitoring and develop and conduct training. There is increased interest from airlines to use eye-tracking technologies in flight simulators. However, much is still unknown about how to best utilize eye-tracking data in pilot training. This paper presents eye-tracking results from a pilot-monitoring training study with 19 pilots. All pilots completed 15 monitoring challenges across four operational scenarios in a B737-700 full flight simulator. In addition, the study investigated the impact of having the flight director engaged or disengaged on the pilot monitoring side in the final approach. It was hypothesized that pilots would focus less on the primary flight display with the flight director off and look more around in the cockpit. To assess this, pilots performed half of the scenarios with the flight director on and half with the flight director off. However, pilots monitoring tended to look less at the primary flight display with the flight director on as indicated by lower Proportion Dwell Times, contrary to the hypothesis. Next, eye-tracking data were analyzed from two monitoring challenges involving waypoint restrictions and two involving extending the flaps at appropriate airspeeds. Pilots that successfully completed the challenges appeared to focus more on areas of interest that contained the most relevant information to successfully complete the challenge. In addition, successful pilots seemed to adapt their monitoring strategy more to the challenge at hand as observed by a distinct shift in focus on either the primary flight display or the navigation display depending on the challenge. Our findings suggest the importance of flexible gaze allocation across specific situations and raise the question whether and to what degree prespecified patterns of eye fixation can be identified and trained.

Nomenclature

χ^2	=	probability distribution, -
p	=	probability, -

*Principal Aerospace Engineer, SimLabs; peter.m.t.zaal@nasa.gov. Associate Fellow AIAA.
†Aerospace Research Engineer, Intelligent Systems Division; thomas.lombaerts@nasa.gov. Associate Fellow AIAA.
‡Senior Research Associate, Human Systems Integration Division; randall.j.mumaw@nasa.gov.
§Senior Researcher, Human Systems Integration Division; dorrit.billman@nasa.gov.
¶Research Associate, Human Systems Integration Division; isabel.c.torronvalverde@nasa.gov.
‖Research Scholar, Human Systems Integration Division; saad.jamal@berkeley.edu.
**Student Trainee, Human Systems Integration Division; megan.c.shyr@nasa.gov.
††Aerospace Technologist, Human Systems Integration Division; michael.s.feary@nasa.gov.

I. Introduction

Reviews of accidents and incidents in commercial aviation suggest that flight crews are sometimes unaware of deviations in basic flight parameters, such as low airspeed [1]. One hypothesis is that pilots may not monitor critical flight parameters adequately. A recent review of the literature on monitoring and on training monitoring identified essential knowledge and skills [2]. The heart of monitoring is a sensemaking cycle that directs attention to address key questions regarding the airplane's current situation, called a Situation Model. This perspective was used to develop training techniques to improve pilot monitoring, especially as it relates to flight path management [3]. Eye tracking can be a useful tool to investigate pilot monitoring [4–6]. Furthermore, there is increased interest from airlines to use eye-tracking technologies to try to observe pilots' monitoring strategies. This information might be used in flight simulator training and to develop future cockpit procedures. Despite this interest, there are still many questions on how to actually process and apply the eye-tracking data for this purpose.

This paper presents an eye-tracking analysis from a flight-director-use and pilot-monitoring study. Pilots flew four simulated scenarios that presented three to four monitoring challenges each. Most challenges threatened meeting flight path targets. Nineteen pilots from a US commercial airline performed each scenario in the role of pilot monitoring starting at top of descent to landing. Unknown to the pilot monitoring, the pilot flying was a confederate. The participants performed two scenarios, then received an hour-long training on monitoring, followed by two additional scenarios that included similar challenges. Eye tracking was used to determine if eye-tracking measures, such as Proportion Dwell Time, were different for pilots successfully completing a monitoring challenge compared to pilots who were unsuccessful. As some airlines have suggested that the presence of the flight director influences pilot monitoring, particularly during the final approach, the study also investigated the impact of having the flight director engaged or disengaged on where pilots looked during this phase of flight.

Section II describes the scenarios of the experiment and Section III the method. Results are presented in Section IV and discussed in Section V. Conclusions are provided in Section VI.

II. Scenarios

The experiment was conducted in a B737-700 full flight simulator. Each pilot participant flew four scenarios, two scenarios before training and two after training. Scenarios 1 and 2 were a descent and arrival into Washington Dulles International Airport (IAD), whereas Scenarios 3 and 4 were a descent and arrival into Las Vegas McCarran International Airport (LAS). All scenarios started just prior to the top of descent (T/D) point and ended either at the initiation of a go-around maneuver or after a successful landing on the runway. Each scenario operated in Instrument Meteorological Conditions to an altitude of approximately 500 ft AGL. Scenario 3 had three monitoring challenges, while the other scenarios had four monitoring challenges. In order to provide pilots with similar challenges before and after training, Scenarios 1 and 3 had a similar structure, as did 2 and 4.

Nineteen first officers participated in the experiment, each in the role of pilot monitoring (PM) for all scenarios. The pilot flying (PF), part of the research team, was a confederate in all scenarios. Each scenario included a few specific challenges that required action from the flight crew, such as an air traffic control (ATC) clearance that was difficult or impossible to meet; the PF entering an inappropriate mode or value, or failing to take appropriate action; the PF directing the PM to take an action that was inappropriate in the given context; or an equipment failure. This paper analyzes eye-tracking data from four representative challenges, two challenges to meet waypoint restrictions and two challenges to extend the flaps at the appropriate time. These challenges were chosen as they had objective start and end times and expected areas of interest for eye tracking.

Pilots each flew two scenarios with the flight director on and two with the flight director off, alternating between scenarios and counterbalanced across participants. In addition to the monitoring challenges, the final approach portion of the flight in Scenarios 2 and 4 was analyzed separately to determine if pilot monitoring and eye tracking were significantly different with the flight director turned on or off. The final approach segment was the portion of the flight between 2,100 ft radio altitude and touchdown on the runway. The remainder of this section is a more detailed description of each monitoring challenge considered in this paper.

A. Waypoint Restrictions

Scenario 1, Challenge 1

In Scenario 1, the aircraft was on a flight plan from Seattle (SEA) to Dulles (IAD) via the GIBBZ3 arrival route with ILS approach to runway 01L. In Challenge 1, the flightcrew was given a descent clearance with an assigned airspeed at

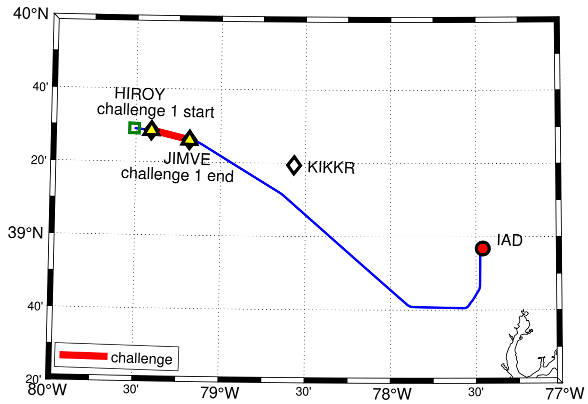

Fig. 1 Map view of Scenario 1.

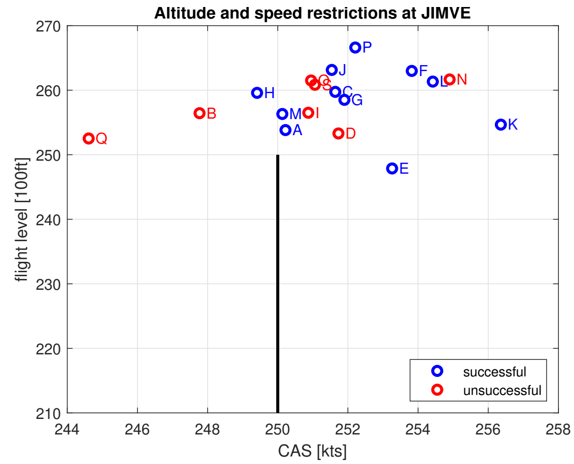

Fig. 2 Individual pilot performance of Scenario 1, Challenge 1

the top of descent (T/D) that was lower than planned. This forces the airplane above the FMS-derived flight path and made it nearly impossible to meet the altitude restrictions at the upcoming waypoint JIMVE with the lower airspeed. A map view of Scenario 1 highlighting Challenge 1 for a representative pilot is provided in Fig. 1. This challenge started at T/D and ended when the aircraft crossed JIMVE. Fig. 2 shows the altitude and speed above JIMVE of every pilot run, and compares it with the waypoint constraints (250 KIAS and altitude between FL210 and 250). The blue markers indicate successful pilots who noticed and mentioned that they were not going to make the restrictions at JIMVE; red markers show the unsuccessful pilots who didn't mention anything about making the waypoint restrictions.

Scenario 4, Challenge 2

In Scenario 4, the aircraft was following a flight plan from New York (JFK) to Las Vegas (LAS) via the TYSSN5 arrival route with ILS approach to runway 26L. In Challenge 2, ATC cleared the airplane to 'cut the corner' directly to waypoint SUZSI. This reduced the number of track miles flown to descend and slow down to satisfy both the speed and altitude requirements at that waypoint. It was hard to simultaneously descend and slow down along this shorter route. Fig. 3 shows a map view of Scenario 4 highlighting Challenge 2. This challenge started when the aircraft was cleared direct to SUZSI and ended when the aircraft crossed SUZSI. Fig. 4 shows the altitude and speed when every flight crew overflew SUZSI, and compares it with the waypoint constraints (210 KIAS and flight level FL100). Again, blue markers represent pilots monitoring who successfully identified not making the waypoint restrictions and red markers represent pilots who didn't.

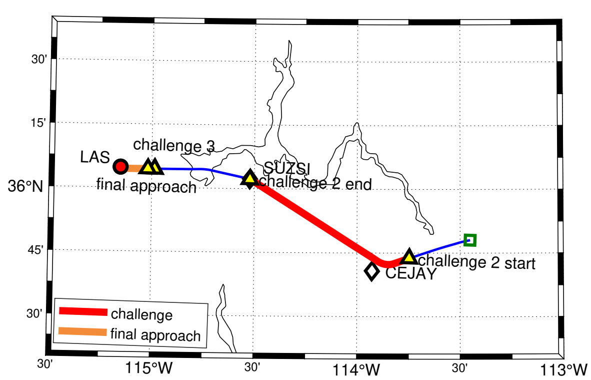

Fig. 3 Map view of Scenario 4.

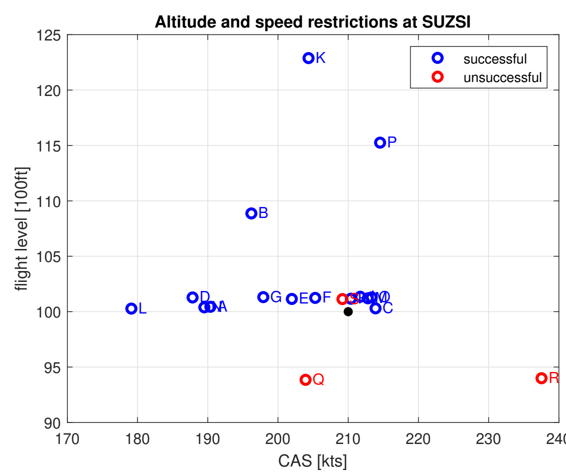

Fig. 4 Individual pilot performance of Scenario 4, Challenge 2.

B. Flap Extension

Scenario 2, Challenge 4

The aircraft was on a flight plan from Boston (BOS) to Dulles (IAD) via the HYPER7 arrival route with RNAV Y approach to runway 19L in Scenario 2. In Challenge 4, as the airplane was slowing for the approach, the PF intentionally called for flaps 25 too early, when the airspeed was still around 190 kts. Flaps 25 should normally be selected around 170 kts. Ideally, the PM is aware of the imminent upper speed limit violation (marked on the airspeed tape) and waits. A map view of Scenario 2 showing the duration of Challenge 4 for a representative pilot is provided in Fig. 5. The challenge started when the PF asked the PM for flaps 25 and ended when the PM moved the flap handle. Fig. 6 shows the buffer to overspeed at flaps 25 for each individual pilot. The zero line represents the maximum flap speed of 170 kts. Positive values mean a positive buffer with respect to the speed limit, which means that flaps 25 were selected at a lower speed than 170 kts. Negative values depict a negative speed buffer and thus an overspeed violation. In general, successful pilots have a positive speed buffer and unsuccessful pilots a negative one. There is one exception. Pilot O did mention the imminent overspeed violation to the PF and held off moving the flap handle for a few seconds, which puts him in the blue group of successful pilots. However, he or she was slightly early in moving the flap handle, which still resulted in a very brief overspeed violation, which lasted only for a few seconds thanks to a significant deceleration rate.

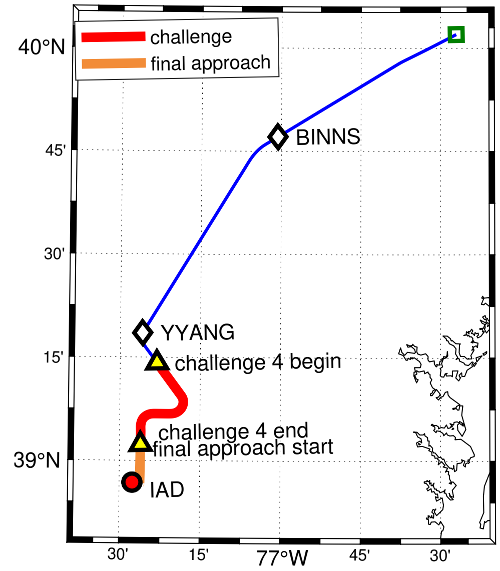

Fig. 5 Map view of Scenario 2.

Scenario 4, Challenge 3

In Scenario 4, the aircraft was following a flight plan from New York (JFK) to Las Vegas (LAS) via the TYSSN5 arrival route with ILS approach to runway 26L. In Challenge 3, in vectoring for the approach, ATC asked the airplane to slow down to 170 kts for spacing and to proceed to the next waypoint and onward towards localizer intercept. The PF intentionally failed to call for flaps 5 as the airplane slows to that flap speed. Fig. 3 shows the map view of Scenario 4 highlighting Challenge 3. The challenge started when the PF entered 170 in MCP speed window and ended when the PM moved the flap handle. For the representative pilot shown here, the duration of Challenge 3 is very short, which results in very closely located challenge start and end markers in Fig. 3. Fig. 6 shows the buffer to minimum speed at flaps 5 for each individual pilot. The zero line represents the minimum flap speed of 170 kts. Positive values mean a positive buffer with respect to the speed limit, which means that flaps 5 were selected at a higher speed than 170 kts. Negative values depict a negative speed buffer and thus a minimum speed limit violation. Just like in the previous challenge, successful pilots have a positive speed buffer and unsuccessful pilots a negative one.

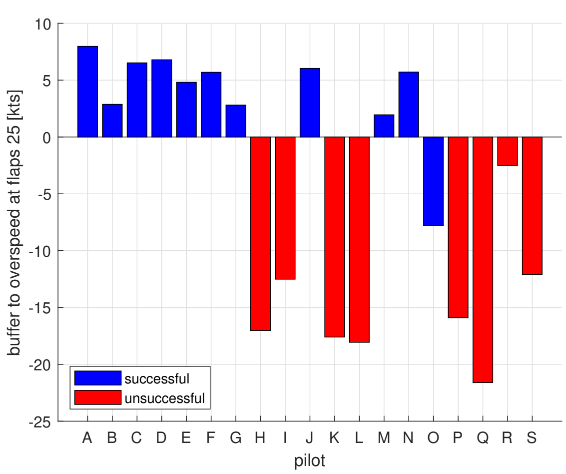

Fig. 6 Individual pilot performance of Scenario 2, Challenge 4

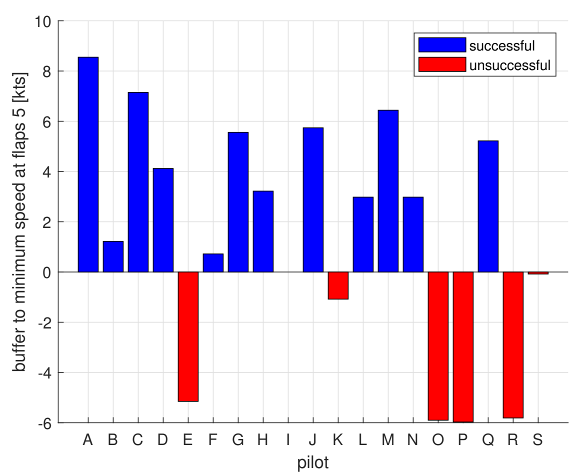

Fig. 7 Individual pilot performance of Scenario 4, Challenge 3.

III. Method

A. Independent Variables

The experimental design used two within-subjects variables: whether a scenario was presented before or after training, and whether the flight director (FD) was present or absent for the PM. One of the scenarios before and one after training had the FD on and the FD was off for the other two. Order was a counterbalancing between-subject factor with four orders used (Table 1). Note that scenarios 1 and 3, and 2 and 4 were similar in nature (see Section II). We intended to run 24 participants with six participants in each order in Table 1 but were able to recruit only 19 participants. Five participants were assigned to Orders 1, 2, and 3 and only four participants were assigned to Order 4.

Table 1 Four between-subject counterbalancing orders of the within-subject variables.

Order 1	Order 2	Order 3	Order 4
Scenario 1/ FD on	Scenario 1 / FD off	Scenario 3 / FD on	Scenario 3 / FD off
Scenario 2/ FD off	Scenario 2 / FD on	Scenario 4 / FD off	Scenario 4 / FD on
Training	Training	Training	Training
Scenario 3/ FD on	Scenario 3 / FD off	Scenario 1 / FD on	Scenario 1 / FD off
Scenario 4/ FD off	Scenario 4 / FD on	Scenario 2 / FD off	Scenario 2 / FD on

B. Apparatus

The study was conducted at the training center of a major US airline, using their training rooms, simulator staff, and simulator. The flight simulator was a CAE 737-700 full-flight simulator used in its standard configuration for the airline. The simulator had a collimated out-the-window visual system; however, pilots were mostly flying in the clouds without any visual references. The motion base was not used during the study. The simulator was set to record 231 variables with a sampling frequency of 6 Hz. The variables included flight variables, control inputs, and cockpit settings and alerts.

The simulator was fitted with a Seeing Machines single-crew configuration Crew Training System. The Crew Training System is a stand-alone single and/or multi-crew eye-tracking solution designed to support flight crew training applications in the Full Flight Simulator (FFS) or Flight Training Device (FTD) environments. The system records crew visual scanning behavior against flight instruments and the cockpit environment. Only one eye-tracking system was installed in front of the right seat where the PM was sitting. The system used a single camera and two separate infrared emitters. The camera was placed above the PFD with the respective infrared (IR) illuminators placed left and right of the camera. The approximate locations of the camera and illuminators are depicted in Fig. 8. The system was calibrated prior to the study but not for each pilot individually. Fifteen eye-tracker variables were collected at 60 Hz. Variables included pilot attention state, eye gaze, and eye-tracker diagnostics and time stamps.

Fig. 8 depicts the eight areas of interest (AoI) that were defined in the eye-tracking software: primary flight display (PFD), navigation display (ND), out-the-window visual (OTW), mode control panel (MCP), electronic flight instrument system (EFIS) settings panel, flight management system (FMS), and upper and lower engine-indicating and crew-alerting system (EICAS) displays.

A single video camera was positioned over the left shoulder of the PM and afforded a view of most of the flight deck interface, with more focus on the PM's instruments. The data from the simulator, audio, video and eye-tracking were all time-synchronized through a network time protocol (NTP) server. Both the simulator and eye tracker logged the synchronized universal coordinated time (UTC). The video was overlaid with the synchronized UTC.

C. Participants

Nineteen first officers (FOs) who were active and current on the 737NG participated in the experiment. Pilots were recruited based on joining the airline in the last five years; we believed that these less-experienced pilots might be more likely to benefit from additional training on monitoring. Pilots typically had previous experience flying airplanes at a regional airline or in the military, and about half had experience as Pilot in Command. Total flight hours ranged from 400 to 14,000. Two pilots had both military and regional carrier flight experience.

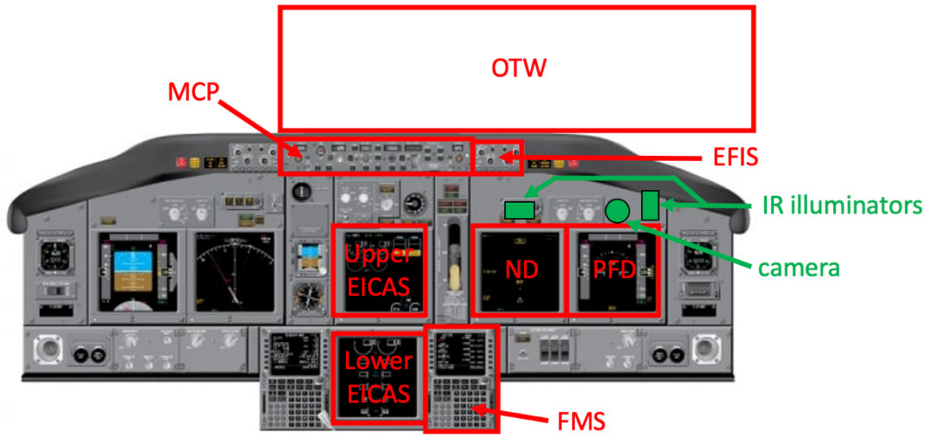

Fig. 8 Areas of interest defined for eye tracking and location of eye-tracking equipment: primary flight display (PFD), navigation display (ND), out-the-window visual (OTW), mode control panel (MCP), electronic flight instrument system (EFIS) settings panel, flight management system (FMS), and upper and lower engine-indicating and crew-alerting system (EICAS) displays.

D. Procedures

The participant was always the PM and was seated in the right seat. The PF, always the captain and in the left seat, and the instructor pilot (IP) were confederates who were familiar with the research goals and scenarios and had practiced the scenarios prior to the sessions. The PF took a more passive role than is normal in line operations and relied heavily on prompts from the PM to fully manage the flight path. As part of the scenarios, the PF made scripted, intentional errors, such as selecting a wrong mode or requesting an inappropriate flap setting. The IP sat behind the flight crew at the simulator controls and initiated and managed the simulator scenario. The IP also issued all ATC clearances.

A session for a single participant was scheduled for 3.5 hours, with 3 hours of available simulator time. There were five phases including two 1-hour simulator sessions plus a short break. First, the participating FO met the Trainer in a briefing room for an individual briefing. Participants were told that the study investigated how pilot activities contribute to safety; that is, the specific study focus on pilot monitoring was not mentioned. The second phase was a simulator session of an hour in which pilots flew two scenarios (see Section II). During each scenario an experiment observer, in coordination with the IP, recorded the start and outcome of each monitoring challenge and any comment made by the PM. Phase three was an hour-long training session that was aimed at improving pilot monitoring using a sensemaking approach [2]. The participant followed a set of slides on a laptop computer, which structured the activities, the information presented, and the interaction with the trainer. Participants were offered a short break before or after the training session. The remaining two scenarios were run in the second hour-long simulator session (phase four of the experiment). These scenarios were performed using the same procedure that was used for the first two scenarios. Finally, in phase five, pilots were debriefed and answered a questionnaire about the study.

E. Dependent Measures

Two dependent measures linked to attention and to situation awareness were derived from the eye-tracking data. Dwell Time was calculated by aggregating time while the gaze vector intersected with a certain AoI during a monitoring challenge. Proportion Dwell Time was the Dwell Time percentage of the total duration of the challenge for a specific pilot. As opposed to Dwell Time, Proportion Dwell Time controls for variation in how long the relevant flight segment took.

It might be that the recency of information sampled and not the amount of time looking at a display is the important link from eye fixations to performance. Thus, we also determined AoI Neglect Latency, the time between moving fixation away from an area of interest and again fixating that area as determined by the intersection of the gaze vector with an AoI [5, 7]. This length of time between attention shifting from a particular AoI and returning to it could be an important indicator of how current the understanding of the situation was as it pertains to information presented only in that AoI.

F. Hypothesis for Flight Director Monitoring

The study had two main objectives. The first objective was to investigate if a novel training strategy using a sensemaking cycle to update the aircraft Situation Model could improve pilot monitoring. Although eye-tracking data were collected for the entire experiment, it was not the emphasis for this objective of the study and no hypotheses about eye-tracking measures were defined beforehand. The second objective was to determine if having the FD on or off in the final approach segment of the flight would affect pilot monitoring. This second focus was of particular interest to the airline involved in the experiment. The airline hypothesized:

- With the flight director engaged on the PFD of the PM, the proportion of time that the pilot fixated the PFD (Proportion Dwell Time) will be higher than when the FD was not engaged.

This hypothesis was based on the professional insight of several instructor pilots of the airline. Because this was a hypothesis specified in advance, inferential tests of significance are appropriate.

G. Data processing

This paper uses the eye-tracking data from the experiment. However, the simulator and video data were used to determine the start and stop times of the different challenges for each pilot and if pilots were successful or unsuccessful solving the challenge. A web-based tool written in Angular was specifically designed for this study to support data integration and visualization of all three data streams. The tool provided animated playback of the simulator data displayed on a representation of the flight deck instruments and synchronized this with eye fixation sequences projected onto the eye-tracking areas of interest (Fig. 8) and with the over-the-shoulder video of pilots' activities. The first step in aligning the data was to resample the simulator data to the sampling rate of the eye-tracking data from 6 Hz to 60 Hz using the synchronized UTC time in both data streams. Next, the video data were aligned in the software tool by applying an offset based on the difference between the UTC time overlaid on the video and the UTC time in the simulator/eye-tracking data.

The start and stop times for each challenge were determined by playing back the data for each scenario using the software tool, and by analyzing the simulator data in MATLAB. Start times for each challenge were defined by the time the PF or the IP introduced the error (e.g. giving an erroneous instruction to the PM) or at a specific point in the flight plan (e.g. T/D). The challenges ended when the PM verbally expressed concerns or performed an action to solve the error, or at a specific point in the flight plan. The definitions of the start and stop times for each challenge are provided in Section II.

The eye-tracking data from two pilots were not used in the analysis. The eye tracker software produced a quality measure which was significantly lower for these pilots. In addition, the eye tracker was not able to detect where the pilot was looking for large portions of the scenarios. In addition, one pilot did not complete scenario 4 of the experiment due to time constraints.

Fixations and saccades were identified using the eye-tracker gaze data using a simple velocity-based algorithm. Using the distance of the pilot to the AoI, the point-to-point velocity between two gaze points was calculated. Each gaze point was then classified as a fixation if the velocity between points was below 100 deg/s and a saccade if it was above this threshold [8].

Finally, to test the hypothesis, we used a linear mixed model analysis, specifically the lmer routine of the lmer4 package in version 4.0.3 of R. This modeling approach can pull out variability from many, diverse factors, including variability from scenarios and from people. The approach can manage missing data effectively and provides a very sensitive method of analysis. Conceptually, we measured the fit of a model that includes a factor of interest as a predictor, here, the FD on/off factor. Then, the fit of that model was compared with the fit of a model which does not include the effect of FD Status but was otherwise identical to determine whether including the effect of FD was statistically significant. The models included the random effect of individual pilot and the fixed effect of scenario. A significance level of 0.05 was used.

IV. Eye-Tracking Results

This section provides the eye-tracking results of the four scenario challenges and final approaches discussed in Section II. Data from 17 of the 19 pilot participants with usable eye-tracking data were included.

A. Flight Director Use

A key hypothesis guiding the study was the possibility that, in the final approach, having the FD on would draw an inappropriately large share of attention (higher Proportion Dwell Time) to the PFD and more specifically, the FD, and reduce attention elsewhere, possibly impacting awareness of other elements of the approach. A prediction about the AoI Neglect Latency might be seen as an extension or refinement of the prediction about Proportion Dwell Time, in the original hypothesis.

Proportion Dwell Time (PDT) with the FD on or off is shown in Table 2. As can be seen, the trend was in the opposite of the predicted direction; that is, a higher PDT was observed when the FD was off, not when on. Using linear mixed models, we tested whether Proportion Dwell Time on the PFD differed significantly with FD on versus off. The trend observed in Table 2 was marginally significant ($\chi^2(1) = 3.65$, $p = 0.056$). In sum, there was a non-significant trend in the opposite direction to that hypothesized, with a longer Proportion Dwell Time on the PFD when the FD is off.

Table 2 PFD Proportion Dwell Time with the flight director on or off in the final approach.

	Proportion Dwell Time, s Mean (Standard Deviation)		
	Scenario 2	Scenario 4	Mean by FD Status
FD Off	0.51 (0.08)	0.47 (0.11)	0.49 (0.10)
FD On	0.44 (0.10)	0.38 (0.13)	0.41 (0.12)
Mean by Scenario	0.48 (0.10)	0.43 (0.13)	0.45 (0.11)

Table 3 PFD Neglect Latency with the flight director on or off in the final approach.

	AoI Neglect Latency, s Mean (Standard Deviation)		
	Scenario 2	Scenario 4	Mean by FD Status
FD Off	1.82 (0.37)	2.31 (0.52)	2.08 (0.51)
FD On	2.14 (0.61)	2.42 (0.58)	2.27 (0.59)
Mean by Scenario	1.98 (0.52)	2.36 (0.53)	2.17 (0.55)

Turning to the effect of AoI Neglect Latency (ANL), Table 3 shows the average neglect latency with and without the FD. The neglect latencies were similar with longer gaps before returning to the PFD when the FD was on (not when off, as predicted). Note that because there were many AoIs, the time spent fixating one area (PDT) and the gap between fixations (ANL) are measuring very different things: one could be high and the other low, or conversely. The effect of FD on the PFD AoI Neglect Latency was tested by comparing the fit of a model with versus a model without the FD factor. The procedure of the statistical test was the same as for Proportion Dwell Time. The mixed model analysis found the effect of the flight director on Neglect Latency was not significant ($\chi^2(1) = 1.23$, $p = 0.27$).

Scenario was not hypothesized in advance to affect either Dwell Time or AoI Neglect Latency. Therefore, the inferential status of significance testing for either PDT or ANL is not the same as testing for the effect of FD status. However, the relatively large contribution of Scenario to the models for each dependent measure prompted assessing the effect of Scenario on each. We used the same analysis procedure as for assessing the effect of FD status. The effect of Scenario was significant and very similar for both Proportion Dwell Time ($\chi^2(1) = 5.73$, $p = 0.018$) and Neglect Latency ($\chi^2(1) = 5.74$, $p = 0.017$). Lower Proportion Dwell Times and higher AoI Neglect Latencies were found for Scenario 4 (approach to KLAS) compared to Scenario 2 (approach to KIAD).

This suggests that there was systematic variability in eye-tracking measures in our data, and thus that our data were sensitive enough to detect some effects. In turn this suggests that the lack of significant effects of FD status was not simply due to very noisy data.

B. Monitoring Challenges

The analysis in this section is not based on hypotheses and is exploratory in nature; therefore, no statistical analyses were conducted. This section presents data from the four monitoring challenges discussed in Section II. Challenge performance data for all 19 pilots is also provided in Section II.

Fig. 9 provides Proportion Dwell Time for all eight AoIs and 17 pilots in every challenge. The bars for a particular pilot do not necessarily add up to 100% as pilots do not look at any AoI for some portions of a challenge or eye-tracking data were not available. Means per AoI across pilots are provided by red dashed lines. The plots on the left provide results for pilots who successfully completed the challenges, while the plots on the right provide results for pilots who were unsuccessful. This division in eye-tracking data was made as we expected successful pilots might allocate their gaze in a more structured or efficient way than unsuccessful pilots allowing them to more quickly detect problems. Note that more pilots were successful than unsuccessful and the number of unsuccessful pilots is particularly small for some challenges. This means one should be cautious drawing strong conclusions from the differences between successful and unsuccessful pilots presented here. Even though pilots performed challenges before and after training, results were not subdivided accordingly as the training did not focus on changing pilots' scan patterns or gaze allocation and initial analysis did not reveal differences in eye tracking before and after training.

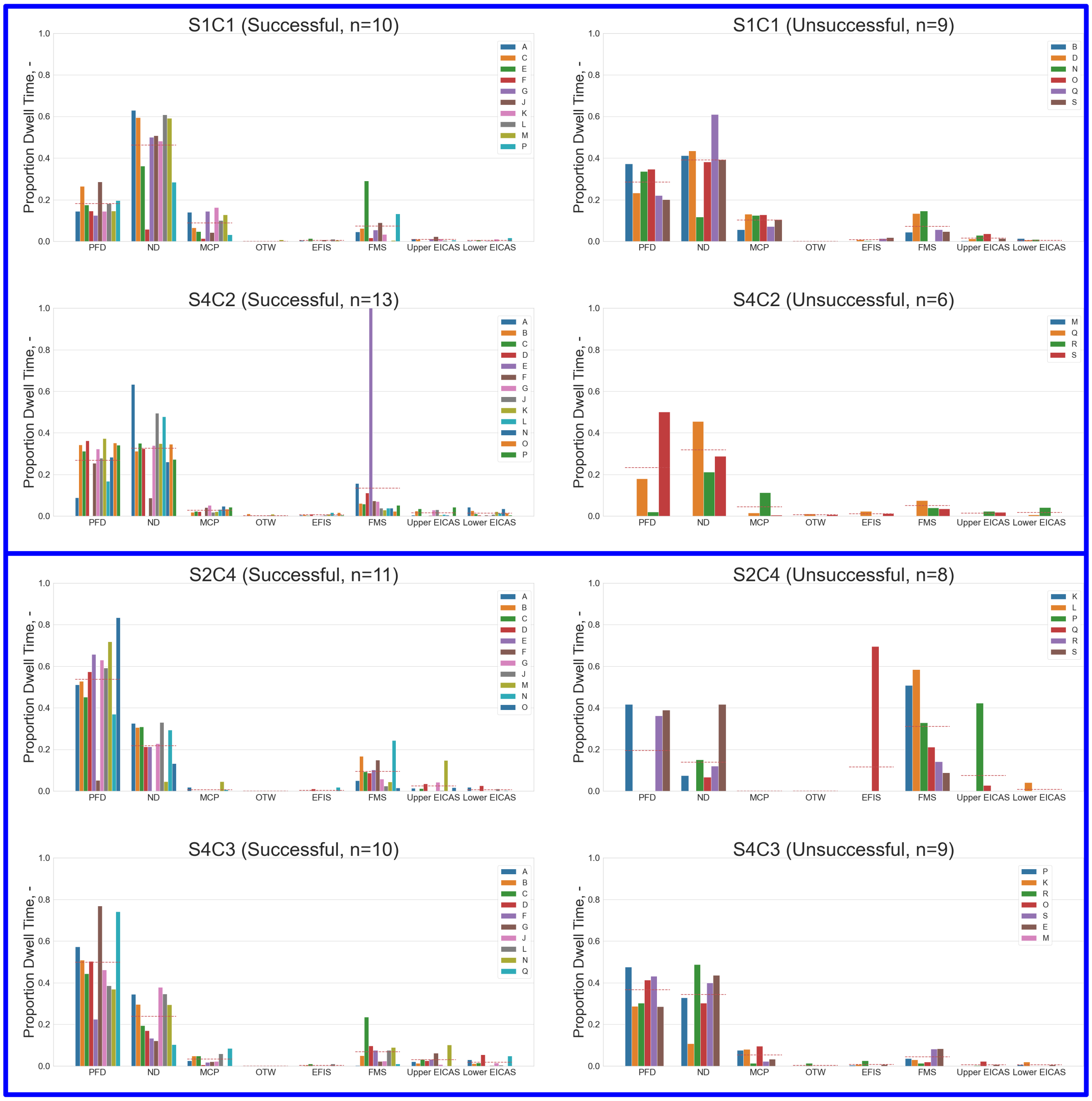

Fig. 9 Proportion Dwell Time for the four monitoring challenges considered in this paper. S1C1 and S4C2 are challenges to meet waypoint restrictions and S2C4 and S4C3 are challenges to extend the flaps at the appropriate time.

The top four plots in Fig. 9 provide PDT for the challenges with waypoint restrictions (Section II.A). Most of the pilots' attention was on the PFD and ND. Pilots had the use of a vertical situation display (VSD) on the ND providing information on their vertical flight path. This is a valuable source of information in the waypoint-restriction challenges and, as such, we were anticipating higher Proportion Dwell Times for the ND. Fig. 9 indicates mean PDT was much higher on the ND compared to all other AoIs for Scenario 1, Challenge 1 (S1C1). However, this is not the case for Scenario 4, Challenge 2 (S4C2). A possible explanation for this is that the altitude restriction was harder to meet for S1C1 compared to S4C2 as indicated by the much larger range in altitudes when crossing the waypoint in Fig. 2 compared to Fig. 4. The airspeed restriction was harder to meet for S4C2 compared to S1C1 which might have resulted in a higher PDT for the PFD, more similar to the PDT for the ND, for S4C2 compared to S1C1.

Proportion Dwell Times were relatively low for all other AoIs. In two cases for S4C2, the PDT for the FMS was 100%. This means that for the entire duration of the challenge the pilot was looking at the FMS. Note that PDT doesn't imply a length of time, so it could have been the challenges for these pilots were shorter. Inspection of the data suggests successful pilots may spend more time on the AoIs most relevant to complete the challenge; that is the ND for S1C1 and the PFD and ND for S4C2.

The bottom four plots of Fig. 9 provide PDT for the challenges concerning flap extensions (Section II.B). For these challenges, flap extension was restricted by airspeed and thus monitoring airspeed on the PFD was most important. Fig. 9 indicates mean PDT was highest on the PFD compared to all other AoIs for both Scenario 2, Challenge 4 (S2C4) and Scenario 4, Challenge 3 (S4C3). The ND is the second most focused on AoI for pilots who successfully completed the challenge, followed by the FMS. Inspection of successful and unsuccessful pilots suggests that unsuccessful pilots focused on the PFD less. Furthermore, unsuccessful pilots focused on the FMS much more in S2C4 and one unsuccessful pilot more on the EFIS settings.

Looking at Fig. 9 as a whole, an interesting observation can be made. Successful pilots seem to adapt their monitoring strategy more to the challenge at hand as observed by a distinct shift in focus on either the PFD or ND between challenges. This does not seem to be the case for unsuccessful pilots for which the PDT distribution between the PFD and ND was more similar between challenges.

Fig. 10 depicts AoI Neglect Latency for all 17 pilots in the four challenges considered here. Means per AoI across pilots are provided by red dashed lines and vertical lines provide the standard deviation for each individual on the AoI. Note that standard deviations are relatively large indicating that Neglect Latencies varied considerably per pilot during the challenges. As pilots focused most on the PFD and ND (see Fig. 9), ANLs for only those AoIs are presented here. Again, plots on the left provide results for successful pilots, while the plots on the right provide results for unsuccessful pilots. The top four plots provide ANL for the challenges with waypoint restrictions (Section II.A). It can be observed that for S1C1, the Neglect Latency for the ND is lower than for the PFD. This does not seem to be the case for S4C2. Fig. 10 indicates that ANL might be slightly higher for unsuccessful than successful pilots.

The bottom four plots of Fig. 10 provide ANL for the challenges concerning flap extensions (Section II.B). PFD and ND Neglect Latencies were similar to each other for both successful and unsuccessful pilots. A very low ANL or even values of zero can be observed for unsuccessful pilots of S2C4. Unsuccessful pilots moved the flap handle right after the PF's request to change flaps instead of waiting for the appropriate airspeed resulting in very short challenge durations. Finally, comparing all challenges might indicate again that successful pilots adapted their monitoring strategy more deliberately depending on the challenge. This trend in ANL is far less clear as compared to that in the PDT in Fig. 9.

V. Discussion

Nineteen pilots participated in a pilot monitoring training experiment. Each participant was the pilot monitoring in four operational scenarios in a B737-700 full flight simulator. Eye-tracking data were collected for the entire experiment; however, eye-tracking data were only the primary focus for testing an hypothesis on flight director use during the final approach for the two scenarios where the aircraft landed (Scenario 2 and 4). During the final approach PMs tended to focus less on the PFD with the FD on (lower Proportion Dwell Time) as compared to when the FD was off. This was a trend in the opposite direction of what was hypothesized; that is, the hypothesis was rejected. No significant effects were found for the AoI Neglect Latency. We cannot identify an obvious explanation for the lower PDT with the FD on. One possible explanation might be that with the FD off pilots had to focus longer on the PFD to get an understanding of the current and future state of the aircraft. Note that, as the PM, participants were not providing any control inputs to the aircraft.

The main focus of the experiment was pilots' ability to solve monitoring challenges before and after training on the sensemaking cycle approach to monitoring, which relies on developing a Situation Model. Each of the four scenarios consisted of three or four challenges. Eye-tracking data from two challenges where pilots had to meet waypoint restrictions and two where pilots had to extend flaps at the appropriate time were analyzed in this paper as they had clear start and end times and expected AoIs for eye tracking. Pilots focused more on the PFD and ND compared to other AoIs for all challenges. Inspection of Figs. 9 and 10 suggests that pilots who successfully completed the challenges focused more on the AoI that contained the most relevant information to successfully complete the challenge: ND for S1C1, ND and PFD for S4C2, and PFD for S2C4 and S4C3. Unsuccessful pilots did not seem to have a clear shift in focus on particular AoIs between different challenges. Note that pilots were not given any instructions about when or where to look at any time during the experiment.

An increasing number of new flight simulators and training devices is equipped with eye-tracking technologies suggesting there will be an increasing interest in whether and how eye tracking can be used effectively in training. Our data did not find, nor was the study designed to discover, specific scan patterns or attention allocation contributing to success on individual monitoring challenges. Our data do suggest that successful pilots allocate their gaze differently for

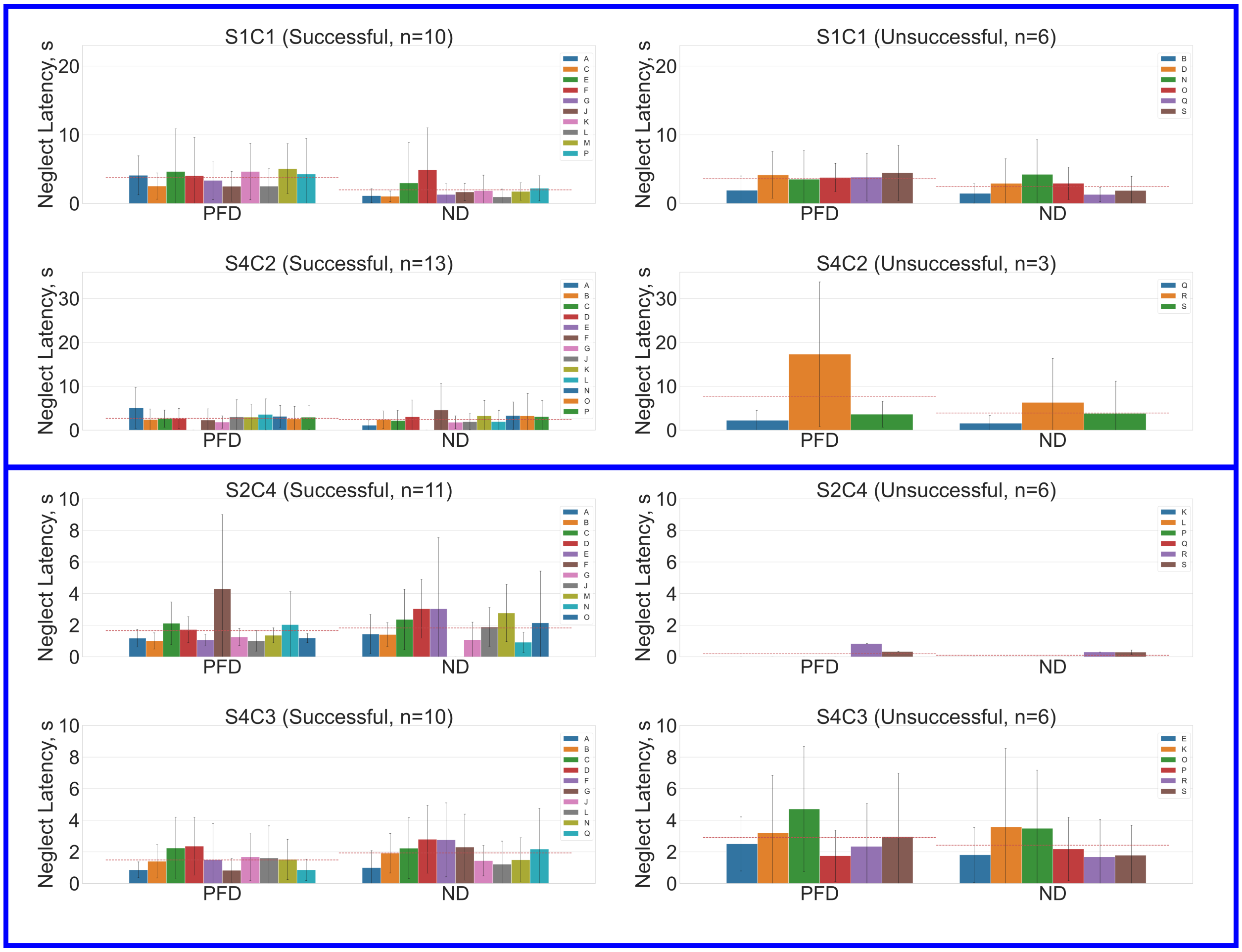

Fig. 10 AoI Neglect Latency for the four monitoring challenges considered in this paper. S1C1 and S4C2 are challenges to meet waypoint restrictions and S2C4 and S4C3 are challenges to extend the flaps at the appropriate time.

different challenges, at least for challenges where performance depends heavily on a specific display. This suggests that part of successfully solving problems is to deliberately look for the information pertaining to those problems, instead of following a standard scan pattern, which is in line with the implications of the sensemaking model of monitoring [7]. Note that very few pilots in our study spent the majority of their time looking at any one AOI. More research would be required to determine whether and where there are patterns of attention distribution that contribute to success on specific types of situations, and what eye-tracking measures are most appropriate to consider.

This paper only investigated the eye-tracking measures with respect to the main AoIs as defined in the eye-tracking software. In follow on research it would be useful to look at specific areas on each AoI such as the speed or altitude tapes on the PFD. While this was technically possible with the data we collected; such AoI analysis would have required the pre-calibration of each individual pilot's eye gaze to ensure high precision eye-tracking data. Given this was not the core focus of this study, and also due to time constraints, the Seeing Machines calibration module was deliberately not supplied for the study to enable sub-AoI analysis. The eye tracker used in this study utilized one camera and two IR illuminators. In general, accurate results were obtained with respect to identifying when pilots were looking at the main AoIs; however, for at least two pilots the eye-tracking data were not sufficiently accurate. One of those pilots was wearing glasses, which can cause reflections that can disrupt the eye-tracking software. Issues like these can possibly be minimized with careful calibration for each individual pilot, and mitigated with further optimization of the camera and infrared installation and orientation, but will certainly need attention before eye tracking can truly be an integral part of pilots' training in flight simulators.

Finally, even though the eye-tracking data were the focus of this paper, it should be emphasized we analyzed three types of data simultaneously: video/audio, simulator, and eye-tracking data. We developed a custom application to integrate and synchronize these data using UTC time stamps. This was a nontrivial, important step to make sense of what pilots were doing and allowed us to separate successful from unsuccessful pilots. Integrating all available data is also an important aspect of utilizing eye tracking for future training and should receive adequate attention.

VI. Conclusions

This paper presents eye-tracking results from a flight-director-use and pilot-monitoring study with 19 pilots. Pilots completed 15 monitoring challenges across four operational scenarios in a B737-700 full flight simulator. For each pilot, the flight director was on for half of the scenarios and off for the other half.

It was hypothesized that turning the flight director off in the final approach would result in pilots focusing less on the primary flight display and more on other areas of interest. Our data disconfirmed this: during two final approach segments, PMs tended to focus less on the primary flight display with the flight director on, not off as hypothesized.

Eye-tracking data from four monitoring challenges were analyzed in this paper: two challenges involving waypoint restrictions and two involving extending the flaps at appropriate airspeeds. Pilots that successfully completed the challenges focused more on the areas of interest that contained the most relevant information to successfully complete the challenge. In addition, successful pilots seem to adapt their monitoring strategy more to the challenge at hand as observed by a distinct shift in focus on either the primary flight display or the navigation display depending on the challenge. In order to effectively utilize eye tracking in future monitoring training, more research is required to investigate what aspects of gaze adapt to a particular context, and what eye-tracking measures are most appropriate to consider.

Acknowledgments

The authors would like to thank the airline for allowing use of their full flight simulator and training facilities. We especially would like to thank the instructors, confederate pilots and simulator engineers who supported the experiment. We would like to thank Seeing Machines for supporting this study with their eye-tracking equipment and expertise. Finally, we are grateful for the 19 pilots who participated in the experiment in their free time.

References

[1] Airplane State Awareness Joint Safety Analysis Team, "Airplane State Awareness Joint Safety Analysis Team Interim Report," Tech. rep., Jun. 2014. URL `https://www.skybrary.aero/bookshelf/books/2999.pdf`.

[2] Billman, D., Mumaw, R. J., and Feary, M., "A model of monitoring as sense-making: Application to flight path management and pilot training," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 64*, 2020.

[3] Billman, D., Zaal, P., Mumaw, R., Lombaerts, T., Torron, I., Jamal, S., and Feary, M., "Training Airline Pilots for Improved Flight Path Monitoring: The Sensemaking Model Framework," *Proceedings of the 21st International Symposium on Aviation Psychology*, 2021, pp. 403–408. https://doi.org/10.5399/osu/1148.

[4] Dill, E. T., Young, S. D., Daniels, T. S., and Evans, E., "Analysis of Eye-Tracking Data During Conditions Conducive to Loss of Airplane State Awareness," *17th AIAA Aviation Technology, Integration, and Operations Conference*, 2017. https://doi.org/10.2514/6.2017-3277.

[5] Dehais, F., Behrend, J., Peysakhovich, V., Causse, M., and Wickens, C. D., "Pilot Flying and Pilot Monitoring's Aircraft State Awareness During Go-Around Execution in Aviation: A Behavioral and Eye Tracking Study," *The International Journal of Aerospace Psychology*, Vol. 27, No. 1-2, 2017, pp. 15–28. https://doi.org/10.1080/10508414.2017.1366269.

[6] Daniels, T. S., Ferguson, C. M., Dangtran, E. T., Korovin, R. M., Kramer, L. J., Evans, E. T., Santiago-Espada, Y., Kiggins, D. J., Etherington, T. J., and Barnes, J. R., "Regarding Pilot Usage of Display Technologies for Improving Awareness of Aircraft System States," *2019 IEEE Aerospace Conference*, 2019, pp. 1–11. https://doi.org/10.1109/AERO.2019.8741948.

[7] Mumaw, R. J., Billman, D., and Feary, M. S., "Analysis of Pilot Monitoring Skills and a Review of Training Effectiveness," Technical Memorandum TM–20210000047, National Aeronautics and Space Administration, Dec. 2020.

[8] Salvucci, D. D., and Goldberg, J. H., "Identifying fixations and saccades in eye-tracking protocols," *Proceedings of the symposium on Eye tracking research & applications - ETRA '00*, ACM Press, 2000. https://doi.org/10.1145/355017.355028.