

## Article

# Multimodal Analysis of Eye Movements and Fatigue in a Simulated Glass Cockpit Environment

Salem Naeeri <sup>1</sup>, Ziho Kang <sup>1,\*</sup>, Saptarshi Mandal <sup>2</sup> and Kwangtaek Kim <sup>3</sup><sup>1</sup> Department of Industrial and Systems Engineering, University of Oklahoma, 202 West Boyd Street, Norman, OK 73019, USA; Salem.M.Naeeri-1@ou.edu<sup>2</sup> Human Factors Engineering Lab, Microsoft, 5069 154th Place NE, Redmond, WA 98052, USA; v-sapmandal@microsoft.com<sup>3</sup> Department of Computer Science, Kent State University, 800 East Summit Street, Kent, OH 44240, USA; kkim@cs.kent.edu

\* Correspondence: zihokang@ou.edu

**Abstract:** Pilot fatigue is a critical reason for aviation accidents related to human errors. Human-related accidents might be reduced if the pilots' eye movement measures can be leveraged to predict fatigue. Eye tracking can be a non-intrusive viable approach that does not require the pilots to pause their current task, and the device does not need to be in direct contact with the pilots. In this study, the positive or negative correlations among the psychomotor vigilance test (PVT) measures (i.e., reaction times, number of false alarms, and number of lapses) and eye movement measures (i.e., pupil size, eye fixation number, eye fixation duration, visual entropy) were investigated. Then, fatigue predictive models were developed to predict fatigue using eye movement measures identified through forward and backward stepwise regressions. The proposed approach was implemented in a simulated short-haul multiphase flight mission involving novice and expert pilots. The results showed that the correlations among the measures were different based on expertise (i.e., novices vs. experts); thus, two predictive models were developed accordingly. In addition, the results from the regressions showed that either a single or a subset of the eye movement measures might be sufficient to predict fatigue. The results show the promise of using non-intrusive eye movements as an indicator for fatigue prediction and provides a foundation that can lead us closer to developing a near real-time warning system to prevent critical accidents.

**Keywords:** multimodal analysis; fatigue; eye movements; eye tracking; psychomotor vigilance task; entropy; expertise; pilot; prediction

**Citation:** Naeeri, S.; Kang, Z.; Mandal, S.; Kim, K. Multimodal Analysis of Eye Movements and Fatigue in a Simulated Glass Cockpit Environment. *Aerospace* **2021**, *8*, 283. <https://doi.org/10.3390/aerospace8100283>

Academic Editor: Michael Schultz

Received: 12 July 2021

Accepted: 22 September 2021

Published: 1 October 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Fatigue is a critical reason for human-error-related aviation accidents [1,2]. A recent review of major airline crashes reported that 48% of aviation crashes were attributed to pilot errors, and approximately 20% of these errors were associated with pilot fatigue [3,4]. Previous studies have identified that high levels of fatigue severely affect a pilots' ability to attend to complex information, detect safety issues, and provide timely responses [5–7]. Therefore, it is important to evaluate and, if possible, also predict pilot fatigue levels so that intervention measures can be implemented on time.

Fatigue can hinder the pilot's ability to stay alert and be attentive during a short-haul flight consisting of multiple consecutive flight missions. Short-haul flights usually include 4–5 legs per day, whereas long-haul flights usually include 20 or more hours of non-stop single leg flight [8]. Limited to short-haul flights, the survey results showed that the number of legs per day, flight duration, and time of day can be factors that increase fatigue [8–10]. In detail, the number of flight legs and duty length (time-on-task) were the most significant factors that increased pilot fatigue in short-haul flights, whereas the time of day

had a weaker impact. Furthermore, prior duty and sleep substantially affected the pilots' fatigue [11]. In detail, the reduction in pilots' prior sleep resulted in the increase in self-rated fatigue and decrease in mean response speed.

Some approaches have focused on observing physiological and subjective data including PVT, electroencephalography (EEG), electrooculogram (EOG), Samn–Perelli fatigue scale (SPS), and the Karolinska sleepiness scale (KSS) [12,13]. The results showed that factors, such as unpredictable duty, number of legs, prolonged duty periods (i.e., time-on-task), insufficient sleep, and circadian disruptions, had substantially influenced pilots' fatigue.

Building a pilot fatigue prediction model is crucial for developing timely alerting or scaffolding methods to prevent fatigue-induced aviation accidents. The effectiveness of a fatigue prediction model depends on the methodology used to evaluate pilot fatigue. For example, intrusive fatigue evaluation methods (requiring the task to be paused for assessing fatigue) will hinder the fatigue prediction model's adaptation for real-time fatigue prediction. Hence, before developing the pilot fatigue prediction model, we explored the limitations of the various pilot fatigue evaluation methods so that we could implement the most appropriate method (or a combination of eye tracking and PVT) in our case. The various pilot fatigue evaluation methods implemented in earlier studies can be broadly categorized into subjective methods [14], objective methods [15–23], and hybrid of subjective and objective methods [11,24–28].

Subjective methods consist of evaluating pilot fatigue through self-assessment scores, for example, fatigue rating and sleepiness scales, that allow us to understand a pilots' opinions and fatigued feelings [14]. Despite their ease of use, these methods might suffer from biased judgment, and it also requires pilots to recollect and write down their self-perceived fatigue level while the experiment is paused [29]. Hence, these methods might affect the accuracy of the fatigue prediction model. Objective methods include psychomotor vigilance tasks (PVTs) [15], EEG [16–18], eye tracking [19,20], or a combination of those methods (e.g., PVT and eye tracking) [21–23].

In detail, PVT evaluates pilot fatigue by measuring the change in their task-based performance [15]. The PVT approach has been proven to be very effective in assessing pilot fatigue [11,24–26]. However, PVT can be intrusive, as it requires the pilots to pause the current task at hand to assess fatigue level. Thus, it affects the natural working environment of a pilot flying an aircraft. Therefore, developing a fatigue prediction model using PVT measures alone might hinder its adaptation for near real-time pilot fatigue prediction that does not require intervention.

The EEG method evaluates pilot fatigue by analyzing their brain waves, and it does not require the task to be paused [16–18]. Evaluation of the brain activities can be a viable approach to assess fatigue, but the EEG device can be intrusive because the electrode cap has to be placed on the pilot's head throughout the task. Thus, it might be challenging to implement this approach for a long duration task, as it impacts the pilots' free movement in their natural working environment [30,31].

The eye tracking method can be used to evaluate fatigue by analyzing pilots' eye movements collected using a small eye tracker placed beneath a monitor or anywhere within the pilots' visual field. Prior studies [19,20] demonstrated that eye tracking methods might effectively measure fatigue levels of pilots, since the pilots do not need to pause their task allowing continuous data collection in real time. Thus, eye tracking provides a viable, unintrusive, and real-time fatigue evaluation method, thereby making it suitable for developing a pilot fatigue prediction model. Furthermore, existing studies combined PVT and eye tracking measures that investigated fatigue for a single take-off and landing task [21,22] or where the participants were only novices [23].

To the best of our knowledge, there has been no research on investigating the following research questions: (a) how fatigue increase might differ based on a pilot's expertise for a multi-leg flight mission; (b) how fatigue levels increase after each flight leg; (c) how eye movement measures are correlated with fatigue levels (measured using PVT that

proved to measure accurate fatigue levels as briefly explained above); (d) whether a fatigue predictive model can be created using only eye movement measures. Furthermore, PVT has several measures (such as reaction times, false alarms, and number of lapses), and there is no research on how to combine those into a single fatigue assessment measure. Motivated by our preliminary research efforts[22,23,28], the present study addresses the questions.

## 2. Background

In this section, an in-depth literature review based on the fatigue evaluation methods within the aviation and other relevant domains is provided below. Table 1 provides a concise summary and classifications based on research topic, evaluation method, expertise, single vs. multiple legs, short vs. long duration flights, and statistical analysis method.

**Table 1.** Classification of existing research in pilot fatigue: Classifications are mostly based on the fatigue evaluation method. The last three listed in the table are studies not related to fatigue but worth mentioning.

Research Related to Pilots' Fatigue	Research Topic	Fatigue Evaluation Method	Expertise	Single or Multiple Take-Off Landings	Short (1~3 hrs.) vs. Long Duration (3+ hrs.) Flight	Statistical Method
[8]	Fatigue	Subjective	Experts	Single	Short and long	Multiple regression and ANOVA
[14]	Fatigue	Subjective	Experts	Single	Long and Short	linear mixed-model
[15]	Workload and fatigue	PVT and Subjective	Experts	Single	Short	Stepwise Regression and correlation
[16]	Reaction time	EEG	Novice	None	Short	Robust linear model
[17]	Mental workload and fatigue	EEG	-	-	-	Literature review
[18]	Fatigue	EEG	Novices	Single	Short	Classification model
[19]	Fatigue	Eye tracking	Experts	Single	Short and Long	ANOVA and linear regression
[20]	Fatigue	Eye tracking	Experts	Single	Short	Pre/Post-Test design
[21]	Fatigue	Eye tracking and PVT	Novices	Single	Short	ANOVA and regression
[22]	Fatigue	Eye tracking and PVT	Novices and experts	Single	Short	Mann-Whitney-Wilcoxon tests
[23]	Fatigue	Eye tracking and PVT	Novices	Multiple	Long	Kruskal-Wallis test
[11]	Fatigue and sustained attention	PVT and Subjective	Experts	Single	Long	Linear mixed model regression
[24]	Fatigue and Performance	PVT and Subjective	Experts	Single	Long and ultra-long	Mixed-model ANOVA
[25]	Fatigue and performance	PVT and Subjective	Experts	Single	Long and ultra-long	ANOVA
[26]	Fatigue	PVT and Subjective	Experts	Single	Short and Long	Statistical/Machine learning model
[27]	Fatigue	EEG	Novice	Single	Long	ANOVA
[28]	Fatigue	Eye tracking and PVT	Experts	Multiple	Long	ANOVA

[32]	Cognitive load	Eye tracking	Experts	Single	Short	Paired t-tests
[33]	Workload	EEG, Eye tracking, and Subjective	Experts	Single	Short	ANOVA and correlation
[34]	Performance	Eye tracking	Novices and experts	Single	Short	ANOVA

### 2.1. PVT Measures and Fatigue Assessment

In the absence of any direct fatigue measurement approach, PVT has been proven to be the most effective and widely used fatigue evaluation method. PVT evaluates fatigue by assessing changes in an individual's performance for a particular button-pressing task when a visual stimulus (e.g., lights) is sequentially presented one by one on a display at random time intervals [35]. Specifically, three measures are evaluated: (a) reaction times (RT), where  $150 \text{ ms} < \text{RT} < 500 \text{ ms}$ ; (b) number of false alarms (tallied when  $\text{RT} < 150 \text{ ms}$ ); (c) number of lapses (tallied when  $\text{RT} > 500 \text{ ms}$ ) [36].

Note that RT increases with the rise in fatigue level. Furthermore, for  $\text{RT} < 150 \text{ ms}$ , the button pressing action is considered a false alarm, which implies that either the onset of the visual stimulus was anticipated or the action was performed without seeing the stimulus. Conversely, the button pressing act having an  $\text{RT} > 500 \text{ ms}$  is counted as a lapse, thus implying a temporary failure of concentration due to the fact of fatigue. Another possible measure, also used by researchers to assess fatigue, is the response speed, which is defined as the inverse of RT (i.e.,  $1/\text{RT}$ ). For example, if RT is 200 ms, then the response speed is  $0.005 \text{ ms}^{-1}$ .

Different researchers have used one or a collection of the four PVT measures mentioned above to evaluate pilot fatigue. For example, the number of lapses was found to increase, whereas response speed decreased with a rise in pilot fatigue levels for short-haul flight operations [15]; a decrease in mean response speed ( $1/\text{RT}$ ) was observed with an increase in time-induced fatigue on pilots in both long [11], ultra-long flights [24], and short flights [25]. Notably, the mean reaction times and the mean number of lapses increased with an increase in pilot fatigue, where pilots performed simulated flying operations under both rested and fatigued conditions [26].

Regarding the fatigue prediction model development, prior studies (which used a hybrid approach of combining PVT and subjective measures) found different results for long- and short-haul flights. For example, for a short-haul flight, the number of lapses was found to be important for pilot fatigue prediction [15]; whereas, for a long-duration flight, pilots' sleep duration was found to be the only significant predictor (in a linear mixed-model regression model) of pilot fatigue [11].

Note that all these previous studies found similar trends, in which response times, number of false alarms, and number of lapses increased as fatigue increased. Response speed (being the inverse of reaction times) decreased with an increase in fatigue. However, to the best of our knowledge, each PVT measure was evaluated separately when we reviewed the previously published papers on PVT. If we used the PVT measures separately, then would we need to develop 3 times the regression models and would not know which one to apply for prediction. Hence, we would need a single unified PVT measure to develop a single prediction model. Therefore, we devised a simple unified measure which is the weighted linear combination of the three basic PVT measures (i.e., reaction times, number of false alarms, and number of lapses). More details are provided in Section 3.

The abovementioned studies, due to the PVT method's implementation, require the current task to be paused, which might not be either favorable or feasible when performing a piloting task. In addition, they include flight operations that only have single take-off and landing tasks for both short- and long-duration flights. Thus, the results obtained

cannot necessarily be transferred to our case of multiple take-offs and landing tasks. Furthermore, these studies did not consider the effect of pilot expertise on their fatigue level.

## 2.2. Eye Movement Measures and Fatigue Evaluation

Eye tracking is nonintrusive and can provide measures such as eye fixation position (or location), duration, pupil dilation, visual scanpath (i.e., the time-order of the eye fixations that occurred on display), saccade, blink, and eyelid closure [19,20,32,37–39], in which eyelid closure slowed as fatigue increased [37,38] and saccadic velocity decreased as fatigue increased after long simulated flights [19,20,32,37–39].

In more detail, previous studies utilizing only the eye tracking method for pilot fatigue evaluation suggested that expert pilots' saccadic movements decreased with an increase in time-induced fatigue for a single take-off and landing operation [19, 20]. Moreover, prior studies that implemented a composite objective approach to fatigue evaluation (i.e., combining both eye tracking and the PVT method) noted that a pilot's pupil diameter increased with a rise in their fatigue level [21]; expert pilots showed faster reaction times and fewer lapses and false starts than novice pilots at higher fatigue levels [22]; expert pilots displayed more frequent eye fixations with a shorter duration than novice pilots as their fatigue level increased [23].

We currently do not know whether similar results can be obtained for a long-duration aircraft piloting task with multiple take-offs and landing operations and whether a fatigue predictive model (i.e., regression model) can be developed for the same. In addition to the traditional measures, eye movement data, especially the saccadic eye movements, can be further processed to evaluate the pilots' overall eye movement transition behavior using visual entropy [23, 33]. Visual entropy is a measure that quantifies the randomness of the eye movement transition behavior, where a large visual entropy value suggests more random eye movement transition behavior over the display and vice-versa. The concept of visual entropy has been adapted from information entropy [40]. Previous studies have used two types of visual entropy measures—transition and stationery—to analyze the impact of fatigue and task complexity on eye movement transition behavior. For example, for a helicopter maneuvering task, expert pilots' transition entropy was found to decrease with an increase in task complexity levels [33]; in the case of a simulated driving study, increase in fatigue levels led to an increase in both the visual entropy measures [41]; in a robotic surgery training task, transition entropy was noted to increase with a rise in perceived workload [42]. Nonetheless, these studies focused on other aspects and did not consider the impact of fatigue.

Before calculating two visual entropy measures, we first needed to evaluate the eye fixation transition probability matrix, also called area of interest (AOI) transition probability using the design principles [40]. In other words, within a display or a field of view, we can define certain important areas as AOIs, then analyze the eye movement characters only using those AOIs. The AOI transition probability matrix was defined as  $\mathbf{P} = [p_{ij}]$ , where  $p_{ij}$  is interpreted as the probability of eye fixation transition from  $i$ th AOI towards the  $j$ th AOI. In other words, the transitions among the AOIs can be investigated using the transition probability matrix. An example of how the AOIs were defined for our research is provided in Section 4.

Visual entropy can be divided into transition entropy and stationary entropy [43]. Transition entropy is calculated using only the data collected during the experiment and is also known as the entropy rate [44], whereas stationary entropy shows to which value it is expected to converge over a very long period. The following Equations (1) and (2) are the most important, which we summarized from [43].

Transition entropy:

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

Stationary entropy:

$$H_s = -\sum_{i \in A} \pi_i \log(\pi_i) \quad (2)$$

where,  $P_{ij} = \frac{n_{ij}}{\sum_{j \in A} n_{ij}}$ ,  $\pi = \pi P$ ,  $i, j \in A$ , and  $\pi$  is the steady stationary distribution (i.e., steady-state vector) associated with AOI transition probability matrix,  $i$  and  $j$  are AOI indexes, and  $A$  is the set of the AOIs. An interesting relationship between  $H_t$  and  $H_s$  is  $H_t \leq H_s$  [45], implying that stationary entropy ( $H_s$ ) can result in a shorter range compared to the range of the transition entropy ( $H_s$ ).

Table 2 shows two extreme numerical examples of eye fixation transition probabilities and the resulting stationary and transition entropy values. Table 2a shows an extreme example of randomness, whereas Table 2b shows an extreme example of concentration (i.e., many eye fixations) on certain transitions from one AOI to another AOI. If we consider eye movement transition matrices in between two extreme examples, the range of the transition entropy can be approximately [0.1, 1.6], whereas the range of the stationary entropy can be approximately [1.6, 2.0] having a shorter range. Since either calculation approach (i.e., (1) or (2)) is viable and the entropy values show relative differences, we chose to investigate both approaches to identify which one might better predict fatigue given that any correlations might exist.

**Table 2.** Example of two different eye fixation transition probability matrices.

TO					TO				
From	A	B	C	D	From	A	B	C	D
A	0	0.33	0.33	0.34	A	0	0.99	0.01	0
B	0.33	0	0.33	0.34	B	0	0	0.99	0.01
C	0.33	0.34	0	0.33	C	0.99	0	0	0.01
D	0.33	0.33	0.34	0	D	0	0.99	0.01	0
$H_t = 1.6$ ; $H_s = 2.0$					$H_t = 0.1$ ; $H_s = 1.6$				
(a)					(b)				

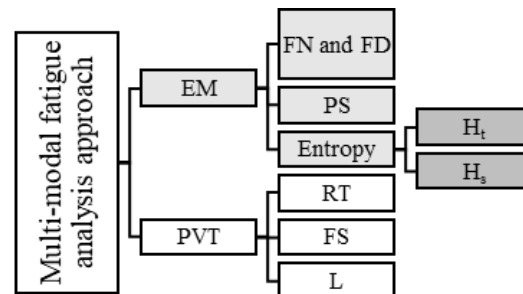
In Table 2, matrix (a) is an extreme example of the eye fixation probabilities based on uniform distribution. The diagonal values were set to zero since we did not consider consecutive eye fixations that occur on the same AOI when calculating the visual entropy. Matrix (b) was an extreme example based on non-uniform eye fixation transition probabilities.

Note that having more eye fixations (leading to more eye movement transitions) does not mean that the entropy is higher. The entropy was calculated based on transition probabilities; therefore, we can have a higher entropy value with a fewer number of eye fixations.

### 3. Methods

Our proposed method consisted of two steps. In the first step, we evaluated pilot fatigue using PVT measures and investigated the correlation between the PVT measures and various eye movement measures. This correlation study helped us to evaluate the eye movement measures' validity in assessing pilot fatigue for the given flight scenario. The second step involved developing a fatigue prediction model using a stepwise regression model where only the normalized eye movement measures were treated as predictor variables. Note that fatigue was assessed by normalizing and aggregating three PVT measures into a single fatigue measure. In other words, we assumed that PVT measures were very accurate in providing fatigue levels based on previous research [21–23], and we were investigating which eye movement measures could be used as effective predictor variables when considering expertise.

Figure 1 represents the two different types of measures including the PVT measures and eye movement (EM) measures. The widely used PVT measures were used as a basis to assess accurate fatigue levels. The EM measures were then investigated to discover which EM measures might be highly correlated with the PVT measures.



**Figure 1.** Measures used to investigate fatigue for a multiphase flight task: FN is eye fixation numbers, FD is eye fixation durations, PS is pupil size, Ht is transition visual entropy, Hs is stationary visual entropy, RT is reaction times, FS is number of false starts, and L is number of lapses.

The detailed analysis steps are as follows:

- Step 1:** Assess fatigue level through PVT after each task. Measures are (a) reaction times (RT), (b) number of false starts (FS), and (c) number of lapses (L).
- Step 2:** Collect eye tracking data, analyzing the data using the context-specific areas of interest (AOIs). Measures are:

- (a) Mean eye fixation number on AOIs;
- (b) Mean eye fixation duration on AOIs;
- (c) Mean pupil size on AOIs;
- (d) Visual entropy (calculation process explained below).

The visual entropy evaluates the amount of randomness associated with the visual scanning strategy of the pilots. Higher visual entropy value means that relatively more randomness (in eye movements) exists. We hypothesized that experts' visual entropy would be lower than that of the novices, meaning that the novices' eye movements might show more randomness, especially when fatigued. Although we followed the procedures provided in [43] to calculate visual entropy, one major difference was that in [43], they used context independent AOIs by dividing an image into equally sized grids, whereas we used context-specific (or context-dependent) AOIs. An example of how we defined the context specific AOIs is provided in the Section 4.

- Step 3:** Plot the relationships of the variables and investigate the correlations between the PVT measures and the eye tracking measures. The measures were all those provided in Steps 1 and 2. Step 3 was needed to first see whether linear correlations could be observed prior to applying multiple regression. In other words, different regression models should be applied based on the relationships. For example, if the relationship among the variables were quadratic, then a quadratic regression should be applied.
- Step 4:** Create a "unified" PVT measure by combining the PVT measures of RT, FA, and L. The unified measure (S) is expressed as follows:

$$S = W_1 \times RT + W_2 \times FA + W_3 \times L \quad (3)$$

where  $W_1 + W_2 + W_3 = 1$ . The weight values can be set to be either the same or different based on the analyst's needs. For example, if a task requires false alarms to be most important factor to consider, then its weight can be increased. In this paper, we assigned the same weight to each factor. Note that we created the "unified" PVT measure to better investigate the relationships between the

PVT measures and the eye tracking measures. Note that normalized RT, FA, and L values were used, meaning that the minimum and maximum values obtained from all the experiment participants were mapped to 0 and 1.

**Step 5:** Discover an optimal regression model that can predict fatigue using one or more eye-tracking measures. Stepwise regression approach was applied (both forward and backward) to discover the optimal regression model. We assumed that the unified PVT measure accurately represented one's fatigue level, and we found eye tracking measures that could predict fatigue level. All eye tracking measures were normalized, meaning that the minimum and maximum values obtained from all the experiment participants were mapped to 0 and 1. The full model and associated variable for the backward regression is:

$$S = \beta_0 + \beta_1 \times FN + \beta_2 \times FD + \beta_3 \times PS + \beta_4 \times TE + \beta_5 \times SE \quad (4)$$

where  $S$  is the unified PVT measure,  $FN$  is number of eye fixations,  $FD$  is eye fixation duration,  $PD$  is pupil dilation,  $TE$  is transition entropy,  $SE$  is steady state entropy,  $\beta_0$  is the model intercept, and  $\beta_i$  (where  $i = 1, \dots, 5$ ) represents the coefficients for eye movement measures. Note that the forward regression model starts with finding a best predictor variable, then adds more and more variables. Both methods, in general, should produce the same outcome.

#### 4. Experiment

A moderate fidelity flight simulation environment was created that involved the short-haul multiphase flight. Details are as follows.

##### 4.1. Participants

A total of twenty pilots participated in the experiment. Ten participants were defined as “novices” who had moderate expertise (pilot experience: mean of 18 months and SD of 2.4), less than 36 months of experience, and at least met the minimum requirements of 40 h of actual or simulated IFR flights. The other ten participants were defined as “experts” who had more expertise (pilot experience: mean of 42 months SD of 4.5), more than 36 months of experience, and expressed they completed substantially more IFR flight hours (at least more than twice) than the minimum requirement of 40 h. Unfortunately, all pilots were not able to exactly recall their IFR flight hours; therefore, the statistics are not provided.

The power analysis indicated that the sample sizes provided reasonable power of 0.91 for the mixed design of within-subjects design related to the tasks and between-subjects design related to the expertise. In addition, other research papers related to evaluating pilots' performance had a mean sample of ten pilots [5,19,34,46,47].

##### 4.2. Apparatus

Microsoft flight simulator software (i.e., FSX) was used for generating the Boeing B-52 aircraft and the flight scenarios. B-52 aircraft was selected to possibly induce more visual attention from the pilots. However, since a moderate fidelity flight simulator was used, piloting a simulated B-52 should be not as difficult as piloting the actual B-52.

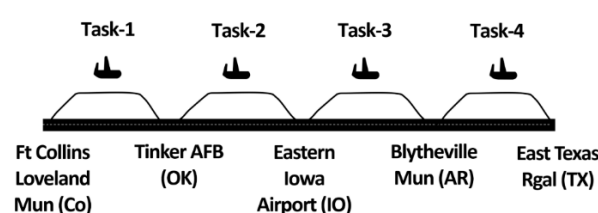
The PVT measures were assessed using the Psychology Experiment Building Language (PEBL) software version 0.13 [48]. Tobi TX 300 eye tracker (having 300 hz data collection rate having 0.5 degrees of visual angle accuracy) and Tobii Studio software was used to collect and process the raw eye tracking data. The I-VT algorithm provided by the Tobii Studio software was applied to calculate the eye fixations. The eye tracking data exported from the software were analyzed using MATLAB and R software. A 21-inch monitor was used for displaying the simulated flight scenarios. A Logitech Extreme 3D Pro Joystick was applied to control the aircraft. A keyboard was used to collect the PVT-related task responses.



### 4.3. Tasks and Procedures

The four consecutive tasks (tasks 1–4, which were equivalent to each leg) are provided in Figure 2. All the participants were instructed to maintain a regular sleep schedule and sleep at 9 pm on the day before the experiment day in order to prevent the possible confounding effect of sleep. The experiment started at 8:30 am and ended around 1:00 pm. At the beginning of the experiment, calibrations were performed to start collecting accurate eye tracking data. Each task lasted for 60 min and involved takeoff, climb, cruise, descent, and landing following the FAA's IFR.

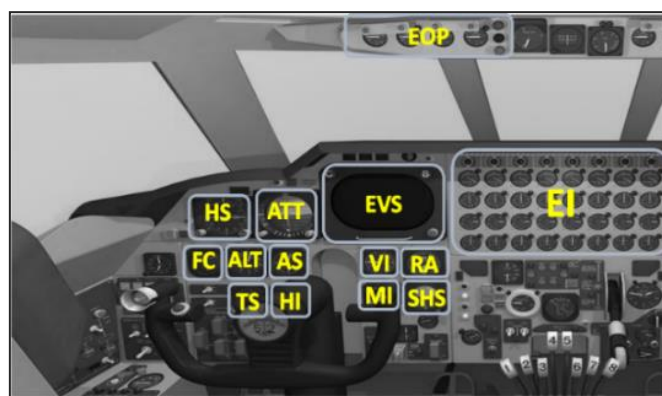
IFR flights mean that the pilot does not have visibility out the window and has to rely on the information obtained from the flight instruments. After completing each flight task, pilots underwent the PVT, which lasted for approximately 5 min (providing 30 stimuli during the 5 min), following the guidelines offered [49]. Therefore, a total of four PVTs were administered for each pilot. Since we used simulator software, the runway configuration was similar among all airports, and no other aircraft were placed on the runway.



**Figure 2.** Four consecutive tasks (without any rest) labeled as tasks 1 through 4. Each task lasted approximately 1 h. The total duration was approximately 4 h.

### 4.4. Measures

The response variables extracted from the PVT task were mean reaction times, mean number of lapses (i.e., number of reaction times greater than 500 milliseconds), and mean number of PVT false starts (i.e., defined as the number of reaction times less than 150 milliseconds). In addition, the unified PVT measure (see Step 4 within the proposed analysis approach above) was calculated using the three PVT measures by assigning equal weights. Context-dependent important AOIs for an IRF flight were identified as shown in Figure 3. The context dependent AOIs were identified based on experts' inputs as well as the collected eye fixation data overlaid on to the visual field of view (see Figure 3) and analyzed using the eye fixations that occurred on those AOIs instead of dividing the whole field of view into AOIs. For example, the front and side windows were not defined as AOIs, since pilots only observed the instruments during the IRF flight. In addition, we verified that the pilots hardly looked through the front and side windows when we analyzed the recorded eye tracking data.



**Figure 3.** Context-specific AOIs that were defined based on the instrument flight rules (IFRs). The AOI names were as follows: engine oil pressure: EOP; engine indicators: EIs; enhanced visual screen:

EVS; attitude indicator: ATT; horizontal situation indicator: HS; flight command indicator: FC; altimeter: ALT; airspeed indicator: AS; true airspeed indicator: TS; heading indicator: HI; vertical velocity indicator: VV; radar altimeter: RA; Mach indicator: MI; standby horizon indicator: SHS. Most of the eye fixations occurred on these AOIs during an IFR flight when we observed the recorded data after the experiments. The response variables related to the eye movements were eye fixation number on the AOIs, eye fixation duration on the AOIs, pupil size, and visual entropy (both transition and stationary entropy).

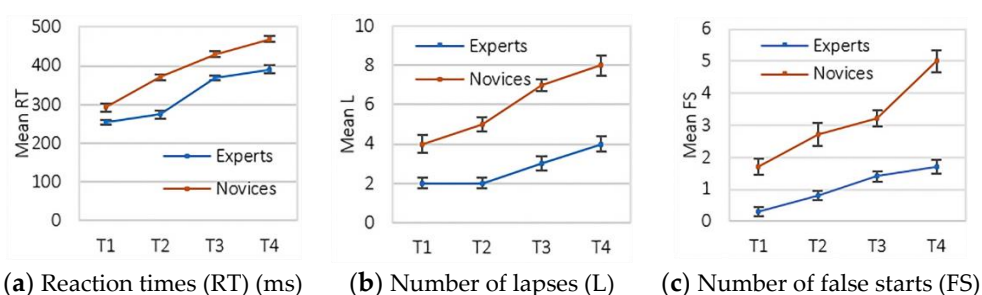
#### 4.5. Data Analysis

Two-way mixed model analysis with repeated measures were applied to consider tasks (i.e., tasks 1–4) and the expertise (i.e., novices vs. experts). After, the relationships among the variables were plotted followed by correlation analysis. After identifying the linear relationships, stepwise regressions were conducted using Equation (4).

### 5. Results

#### 5.1. PVT Measures

Descriptive statistics (i.e., means and standard errors) are plotted in Figure 4. In detail, experts showed faster mean reaction time, lower mean number of lapses, and lower mean number of false starts compared to the novices. In detail, mean reaction time of the experts ( $M = 321.6$  ms (or milliseconds),  $SE = 8.3$  ms) were lower than that of the novices ( $M = 390.3$  ms,  $SE = 8.2$  ms) considering all tasks together. Similarly, expert pilots ( $M = 2.4$ ,  $SE = 0.3$ ) had a lower mean number of lapses than that of the novice pilots ( $M = 5.6$ ,  $SE = 0.4$ ), and experts ( $M = 1.05$ ,  $SE = 0.2$ ) had a lower mean number of false starts compared to novices ( $M = 3.15$ ,  $SE = 0.3$ ). Furthermore, Figure 4 shows that all three PVT measures showed increasing trends from Task 1 up to Task 4 for both the expert and novice pilots.



**Figure 4.** Means and standard errors of the PVT measures.

The mixed-model (i.e., mix of between-subjects design of expertise and within-subjects design of task) analysis results are provided in Table 3. In short, significant differences existed between the experts and novices for all three PVT measures ( $p < 0.001$ ), and significant differences existed among the four tasks for all PVT measures ( $p < 0.001$ ). No outliers were found in the data. Statistical assumptions (i.e., normality and equal variance) were not violated when analyzing the data.

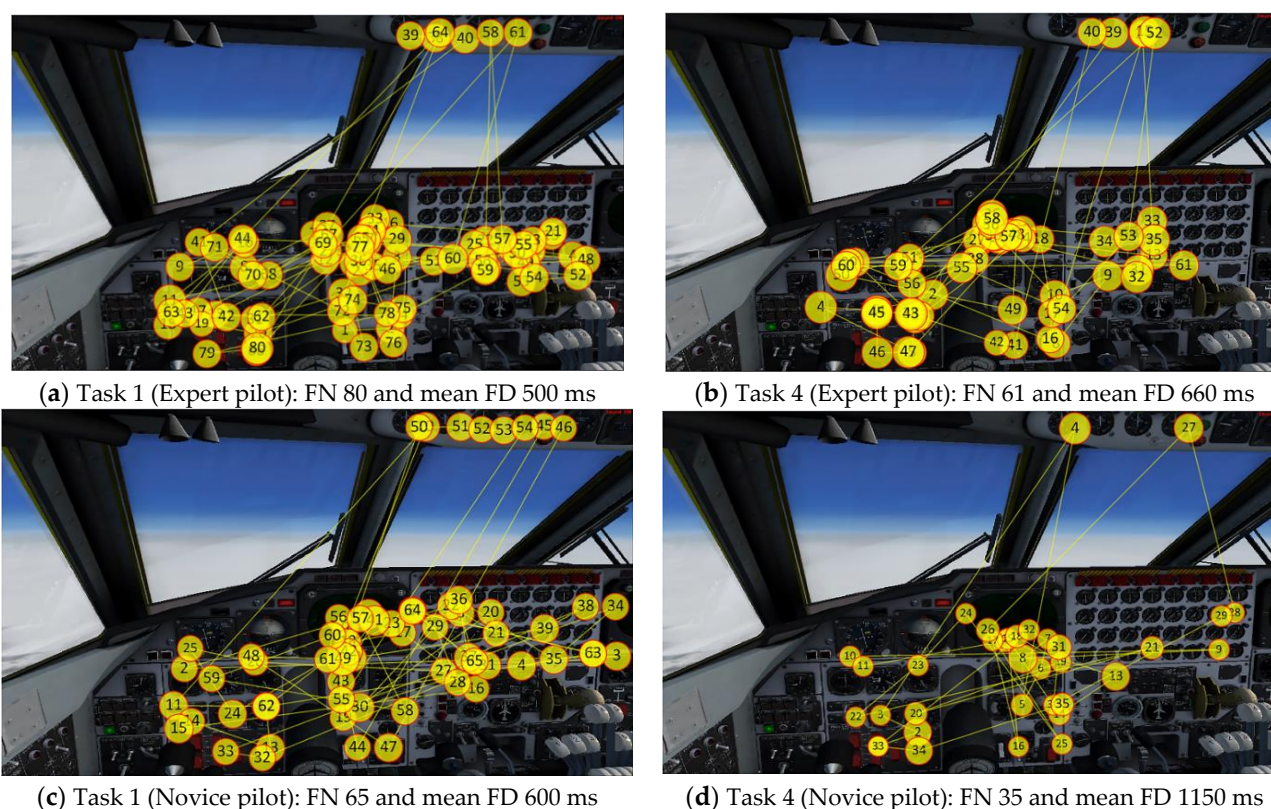
**Table 3.** Results of the mixed model analysis of variance on PVT measures: Exp is expertise factor (experts vs. novices) related to the between-subjects design, and task is the task factor (tasks 1, 2, 3, and 4) related to the within-subjects design.

	Between-Subjects			Within-Subjects		
	F (1,18)	p	$\eta_p^2$	F (3,54)	p	$\eta_p^2$
Reaction time (RT)						
Exp	82.45	<0.001	0.8			
Task				177	<0.001	0.91
Exp $\times$ Task				5.26	<0.003	0.23

Lapse (L)						
Exp	104.7	<0.001	5.26			
Task				35.11	<0.001	0.66
Exp × Task				4.67	<0.001	0.21
False start (FS)						
Exp	90.72	<0.001	0.83			
Task				39.87	<0.001	0.69
Exp × Task				6.99	<0.001	0.28

## 5.2. Eye Movement Measures

Figure 5 shows examples of the visual scanpaths for an expert pilot and a novice pilot for the duration of 40 s during Task 1 (first leg) and Task 4 (last leg). The 40 s duration was selected to show enough eye movements in the figure but also to avoid clutter by showing too many eye movements.



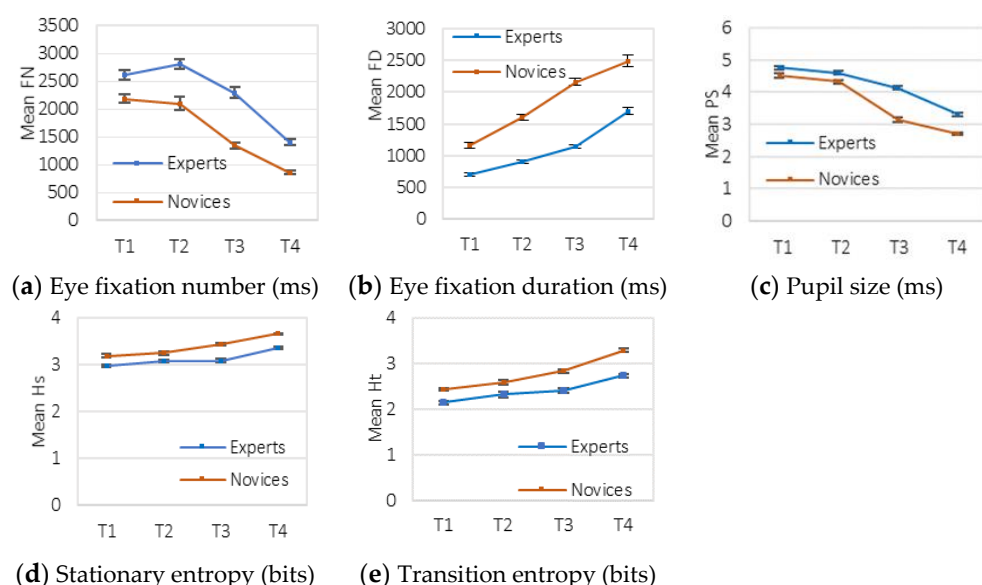
**Figure 5.** Examples of visual scanpaths of an expert and a novice pilot: The yellow circles represent the eye fixations where the numbers represent its index. The yellow lines represent the saccades. The size of the eye fixation circles have been kept at a fixed size for visual clarity. In addition, only 40 s of data are provided for each sample. FN is the eye fixation number, and FD is the eye fixation duration.

In Figure 5, the expert pilot has more eye fixation numbers as compared to the novice for both the tasks. But the expert had lower mean eye fixation duration than the novice. More importantly, the expert showed lower eye fixation numbers in Task 4 than Task 1; however, the mean eye fixations duration higher more in Task 4 than Task 1 (see Figure 5a,b). A similar result was observed for the novice pilot also (see Figure 5c,d).

Descriptive statistics for the overall tasks are plotted in Figure 6. Figure 6a shows that the mean eye fixation number decreased over the course of the flight (i.e., Task 1~Task 4) for both expert and novice pilots.

Overall, the mean eye fixation number trended downwards as the task number increased. Moreover, expert pilots showed higher mean number of eye fixations than those

of the novice pilots for all tasks. On the other hand, mean eye fixation duration showed an increasing trend over the course of the flight for both experts and novices (see Figure 6b). Moreover, novice pilots had higher mean eye fixation duration than experts across all tasks. The pupil size also followed a decreasing trend over the course of the flight. Figure 6c shows that for Tasks 1 and 2, the difference between mean pupil sizes of novice and expert was small. However, this difference increases for Tasks 3 and 4. In addition, the rate of decrease of pupil size was higher for novices than experts.



**Figure 6.** Means and standard errors of the eye movement measures.

In addition, Figure 6d,e shows the stationary entropy ( $H_s$ ) and transition entropy ( $H_t$ ) for four different tasks for both expert and novice pilots. Both stationary entropy and transition entropy showed an increasing trend for both novice and expert pilots as the task index increased (i.e., from Task 1 to Task 4); however, the rate of increase of transition entropy was more prominent for both groups compared to stationary entropy. For both transition and stationary entropies, the novice pilots showed higher values than those of the expert pilots.

The results from the mixed-model analysis show significant effects of both pilot experience and task number, and their interactions on all eye movement measures (see Table 4). The results from the one-way repeated measures analysis of variance show that task number significantly affected the eye movement measures (see Table 5).

**Table 4.** Mixed-model analysis on eye movement measures.

	Between-Subjects			Within-Subjects		
	F (1,18)	p	$\eta_p^2$	F (3,54)	p	$\eta_p^2$
Eye fixation number (FN)						
Exp	72.41	<0.001	0.80			
Task #				157.12	< 0.001	0.89
Exp $\times$ Task #				5.25	< 0.003	0.22
Eye fixation duration (FD)						
Exp	459.9	<0.001	0.96			
Task #				168.75	<0.001	0.90
Exp $\times$ Task #				7.51	<0.001	0.29
Pupil size (PS)						
Exp	101.89	<0.001	0.85			

Task #				408.79	<0.001	0.96
Exp × Task #				21.24	<0.001	0.54
Transition entropy ( $H_t$ )						
Exp	210.88	<0.001	0.92			
Task #				200.75	<0.001	0.92
Exp × Task #				9.15	<0.001	0.34
Stationary entropy ( $H_s$ )						
Exp	119.11	<0.001	0.87			
Task #				75.99	<0.001	0.81
Exp × Task #				3.85	<0.014	0.18

**Table 5.** Results of the one-way repeated measures analysis of variance on eye movements measures in which the task number (tasks 1, 2, 3, and 4) is the factor.

DV	Experts		Novices	
	F (3,27)	p	F (3,27)	p
Eye fixation number (FN)	64.66	<0.001	107.72	<0.001
Eye fixation duration (FD)	151.37	<0.001	71.98	<0.001
Pupil size (PS)	160.11	<0.001	264.57	<0.001
Transition entropy ( $H_t$ )	125.02	<0.001	98.08	<0.001
Stationary entropy ( $H_s$ )	42.39	<0.001	38.59	<0.001

### 5.3. Correlation Results

The increasing and decreasing trends were quantified through the correlation analysis shown in Table 6. In general, the correlation values were high. In addition, the correlation values were higher for novices than experts. In detail, all three PVT measures (i.e., reaction times, number of lapses, and number of false starts) showed positive correlation with eye fixation duration and both the entropy measures, whereas they were negatively associated with eye fixation number and pupil size. Thus, the association between the PVT measures and eye movement measures suggested that the latter might be alternatively used to replace PVT variables in predicting fatigue levels.

**Table 6.** Correlations among PVT measures and eye movement measures.

	Expert			Novice		
	RT	L	FS	RT	L	FS
FN	−0.69	−0.61	−0.49	−0.84	−0.73	−0.76
FD	0.76	0.68	0.61	0.88	0.78	0.74
PS	−0.81	−0.56	−0.53	−0.86	−0.78	−0.70
$H_t$	0.75	0.63	0.63	0.79	0.68	0.70
$H_s$	0.65	0.56	0.59	0.73	0.64	0.62

### 5.4. Regression Models

After identifying that high correlation exist among the variables, two types of regression models (i.e., full models and optimized models) were investigated as provided below. Note that we chose to conduct the regression analysis using the unified PVT measure instead of using each measure separately. Details of the reason are provided in Section 2.1 and the procedure is provided in Section 3 (Step 4).

- (1) Multiple linear regression results: The multiple linear regression analysis using the unified PVT measure ( $S$ ) and all eye movement measures resulted in regression models provided in Equations (5) and (6). The full model of the novice pilots resulted in the overall model fit of adjusted  $R^2 = 0.85$  and  $AIC = -177.04$ . Whereas, the full



model of the expert pilots' group, resulted in the overall model fit of adjusted  $R^2 = 0.66$  and  $AIC = -145.12$ .

Novice pilots (full model):

$$S = 0.46 + (0.42 \times FD) - (0.27 \times FN) - (0.08 \times PS) + (0.42 \times H_t) - (0.40 \times H_s) \quad (5)$$

Expert pilots (full model):

$$S = 0.31 + (0.33 \times FD) - (0.046 \times FN) - (0.07 \times PS) + (0.25 \times H_t) - (0.007 \times H_s) \quad (6)$$

- (2) Stepwise regression results: The results of the stepwise regressions are provided in Equations (7) and (8). For the novice pilots, to predict  $S$ , eye fixation duration ( $\beta = 0.45, p < 0.05$ ), eye fixation number ( $\beta = -0.31, p < 0.05$ ), transition entropy ( $\beta = 0.42, p < 0.05$ ), and stationary entropy ( $\beta = 0.38, p < 0.05$ ) were found to be significant with an overall model fit of  $R^2 = 0.84$  and  $AIC = -178.76$ . For expert pilots, only eye fixation duration ( $\beta = 0.65, p < 0.05$ ) was found significant with a lower model fit  $R^2 = 0.64$  and  $AIC = -151.15$ .

Novice pilots (optimized model):

$$S = 0.43 + (0.45 \times FD) - (0.31 \times FN) + (0.42 \times H_t) - (0.38 \times H_s) \quad (7)$$

Expert pilots (optimized model):

$$S = -0.23 + (0.65 \times FD) \quad (8)$$

Steps of the stepwise regression analysis are provided in Table 7. We conducted both backward and forward methods which resulted in obtaining the same model. Therefore, we only provided the backward steps in Table 7.

**Table 7.** Stepwise regression (backward) results with unified PVT measure as response and eye movement measures as predictors for both expert and novice pilots.

Variables	Expert					Novice	
	Step I	Step II	Step III	Step IV	Step V	Step I	Step II
(Constant)	0.31	0.31	0.29	0.20	0.23	0.46	0.43
FD	0.33	0.33	0.35	0.40	0.65	0.42	0.45
FN	-0.05	-0.05				-0.27	-0.31
PS	-0.07		-0.09			-0.08	0.42
HT	0.25	-0.07	0.25	0.29		0.42	-0.38
HS	-0.007	0.25				-0.40	
Adjusted R	0.61	0.62	0.63	0.64	0.64	0.83	0.84
F	13.26 *	17.06 *	23.32 *	35.61 *	68.72 *	39.17 *	50.0 *
AIC	-145.12	-147.1	-149.02	-150.78	-151.15	-177.03	-178.76

\*  $p < 0.001$ .

Note that it is possible to have positive correlation but obtain a negative regression coefficient in a multiple regression model due to the effect of other variables [50].

## 6. Discussion

In summary, the increase in fatigue was verified through the PVT measures of reaction time, number of lapses, and number of false starts, and the results accord with many previous research efforts in aviation [11,15,21–26]. The results allowed us to devise a unified PVT measure of combining the three measures to quantify a fatigued state as a single point. In addition, as fatigue increased, eye fixation duration increased, visual entropies (i.e., transition and stationary) increased, eye fixation number decreased, and pupil size decreased. The results enabled us to discover viable fatigue prediction models in a multi-leg flight based on expertise and using eye movement measures.

We learned that, unlike novices, the expert pilots had a greater number of eye fixations and shorter eye fixation duration on the context dependent AOIs throughout the flight mission. The results accord with a previous research [34], in which expert pilots fixated upon more instruments and spent less time viewing each individual instrument compared to the novice pilots. We believe that, limited to the piloting task, more eye fixations might indicate more active information processing, whereas longer eye fixation might indicate the pilot needing more time to focus and process the information of interest.

Furthermore, in the case of the pupil size, it became progressively smaller (for both expert and novices) as fatigue increased over the course of the flight mission. The results are similar to existing research [39], which also reported significantly smaller pupil size with increased pilot fatigue. What we have additionally discovered is that the experts' pupil size remained relatively larger compared to the novice pilots as the task number increased (see Figure 6c in the Results section). The size of the pupil varies with a person's state of arousal, with an increase in arousal level resulting in dilation [51]. Therefore, it seems that the novices were affected more by fatigue, meaning that the experts were able to keep the arousal state better than the novices.

Regarding the visual entropy outputs, both the stationary entropy and the transition entropy increased with higher fatigue levels. One possible reason might be that, with higher fatigue levels, pilots' visual search strategy became more random in nature resulting in higher entropy values. Expert pilots showed significantly lower visual entropies (both stationary entropy and transition entropy) compared to novice pilots (see Figure 6d,e in the Results section), indicating that the experts might have applied more overall organized (less random) visual search strategies that can reduce fatigue. Note that we have introduced the concept of entropy to better develop the fatigue prediction model and have not considered characterizing and classifying the visual search strategies. The analysis of the visual search strategies is out of the scope of this paper and requires an in-depth follow-up research.

The regression results show that, depending on the level of expertise of the pilot (experts vs. novices), a different set of eye tracking measures can be used for predicting fatigue. Furthermore, limited to our experiment conditions in a multiphase consecutive flight mission, the optimized models show that some eye movement measures can be more effective at predicting fatigue than others. Specifically, when observing the optimized models (equations 7 and 8), eye fixation duration was a significant predictor variable for both pilot groups, whereas the eye fixation number and the visual entropies can be additionally useful when assessing the fatigue of the novice pilots.

In addition, note that only eye fixation duration (FD) was sufficient in the optimized model for the expert pilots even though the eye fixation numbers (FN) seemed to be equally highly correlated. The reason is that FN was highly correlated with reaction times (RT) and number of lapses (L), but not highly correlated with false starts (FS). Since we used the unified PVT measure that considers all three measures of RT, L, and FS, the step-wise regression analysis resulted in not including FN as a predictor and only using FD was sufficient, in which the results are limited to our experiment conditions. We do not recommend the stakeholders to simply use the computed coefficients and the predictors in their unique environment. We do believe that the stakeholders could benefit by applying our developed research methods. We would be very interested in any insight other researchers could provide if they obtained similar or different regression models.

In detail, the results make us question whether we should only use the optimized models to predict fatigue. The important contribution of this research is that we were able to discover that all eye movement measures introduced in this paper are somewhat strongly correlated with fatigue, and some eye movement measures might better predict fatigue over other measures. The optimized models can vary based on individual differences, experiment settings, and/or the flight task types. Therefore, we recommend that the proposed research approach should be used as a foundation that can be further customized based on individual needs and flight environment.

Furthermore, it will be possible to use each eye movement measure separately or in different combinations to provide multiple evidence (or accumulated evidence) to better detect and verify fatigue levels. To the best of our knowledge, the general guidelines are recommending a break after piloting an aircraft for a certain number of hours or legs. The multimodal analysis approach provided in this paper can be utilized to develop near real-time fatigue detection models that can be used as a tool to manage fatigue-related risk by proactively detecting fatigue of pilots.

## 7. Limitations and Future Research

One of the limitations of this research is that we collected and analyzed the data based on each flight phase (or leg) rather than across a continuous flow of time. The reason that we chose the above option was to compare the eye movement measures against the discrete PVT measures. Therefore, future research involves devising methods to evaluate fatigue continuously using only eye movement measures. The continuous evaluation can be done based on time (i.e., seconds, minutes, hours) or based on detailed events during take-off, cruising, and landing.

Another limitation is in defining an expert and a novice. How to define an expert and a novice has always been an issue raised by the research communities in all applications. Although we have used the thresholds based on the inputs of the flight instructors, opinions can differ, and unfortunately, the participants were not able to precisely recall their IFR flight hours. However, we believe that our classification of the participants into two groups was somewhat successful, as we did obtain distinctive differences between the two groups. We are planning to apply a set of carefully constructed criteria for follow-up research.

In addition, this research is concentrated on providing aggregated outputs. It is possible that individual differences can exist. Therefore, future research involves investigating whether individual eye movement characteristics, especially the individual's visual scanning patterns, differ as fatigue levels increase. The analysis of the visual scanning patterns involves developing algorithms to effectively characterize and compare those differences.

In terms of the methodology, we have proposed the concept of the unified PVT measure, but more in-depth analysis is required on how to assign an optimal weight to each PVT measure. In this research, we assumed equal weights, but our assumption might be incorrect. Discovering an optimal weight value for each PVT measure is a challenging task which can be investigated through various algorithms and associated sensitivity analyses. We are currently working on how to improve the regression models by developing appropriate algorithms that can find optimal weight values.

In addition, we had assumed that the initial fatigue levels of all the participants should be somewhat similar since the sleep and experiment time were controlled to the best of our abilities. In our future research, baseline measurements of initial fatigue should be obtained before the experiment is conducted.

The reason for the significant differences in terms of the PVT measures seems that the experts might have developed more effective visual scanning strategies to reduce fatigue, and more in-depth analysis on the visual scanning strategies will be needed as future research. In more detail, the visual scanpaths were analyzed using the concept of visual entropy in this research; however, the visual scanpaths can be also characterized and classified based on the concept of visual groupings [52] or graph theory [53], among many others that we have published. We are currently investigating viable options, including machine learning [54], to better characterize and classify the visual scanning behaviors that can be used to predict fatigue.

Finally, this research can be used as a foundation to further develop near real-time fatigue detection models that can be used to alert the stakeholders and provide scaffolding options to the pilots, but we currently do not know what the threshold should be to trigger such alerts or the scaffolding options. If we could identify the possible thresholds, then



the alerting and scaffolding options can be used in conjunction with the Boeing Alertness model [55], currently used to develop regulations for duty time limitations. Note that the Boeing Alertness model cannot definitively answer whether the work schedule is acceptable and safe [56], and the fatigue prediction approaches provided in this research might be able to provide a solution, possibly tailored to each pilot.

**Author Contributions:** Conceptualization of the research topic and the methodology were developed by S.N. and Z.K. Experiment scenarios were designed by S.N. Data were collected by S.N. Data analysis approaches were devised by S.N. and Z.K. Data analysis was performed by S.N. and S.M. Data analysis results were validated by Z.K. and K.K. Original draft was prepared by S.N. and S.M. Final draft was prepared by Z.K. and K.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This material was based upon work supported by the National Science Foundation under Grant No. 1943526. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the University of Oklahoma. Approved protocol code is 7325.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the participants to publish this paper.

**Data Availability Statement:** Data are available by contacting the corresponding author, Zihong Kang.

**Acknowledgments:** We sincerely thank the aircraft pilots at the University of Oklahoma who participated in this simulated flight experiment.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Goode, J.H. Are pilots at risk of accidents due to fatigue? *J. Saf. Res.* **2003**, *34*, 309–313, doi:10.1016/s0022-4375(03)00033-1.
2. Wiegmann, D.A.; Shappell, S.A. The Human Factors Analysis and Classification System (HFACS). In *A Human Error Approach to Aviation Accident Analysis*; Routledge: Burlington, VT, USA, 2003; pp 45–71.
3. Li, G.; Baker, S.P.; Grabowski, J.G.; Rebok, G.W. Factors associated with pilot error in aviation crashes. *Aviat. Space, Environ. Med.* **2001**, *72*, 52–58.
4. Oster, C.V.; Strong, J.S.; Zorn, C.K. Analyzing aviation safety: Problems, challenges, opportunities. *Res. Transp. Econ.* **2013**, *43*, 148–164, doi:10.1016/j.retrec.2012.12.001.
5. Hartzler, B.M. Fatigue on the flight deck: The consequences of sleep loss and the benefits of napping. *Accid. Anal. Prev.* **2014**, *62*, 309–318, doi:10.1016/j.aap.2013.10.010.
6. Dismukes, R.K. Effects of Acute Stress on Aircrew Performance: Literature Review and Analysis of Operational Aspects. Available online: [https://human-factors.arc.nasa.gov/publications/NASA\\_TM\\_2015\\_218930-2.pdf](https://human-factors.arc.nasa.gov/publications/NASA_TM_2015_218930-2.pdf) (accessed on 11 July 2021).
7. Lee, S.; Kim, J.K. Factors contributing to the risk of airline pilot fatigue. *J. Air Transp. Manag.* **2018**, *67*, 197–207, doi:10.1016/j.jairtraman.2017.12.009.
8. Bourgeois-Bougrine, S.; Carbon, P.; Gounelle, C.; Mollard, R.; Coblenz, A. Perceived fatigue for short- and long-haul flights: a survey of 739 airline pilots. *Aviat. Space Environ. Med.* **2003**, *74*, 1072–1077.
9. Powell, D.M.C.; Spencer, M.B.; Holland, D.; Broadbent, E.; Petrie, K.J. Pilot fatigue in short-haul operations: Effects of number of sectors, duty length, and time of day. *Aviat. Space Environ. Med.* **2007**, *78*, 698–701.
10. Powell, D.; Spencer, M.B.; Holland, D.; Petrie, K.J. Fatigue in two-pilot operations: implications for flight and duty time limitations. *Aviat. Space Environ. Med.* **2008**, *79*, 1047–1050, doi:10.3357/ASEM.2362.2008.
11. Petrilli, R.M.; Roach, G.; Dawson, D.; Lamond, N. The Sleep, Subjective Fatigue, and Sustained Attention of Commercial Airline Pilots during an International Pattern. *Chrono-Int.* **2006**, *23*, 1357–1362, doi:10.1080/07420520601085925.
12. Samel, A.; Wegmann, H.M.; Vejvoda, M. Jet lag and sleepiness in aircrew. *J. Sleep Res.* **1995**, *4*, 30–36.
13. Honn, K.A.; Satterfield, B.C.; McCauley, P.; Caldwell, J.L.; Van Dongen, H.P. Fatiguing effect of multiple take-offs and landings in regional airline operations. *Accid. Anal. Prev.* **2016**, *86*, 199–208, doi:10.1016/j.aap.2015.10.005.

14. Van Drongelen, A.; Van Der Beek, A.J.; Hlobil, H.; Smid, T.; Boot, C.R. Development and evaluation of an intervention aiming to reduce fatigue in airline pilots: design of a randomised controlled trial. *BMC Public Heal* **2013**, *13*, 776, doi:10.1186/1471-2458-13-776.
15. Arsintescu, L.; Chachad, R.; Gregory, K.B.; Mulligan, J.B.; Flynn-Evans, E.E. The Relationship between Workload, Performance and Fatigue in a Short-Haul Airline. *Chronobiol. Int.* **2020**, *37*, 1492–1494.
16. Binias, B.; Myszor, D.; Palus, H.; Cyran, K.A. Prediction of Pilot's Reaction Time Based on EEG Signals. *Front. Neuroinform.* **2020**, *14*, 6, doi:10.3389/fninf.2020.00006.
17. Borghini, G.; Astolfi, L.; Vecchiato, G.; Mattia, D.; Babiloni, F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* **2014**, *44*, 58–75, doi:10.1016/j.neubiorev.2012.10.003.
18. Dehais, F.; Dupres, A.; Di Flumeri, G.; Verdiere, K.; Borghini, G.; Babiloni, F.; Roy, R. Monitoring Pilot's Cognitive Fatigue with Engagement Features in Simulated and Actual Flight Conditions Using an Hybrid FNIRS-EEG Passive BCI. In Proceedings of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 7–10 October 2018.
19. Di Stasi, L.L.; McCamy, M.B.; Martinez-Conde, S.; Gayles, E.; Hoare, C.; Foster, M.; Catena, A.; Macknik, S.L. Effects of long and short simulated flights on the saccadic eye movement velocity of aviators. *Physiol. Behav.* **2016**, *153*, 91–96, doi:10.1016/j.physbeh.2015.10.024.
20. Diaz-Piedra, C.; Rieiro, H.; Suárez, J.; Rios-Tejada, F.; Catena, A.; Di Stasi, L.L. Fatigue in the military: towards a fatigue detection test based on the saccadic velocity. *Physiol. Meas.* **2016**, *37*, N62–N75, doi:10.1088/0967-3334/37/9/n62.
21. Wu, X.; Wanyan, X.; Zhuang, D. Pilot's visual attention allocation modeling under fatigue. *Technol. Heal. Care* **2015**, *23* (Suppl. S2), S373–S381, doi:10.3233/THC-150974.
22. Naeeri, S.; Kang, Z. Exploring the Relationship between Pilot's Performance and Fatigue When Interacting with Cockpit Interfaces. In Proceedings of the 2018 IISE Annual Conference, Orlando, FL, USA, 19–22 May 2018.
23. Naeeri, S.; Mandal, S.; Kang, Z. Analyzing Pilots' Fatigue for Prolonged Flight Missions: Multimodal Analysis Approach Using Vigilance Test and Eye Tracking. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Seattle, WA, USA, 28 October–1 November 2019; Volume 63, pp. 111–115.
24. Gander, P.H.; Signal, T.L.; Berg, M.V.D.; Mulrine, H.M.; Jay, S.M.; Mangie, C.J. In-flight sleep, pilot fatigue and Psychomotor Vigilance Task performance on ultra-long range versus long range flights. *J. Sleep Res.* **2013**, *22*, 697–706, doi:10.1111/jsr.12071.
25. Gander, P.H.; Mulrine, H.M.; Berg, M.V.D.; Smith, A.A.T.; Signal, T.L.; Wu, L.J.; Belenky, G. Pilot Fatigue: Relationships with Departure and Arrival Times, Flight Duration, and Direction. *Aviat. Space Environ. Med.* **2014**, *85*, 833–840, doi:10.3357/ase.3963.2014.
26. Thomas, L.C.; Gast, C.; Grube, R.; Craig, K. Fatigue Detection in Commercial Flight Operations: Results Using Physiological Measures. *Procedia Manuf.* **2015**, *3*, 2357–2364, doi:10.1016/j.promfg.2015.07.383.
27. Caldwell, J.A.; Hall, K.K.; Erickson, B.S. EEG Data Collected from Helicopter Pilots in Flight Are Sufficiently Sensitive to Detect Increased Fatigue from Sleep Deprivation. *Int. J. Aviat. Psychol.* **2002**, *12*, 19–32.
28. Naeeri, S.; Kang, Z.; Mandal, S. Exploring the effect of fatigue on pilot performance during single and multi-takeoffs and landings flight missions. In Proceedings of the 7th Annual World Conference of the Society for Industrial and Systems Engineering, Binghamton, NY, USA October 11–12, 2018; Volume 7, pp. 174–181.
29. Millar, M. "Measuring Fatigue Overview," in Asia – Pacific, ICAO / IATA / IFALPA FRMS Seminar, Bangkok, 2012.
30. Bleichner, M.G.; Debener, S. Concealed, Unobtrusive Ear-Centered EEG Acquisition: cEEGrids for Transparent EEG. *Front. Hum. Neurosci.* **2017**, *11*, 163, doi:10.3389/fnhum.2017.00163.
31. Bodala, I.P.; Li, J.; Thakor, N.V.; Al-Nashash, H. EEG and Eye Tracking Demonstrate Vigilance Enhancement with Challenge Integration. *Front. Hum. Neurosci.* **2016**, *10*, 273, doi:10.3389/fnhum.2016.00273.
32. Lu, T.; Lou, Z.; Shao, F.; Li, Y.; You, X. Attention and Entropy in Simulated Flight with Varying Cognitive Loads. *Aerosp. Med. Hum. Perform.* **2020**, *91*, 489–495, doi:10.3357/amhp.5504.2020.
33. Diaz-Piedra, C.; Rieiro, H.; Cherino, A.; Fuentes, L.J.; Catena, A.; Stasi, L.L.D. The Effects of Flight Complexity on Gaze Entropy: An Experimental Study with Fighter Pilots. *Appl. Ergon.* **2019**, *77*, 92–99.
34. Bellenkes, A.H.; Wickens, C.D.; Kramer, A. Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. *Aviat. Space Environ. Med.* **1997**, *68*, 569–579.
35. Thorne, D.R.; Johnson, D.E.; Redmond, D.P.; Sing, H.C.; Belenky, G.; Shapiro, J.M. The Walter Reed palm-held psychomotor vigilance test. *Behav. Res. Methods* **2005**, *37*, 111–118, doi:10.3758/bf03206404.
36. Lee, I.-S.; Bardwell, W.A.; Ancoli-Israel, S.; Dimsdale, J.E. Number of Lapses during the Psychomotor Vigilance Task as an Objective Measure of Fatigue. *J. Clin. Sleep Med.* **2010**, *6*, 163–168, doi:10.5664/jcsm.27766.
37. Dinges, D.F.; Mallis, M.M.; Maislin, G.; Powell, J.W. *Evaluation of Techniques for Ocular Measurement as an Index of Fatigue and as the Basis for Alertness Management*; National Highway Traffic Safety Administration: Washington, DC, USA, 1998.
38. Dinges, D.F.; Maislin, G.; Brewster, R.M.; Krueger, G.P.; Carroll, R.J. Pilot test of fatigue management technologies. *Transp. Res. Rec.* **2005**, *1922*, 175–182.
39. LeDuc, P.A.; Greig, J.L.; Dumond, S.L. *Self-Report and Ocular Measures of Fatigue in U.S. Army Apache Aviators Following Flight* Army Aeromedical Research Lab: Fort Rucker, AL, USA, 2005.
40. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.

41. Shiferaw, B.; Downey, L.; Westlake, J.; Stevens, B.; Rajaratnam, S.; Berlowitz, D.J.; Swann, P.; Howard, M.E. Stationary gaze entropy predicts lane departure events in sleep-deprived drivers. *Sci. Rep.* **2018**, *8*, 1–10, doi:10.1038/s41598-018-20588-7.
42. Wu, C.; Cha, J.; Sulek, J.; Zhou, T.; Sundaram, C.P.; Wachs, J.; Yu, D. Eye-Tracking Metrics Predict Perceived Workload in Robotic Surgical Skills Training. *Hum. Factors: J. Hum. Factors Ergon. Soc.* **2020**, *62*, 1365–1386, doi:10.1177/0018720819874544.
43. Krejtz, K.; Szmidt, T.; Duchowski, A.T.; Krejtz, I. Entropy-Based Statistical Analysis of Eye Movement Transitions. In Proceedings of the Symposium on Eye Tracking Research and Applications, Safety Harbor, FL, 26–28 March 2014.
44. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons, Inc.: New York, NY, USA, 2006.
45. Han, X.; Shao, Y.; Yang, S.; Yu, P. Entropy-Based Effect Evaluation of Delineators in Tunnels on Drivers' Gaze Behavior. *Entropy* **2020**, *22*, 113, doi:10.3390/e22010113.
46. Gateau, T.; Durantin, G.; Lancelot, F.; Scannella, S.; Dehais, F. Real-Time State Estimation in a Flight Simulator Using fNIRS. *PLoS ONE* **2015**, *10*, e0121279, doi:10.1371/journal.pone.0121279.
47. Kirby, C.E.; Kennedy, Q.; Yang, J.H. Helicopter pilot scan techniques during low-altitude high-speed flight. *Aviat. Space Environ. Med.* **2014**, *85*, 740–744, doi:10.3357/asem.3888.2014.
48. Mueller, S.T.; Piper, B.J. The Psychology Experiment Building Language (PEBL) and PEBL Test Battery. *J. Neurosci. Methods* **2014**, *222*, 250–259, doi:10.1016/j.jneumeth.2013.10.024.
49. Lim, J.; Dinges, D.F. Sleep Deprivation and Vigilant Attention. *Ann. N. Y. Acad. Sci.* **2008**, *1129*, 305–322, doi:10.1196/annals.1417.002.
50. Falk, R.F.; Miller, N.B. *A Primer for Soft Modeling*; University of Akron: Akron, OH, USA, 1992.
51. Bradley, M.M.; Miccoli, L.; Escrig, M.A.; Lang, P.J. The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* **2008**, *45*, 602–607, doi:10.1111/j.1469-8986.2008.00654.x.
52. Kang, Z.; Landry, S.J. An Eye Movement Analysis Algorithm for a Multielement Target Tracking Task: Maximum Transition-Based Agglomerative Hierarchical Clustering. *IEEE Trans. Hum. Mach. Syst.* **2015**, *45*, 13–24.
53. Mandal, S.; Kang, Z. Using Eye Movement Data Visualization to Enhance Training of Air Traffic Controllers: A Dynamic Network Approach. *J. Eye Mov. Res.* **2018**, *11*, 1–20, doi:10.16910/jemr.11.4.1.
54. Gao, X.-Y.; Zhang, Y.-F.; Zheng, W.-L.; Lu, B.-L. Evaluating Driving Fatigue Detection Algorithms Using Eye Tracking Glasses. In Proceedings of the 7th International IEEE/EMBS Conference on Neural Engineering (NER), Montpellier, France, 22–24 April 2015.
55. Powell, D.M.C.; Spencer, M.B.; Petrie, K.J. Comparison of In-Flight Measures with Predictions of a Bio-Mathematical Fatigue Model. *Aviat. Space Environ. Med.* **2014**, *85*, 1177–1184, doi:10.3357/asem.3806.2014.
56. Williamson, A.; Lombardi, D.A.; Folkard, S.; Stutts, J.; Courtney, T.; Connor, J. The link between fatigue and safety. *Accid. Anal. Prev.* **2011**, *43*, 498–515, doi:10.1016/j.aap.2009.11.011.