Towards Intelligent Flight Deck – A Preliminary Study of Applied Eye Tracking in The Predictions of Pilot Success and Failure During Simulated Flight Takeoff

Bo Fu*, Peter Gatsby, Angelo Ryan Soriano, Kayla Chu, and Nicolas Guardado Guardado

California State University Long Beach, 1250 Bellflower Boulevard, USA

Abstract

Flight instruments provide pilots with many different indications, where a significant amount of information is processed via the human eyes. As such, there is an opportunity to capture knowledge from pilots' eye movements in an effort to inform machine learning models in the prediction of pilot success and failure. In the event of inferring a potential failure, an intelligent flight deck can then warn and alert the crew to take early action in order to prevent further errors and evade disasters. To this end, this paper presents a preliminary case study in a simulated flight take-off environment to investigate the feasibility of utilizing pilots' eye gaze when predicting pilot success and failure. The preliminary results indicate that accurate predictions of pilot performance can be achieved with at least 80% accuracy after only 20 seconds of analyzing pilots' gaze, highlighting the potential of realizing intelligent flight decks in the future. This finding suggests that it is likely feasible to support human-machine teaming in aviation, whereby the next generation of aircrafts can leverage artificial intelligence in the development of intelligent flight decks to optimize safety and pilot performance.

Keywords

Eye tracking, intelligent systems, human-machine teaming

1. Introduction

With automation being one of the main contributing factors of decreased aviation accidents over the last few decades, a drawback of automation is the over reliance and over confidence in automated systems that have potentially led to a lack of manual and active monitoring from the flight crew, whereby 60-80% of aviation accidents are reported to be caused by human error [1]. Effective and efficient monitoring plays a critical role in aviation safety particularly during dynamic phases such as takeoff and landing, where accurately observing various flight instruments and a pilot's ability to integrate many different sources of readings and visual cues remain a core component in decision-making. With the majority of such visual information being processed by the human eyes, there is an opportunity to investigate the feasibility of eye gaze in human-centered flight deck designs. Given that eye movements are relatively unique to an individual and are inherently unbiased and hyper-personalized much like fingerprints, eye

Proceedings of the 1st International Workshop on Designing and Building Hybrid Human–AI Systems (SYNERGY 2024), Arenzano (Genoa), Italy, June 03, 2024.

* Corresponding author.

Bo.Fu@csulb.edu; {peter.gatsby01; angeloryan.soriano; kayla.chu; nicolas.guardadoguardado}@student.csulb.edu

0000-0001-9874-9551 (B. Fu); 0009-0003-8345-8032 (A. R. Soriano)

tracking may potentially provide a useful source of input that can be leveraged by intelligent systems.

To this end, this paper presents a case study of applied eye tracking in the context of human-machine teaming in aviation. More specifically, we demonstrate the feasibility of utilizing eye gaze and established classification models to predict pilot success and failure in the realization of future intelligent flight decks. It is envisioned that the next generation of aircrafts will utilize knowledge captured from pilots' physiological data such as eye movements in the detection of potential human errors to support early warnings and early actions in the prevention of aviation accidents. In order to contribute to research in "how to" alert the crew, we must know "when" to alert as a first step. As such, this paper presents a proof-of-concept case study involving 17 participants taking off in a Cessna 172 aircraft on the X-Plane 11 simulator in a controlled lab environment, demonstrating the feasibility of gaze-based pilot predictions with success as early as a few seconds after task initiation. The experimental results show that without specialized training models or configurations, a number of established classifiers yield 80% and higher accuracies in their predictions of pilot success and failure after only 20 seconds of analyzing eye gaze. This result showcases the feasibility of applied eye tracking in this particular use case and is highly motivating for the development of more sophisticated systems in the future. Furthermore, we report on a number of influential gaze measures that are most indicative of a pilot's success such as saccadic lengths and their variation, scan area and scanpath length.

2. Related Work

The Federal Aviation Administration recently established a requirement to incorporate enhanced training in pilot monitoring in existing air career training programs [2], and the European Bureau Enquêtes-Accidents recommended the use of eye tracking to improve studies of pilot monitoring [3]. This section provides a brief summary of selected work that focuses on gaze-based intelligence in the context of human-machine cooperation, more extensive reviews of eye tracking in aviation can be found in [4-6].

A prior study [7] has demonstrated the benefits of leveraging knowledge gained from eye movements in go-around maneuvers, where human errors such as critical flight-path deviations during cross-checking can be detected and can be useful to inform crew members in the subsequent training of visual attention allocation to improve safety. In the context of training pilots in commercial aviation [8], eye movement analysis has been shown to assist instructors to gain better knowledge in the identification of the specific types of errors made by the student pilots. This knowledge is helpful for the instructors to identify student pilots who may be struggling in the program and to adapt the program with more personalized and targeted training components. An eye tracking study involving 8 pilots to train linear support vector machine models was presented in [9] showing accuracies to classify flight phases can be improved from 51.6% to 64.1% when utilizing gaze patterns. In the context of identifying novice vs. expert pilots, a prior study [10] involving 16 licensed pilots and 16 novices found that visual information acquisition, gaze dispersion, and gaze patterns can be utilized to classify pilot expertise level. In an effort to assist flight instructors with better assessment of student pilots' visual attention spent during practice, visualizations of the student pilot's eye gaze have been developed to enhance evaluation during pilot training programs [11]. An eye tracking analysis of 10 licensed pilots [12] in a simulated takeoff scenario found that different scanning patterns employed by the pilots can lead to varying levels of success. Evidentially, eye tracking and gaze analytics can provide helpful information when assessing pilot performance and expertise, we therefore hypothesize that eye gaze can be utilized as a form of input (in addition to acting as an evaluation tool) to inform intelligent systems. More specifically, if it is possible to quantify a pilot's visual attention and implicitly infer this person's level of engagement and performance in a particular task, it may then be possible to implement intervention in the event of a predicted pilot failure. Given that notable efforts have focused on research that uses gaze in pilot assessment and personalized training, there is an opportunity to advance the state of the art in gaze-based intelligent systems for the flight deck. In order to realize the notion of system intervention, the first step is to recognize *when* to intervene, before determining *how* to intervene. As such, the contribution of this paper lies in the proposed approach to potentially solving *when* an intelligent flight desk should mediate, as we add to the existing body of knowledge regarding the feasibility of applied eye tracking in the predictions of pilot success and failure, through a simulated flight scenario.

3. Experimental Setup

The use case study shown in this paper simulates the necessary conditions in a controlled environment to collect eye gaze data from a mixture of novices and experts, whereby gaze-based predictive findings may inform the design and development of future intelligent flight decks such as when a user interface adaptation may be warranted. The participants' eye gaze was collected using an eye tracker throughout the take-off, which were then processed through off-the-shelf machine learning classifiers to predict pilot success and failure. The flight deck view shown to the participants remained static throughout the experiment, no dynamic areas of interest, or mouse to hold and drag the screen were allowed. Participants simply gazed at displays as how they normally would in the flight deck.

We recruited a total of 17 participants in the study, including 7 true novices with no prior experience using X-Plane (with backgrounds in healthcare, civil engineering, and education) and 10 licensed expert pilots (with private, commercial, and airline transport pilot licenses). We recognize there are diverse skill levels in the participant pool, and that ideally, a sample with only licensed pilots may ground the findings with more generalizability. Though in the context of aviation safety, detecting task failures are of greater importance than successes. As such, this sample population containing true novices provides us with the necessary experimental conditions, i.e., task failures, to enable the investigation on the feasibility of the proposed gaze-based predictions of task performance.

The participants were asked to complete a takeoff in a Cessna 172 aircraft on the X-Plane 11 simulator. Prior to a study session, novice participants completed a one-on-one hour-long tutorial. The tutorial includes explanations of the instruments in the flight deck, the controls and functions of the physical equipment, as well as step-by-step instructions on how to take off and level a Cessna 172 aircraft. Prior to beginning the task, each participant completed a 9-point calibration with the eye tracker to maximize the accuracy of the data collected. The distance from a participant to the eye tracker is approximately 65cm. The eye tracker is positioned approximately 40cm below the eye level and is angled upwards towards the participant's face.

During a study session, on-screen messages were displayed by X-Plane instructing participants to initiate takeoff by gradually increasing maximum throttle while simultaneously using the rudder pedals to maintain the aircraft's alignment with the runway centerline. When the airspeed indication showed a speed of 55 knots, the participant would pull back the flight yoke to lift the aircraft off the ground. Upon leaving the ground, the participant would need to maintain the best rate of climb in a Cessna 172 at 74 knots. If the airspeed was greater than 74 knots, the participant would need to raise the nose of the aircraft to reduce airspeed. If the airspeed is less than 74 knots, the participant would need to lower the nose of the aircraft to increase airspeed. Once the aircraft has successfully reached an altitude of 1,500 feet, the participant would need to reduce pitch and power to level off and sustain that altitude to complete the task and conclude the simulation. Eye gaze from each participant was collected throughout the study session from task initiation until task completion. Figure 1 shows the physical setup of the following:

- The X-Plane 11 flight simulator ran on a Dell XPS workstation and a 24" full HD Monitor (1920×1080) at 60Hz with 8ms (gray-to-gray normal)/5ms (gray-to-gray fast) response time.
- The GazePoint GP3 HD eye tracker with 9-point calibration, 150Hz sampling rate, 0.5-1.0 degree of visual angle accuracy, 35cm (horizontal) x 22cm (vertical) movement, and ±15cm range of depth movement.
- The Logitech G Pro flight yoke, which controls an aircraft's ailerons and allows the pilot to adjust the roll and pitch of an aircraft in the simulation.
- The Logitech G Pro throttle quadrant, which powers the aircraft with a two-lever design (excluding the blue knob) was used in the study to simulate throttle (i.e., the black knob delivers full power when pushed upwards) and mixture (i.e., the red knob is fully activated when fully pushed in).
- The Logitech G Pro rudder pedals, which control the aircraft's rudder at the rear to adjust yaw in flight. Operated by feet, a pilot would push the left rudder pedal to turn left and push the right rudder pedal to turn right to control the aircraft in the simulation. The rudder plays a vital role during takeoff and landing when the aircraft is moving at low speed and needs extra assistance to maintain a straight and stable trajectory.

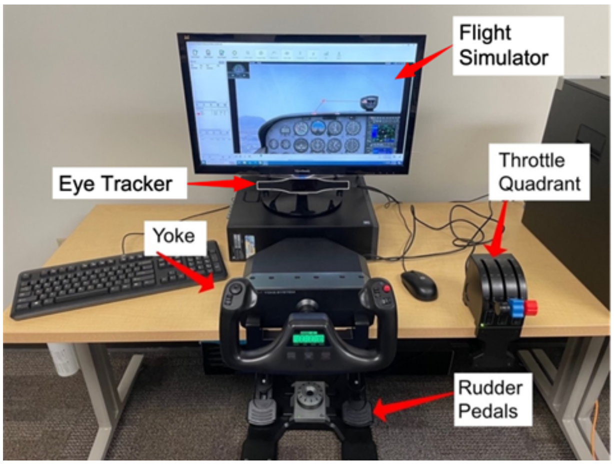
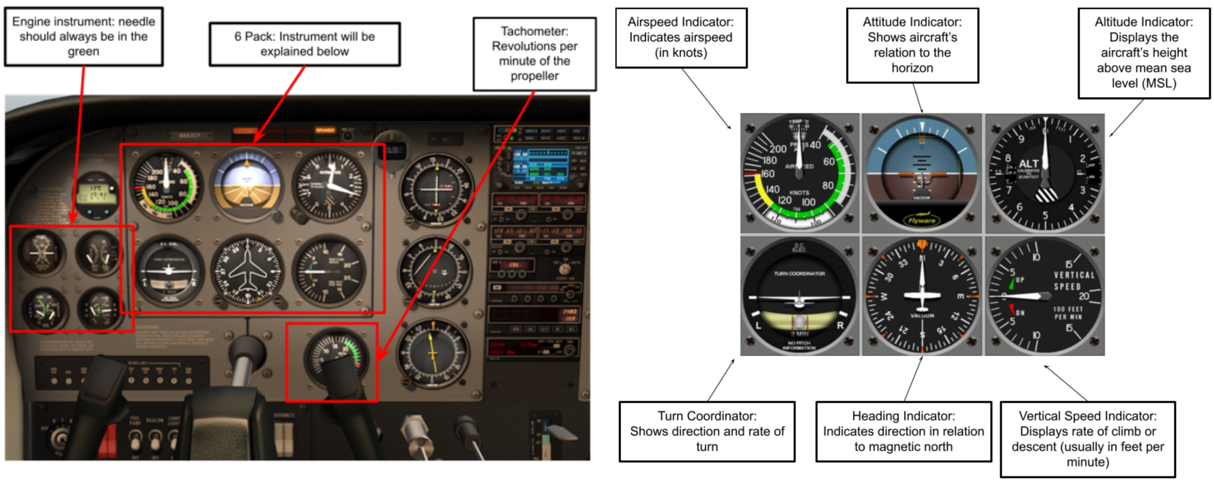


Figure 1: Physical configurations of the Logitech throttle quadrant, yoke, and rudder pedals on a X-Plane 11 simulator in front of a GazePoint GP3 HD eye tracker. Keyboard and mouse are used to complete the tutorial only, and are not used during task completion.

Figure 2 illustrates the flight deck view (Fig. 2a) and the main flight instruments (Fig. 2b) needed to takeoff a Cessna 172 aircraft on the X-Plane 11 simulator. The main flight instruments include the airspeed indicator, attitude indicator, altimeter, turn coordinator, heading indicator, and vertical speed indicator. The airspeed indicator presents the aircraft's speed in knots, while the attitude indicator depicts the aircraft's orientation relative to the horizon. The altitude indicator conveys the aircraft's elevation above mean sea level, and the heading indicator communicates the aircraft's direction in relation to magnetic north. The vertical speed indicator shows the climb or descent rate, typically measured in feet per minute.



a) Flight Deck View b) The six pack

Figure 2: Physical configurations of the Logitech throttle quadrant, yoke, and rudder pedals on an X-Plane 11 simulator in front of a GazePoint GP3 HD eye tracker.

4. Results & Findings

Based on the raw gaze data provided by the eye tracker, we determined various descriptive gaze measures for each participant using established eye tracking metrics [13]. A summary is presented in Table 1. *Fixations* refer to those moments where the eyes are relatively stationery, and *saccades* refer to the quick eye movements between successive fixations. Descriptive gaze measures included the number of fixations and saccades, saccadic lengths (measured in px between successive fixations) and duration (measured in ms), convex hull area (measured in px^2 of the region that all fixations fall within), and the overall scanpath length (measured in px of all saccades found in an interaction) captured during an interaction, which are typically considered as information search activities. We also generated fixation durations (measured in ms) and saccade-to-fixation ratios (the proportion a pilot spent on searching versus processing information), which are typically considered as indicators of information processing activities. Lastly, we determined pupil dilations (measured in mm), eye movement directions (measured in degrees) as gradients of a saccade respective to the horizontal axis (absolute angles) and consecutive saccades (relative angles), which are typically considered as indicators of cognitive workload.

These descriptive gaze measures captured from each participant is then passed to a set of off-the-shelf classification models using the WEKA toolkit [14]. Since the task completion time varies among the participants and the predictions should not be based on one or two individuals (e.g., some participants took less than 5 minutes to complete the takeoff, while others took 15 minutes or more), we therefore analyzed the first 4 minutes of the eye gaze data because all participants were still actively engaged in the task at that time (the average time on task was approximately 6 minutes and 41 seconds). A performance score for each participant is automatically generated by the X-Plane 11 upon task completion that ranges between 0 and 100 (0 being complete failure and 100 being complete success). It is not possible to configure the specific parameters for the calculation of this performance score, though we speculate it may be determined based on a combination of factors such as the aircraft's heading, speed, altitude. We used a median split to categorize successful participants (as those above the median score) and unsuccessful participants (as those below the median score). The goal of the classifications is to predict whether a participant will fall into the successful group or unsuccessful group based on their gaze. We tested over 50 classification models in WEKA such as Decision Trees, Support Vector Machines, Neural Networks, and Logistic Regression, using stratified 10-fold cross validation for model evaluation and Bonferroni-corrected t-tests for statistical testing. In this paper, we report those classifiers with the highest prediction accuracies, against the Zero Rule classifier as the baseline that predicts the majority. We recognize that there are other classification models that may be chosen as the baseline, though since the goal is to validate the soundness of gaze-based predictions on pilot performance, and not to emphasize on creating the most effective learning models, this baseline is sufficient for the purpose of this case study. The overall aim to choose off-the-shelf classifiers with default configurations is to demonstrate that the proposed approach to predict pilot performance using eye gaze can succeed without specialized bespoke learning algorithms and thus likely to perform well at scale in real world applications.

Table 1
Descriptive Gaze Measures

Gaze Measure	Definition
Fixation Count	The number of fixation points found in an interaction.
Saccade Count	The number of saccades found in an interaction.
Saccadic Length (px)	The distance between successive fixations.
Saccade Duration (ms)	The duration of a saccade.
Convex Hull (px^2)	The size of an area defined by the bounding fixations found in an interaction.
Scanpath Length (px)	The distance of all saccades found in an interaction.
Fixation Duration (ms)	The duration of a fixation.
Saccade-to-Fixation Ratio	The sum of saccadic duration divided by the sum of fixation duration.
Pupil Dilation (mm)	The widening of the pupils.
Absolute Saccadic Angle (°)	The gradient of a saccade with respective to the horizontal axis.
Relative Saccadic Angle (°)	The gradient of two consecutive saccades.

We analyzed eye gaze every 20 seconds and chose this time interval empirically based on experimental trials of other frequencies. This approach aims to simulate an implementation of real-time predictive gaze analytics, as participants interact with various flight instruments to complete the takeoff and their eye gaze data would continue to grow in timed intervals. Figure 3 presents the top-performing classification models in the prediction of pilot success and failure, including Bayesian Network, Logistic Regression, Logit Boost, Random Committee of Decision Trees, and Hoeffding Tree that achieved prediction accuracies ranging from 67-80% compared to the baseline (at 35-65%). Notably, a number of classifiers achieved high prediction accuracies (up to 80%) after only 20 seconds of analyzing a participant's eye gaze. This result suggests that discerning eye movement differences between those successful and unsuccessful pilots may be captured and useful for predictions early on during an interaction. Furthermore, we performed Pearson correlation coefficient to identify those entities that are most influential to further improve prediction accuracy, including the median relative angles, standard deviation of saccadic lengths, and maximum saccade duration based on weak-high correlations (where r values were greater than 0.25). Figure 4 presents the predictions accuracies based on just the influential gaze measures (as outlined in Table 1), where the overall peak accuracy was improved to 83.5% using the same classifiers. Lastly, as shown in Figure 3 and 4, the prediction accuracies peaked several times after approximately 20 seconds, 2 minutes, and again after 4 minutes for a number of classifiers. We hypothesize that with the runway roll down taking approximately 20 seconds, the lift off and climb taking around 2 minutes, where most pilots would have begun to cruise after roughly 4 minutes, these peak values may be grounded in the differing gaze behaviors at these significant stages of a takeoff. In other words, there may be fewer discerning gaze patterns amongst the participants between these key stages of the task, leading to increased challenges when yielding more accurate predictions at other time intervals.

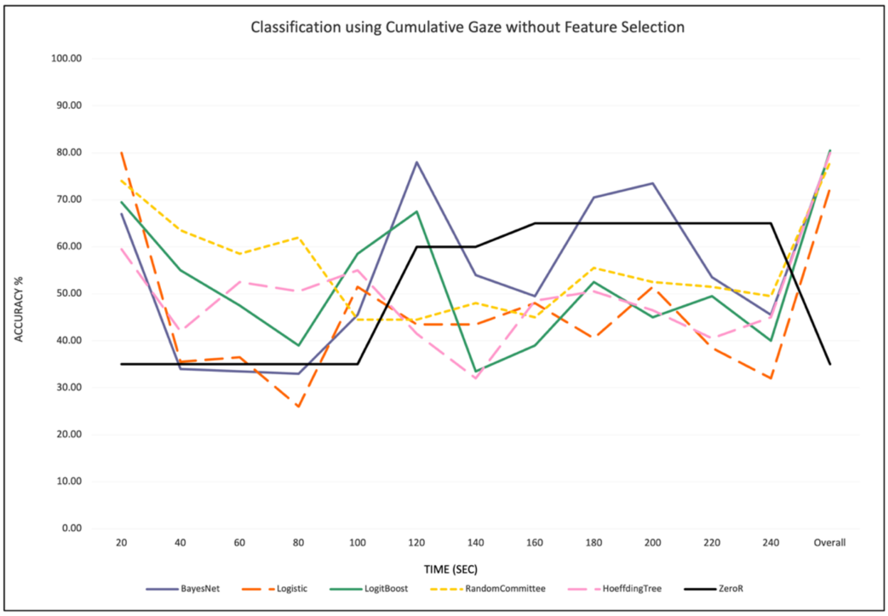


Figure 3: Predictions of task success and failure over time.

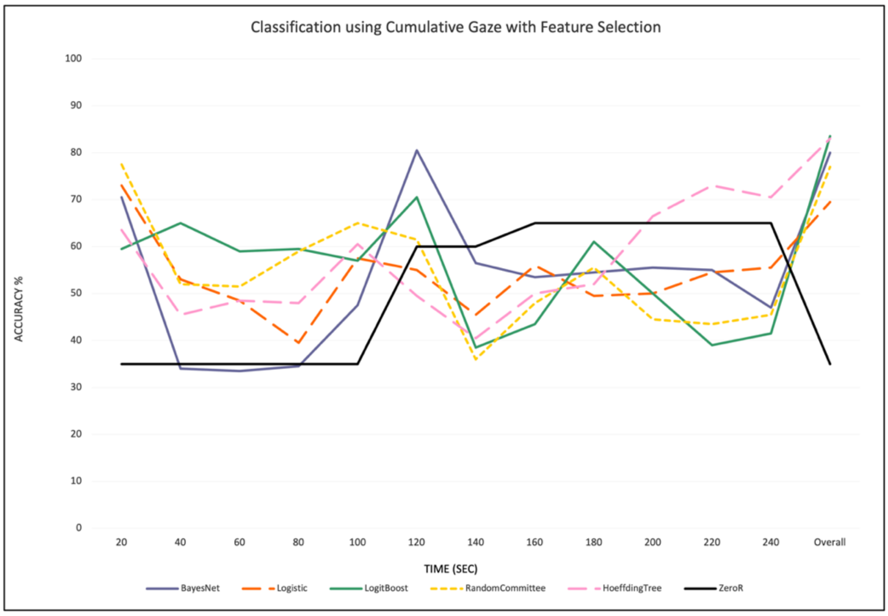


Figure 4: Predictions of task success and failure over time with feature selection.

Notably, as shown in Table 2, a combination of several gaze features was found to be associated with higher takeoff success, including less dispersed and longer saccadic length on average, longer lengths of one's entire scanpath, and larger areas scanned during the given task. This finding suggests that those pilots who were more successful in the given task exhibited more thorough and consistent scanning patterns, completed more search activities, as well as having scanned larger areas of the visual scene (e.g., outside views of the aircraft in addition to flight instruments). These influential gaze measures further suggest that it may be possible to evaluate a pilot's performance in real time by observing their gaze tendencies and make predictions on their potential failures to allow the integration of alert and warning systems in gaze-based intelligent flight decks.

Table 2
Influential Gaze Measures in Pilot Success and Failure Prediction

Style Tag	Frequency Comments
StDev of Saccadic Lengths	*low* value $\rightarrow$ *high* success
Mean Saccadic Length	*low* value $\rightarrow$ *high* success
Scanpath Length	*low* value $\rightarrow$ *high* success
Convex Hull Area	*low* value $\rightarrow$ *high* success

5. Conclusions & Future Work

This research investigates the feasibility of applied eye tracking in enhanced decision-making towards future intelligent flight deck. We demonstrate a proof-of-concept implementation of gaze analytics as helpful indicators to inform classification models in the predictions of pilot success and failure. It is envisioned that future intelligent flight decks will alarm the crew in the event of inferred failure during critical tasks such as takeoff and landing, where early warnings and early actions can be taken to prevent catastrophic events. The case study shown in this paper aims to demonstrate how accurate predictions of pilot success and failure may be generated with a number of off-the-shelf classification models without specialized configurations (after only a few seconds into an interaction, which can also be further improved over time), and eye gaze may serve as helpful resources when inferring success and failure. These findings are highly motivating for the development of more sophisticated predictive algorithms and more advanced user models to reveal innovative means to detect potential human errors based on a pilot's eye gaze in the flight deck.

It may be necessary to note that the findings shall be interpreted within the limitations of the study, whereby the behaviors of the participants may be grounded in their distinct piloting skills. For example, it may be possible that a participant with extensive experience playing video games could potentially perform well on an at-home simulator, but not as well in real-world settings. Also, gaze data from the first few minutes in the given task may contain discerning differences that become less or more prominent at later stage of the interaction. These factors have not been studied extensively and may be further investigated in future experiments. While the case study shown in this paper presents a first step towards realizing intelligent flight decks in the future, there are a number of research opportunities that may be explored in future work. Firstly, the participant pool contained a mixture of licensed pilots and true novices, while this sample is sufficient for the purpose of this preliminary study (i.e., providing the necessary conditions where both successes and failures are present in the scenario to validate gaze-based predictions of pilot performance), a sample population containing solely licensed pilots would likely improve the generalizability of the results and findings. Secondly, we used examples of basic machine learning techniques, and future work may focus on utilizing more sophisticated models to further increase prediction accuracies. Finally, in order to realize intelligent flight decks, while it is an essential step to first recognize those potential moments of intervention (i.e., *when* to provide assistance), it is equally important to determine the subsequent actions to implement (i.e., *how* to intervene). This paper presents a feasible solution answering the question of *when* to potentially intervene, but we have not tested means that may be effective in determining *how* to intervene. There are a number of research opportunities to gauge run time error prevention through system automation, fault tolerance, and alerti mechanisms that engage the pilot.

References

[1] S. Shappell, C. Detwiler, K. Holcomb, C. Hackworth, A. Boquet, D. Wiegmann, Human Error and Commercial Aviation Accidents: A Comprehensive, Fine-Grained Analysis Using HFACS, 2006. URL: https://commons.erau.edu/publication/1218

[2] Federal Aviation Administration (FAA). 14 CFR Part 121 Qualification, Service, and Use of Crewmembers and Aircraft Dispatchers; Final Rule; Federal Register: Washington, DC, USA, 2013; Volume 78, Issue 218.

[3] British European Airways (BEA). Study on Aeroplane State Awareness during Go-Around. 2013.

[4] S. Peißl, C. D. Wickens, R. Baruah, Eye-Tracking Measures in Aviation: A Selective Literature Review, The International Journal of Aerospace Psychology, 28:3-4, 98-112, 2018.

[5] D. Martinez-Marquez, S. Pingali, K. Panuwatwanich, R. A. Stewart, S. Mohamed, Application of Eye Tracking Technology in Aviation, Maritime, and Construction Industries: A Systematic Review. Sensors 21, 4289, 2021.

[6] L. Mengtao, L. Fan, X. Gangyan, H. Su, Leveraging eye-tracking technologies to promote aviation safety - A review of key aspects, challenges, and future perspectives, Safety Science, Volume 168, 106295, ISSN 0925-7535, 2023.

[7] F. Dehais, J. Behrend, V. Peysakhovich, M. Causse, C. D. Wickens, 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, pp.15-28, 2017. DOI: 10.1080/10508414.2017.1366269.

[8] D. Rudi, P. Kiefer, M. Raubal, The instructor assistant system (iASSYST) - utilizing eye tracking for commercial aviation training purposes, Ergonomics, vol.63, no.1, pp.61-79, 2020, PMID: 31646952. DOI: 10.1080/00140139.2019.1685132.

[9] V. Peysakhovich, W. D. Ledegang, M. M.J. Houben, E. L. Groen, Classification of flight phases based on pilots' visual scanning strategies. In 2022 Symposium on Eye Tracking Research and Applications (ETRA '22), ACM, New York, NY, USA, 2022.

[10] C. Lounis, V. Peysakhovich, M. Causse, Visual scanning strategies in the flight deck are modulated by pilots' expertise: A flight simulator study. PLoS ONE 16(2): e0247061, 2021.

[11] D. Rudi, P. Kiefer, M. Raubal, Visualizing Pilot Eye Movements for Flight Instructors. In ETVIS'18: 3rd Workshop on Eye Tracking and Visualization, ACM, New York, NY, USA, 2018.

[12] B. Fu, A. R. Soriano, K. Chu, P. Gatsby, N. G. Guardado, Modelling Visual Attention for Future Intelligent Flight Deck - A Case Study of Pilot Eye Tracking in Simulated Flight Takeoff, In: Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, ACM, New York, NY, USA, 2024.

[13] J. H. Goldberg, X. P. Kotval, Computer interface evaluation using eye movements: methods and constructs, International Journal of Industrial Ergonomics, 24(6), 631-645, 1999.

[14] University of Waikato, Weka, version 1.0, 2023. URL: https://www.cs.waikato.ac.nz/ml/weka/