Journal of Aviation/Aerospace Education & Research

Volume 31 | Number 2 Article 2

2022

Investigation of Pilots' Visual Entropy and Eye Fixations for Simulated Flights Consisted of Multiple Take-Offs and Landings

Salem M. Naeeri Ph.D.
University of Oklahoma, Norman, Salem.M.Naeeri-1@ou.edu

Ziho Kang Ph.D.
University of Oklahoma, Norman, zihokang@ou.edu

Ricardo Palma Fraga M.S.
University of Oklahoma, Norman, rpalmafr@ou.edu

Introduction

Analysis of eye movements has enhanced our capacity to better understand an operator's cognitive and emotional state, as well as behaviors and interactions in complex and dynamic domains (e.g., Martinez-Marquez et al., 2021). Recent advances in eye tracking technology have created non-intrusive methods to collect eye movements. Additionally, eye tracking devices can be used without the risk of physical strain while at the same time not interrupting or intruding on the task at hand (Richardson & Spivey, 2008; Wang et al., 2018).

Eye-tracking has been widely used, especially in the domain of aviation, due to the continuous interactions between individuals, such as pilots or air traffic controllers, and the environment. Some examples include but are not limited to: (1) Exploring the effects of simulated air traffic complexity on cognitive workload via eye movement characteristics such as eye movement fixations and saccades (Marchitto et al., 2016); (2) Investigating the situational awareness and visual attention of air traffic controllers when task load increases (Friedrich et al., 2018); (3) Examining the automation monitoring strategies of commercial pilots in a B-747-400 simulator by investigating whether pilots fixated on key areas of interest (AOIs) on the dashboard (Sarter et al., 2007); (4) Characterizing the visual search and conflict mitigation strategies of en-route air traffic controllers by analyzing their time-ordered eye movement fixations and saccades (Palma Fraga et al., 2021); (5) Conducting multimodal analysis on pilot fatigue using vigilance tests and eye tracking measures (Naeeri et al., 2019); (6) Evaluating the performance of en-route air traffic controllers via their visual groupings (Kang & Landry, 2015); (7) Using time-ordered visual scanpaths to increase learning performance on an air traffic control task (Kang & Landry, 2014); (8) Visualizing the expert en-route air traffic controllers' eye movement characteristics that might be used for training (Mandal & Kang, 2018).

Historically, pilots experience higher sustained cognitive activity levels and high attentional demands due to the instrumentation's intricacy, along with the complexity (and rate of occurrence) of takeoff and landing procedures (Gartner & Murphy, 1976; Noy et al., 2011; Zaslona et al., 2018), which can manifest as fatigue (e.g., Honn et al., 2016). For example, Bourgeois-Bougrine et al. (2003) carried out questionnaires with commercial airline pilots, in which 53% of the respondents reported that fatigue was caused by prolonged duty periods, defined as four to five multi-segment flights. Among novice groups of pilots, such as collegiate aviation students, fatigue can be an important safety hazard. In an online survey targeted at college aviation pilots carried out by Mendonca et al. (2019), where 68% of respondents had less than 250 flight hours, 51% of participants had, at least sometimes, "proceeded with flight activities despite being extremely tired" and 78% of those pilots overlooked "mistakes during flight training due to impaired judgement and situational awareness due to fatigue" (p. 20).

Analyzing eye movements might help quantify and better understand how pilots' visual attentiveness (i.e., the ability to prepare for, select, and maintain awareness of specific locations, objects, or attributes of the visual scene) and fatigue change over time during prolonged flight missions. Prior research has focused on investigating visual information processing impairments during long simulated flights (e.g., Rosa et al., 2020). For example, Russo et al. (2005) describes how, after 19 hours of continuous wake, U.S. Air Force pilots had significant omission error rates in a visual perception task (which consisted of attending to a light stimulus in the instrument panel) while piloting a C-141 simulator during an air refueling task. Other studies have presented evidence towards some eye movement metrics, such as blink amplitude, having some predictive power to changes in performance due to fatigue during long simulated flight (Morris & Miller, 1996). More recently, the work of Di Stasi et al. (2016) found significant decreases in eye movement saccadic peak velocity of pilots in a long-simulated flight compared to pilots in a short-simulated flight. In the future, we might be able to develop non-intrusive monitoring systems that leverage eye movements to oversee, for example, a pilot's visual attentiveness throughout various flight phases (e.g., landings, takeoffs) in prolonged flight missions. Such a tool might be capable of enhancing aviation safety (Borghini et al., 2014; Reis et al., 2013).

Visual attention on areas of interest (AOIs), defined via top-down subject-matter expertise, can be tracked and studied through eye movement measures. Due to the enormous number of publications in eye-tracking research, only a few representative examples are provided below. Essential eye movements consist of eye fixations and their respective durations on AOIs (Goldberg & Kotval, 1999; Noton & Stark, 1971). Generally, the temporal threshold to determine whether an eye fixation occurred is between 50 ms and 100 ms (Cristino et al., 2010; Goldberg & Kotval, 1999; Noton & Stark, 1971). The eye fixation duration is the entire duration of a single eye fixation. In addition, we can combine these variables to obtain visual scanpaths: a combination of time-ordered eye fixations and saccades (Goldberg & Kotval, 1999; Noton & Stark, 1971); however, it is difficult to quantify and analyze the information embedded in visual scanpaths due to their complexity (both spatially & temporally), as well as due to the inherent variability that exists between individuals (Kang & Landry, 2015).

One viable approach presented in the literature has been to quantify the characteristics of attentional spread (or randomness) of a visual scanpath via measures of visual entropy – which are sensitive to increases in fatigue (Krejtz et al., 2014; Naeeri et al., 2019). Visual entropy, also known as gaze entropy, provides the randomness associated with a visual scanpath. If the probability of an eye fixation transition from one AOI to all other AOIs is equally likely, then one can infer complete randomness from the visual scanning applied (e.g., think of rolling double numbers in dice). Visual entropy values are calculated from the transition probability matrix of eye movement data obtained from a visual scanpath, and the two types of visual entropy, transition, and stationary entropy, can be computed. The latter quantifies the long-term spatial distribution of a gaze pattern, while the former measures the complexity associated with the pattern (Jeong et al., 2019). The mathematical models of the two types of visual entropy, from Krejtz et al. (2014), are provided below. The transition entropy (H_t) (Equation 1) is obtained using the collected data. On the other hand, the stationary entropy (H_S) (Equation 2) is obtained from deriving stationary distributions (π), meaning that we can estimate the converging visual entropy value over a theoretically infinite period.

$$H_t = -\sum_{i \in A} \pi_i \sum_{i \in A} p_{ij} \log\left(p_{ij}\right) \ , i \ \neq j \qquad (1)$$

$$H_S = \ -\sum_{i \in A} \pi_i \log(\pi_i) \qquad (2)$$

where, $P_{ij} = \frac{n_{ij}}{\sum_{j \in A}(n_{ij})}$, $\pi = \pi P$, $i, j \ \in \ A$. In this set of equations, π depicts the steady stationary distribution of transition probability matrix from AOI i to AOI j, where these belong to the set A of all AOIs. Note that the details of the equations are provided in Naeeri et al. (2021).

Throughout the literature, various methodologies have been developed and applied to measure mental fatigue, including both subjective and objective techniques. The former consists of evaluating fatigue through self-assessment scores, for example, the Samn–Perelli fatigue scale (SPS) and Karolinska sleepiness scale (KSS) (Honn et al., 2016; Samel et al., 1995), which allow us to directly collect pilot feedback on fatigue (van Drongelen et al., 2013). It's important to note that the self-assessments mentioned above can have bias due to the subjectivity associated with them. On the other hand, the latter set of methods, such as the psychomotor vigilance test (PVT) (e.g., Arsintescu et al., 2020; Gander et al., 2013), electroencephalogram (EEG) (e.g., Binias et

DOI: https://doi.org/10.15394/jaaer.2022.1920

al., 2020; Borghini et al., 2014), and eye tracking (e.g., Di Stasi et al., 2016), may possess a different set of conditions that need to be addressed. For instance, in the case of the PVT, the flying task may need to be paused in order to carry out the assessment. The EEG method, which evaluates pilot fatigue by analyzing brain wave data collected through an electrode cap, may cause the pilots to become uncomfortable if the device is worn for a long period of time. On the other hand, eye tracking can be a viable alternative, in which the small device does not have physical contact with the participant and can be used for substantially long periods. Additionally, the data collected by the eye tracking system can be collected without interrupting the task.

Prior preliminary research in prolonged aircraft flight reported increased visual tiredness (i.e., entropy). Eye fixation duration increased, while eye fixation numbers decreased, and self-reported mental fatigue increased (Naeeri et al., 2019; Naeeri & Kang, 2018). Visual entropy measures have also been employed to detect operator impairment in visually demanding tasks in other complex domains, such as (1) in pilot helicopter performance (Diaz-Piedra et al., 2019); (2) in driving performance (Jeong et al., 2019; Schieber & Gilland, 2008; Shiferaw et al., 2018). In addition, preliminary research investigated the effect of fatigue on the multi-segment flight task utilized in the current study, which uses the psychomotor vigilance test (PVT) (Naeeri et al., 2021). The analysis reported higher fatigue levels as the pilots progressed through the flight legs. The present work differs in that the analysis of the eye movement characteristics were not conducted based on specific flight phases, such as takeoff and landing within each flight leg, and that the PVT was not conducted for each flight phase.

Therefore, if we could quantify the pilot’s eye movement characteristics, we could better understand the factors that affect a pilot’s performance, such as mental fatigue from multiple flight legs, particularly during critical phases of flight. There has been little research on fatigue in pilots and phase complexity (e.g., takeoff and landing), although most of the current commercial flight pattern involves multi-takeoffs and landing. The present work expands upon this literature by narrowing the scope to the impacts mental fatigue places on pilots, quantified via eye-tracking, during specific phases: takeoff, climbs, cruises, descends, and landings.

Methods

Participants

A total of 20 licensed pilots were recruited from the Department of Aviation at the University of Oklahoma, each certified with instrument rating (i.e., rated to fly solely via aircraft instrumentation). We defined the pilots as “novices” and “experts” based upon reported flying experience (in months), since they were not able to precisely recall their IFR flight hours. In more detail, “novices” consisted of pilots with less than 36 months of flying experience and “experts” were those participants with more than 36 months. A total of 10 pilots were novices (experience: $\mu = 18$ months; $\sigma = 2.4$) and the remaining 10 pilots were experts (experience: $\mu = 42$ months; $\sigma = 4.5$). The experts had approximately double, or more, flight experience compared to the novices. Novices' age ranged between 21-29, and the experts' age ranged between 28-36. All the pilots had normal hearing and vision.

A power analysis indicated that a sample size of 20 participants provided a reasonable power of 0.91. In addition, since the recruitment of pilots is a challenging task, other existing literature (Bellenkes et al., 1997; Di Stasi et al., 2016; Gateau et al., 2015; Hartzler, 2014) have used an average of 10 pilots to evaluate pilot performance. Note that acquiring instrument rating requires 40 hours of flying under IFR conditions on simulated or actual environments.

DOI: https://doi.org/10.15394/jaaer.2022.1920

Apparatus

Moderate-fidelity Microsoft Flight Simulator (MFS) was used for the experiment. The participants flew a simulated B-52 aircraft. Logitech Extreme 3D Pro Joystick was used to control the simulated aircraft. Tobii TX300 eye tracker, which uses near-infrared diodes to generate reflection patterns on the cornea of the eyes, was used to collect the eye-tracking data. The sampling frequency was 300 Hz with a visual angle accuracy of 0.5 degrees. A 24-inch monitor was used to display the simulation, and Tobii TX300 was placed beneath the monitor.

Task

Pilots were tasked with safely operating the simulated B-52 following instrument flight rules (IFR) for 4 consecutive flight legs without any breaks in-between (see Figure 1). Each flight leg lasted approximately one-hour, for a total of four hours, and were composed of five phases of flight (see Figure 2): (1) takeoff; (2) climb; (3) cruise; (4) descend; (5) landing. All airports had similar runway configurations and consisted of regional general aviation airports. No traffic nor additional aircraft were implemented into the simulation to reduce the complexity of the experiment. The environmental conditions consisted of heavy fog across all flight legs in order to ensure compliance to IFR protocols.

Figure 1

A Visual Representation of the Four Consecutive Flight Legs

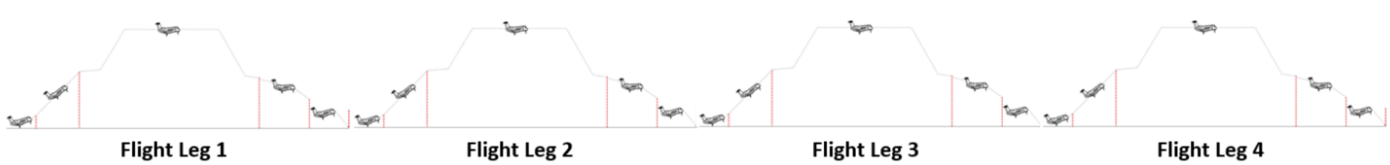
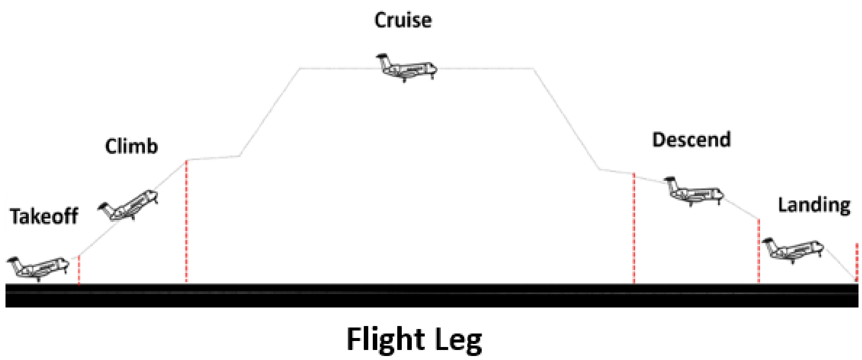


Figure 2

Phases Involved in Each Flight Leg



Procedure

The participants received training for two hours on average on how to use the simulation software and on how to fly the B-52 aircraft before initiating the actual experiment. The participants were instructed to maintain a regular sleep schedule prior to the day of the experiment to prevent possible confounding effects from an irregular sleep cycle. The experiment started at 8:30 a.m. and concluded at approximately 1:00 p.m. During the multi-segment flight experiment, a pilot flew a total of four flight legs, and each leg lasted for about an hour. All participants were provided with the same instructions prior to starting the experiment. At the beginning of the experiment, the participant's eye movements were calibrated. In more detail, a 9-point calibration process was held, in which the process consists of the participant focusing on 9 different (x, y) coordinates on the display. The eye tracking data were collected throughout the multi-segment flight experiment.

Eye tracking measures

The eye tracking measures were:

- Average eye fixation numbers among all AOIs (see Figure 3)

DOI: https://doi.org/10.15394/jaaer.2022.1920

- Average eye fixation durations among all AOIs
- Transition gaze entropy
- Stationary gaze entropy

We verified that most eye fixations occurred on the AOIs defined in Figure 3 by examining all the collected eye fixation data during the experiment; therefore, we did not analyze the eye fixations that occurred outside the defined AOIs.

Figure 3

Pilot's Field of View: Numbers Indicate the 14 Displays that the Pilot Observes During IFR Flight

Note. The areas of interest (AOI) used for the eye movement data analysis are provided in Figure 2. The AOIs consist of the displays that the pilot should observe during IFR flight rules. These include, in the order seen in Figure 2: (1) engine oil pressure; (2) horizontal situation indicator; (3) attitude indicator; (4) enhanced visual screen; (5) engine indicators; (6) flight command indicator; (7) altimeter; (8) airspeed indicator; (9) true airspeed indicator; (10) heading indicator; (11) vertical velocity indicator; (12) radar altimeter; (13) Mach indicator; (14) standby horizon indicator.

Data Analysis

In this paper, our analyses were concentrated on discovering the eye movement characteristics (i.e., static and transition entropies, eye movement durations, and fixations) between experts and novices based on the flight legs, as well as the associated flight phases within each flight leg. Descriptive statistics were plotted, along with the regression results, followed by the statistical inference, including the analysis of the effect of the factors and post-hoc analysis through pairwise comparisons. The collected data were analyzed using SPSS software (version 25) and R.

In more detail, to analyze the effect of the independent variables (i.e., flight legs, nested flight phases, and pilot expertise), a cross-nested mixed three-way ANOVA (Equation 3) was used for each dependent variable. The cross-nested structure comes from the fact that each flight phase is nested within a flight leg and crossed with the pilots' expertise category. A Tukey post-hoc test was used to identify significant differences between the flight legs and phases. Assumptions of normality, homogeneity of covariance, and linearity were satisfied. A significance level of 0.05 was applied for statistical analysis.

$$Y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \rho_k + (\alpha\rho)_{ik} + (\beta\rho)_{jk(i)} + \varepsilon_{ijkl} \tag{3}$$

where, μ indicates the grand mean of the response, α_i represents the flight leg $i \in \{1, 2, 3, 4\}$, $\beta_{j(i)}$ the flight phase $j \in \{1, 2, 3, 4, 5\}$ at flight leg i, ρ_k the participant's expertise category $k \in \{1, 2\}$, $(\alpha\rho)_{ik}$ the interaction effect between the flight leg i and expertise level k, $(\beta\rho)_{jk(i)}$ the interaction effect between the flight phase j and expertise level k, and lastly, ε_{ijkl} contains the random error.

DOI: https://doi.org/10.15394/jaaer.2022.1920

Results

Overall descriptive statistics (e.g., means and standard deviations) related to flight legs and expertise are provided in Table 1, followed by a statistical inferential analysis of each dependent variable.

Table 1

Mean and Standard Deviations of the Four Measurements

Flight leg	Expertise	Transition entropy (Ht)	Stationary entropy (Hs)	Eye fixation durations (FD)	Eye fixation numbers (FN)
flight leg 1	Experts	2.1±0.2	2.9 ±0.2	73.3± .7	278.4±84.4
	Novices	2.8± 0.2	3.2±0.2	115.6±.7	208.9±56.8
flight leg 2	Experts	2.3±0.2	3.1±0.2	87.8±21	291.8±81.8
	Novices	2.4±0.2	3.3±0.2	154.8±40	206.2±67.3
flight leg 3	Experts	2.4±0.2	3.1±0.2	114.1±31.4	252.9±89.8
	Novices	2.6±0.2	3.5±0.2	201.6±52.2	123.8±37.7
flight leg 4	Experts	2.7±0.2	3.3±0.2	175.4±53.5	135.3±37.2
	Novices	3.3±0.2	3.6±0.1	247.0±77.8	83.02±26.7

Transition Visual Entropy

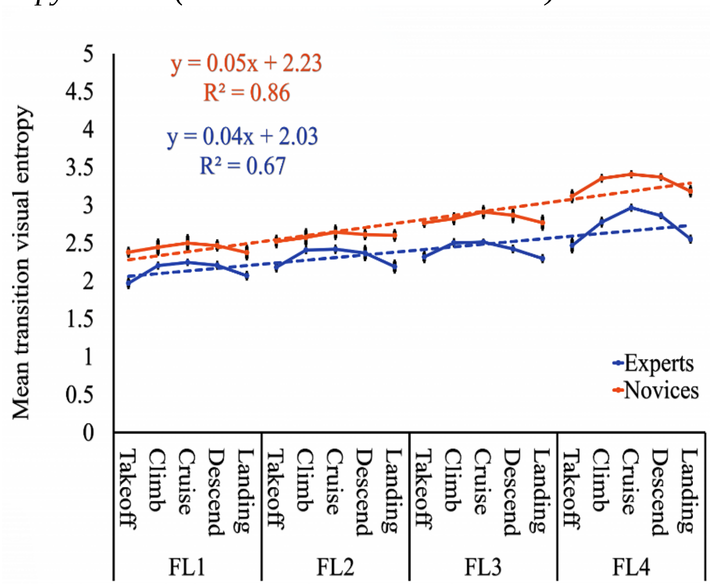
Figure 4 shows the trends of the transition visual entropy based on flight legs, phases, and expertise. An increasing linear trend can be observed based on the increase of the flight leg (FL1 to FL4). Within each flight, transition visual entropies create a concave shape, meaning that the entropy–lack of alertness is lower during takeoff and landing phases, whereas the entropies are higher during the climb, cruise, and descend phases. Finally, experts consistently maintained lower entropies (when they observed the AOIs) than novices.

The effect of expertise ($F(1, 360) = 471.13$, *p-value* < 0.001) and flight legs ($F(3, 360) = 299.16$, *p-value* < 0.001) were significant, yielding an effect size of 0.56 and 0.70. The effect of the flight phase (takeoff vs. climb vs. cruise vs. descend vs. landing) was significant ($F(16, 360) = 9.12$, *p-value* < 0.001), showing an effect size of 0.23. The interaction effect between expertise and flight legs was significant (F (3, 360) = 13.65, *p-value* < 0.001) with an effect size of 0.10. The interaction effect between expertise and flight phases was not significant (F (3, 360) = 1.16, *p-value* = 0.31) with an effect size of 0.04.

Post hoc analysis of the flight legs, using the Tukey test, showed that the transition visual entropy for flight leg 4 (i.e., FL4) was significantly higher (with *p-value* < .001) than those of the other three flight legs (i.e., FL1, FL2, and FL3). The Tukey post hoc test results for novices and experts are provided in Tables 2 and 3, respectively. Overall, the experts have a lot more significantly different cases compared to those of the novices. Significant differences (highlighted in red) were prominent among most of the flight legs.

Figure 4

Transition Visual Entropy Results (Means and Standard Errors)



Note. FL_i shows the i^{th} flight leg, where, $i \in \{1,2,3,4\}$. Each flight leg involved five phases (Takeoff, Climb, Cruise, Descend, and Landing). The regression line coefficient between fixation duration change rate and flight phases was 0.67 for experts and 0.86 for novices.

Table 2

Tukey Post Hoc Analysis of the Phases for Each Flight Leg: Experts Only

Pairwise comparisons		Fight leg 1	Fight leg 2	Fight leg 3	Fight leg 4
Takeoff vs.	Climb	.007	.002	.009	.020
	Cruise	< .001	.003	.009	< .001
	Descend	.005	.085	.072	< .001
	Landing	.195	.936	.716	.301
Climb vs.	Cruise	.251	.770	.782	< .001
	Descend	.938	.564	.161	.100
	Landing	.013	.017	< .001	.012
Cruise vs.	Descend	.185	.481	.047	.017
	Landing	.003	.017	< .001	< .001
Descend vs.	Landing	.002	.109	.008	< .001

Table 3

Tukey Post Hoc Analysis of the Phases for Each Flight Leg: Novices Only

Pairwise comparisons		Fight leg 1	Fight leg 2	Fight leg 3	Fight leg 4
Takeoff vs.	Climb	.412	.537	.632	.006
	Cruise	.100	.290	.130	.004
	Descend	.121	.294	.144	.006
	Landing	1.000	.950	. 600	.138
Climb vs.	Cruise	.217	.624	.569	.139
	Descend	.511	.824	.725	.461
	Landing	.581	.631	.798	.015
Cruise vs.	Descend	.504	.631	.693	.188
	Landing	.014	.415	.611	.005
Descend vs.	Landing	.364	.416	.829	.007

Stationary Visual Entropy

Figure 5 shows the trends of the stationary visual entropy based on flight legs, phases, and expertise. Similar to the transition visual entropy results, an increasing linear trend can be observed based on the increase of the flight leg (FL1 to FL4), but a smaller positing slope is observed. Similarly, within each flight leg, transition visual entropies sometimes create a concave shape, but those concave shapes are much less prominent when compared with the results of the transition visual entropy. However, it is clear that the experts' visual entropy was lower than the novices.

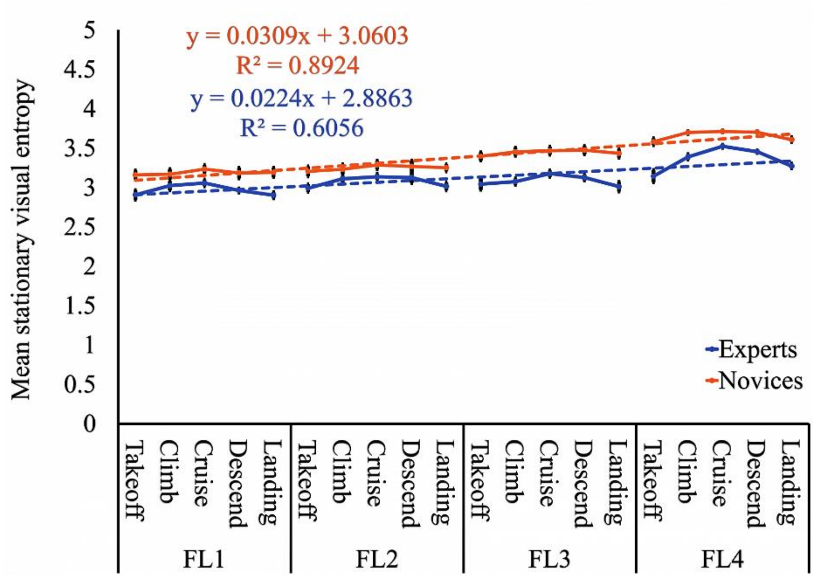
The effects of expertise (F (1, 360) = 277.04, *p-value* < 0.001) and flight legs (F (3, 360) = 138.34, *p-value* < 0.001) were significant, with effect sizes of 0.43 and 0.54, correspondingly. The effect of the phase (takeoff vs. climb vs. cruise vs. descend vs. landing) was significant (F (16, 360) = 3.45, *p-value* < 0.001) with an effect size of 0.13. The interaction effect between expertise and flight legs was significant (F (3, 360) = 7.0, *p-value* < 0.001), yielding an effect size of 0.05. The interaction effect between expertise and flight phases was not significant (F (3, 360) = 0.83, *p-value* = 0.65) with an effect size of 0.035.

Post hoc analysis using the Tukey test showed that the stationary visual entropy for flight leg 4 (i.e., FL 4) was significantly higher than those of the other three flight legs (i.e., FL1, FL2, and FL3) (*p-value* < .001). The Tukey post hoc test results for novices and experts are provided in Tables 4 and 5. Significant differences are not prominent for flight legs 1, 2, and 3. Most of the significant differences (highlighted in red) are found for flight leg 4.

https://commons.erau.edu/jaaer/vol31/iss2/2
DOI: https://doi.org/10.15394/jaaer.2022.1920

Figure 5

Stationary Visual Entropy Results (Means and Standard Errors)



Note. FL_i shows the i^{th} flight leg, where, $i \in \{1,2,3,4\}$. Each flight involved five phases (Takeoff, Climb, Cruise, Descend, and Landing). The regression line coefficient between fixation duration change rate and phases was 0.61 for experts and 0.89 for novices. The ranges of some standard errors were short and are covered by the dots.

Table 4

Tukey Post Hoc Analysis of the Phases for Each Flight Leg: Experts Only

Pairwise comparisons		Fight leg 1	Fight leg 2	Fight leg 3	Fight leg 4
Takeoff vs.	Climb	.166	.046	.650	.028
	Cruise	.042	.047	.076	< .001
	Descend	.394	.125	.356	.004
	Landing	.981	.762	.725	.080
Climb vs.	Cruise	.395	.580	.017	.002
	Descend	.390	.759	.122	.090
	Landing	.076	.083	.283	.070
Cruise vs.	Descend	.108	.900	.267	.026
	Landing	.020	.038	.014	<.001
Descend vs.	Landing	.211	.178	.084	<.001

Table 5

Tukey Post Hoc Analysis of the Phases for Each Flight Leg: Novices Only

Pairwise comparisons		Fight leg 1	Fight leg 2	Fight leg 3	Fight leg 4
Takeoff vs.	Climb	.906	.423	.459	.009
	Cruise	.303	.135	.093	.014
	Descend	.763	.488	.333	.007
	Landing	.621	.421	.666	.520
Climb vs.	Cruise	.350	.431	.812	.583
	Descend	.847	.699	.731	.854
	Landing	.767	.741	.813	.016
Cruise vs.	Descend	.457	.850	.913	.563
	Landing	.546	.504	.658	.022
Descend vs.	Landing	.919	.849	.272	.017

Eye Fixation Durations

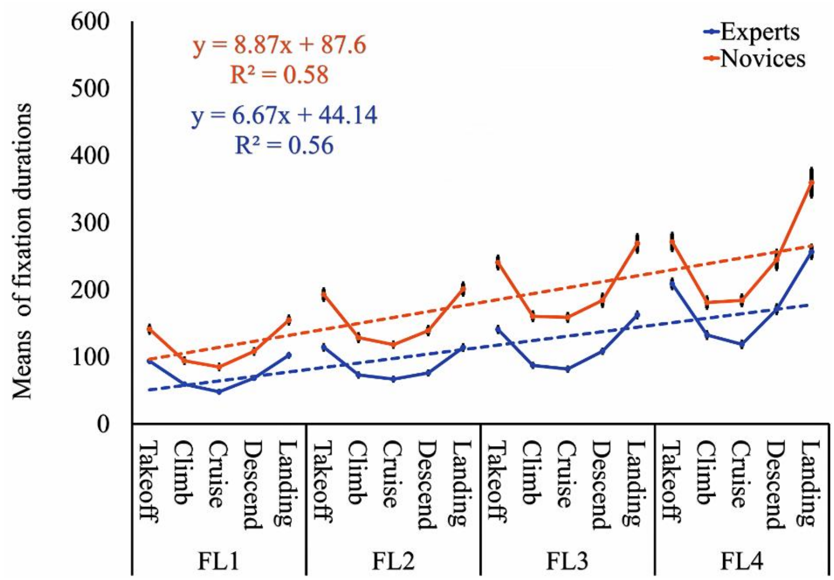
Figure 6 shows the trends of the eye fixation durations based on flight legs, phases, and expertise. Similar to visual entropy results, an increasing linear trend can be observed based on the increase of the flight leg (FL1 to FL4). However, within each flight leg, the eye fixation durations create a convex shape (instead of the concave shape observed from visual entropies), meaning that the eye fixation duration is higher (i.e., more focused) during takeoff and landing, and the eye fixation duration is lower during cruising. Finally, experts consistently maintained shorter eye fixation duration–they were quicker at recognizing AOI than novices.

The effect of expertise was significant ($F(1, 360) = 946.22$, *p-value* < 0.001), yielding an effect size of 0.72. The effect of flight legs was significant ($F(3, 360) = 550.23$, *p-value* < 0.001), resulting in an effect size of 0.82. The effect of the phase (takeoff vs. climb vs. cruise vs. descend vs. landing) was significant ($F(16, 360) = 81.28$, *p-value* < 0.001), with an effect size of 0.78. The interaction effect between expertise and flight legs was also significant ($F(3, 360) = 17.85$, *p-value* < 0.001), yielding an effect size of 0.13. The interaction effect between expertise and flight phases was significant ($F(3, 360) = 2.5$, *p-value* < 0.001), showing an effect size of 0.1.

Post hoc analysis regarding the flight legs using the Tukey test showed that the eye fixation durations for flight leg 4 (i.e., FL4) were significantly higher than those of the other three flight legs (i.e., FL1, FL2, and FL3) (*p-value* < .001). The Tukey post hoc test results for novices and experts are provided in Tables 6 and 7. Significant differences (highlighted in red) were prominent among most of the flight legs.

Figure 6

Eye Fixation Durations Results (Means and Standard Errors)



Note. FL_i shows the i^{th} flight leg, where, $i \in \{1,2,3,4\}$. Each flight leg involved five phases (Takeoff, Climb, Cruise, Descend, and Landing). The regression line coefficient between the fixation durations and phases was 0.56 for experts and 0.58 for novices.

Table 6

Tukey Post Hoc Analysis of the Phases for Each Flight Leg: Experts Only

Pairwise comparisons		Fight leg 1	Fight leg 2	Fight leg 3	Fight leg 4
Takeoff vs.	Climb	.012	<.001	.013	.002
	Cruise	<.001	.113	.832	.015
	Descend	.013	<.001	<.001	<.001
	Landing	.841	.210	.060	.619
Climb vs.	Cruise	<.001	.002	.015	.004
	Descend	.131	<.001	.938	.100
	Landing	<.001	<.001	.002	<.001
Cruise vs.	Descend	<.001	.003	.062	.002
	Landing	<.001	.039	.572	<.001
Descend vs.	Landing	<.001	.002	<.001	<.001

Table 7

Tukey Post Hoc Analysis of the Phases for Each Flight Leg: Novices Only

Pairwise comparisons		Fight leg 1	Fight leg 2	Fight leg 3	Fight leg 4
Takeoff vs.	Climb	.005	<.001	<.001	.024
	Cruise	<.001	.166	.249	.893
	Descend	.011	<.001	<.001	.081
	Landing	.056	.218	.944	.111
Climb vs.	Cruise	<.001	<.001	.005	.072
	Descend	.411	.395	.214	.460
	Landing	<.001	<.001	<.001	.008
Cruise vs.	Descend	<.001	.008	.004	.067
	Landing	<.001	.094	.124	.431
Descend vs.	Landing	<.001	<.001	<.001	.030

Eye Fixation Numbers

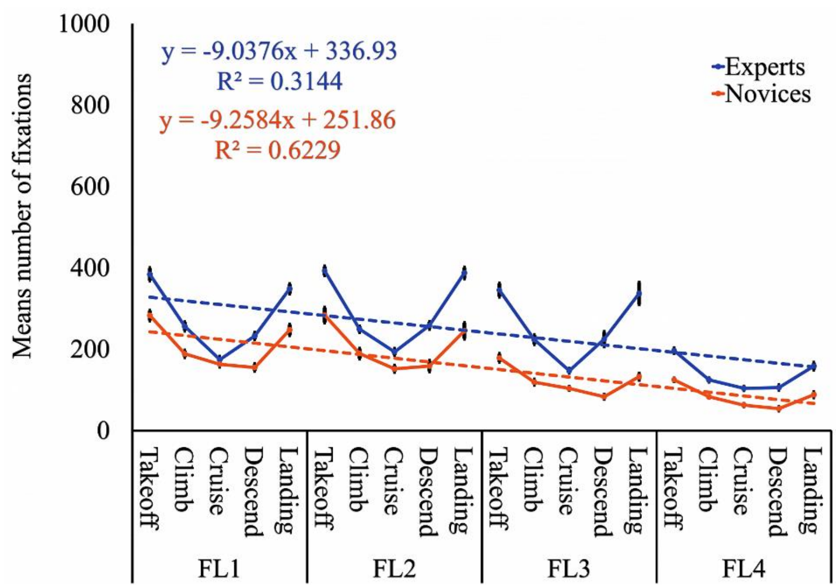
Figure 7 shows the trends of the eye-fixation numbers based on flight leg, phases, and expertise. A linear decreasing trend (instead of the increasing linear trend observed for other measures) can be observed based on the increase of the flight leg (FL1 to FL4). Each flight leg eye number of fixations creates a convex shape similar to the eye fixation durations. Experts consistently maintained a higher eye number of fixations than novices.

The effect of expertise was significant (F (1, 360) = 669.98, *p-value* < 0.001), yielding an effect size of 0.65. The effect of flight leg was also significant (F (3, 360) = 369.59, *p-value* < 0.001), resulting in an effect size of 0.75. The effect of the phase (takeoff vs. climb vs. cruise vs. descend vs. landing) was significant (F (16, 360) = 63.43, *p-value* < 0.001), yielding an effect size of 0.74. The interaction effect between expertise and flight legs was significant (F (3, 360) = 24.04, *p-value* < 0.001) with an effect size of 0.17. The interaction effect between expertise and flight phases was significant (F (3, 360) = 7.65, *p-value* < 0.001), showing an effect size of 0.25.

Post hoc analysis of the flight legs using the Tukey test showed that the eye fixation numbers for flight leg 4 (i.e., FL4) were significantly lower than those of the other three flight legs (i.e., FL1, FL2, and FL3) (*p-value* < .001). The Tukey post hoc test results are provided in Tables 8 and 9. Significant differences (highlighted in red) were prominent among most of the flight legs.

Figure 7

Eye Numbers of Fixations Results (Means and Standard Errors)



Note. FL_i shows the i^{th} flight leg, where, $i \in \{1,2,3,4\}$. Each flight leg involved five phases (Takeoff, Climb, Cruise, Descend, and Landing). The regression line coefficient between the number of fixations and phases was 0.31 for experts and 0.63 for novices.

https://commons.erau.edu/jaaer/vol31/iss2/2
DOI: https://doi.org/10.15394/jaaer.2022.1920

Table 8

Tukey Post Hoc Analysis of the Phases for Each Flight Leg: Experts Only

Pairwise comparisons		Fight leg 1	Fight leg 2	Fight leg 3	Fight leg 4
Takeoff vs.	Climb	<.001	<.001	<.001	<.001
	Cruise	<.001	.002	<.001	<.001
	Descend	<.001	<.001	<.001	.002
	Landing	.779	.196	.265	.901
Climb vs.	Cruise	<.001	.450	.039	.007
	Descend	.196	.198	.935	.665
	Landing	<.001	<.001	<.001	.003
Cruise vs.	Descend	.004	.292	.038	.227
	Landing	<.001	<.001	<.001	<.001
Descend vs.	Landing	<.001	<.001	<.001	.005

Table 9

Tukey Post Hoc Analysis of the Phases for Each Flight Leg: Novice Only

Pairwise comparisons		Fight leg 1	Fight leg 2	Fight leg 3	Fight leg 4
Takeoff vs.	Climb	<.001	<.001	<.001	<.001
	Cruise	<.001	.070	.197	<.001
	Descend	<.001	<.001	<.001	.001
	Landing	.376	.679	.179	.792
Climb vs.	Cruise	.122	.957	.650	<.001
	Descend	.564	<.001	<.001	.215
	Landing	.006	<.001	.119	.161
Cruise vs.	Descend	.625	.530	.170	.563
	Landing	<.001	.015	.648	.044
Descend vs.	Landing	<.001	<.001	.013	.081

Discussion

The results can be interpreted from three aspects: expertise, consecutive repetitive flight legs (i.e., flight legs 1 - 4), and the flight phases nested within the flight legs.

First, significant differences in eye movement characteristics were found between experts and novices. The latter group showed significantly higher visual laxness or entropy (both transition and stationery) and means of eye fixation durations on the areas of interest (AOIs) when compared to the former. On the other hand, novices showed a significantly lower mean of eye fixation numbers on the AOIs than the experts; this may result from experts applying a more visually attentive behavior. For example, looking back and forth at a set of instruments, observing them more often, and acquiring information more quickly when compared to novices throughout the flight legs. In more detail, the visual entropies show that the experts may not focus on less relevant AOIs, as opposed to novices, who tended to have more scattered, and therefore random, eye movements on all the AOIs. Especially, experts' eye fixation numbers were roughly twice those of the novices, and in addition, their eye fixation durations were roughly half of those of the novices. Finally, note the decreased visual search effectiveness with flight legs sequences' progression (Bellenkes et al., 1997). In other words, it seems that more time was required to read the necessary information to carry out the flight legs.

Second, significant differences were found between flight legs, and they can be summarized as follows: as the flight leg number increased, both visual entropies (transition and stationery) and the mean of eye movement durations increased. On the other hand, both groups' mean eye movement fixation numbers decreased. In other words, as the number of flight legs increased, the pilots’ eye movements became less focused (i.e., more random) and had to visually dwell longer on the indicators to extract the necessary information. These results may be attributed to the negative impact of mental fatigue.

Finally, many cases found significant differences among the phases nested within each flight leg. The results show that both experts' and novices' visual entropies were relatively higher during climb, cruise, and descend phases, whereas those were relatively lower during the takeoff and landing phases. The mean eye movement fixation and durations on the AOIs were higher during takeoffs and landings and lower during climbs, cruises, and descends. The reason seems to be that both groups continue to perform rapid cross-checks of heading and attitude indicators during the takeoff and landing phases needed to complete the flight legs. Such actions will lower the visual entropy since few indicators are observed more often when compared to other indicators.

The pilot's intensive back and forth visual attention to those indicators could increase the eye movement fixation on the AOIs. It may seem that the pilots were more freely observing other areas not defined as the AOIs during the climb, descend, and especially cruise phases, which led to fewer eye movement fixations. Similarly, the reason that the eye movement durations are lower during the climb, cruise, and descend phases, compared to takeoffs and landings, may be due to the pilot's heightened visual focus on the instruments during takeoff and landing, and they might have more freely observed the non-instrument areas during the other phases.

Limited to visual entropy, the results accord with those of Diaz-Piedra et al. (2019), in which they found that visual entropy decreased during an emergency. In detail, Diaz-Piedra and colleagues suggested that "attentional tunneling" can occur during an emergency, meaning that the pilots focused on a few important indicators during an emergency. We believe that the takeoff and landing phases align more with the emergency than the other phases, resulting in a decreased trend in visual entropies.

Limitations and Future Research

Our research provides a foundation to delve deeper into the analysis of the visual scanning behaviors of expert and novice pilots, but several limitations will be addressed in future research.

The time for each phase was not equal, meaning the cruise flight phase time was the longest. The cruise time was not shortened to make the experimental design as realistic as possible. Therefore, we plan to divide cruise time into multiple segments to see whether the time duration factor within the cruise phase might affect the results.

The majority of the analysis carried out focused on aggregating AOIs and analyzing them as a singular AOI to identify differences between the novices and experts. Nonetheless, such an approach might be considered a limitation, as some information may be lost when the eye movement fixations and durations on all AOIs are treated as one, such as individual differences between participants. Thus, future research involves exploring differences in the mean eye-fixation numbers and durations between novices and experts while considering each AOI separately. Furthermore, incorporating the analysis of the time-ordered visual scanpaths could facilitate the exploration and identification of scanning patterns used by participants.

Another limitation is the fidelity of the simulator computer-software and apparatus. We used a moderate-fidelity simulator (i.e., Microsoft Flight Simulator), and we do not know whether similar results can be obtained if a high-fidelity simulator, such as an FAA-approved level 6 Flight Training Device (FTD), was used in the study. Therefore, future research efforts include collaborating with organizations, such as the Aviation department at the University of Oklahoma, Embry-Riddle Aeronautical University, or the Federal Aviation Administration's Civil Aeronautical Medical Center, to incorporate such high-fidelity apparatus and environment in an eye tracking study. In addition, further studies are needed to understand better how other factors, such as aircraft size and type (e.g., C-172), might differently affect the eye movement characteristics of pilots.

https://commons.erau.edu/jaaer/vol31/iss2/2
DOI: https://doi.org/10.15394/jaaer.2022.1920

An additional limitation in the study comes from the ad-hoc manner in which participants were classified into novice and expert groups. More commonly, in the literature, the metric “flight hours” is used as the primary criterion to identify expertise (e.g., total flight hours, total flight hours in a specific aircraft). In our case, the participants were not able to precisely recall their total IFR flight hours. Although their flying experience “in months” was used as a surrogate criterion for their expertise, we were able to obtain clear differences in their eye movement measures. In addition, no participants had experience flying the simulated B-52 aircraft; thus, we did not have to consider the possible confounding effect of prior experience on flying the B-52 simulated aircraft. Therefore, future research involves finding multiple supporting evidences that might more clearly indicate the expertise of a pilot.

Finally, the current research can be expanded into multimodal research. For example, eye movements, brain activities, haptic interactions, and voice communications can be analyzed together to understand better the pilots’ behavior that could be used for more effective accident prevention and training.

Conclusion

Overall, the present work contributes to flight safety research by addressing the gap between mental fatigue impact on expert and novice pilots during specific phases in multiple flight legs scenarios. The results indicate that both groups showed significant differences in eye movement characteristics, driven by the number of flight legs and the phases. Experts focus on operationally relevant AOIs, with higher numbers of eye movements, lower durations, and lower visual entropies. On the other hand, the novices’ eye movements were scattered throughout all AOIs, used fewer eye movement fixations, with longer eye movement durations.

This research also found that the impact of mental fatigue on eye movement characteristics increased with the number of flight legs. The pilots' eye movements became less focused, and the eye movement duration was longer, as more time was needed to extract the necessary information from the environment. Lastly, phases also affected the visual entropy for novices and experts, as both showed higher visual entropies during climbs, cruises, and descends and higher mean eye movement fixations on takeoffs and landings. These results may be attributed to the impact of mental fatigue on the participants, exacerbated by flight length and complexity. In conclusion, the continuous monitoring of pilots' eye movement characteristics would detect anomalies that could enable near real-time feedback to the crew, prompting counteractions and preventing fatal errors and accidents, that can be implemented in real-life operations and training.

Acknowledgement

This material was based upon work supported by the National Science Foundation under Grant 1943526. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Arsintescu, L., Chachad, R., Gregory, K. B., Mulligan, J. B., & Flynn-Evans, E. E. (2020). The relationship between workload, performance and fatigue in a short-haul airline. *Chronobiology International*, *37*(9–10), 1492–1494. https://doi.org/10.1080/07420528.2020.1804924

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine*, *68*(7), 569–579.

Binias, B., Myszor, D., Palus, H., & Cyran, K. A. (2020). Prediction of pilot’s reaction time based on EEG signals. *Frontiers in Neuroinformatics*, *14*, 6. https://doi.org/10.3389/FNINF.2020.00006

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue, and drowsiness. *Neuroscience and Biobehavioral Reviews*, *44*, 58–75. https://doi.org/10.1016/j.neubiorev.2012.10.003

Bourgeois-Bougrine, S., Carbon, P., Gounelle, C., Mollard, R., & Coblent, A. (2003). Perceived fatigue for short- and long-haul flights: A survey of 739 airline pilots. *Aviation, Space and Environmental Medicine*, *74*, 1072–1077.

Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, *42*(3), 692–700. https://doi.org/10.3758/BRM.42.3.692

Di Stasi, L. L., McCamy, M. B., Martinez-Conde, S., Gayles, E., Hoare, C., Foster, M., Catena, A., & Macknik, S. L. (2016). Effects of long and short simulated flights on the saccadic eye movement velocity of aviators. *Physiology & Behavior*, *153*, 91–96. https://doi.org/10.1016/j.physbeh.2015.10.024

Diaz-Piedra, C., Rieiro, H., Cherino, A., Fuentes, L. J., Catena, A., & Di Stasi, L. L. (2019). The effects of flight complexity on visual entropy: An experimental study with fighter pilots. *Applied Ergonomics*, *77*(9), 92–99. https://doi.org/10.1016/j.apergo.2019.01.012

Friedrich, M., Biermann, M., Gontar, P., Biella, M., & Bengler, K. (2018). The influence of task load on situation awareness and control strategy in the ATC tower environment. *Cognition, Technology & Work*, *20*(2), 205–217. https://doi.org/10.1007/s10111-018-0464-4

Gander, P. H., Signal, T. L., van den Berg, M. J., Mulrine, H. M., Jay, S. M., & Mangie, J. (2013). In-flight sleep, pilot fatigue and Psychomotor Vigilance Task performance on ultra-long range versus long range flights. *Journal of Sleep Research*, *22*(6), 697–706. https://doi.org/10.1111/jsr.12071

Gartner, W. B., & Murphy, M. R. (1977). Pilot workload and fatigue: A critical survey of concepts and assessment techniques [Abstract]. *Applied Ergonomics*, *8*(4), 238. https://doi.org/10.1016/0003-6870(77)90183-1

Gateau, T., Durantin, G., Lancelot, F., Scannella, S., & Dehais, F. (2015) Real-Time state estimation in a flight simulator using fNIRS. *PLoS ONE*, *10*(3). https://doi.org/10.1371/journal.pone.0121279

Goldberg, J. H., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, *24*(6), 631–645. https://doi.org/10.1016/s0169-8141(98)00068-7

Hartzler, B.M. (2014). Fatigue on the flight deck: The consequences of sleep loss and the benefits of napping. *Accident Analysis & Prevention*, *62*, 309–318. https://doi.org/10.1016/j.aap.2013.10.010

Honn, K. A., Satterfield, B. C., McCauley, P., Caldwell, J. L., & Van Dongen, H. P. A. (2016). Fatiguing effect of multiple takeoffs and landings in regional airline operations. *Accident Analysis & Prevention*, *86*, 199–208. https://doi.org/10.1016/j.aap.2015.10.005

Jeong H., Liu Y., & Kang. Z. (2019). Driver glance behaviors and scanning patterns: Applying static and dynamic glance measures to the analysis of curve driving with secondary tasks. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *29*(6), 437-446. https://doi.org/10.1002/hfm.20798

Kang, Z., & Landry, S. J. (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Human Factors: The Journal of Human Factors and Ergonomics Society*, *56*(1), 1150-1162. https://doi.org/10.1177/0018720814523066

Kang, Z., & Landry, S. J. (2015). An eye movement analysis algorithm for a multielement target tracking task: Maximum transition-based agglomerative hierarchical clustering. *IEEE Transactions on Human-Machine Systems*, *45*(1), 13–24. https://doi.org/10.1109/THMS.2014.2363121

Krejtz, K., Szmidt, T., Duchowski, A. T., & Krejtz, I. (2014). Entropy-based statistical analysis of eye movement transitions. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 159-166. https://doi.org/10.1145/2578153.2578176

Mandal, S., & Kang. Z. (2018). Using eye movement data visualization to enhance training of air traffic controllers: A dynamic network approach. *Journal of Eye Movement Research*, *11*(4), 1-20. https://doi.org/10.16910/jemr.11.4.1

Marchitto, M., Benedetto, S., Baccino, T., & Cañas, J. J. (2016). Air traffic control: Ocular metrics reflect cognitive complexity. *International Journal of Industrial Ergonomics*, *54*, 120-130. https://doi.org/10.1016/j.ergon.2016.05.010

Martinez-Marquez D., Pingali S., Panuwatwanich K., Stewart R. A., & Mohamed, S. (2021). Application of eye tracking technology in aviation, maritime, and construction industries: A systematic review. *Sensors*, *21*(13). https://doi.org/10.3390/s21134289

Mendonca, F. A., Keller, J., & Lu, C. (2019). Fatigue identification and management in flight training: An investigation of collegiate aviation pilots. *International Journal of Aviation, Aeronautics, and Aerospace*, *6*(5). https://doi.org/10.15394/ijaaa.2019.1365

Morris, T. L., & Miller, J. C. (1996). Electrooculographic and performance indices of fatigue during simulated flight. *Biological Psychology*, *42*(3), 343–360. https://doi.org/10.1016/0301-0511(95)05166-x

Naeeri, S., & Kang, Z. (2018). Exploring the relationship between pilot's performance and fatigue when interacting with cockpit interfaces. In K. Barker, D. Berry, & C. Rainwater (Eds.), *Proceedings of the 2018 IISE Conference* (pp. 1498-1503). Institute of Industrial and Systems Engineers.

Naeeri, S., Mandal, S., & Kang, Z. (2019). Analyzing pilots' fatigue for prolonged flight missions: Multimodal analysis approach using vigilance test and eye tracking. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*,*63*(1), 111-115. https://doi.org/10.1177/1071181319631092

Naeeri, S., Kang, Z., Mandal, S., & Kim, K. (2021). Multimodal analysis of eye movements and fatigue in a simulated glass cockpit environment. *Aerospace*, *8*(10), 1-19. https://doi.org/10.3390/aerospace8100283

Noton, D., & Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science*, *171*(3968), 308–311. https://doi.org/10.1126/science.171.3968.308

Noy, Y. I., Horrey, W. J., Popkin, S. M., Folkard, S., Howarth, H. D., & Courtney, T. K. (2011). Future directions in fatigue and safety research. *Accident Analysis & Prevention*, *43*(2), 495–497. https://doi.org/10.1016/j.aap.2009.12.017

Palma Fraga, R., Kang, Z., Crutchfield, J. M., & Mandal, S. (2021). Visual search and conflict mitigation strategies used by expert en route air traffic controllers. *Aerospace*, *8*(7), 170. https://doi.org/10.3390/aerospace8070170

Reis, C., Mestre, C., & Canhão, H. (2013). Prevalence of fatigue in a group of airline pilots. *Aviation, Space, and Environmental Medicine*, *84*(8), 828–833. https://doi.org/10.3357/ASEM.3548.2013

Richardson, D., & Spivey, M. (2008). Eye tracking: Characteristics and methods. In G. Wnek & G. Bowlin (Eds.), *Encyclopedia of Biomaterials and Biomedical Engineering* (2nd ed., pp. 1028-1032). CRC Press. https://doi.org/10.1201/9780429154065

Rosa, E., Eiken, O., Grönkvist, M., Kölegård, R., Dahlström, N., Knez, I., Ljung, R., Jönsson, F. U., & Willander, J. (2020). Effects of fatigue on cognitive performance in long-duration simulated flight missions. *Aviation Psychology and Applied Human Factors, 10*(2), 82-93. https://doi.org/10.1027/2192-0923/a000191

Russo, M., Sing, H., Kendall, A. P., Johnson, D. E., Santiago, S., Escolas, S., Holland, D., Thorne, D., Hall, S., Redmond, D., & Thomas, M. (2005). Visual perception, flight performance, and reaction time impairments in military pilots during 26 hours of continuous wake: Implications for automated workload control systems as fatigue management tools. In *Strategies to maintain combat readiness during extended* *deployments – A human systems approach* (pp. 27-1-27-16). North Atlantic Treaty Organization Research and Technology Organization.

Samel, A., Wegmann, H. M., & Vejvoda, M. (1995). Jet lag and sleepiness in aircrew. *Journal of Sleep Research*, *4*(S2), 30–36. https://doi.org/10.1111/j.1365-2869.1995.tb00223.x

Sarter, N. B., Mumaw, R. J., & Wickens, C. D. (2007). Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data. *Human Factors*, *49*(3), 347-357. https://doi.org/10.1518/001872007X196685

Schieber, F., & Gilland, J. (2008). Visual entropy metric reveals differences in drivers' eye gaze complexity across variations in age and subsidiary task load. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *52* (23), 1883–1887. https://doi.org/10.1177/154193120805202311

Shiferaw, B. A., Downey, L. A., Westlake, J., Stevens, B., Rajaratnam, S. M. W., Berlowitz, D. J., Swann, P., & Howard, M. E. (2018). Stationary visual entropy predicts lane departure events in sleep-deprived drivers. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-20588-7

van Drongelen, A., van der Beek, A. J., Hlobil, H., Smid, T., & Boot, C. R. L. (2013). Development and evaluation of an intervention aiming to reduce fatigue in airline pilots: Design of a randomized controlled trial. *BMC Public Health*, *13*(1). https://doi.org/10.1186/1471-2458-13-776

Wang, Y., Zhai, G., Zhou, S., Chen, S., Min, X., Gao, Z., & Hu, M. (2018). Eye fatigue assessment using unobtrusive eye tracker. *IEEE Access*, *6*, 55948–55962. https://doi.org/10.1109/ACCESS.2018.2869624

Zaslona, J. L., O'Keeffe, K. M., Signal, T. L., & Gander, P. H. (2018). Shared responsibility for managing fatigue: Hearing the pilots. *PLoS ONE*, *13*(5), 1–11. https://doi.org/10.1371/journal.pone.0195530