frontiers in Psychology

ORIGINAL RESEARCH
published: 20 April 2021
doi: 10.3389/fpsyg.2021.555446

Correlation Between Physiological and Performance-Based Metrics to Estimate Pilots' Cognitive Workload

*P. Archana Hebbar[1,2], Kausik Bhattacharya[1], Gowdham Prabhakar[1], Abhay A. Pashilkar[2] and Pradipta Biswas[1]**

[1] I3D Lab, Centre for Product Design and Manufacturing, Indian Institute of Science (IISc), Bengaluru, India, [2] Council of Scientific & Industrial Research (CSIR)-National Aerospace Laboratories, Bengaluru, India

This paper discusses the utilization of pilots' physiological indications such as electroencephalographic (EEG) signals, ocular parameters, and pilot performance-based quantitative metrics to estimate cognitive workload. The study aims to derive a non-invasive technique to estimate pilot's cognitive workload and study their correlation with standard physiological parameters. Initially, we conducted a set of user trials using well-established psychometric tests for evaluating the effectiveness of pupil and gaze-based ocular metrics for estimating cognitive workload at different levels of task difficulty and lighting conditions. Later, we conducted user trials with the NALSim flight simulator using a business class Learjet aircraft model. We analyzed participants' ocular parameters, power levels of different EEG frequency bands, and flight parameters for estimating variations in cognitive workload. Results indicate that introduction of secondary task increases pilot's cognitive workload significantly. The beta frequency band of EEG, nearest neighborhood index specifying distribution of gaze fixation, L1 Norm of power spectral density of pupil diameter, and the duty cycle metric indicated variations in cognitive workload.

Keywords: flight simulator, EEG, ocular parameters, pupil dilation, saccades, cognitive load

OPEN ACCESS

Edited by:
Varun Dutt,
Indian Institute of Technology Mandi, India

Reviewed by:
Peter Thomas,
University of Hertfordshire, United Kingdom
Pratik Chaturvedi,
Defence Terrain Research Laboratory (DRDO), India

***Correspondence:**
Pradipta Biswas
pradipta@iisc.ac.in

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *24 April 2020*
Accepted: *03 March 2021*
Published: *20 April 2021*

Citation:
Hebbar PA, Bhattacharya K, Prabhakar G, Pashilkar AA and Biswas P (2021) Correlation Between Physiological and Performance-Based Metrics to Estimate Pilots' Cognitive Workload. Front. Psychol. 12:555446. doi: 10.3389/fpsyg.2021.555446

INTRODUCTION

It is well-known that pilot's cognitive workload has an impact on performance and, in turn, on flight safety. When workload is high, pilots pay less attention to the task at hand and their performance deteriorates due to narrowing of attention (Wanyan et al., 2011). Sufficiently low workload causes boredom, resulting in degradation in performance (Yerkes and Dodson, 1908). Designers need to consider these constraints for optimizing any pilot vehicle interface (PVI) designs. This is possible with an automatic estimation of pilot's cognitive workload. Other potential areas where pilot's cognitive workload estimation could be beneficial are:

a. to design adaptive automation strategies based on human performance envelope (Thomas et al., 2015; Biella et al., 2017).
b. to provide a basis for ergonomic design evaluation of aircraft cockpit display interface (Zongmin et al., 2014).
c. to reason the causes of performance degradation for certain flight demands (Lee, 2010).
d. to establish the performance limits for an aircraft with poor handling qualities (Harper and Cooper, 1986).
e. to assist instructors in creating a sophisticated pilot assessment methodology (Ryffel et al., 2019; Rudi et al., 2020).

However, as flying an aircraft is a complex task, we need to consider innumerable physiological and psychological factors while evaluating pilot's cognitive workload. Even though subjective methods like NASA TLX questionnaire or Cooper Harper ratings are more prevalent in practice, researchers have proven that physiological variables are more sensitive for estimating cognitive workload (Causse et al., 2015; Trejo et al., 2015; Li et al., 2016). Several researchers have been exploring different physiological measures to quantify pilot's cognitive workload. Sharma et al. (2012) estimated pilot's cognitive workload using a spatial disorientation simulator and measured heart rate, respiratory rate, and galvanic skin responses. The study shows that physiological measures provide more valuable instantaneous information than subjective measures. Another finding of the study is that cardiac activity is a useful measure of cognitive processes. In another study, Othman and Romli (2015) employed multi-index evaluation to estimate cognitive workload, where the percentage of mean pupil dilation was evaluated along with subjective methods. According to a recent study by Mohanavelu et al. (2020), the effect of varying visibility conditions on pilots' cognitive demands could be evaluated through HRV features, pilot performance measures, and subjective assessment methods. It was found that even though pilots' performance scores were similar, the physiological measures were statistically significant. A comparative study by Gentili et al. (2014) revealed that when compared to EEG, HRV was less sensitive to variations in cognitive workload.

Hence, it is evident that researchers have extensively explored various psychophysiological measures such as brain-related measures (ERP, EEG, MEG, and brain metabolism), ocular measures (fixations, scan path, blinks, and pupil diameter), cardiac measures (HRV), and facial expression measures. However, there are few studies that correlate the different independent physiological and pilot performance-based parameters. There are even fewer studies that examine multiple measurement methods in a controlled experimental environment. This research work broadly covers the following objectives:

1. To ascertain the robustness of the proposed ocular parameters to distinguish variations in cognitive workload.
2. To design and conduct a realistic user experimental study using a flight simulator that simulates real-life flight environment as encountered by pilots.
3. To study various modalities of cognitive workload estimation and understand the significance of secondary tasks on pilot's cognitive workload.
4. To find relation among physiological measures such as eye gaze and EEG-based measurement and flying performance-based measures, and report differences among them.

We conducted two different user studies with the help of participants from our university. Ethical approval was taken from the Institute's Ethics committee for undertaking eye gaze tracking-based user studies. Written informed consent was also obtained from the participants for the publication of any potentially identifiable images and data used in this study. Our first study investigated differences in values of ocular metrics for standard psychometric tests in the laboratory to establish the robustness of the metrics to differentiate cognitive workload. In the second study, we conducted 36 flight simulator experiments with 12 participants. These trials were conducted for three different task scenarios. Participants' ocular parameters, EEG band power variations, and their flying performance parameters were recorded and analyzed. In total, we investigated 11 independent metrics to measure cognitive workload. This includes two pupil dilation-based ocular metrics, two gaze-based ocular metrics, variations in the median values of five different EEG frequency bands, and two pilot performance-based metrics. However, due to the limited resources available, other physiological measures such as heart rate variability and facial expression recognition could not be evaluated.

The choice of flight scenario was based on the recent incident and accident survey. A recent study by Boeing (2018)shows that taxing, climbing, approach, and landing are critical phases of civil aircraft flight. We have developed flight scenarios for taxing, take-off, and climb segments in the experimental flight simulator study reported in this paper. As the participants were non-pilots, scenarios were designed to increase task difficulty levels, starting from simple take-off, then with more monitoring and control requirements and additional secondary tasks. Results suggest that introducing the secondary task causes a significant increase in pilot's cognitive workload. This is observed in all the three estimated metrics, namely, EEG, ocular, and pilot performance-based metrics. Correlation between different parameters is explained in detail in section EEG Signal Analysis. Analysis results indicate a positive correlation among the three metrics.

This paper is organized as follows: The next section gives details of the literature survey of relevant research work in the field of cognitive workload estimation. Section User Study on Psychometric Tests presents the results of the comparison of ocular parameters for psychometric tests. Section Flight Simulator Study discusses the structure of the flight simulation experiments, followed by analysis of results. Section General Discussions deliberates on the results and the implications thereof. Section Conclusions concludes the results and discusses the future course of action.

RELATED RESEARCH

There is a plethora of research articles that discuss cognitive workload measurement methodologies. In this study, we have dealt with three types of cognitive workload measures: EEG-based, ocular parameter-based, and flying performance-based metrics. Accordingly, our discussion in this section has been limited to the above methods only.

Electroencephalogram (EEG) Signals

EEG is the measurement of brain's electrical activity. EEG signals are recorded through the EEG electrodes placed on the participant's scalp surface. Several studies validate that EEG power in different frequency bands is sensitive to changes in cognitive demand (Gevins et al., 1997; Petkar et al., 2009; Antonenko et al., 2010; Pavlov and Kotchoubey, 2017; Friedman et al., 2019). In a similar study, Cheng and Hsu (2011) estimated workers' fatigued state using EEG signal measurement. The study found out that an increased EEG activity in the theta band

indicates decreased levels of attention. Borghini et al. (2012) introduced an EEG-based cerebral workload index to detect the driver's mental efforts during different levels of difficulty. This method was based on the estimation of increase in EEG power spectra. Schrauf et al. (2011) described EEG alpha spindles and alpha band power to be indicators of the driver's task performance during secondary auditory tasks. These findings in general suggest that EEG signal power levels are strong indicators of variations in cognitive workload.

Ocular Parameters

Eye-tracking is a well-researched area of study for measuring cognitive workload (Hess, 1975; Kramer, 1991; Hyönä et al., 2003; Palinko et al., 2010; Babu et al., 2019). Ocular parameter-based measures for cognitive workload measurement can be categorized as pupil dynamics-based and fixation-based measures.

Studies suggest that the pupil dilates more with increase in cognitive workload (Marshall, 2007; Biswas and Langdon, 2015). Demberg and Sayeed (2016) study provides evidence of higher rates of rapid pupil dilations for more difficult task conditions. Prabhakar and Biswas (2018) study discussed evidence of using velocity of saccadic intrusion (SI) to detect the distraction of automobile drivers. The study also discussed the application of pupil dilation and fixation duration metrics for estimating cognitive workload. In a similar study, Abadi and Gowen (2004) used SI and micro-saccade rates to estimate cognitive workload. In another study, Xu et al. (2011) used non-intrusive remotely mounted eye trackers to measure variations of pupillary responses with cognitive workload. The study proved that pupil tracking is effective even with varying luminance conditions.

The distribution pattern of eye fixations is another proven cognitive load measure (Di Nocera et al., 2007). The visual scanning patterns in nominal environments tend to be deterministic and repetitive at regular intervals. The order of visual scanning tends to be more random with increase in cognitive workload. De Nocera suggested a widely used distance indicator called Nearest Neighbor Index (NNI) as a sensitive measure to perceive cognitive workload.

Performance-Based Methods

Performance-based methods are indirect measures of cognitive workload. They are based on the assumption that an increase in task difficulty results in deterioration of performance, which increases the pilot's cognitive workload (or reduces the working memory capacity) (Wei et al., 2014). The simplest of these methods is the time domain statistical methods such as root mean squared error (RMSE), standard deviation of error, number of deviations outside tolerance, and computation of reaction time (Reising et al., 1995). For example, Smith and Caldwell (2004) conducted exhaustive simulated flight experiments to study pilot fatigue using RMSE. According to Ebbatson et al. (2007) how a pilot operates his/her control is also an indication of workload. Authors used power spectral density and autocorrelation coefficient of the control column data to infer pilots' control strategy.

Cognitive workload experienced by the participants is also indicated through his/her inceptor control strategy. Two such measures of pilots' efforts are the duty cycle (DC) and aggressiveness (Shepherd et al., 2009). Aggressiveness is the rate of change of inceptor control movements. DC indicates the percentage of time a participant controls his/her input on the inceptor. Hanson et al. (2014) have observed that an increase in aggressiveness and DC is an indicator of increased pilot workload.

To summarize, a variety of physiological and performance-based methods have been defined, tested, and validated to quantify pilot's cognitive workload. Correlation between different methods has also been reported in the literature. Bodala et al. (2016) inferred a positive correlation between pupil saccadic velocity and EEG theta frequency amplitude with increasing task difficulty. In another interesting study conducted by Scharinger et al. (2015) on investigation of working memory on reading comprehension, the authors concluded that pupil dilation dynamics functions as a global workload measure that includes motivational cognitive workload aspects. However, as per the authors, EEG band power is a more promising measure for identifying variations in cognitive processes. However, in another study by Borys et al. (2017), the authors reveal eye movement measures to be a good indicator of cognitive workload. The authors could not establish a significant relation between EEG and cognitive measures.

Hence, there have been such initial studies reporting the correlation between EEG and pupil dilation data analysis in basic research. However, to the best of the author's knowledge, there is limited published research carried out to investigate the relation between physiological parameters with flying performance-based parameters such as aggressiveness and DC. One of the aims of this study is to derive a correlation between the abovementioned parameters.

USER STUDY ON PSYCHOMETRIC TESTS

In this section, we describe a user study that was conducted to validate if L1 Norm of Spectrum (L1NS), Standard Deviation of Pupil (STDP), Low Pass Filter (LPF) of pupil diameter saccade rate, fixation rate, and median SI velocity can distinguish between different cognitive workloads of participants caused by task difficulty. Detailed description of metrics and their implementation can be found in Prabhakar et al. (2020). We used psychometric tests like the N-back test and arithmetic questions to assess the increase in participants' cognitive workload with increased task difficulty. We chose these tests as they were associated with working memory load (Marshall, 2007; Tokuda et al., 2011). Since the pupil dilation is sensitive (Beatty and Lucero-Wagoner, 2000; Vrzakova and Bednarik, 2012) to ambient light variation, we evaluated both the N-back test and arithmetic test in dark rooms as well as varying light conditions in the same room. While we evaluated the N-back test in both auditory and visual presentations, an arithmetic test was conducted only in auditory presentation.

We hypothesized that L1NS, STDP, LPF, saccade rate, fixation rate, and median SI velocity

1. are robust to ambient light variations

2. can be used to distinguish different levels of cognitive workload with respect to change in task difficulty of visual and auditory tasks.

Participants

We collected data from 21 participants (16 male and 5 female) with an average age of 26 years from our university. We chose participants randomly such that the group had a mixture of people wearing and not wearing prescription lenses. Participants wearing lenses had either spherical or cylindrical or both types of powers.

Materials

We collected data using Tobii Pro Glasses 2. We affixed two ambient light sensor modules, one sensor on either side of the glass frame, to capture illumination variations on both eyes independently (**Figure 1**). We used a Dell 17″ monitor to display numbers for visual N-Back and a Logitech keyboard to press the space bar to respond to the N-back test. We also used a Bose SoundLink speaker for an auditory cue in the auditory N-back test.

Design

We undertook the following three tests:

1. Auditory N-back Test
2. Visual N-back Test
3. Auditory Arithmetic Test

The auditory tests were carried out in the dark as well as in dynamically varying light conditions. The room illuminance was varied from 0 to 150 lux by turning on and off a set of lights. The variation of illuminance was randomized.

N-Back Test

The N-back test had three levels of difficulties, viz., 1-back, 2-back, and 3-back. Participants were shown/spelled one stimulus (sequence of one-digit numbers from 1 to 9) in intervals of 2 s and had to press the space bar if the current stimulus matches the previous one (1-back), or second previous (2-back), or third previous (3-back). The N-back test levels were randomized to avoid the order effect. We developed software (Bjäreholt, 2014) to spell out/visually display numbers in N-back and to log the response from participants with a local time stamp.

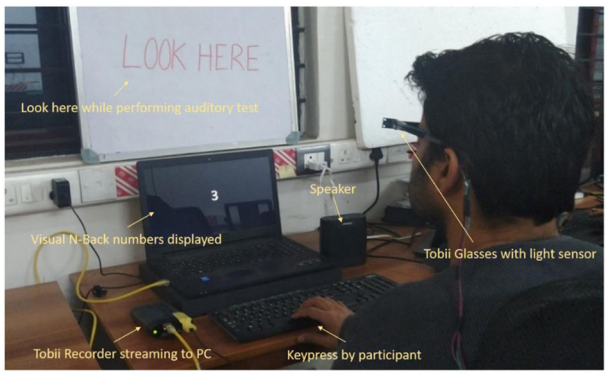

FIGURE 1 | Participant performing the visual N-back test.

Arithmetic Test

The arithmetic test had three levels, viz., easy, medium, and difficult. We developed a tool using python to read out questions using the Text-to-Speech engine in an arithmetic test. We recorded participants' response using the following steps:

1. Software read out all questions loudly.
2. Participant answered to questions loudly.
3. Instructor checked the answer and pressed the right/wrong key to log the event.

The difficulty levels were randomized to avoid the order effect.

Procedure

Participants were asked to wear the Tobii glass affixed with light sensor modules. They were instructed to look at a poster pasted on the wall in front of them and to concentrate on the auditory task given to them. They were asked neither to close their eyes and nor to look around during answering the questions such that the tracker always detected eyes. Participants were explained about the N-back task and arithmetic task. They could practice the 1-back test before the actual trial in order to avoid the learning effect. The time stamps from logged events were used to synchronize the pupil/gaze data corresponding to the start and stop of N-back tests and arithmetic tests. We calculated L1NS, STDP, LPF, saccade rate, fixation rate, and median SI velocity corresponding to events. We checked if these metrics were high for 3-back compared to 2-back and to 1-back. We also checked if these metrics were high for difficult compared to medium and to easy arithmetic levels.

Result

Performance of Tests

We measured performance of the tests as accuracy calculated from the confusion matrix as described in **Table 1**. The accuracy of the N-back test is calculated as

$$\text{Accuracy} = \frac{correct + avoid}{correct + wrong + avoid + missed}$$

TABLE 1 | Performance of the N-back test in terms of accuracy.

	1-Back/Easy	2-Back/Medium	3-Back/Difficult
Auditory N-back dark room	0.961 (0.073)	0.962 (0.059)	0.874 (0.122)
Auditory N-back dynamic light room	0.972 (0.084)	0.948 (0.090)	0.889 (0.121)
Visual N-back	0.985 (0.041)	0.942 (0.071)	0.891 (0.132)
Auditory Arithmetic dark room	0.992 (0.036)	0.905 (0.135)	0.770 (0.207)
Auditory Arithmetic dynamic light room	0.968 (0.067)	0.937 (0.134)	0.730 (0.318)

and accuracy of the arithmetic test is calculated as

$$\text{Accuracy} = \frac{correct}{correct + wrong}$$

As the groups did not follow normality, we performed signed-rank test for each pair and found that accuracy of 3-Back/Difficult was significantly ($p < 0.05$) less than that of 1-Back/Easy for all the tests. The accuracy of 3-back/Difficult was significantly ($p < 0.05$) less than 2-Back/Medium for the auditory N-back dark room and both arithmetic tests. Accuracy of 2-back/Medium was significantly ($p < 0.05$) less than 1-back/Easy for visual N-back and auditory arithmetic dark room.

TABLE 2 | Repeated measure one-way ANOVA for each metric with effect size.

L1NS Right eye	$F_{(2,19)} = 6.419, p < 0.05, \eta^2 = 0.403$
L1NS Left eye	$F_{(2,19)} = 33.964, p < 0.05, \eta^2 = 0.781$
STDP Right eye	$F_{(2,19)} = 7.849, p < 0.05, \eta^2 = 0.452$
STDP Left eye	$F_{(2,19)} = 29.408, p < 0.05, \eta^2 = 0.756$
LPF Left eye	$F_{(2,19)} = 30.718, p < 0.05, \eta^2 = 0.764$

Visual N-Back (Pupil Dilation)

A repeated measure one-way ANOVA for metrics in Visual N-back is described in **Table 2**.

We found that L1NS and STDP of both eyes were significantly (t-test: $p < 0.05$) higher for 3-back than for 1-back. Similarly, 3-back was significantly (t-test: $p < 0.05$) higher than 2-back. We also found that LPF of the left eye was significantly (t-test: $p < 0.05$) higher for 3-back than for 1-back and higher for 3-back than for 2-back. We did not find any significant difference for saccade rate, fixation rate, and median SI velocity. A comparison graph of L1NS for visual N-back is given in **Figure 2**.

Auditory N-Back Dark Room (Pupil Dilation)

A repeated measure one-way ANOVA for metrics in Auditory N-back darkroom is described in **Table 3**.

We found that L1NS and STDP of both eyes, as well as LPF of the left eye, were significantly (t-test: $p < 0.05$) higher for 3-back than for 1-back. We did not find significant difference in saccade rate, fixation rate, and median SI velocity. A comparison graph of L1NS for auditory N-back in the darkroom is shown in **Figure 2**.

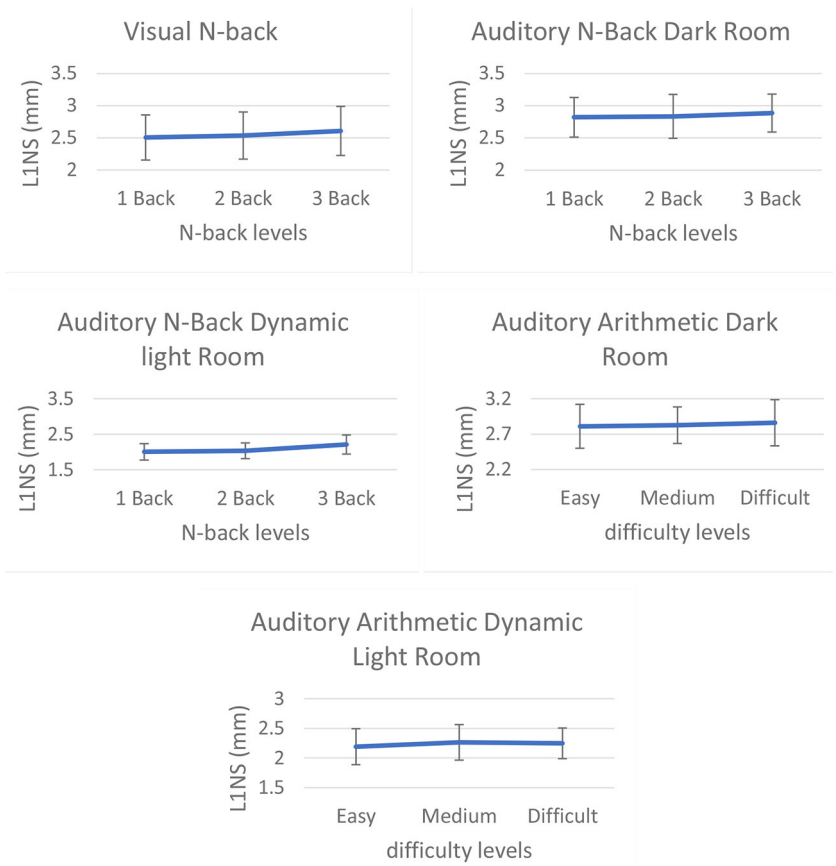

FIGURE 2 | L1NS of the right eye for (from top left) visual N-back, auditory N-Back dark room, auditory N-back dynamic lightroom, auditory arithmetic dark room, and auditory arithmetic dynamic lightroom.

Auditory N-Back Dynamic Light Room (Pupil Dilation)

A repeated measure one-way ANOVA for metrics in Auditory N-back dynamic lightroom is described in **Table 4**.

We found that L1NS, STDP, and LPF of both eyes were significantly (t-test: $p < 0.05$) higher for 3-back than for 1-back. Similarly, 3-back was significantly (t-test: $p < 0.05$) higher than 2-back. We did not find a significant difference for saccade rate, fixation rate, and median SI velocity. A comparison graph of L1NS for auditory N-back in the dynamically lit room is shown in **Figure 2**.

Auditory Arithmetic Dark Room (Pupil Dilation)

A repeated measure one-way ANOVA for metrics in Arithmetic darkroom is described in **Table 5**.

We found no significant differences for L1NS and STDP of both eyes. LPF of both eyes were significantly (t-test: $p < 0.05$) higher for 3-back than for 1-back. Similarly, 3-back was significantly (t-test: $p < 0.05$) higher than 2-back. We did not find a significant difference for saccade rate, fixation rate, and median SI velocity. We showed a comparison graph of L1NS for the auditory arithmetic test in the darkroom in **Figure 2**.

Auditory Arithmetic Dynamic Light Room (Pupil Dilation)

A repeated measure one-way ANOVA for metrics for Arithmetic test in dynamic lightroom is described in **Table 6**.

We found that L1NS and STDP of both eyes were significantly (t-test: $p < 0.05$) higher for 2-back than for 1-back. We also found that the LPF of both eyes was significantly (t-test: $p < 0.05$) higher for 3-back than for 1-back. We did not find a significant difference for saccade rate, fixation rate, and median SI velocity. We showed a comparison graph of L1NS for the auditory arithmetic test in the dynamic lightroom in **Figure 2**.

Interaction Effect

We performed a repeated measure two-way ANOVA on metric values for factors like light, presentation, task type, and task difficulty and reported the metrics that showed a significant interaction effect between respective factors in **Table 7** (tests of within-subjects effects) and **Table 8** (multivariate tests). The factors and their levels are listed below.

1. Darkroom vs. dynamic lightroom (factors: light and task difficulty)
   a. Darkroom (Auditory N-back) vs. dynamic lightroom (Auditory N-back)
   b. Darkroom (Auditory Arithmetic) vs. dynamic lightroom (Auditory Arithmetic)
2. Auditory Arithmetic vs. Auditory N-back (factors: task type and task difficulty)
   a. Auditory Arithmetic (Darkroom) vs. Auditory N-back (Darkroom)
   b. Auditory Arithmetic (Dynamic lightroom) vs. Auditory N-back (Dynamic lightroom)
3. Auditory N-back vs. Visual N-back (factors: presentation and task difficulty)
   a. Auditory N-back (Darkroom) vs. Visual N-back
   b. Auditory N-back (Dynamic lightroom) vs. Visual N-back

For Auditory N-back (Dynamic lightroom) vs. Visual N-back, LPF Right violated sphericity assumption, and we found a significant interaction between the factors using Greenhouse–Geisser as $F_{(1.453,29.059)} = 5.094$, $p < 0.05$, $\eta^2 = 0.203$.

TABLE 3 | Repeated measure one-way ANOVA for each metric with an effect size.

L1NS Right eye	$F_{(2,19)} = 8.155$, $p < 0.05$, $\eta^2 = 0.462$
L1NS Left eye	$F_{(2,19)} = 7.813$, $p < 0.05$, $\eta^2 = 0.451$
STDP Right eye	$F_{(2,19)} = 5.91$, $p < 0.05$, $\eta^2 = 0.384$
STDP Left eye	$F_{(2,19)} = 15.842$, $p < 0.05$, $\eta^2 = 0.625$
LPF Left eye	$F_{(2,19)} = 18.088$, $p < 0.05$, $\eta^2 = 0.656$

TABLE 4 | Repeated measure one-way ANOVA for each metric with an effect size.

L1NS Right eye	$F_{(2,19)} = 24.961$, $p < 0.05$, $\eta^2 = 0.724$
L1NS Left eye	$F_{(2,19)} = 43.017$, $p < 0.05$, $\eta^2 = 0.819$
STDP Right eye	$F_{(2,19)} = 29.461$, $p < 0.05$, $\eta^2 = 0.756$
STDP Left eye	$F_{(2,19)} = 39.767$, $p < 0.05$, $\eta^2 = 0.807$
LPF Left eye	$F_{(2,19)} = 38.847$, $p < 0.05$, $\eta^2 = 0.804$
LPF Right eye	$F_{(2,19)} = 28.797$, $p < 0.05$, $\eta^2 = 0.752$

TABLE 5 | Repeated measure one-way ANOVA for each metric with effect size.

LPF Left eye	$F_{(2,19)} = 7.657$, $p < 0.05$, $\eta^2 = 0.446$
LPF Right eye	$F_{(2,19)} = 6.280$, $p < 0.05$, $\eta^2 = 0.398$

TABLE 6 | Repeated measure one-way ANOVA for each metric with effect size.

L1NS Right eye	$F_{(2,18)} = 4.928$, $p < 0.05$, $\eta^2 = 0.354$
L1NS Left eye	$F_{(2,19)} = 5.966$, $p < 0.05$, $\eta^2 = 0.386$
STDP Right eye	$F_{(2,18)} = 4.790$, $p < 0.05$, $\eta^2 = 0.347$
STDP Left eye	$F_{(2,19)} = 4.595$, $p < 0.05$, $\eta^2 = 0.326$
LPF Left eye	$F_{(2,18)} = 7.662$, $p < 0.05$, $\eta^2 = 0.460$
LPF Right eye	$F_{(2,18)} = 6.648$, $p < 0.05$, $\eta^2 = 0.425$

Discussion

Our study confirmed the decrease in performance with increase in task difficulty (Granholm et al., 1996; Tokuda et al., 2011). We observed that L1NS, STDP, and LPF increased with increase in task difficulty, consistent with the study reported by Coulacoglou and Saklofske (2017). In all the cases, we observed that the parameter corresponding to a difficult task (3-Back and difficult arithmetic) was significantly higher than that corresponding to an easy task (1-Back and easy arithmetic). The intermediate task difficulty did not have a significant effect on all parameters. This might be because of the overlapping region of cognitive workload present in the 2-back test due to the transition of difficulty levels from 1-back to 3-back tests. Some participants would have found 2-back level as easy, and some would have found it difficult. Similarly, an overlapping region might be present in medium-level arithmetic questions. We found relatively large

TABLE 7 | Tests of within-subjects effects.

Interacting factors	Levels of factors	Metrics	Tests of within-subjects Effects (sphericity assumed)
Light vs. Task Difficulty	Dark room (Auditory N-back) vs. dynamic light room (Auditory N-back)	STDP Left	$F_{(2,40)}$ =19.369, $p < 0.05$, $\eta^2 = 0.492$
		STDP Right	$F_{(2,40)}$ =28.784, $p < 0.05$, $\eta^2 = 0.309$
		L1NS Left	$F_{(2,40)}$ =20.654, $p < 0.05$, $\eta^2 = 0.508$
		L1NS Right	$F_{(2,40)}$ =8.215, $p < 0.05$, $\eta^2 = 0.291$
		LPF Left	$F_{(2,40)}$ =18.881, $p < 0.05$, $\eta^2 = 0.486$
		LPF Right	$F_{(2,40)}$ =8.303, $p < 0.05$, $\eta^2 = 0.293$
	Dark room (Auditory Arithmetic) vs. dynamic light room (Auditory Arithmetic)	STDP Left	$F_{(2,40)}$ =3.708, $p < 0.05$, $\eta^2 = 0.156$
		L1NS Left	$F_{(2,40)}$ =3.394, $p < 0.05$, $\eta^2 = 0.145$
Task Type vs. Task Difficulty	Auditory Arithmetic (Dynamic light room) vs. Auditory N-back (Dynamic light room)	STDP Left	$F_{(2,40)}$ =18.229, $p < 0.05$, $\eta^2 = 0.477$
		STDP Right	$F_{(2,38)}$ =13.501, $p < 0.05$, $\eta^2 = 0.415$
		L1NS Left	$F_{(2,40)}$ =20.832, $p < 0.05$, $\eta^2 = 0.51$
		L1NS Right	$F_{(2,38)}$ =15.238, $p < 0.05$, $\eta^2 = 0.445$
		LPF Left	$F_{(2,38)}$ =17.889, $p < 0.05$, $\eta^2 = 0.485$
		LPF Right	$F_{(2,38)}$ =12.496, $p < 0.05$, $\eta^2 = 0.397$
Presentation vs. Task Difficulty	Auditory N-back (Dark room) vs. Visual N-back	STDP Left	$F_{(2,40)}$ =8.348, $p < 0.05$, $\eta^2 = 0.294$
		L1NS Left	$F_{(2,40)}$ =9.381, $p < 0.05$, $\eta^2 = 0.319$
		LPF Left	$F_{(2,40)}$ =8.855, $p < 0.05$, $\eta^2 = 0.307$
	Auditory N-back (Dynamic light room) vs. Visual N-back	STDP Left	$F_{(2,40)}$ =3.719, $p < 0.05$, $\eta^2 = 0.157$
		STDP Right	$F_{(2,40)}$ =4.979, $p < 0.05$, $\eta^2 = 0.199$
		L1NS Left	$F_{(2,40)}$ =3.444, $p < 0.05$, $\eta^2 = 0.147$
		L1NS Right	$F_{(2,40)}$ =4.807, $p < 0.05$, $\eta^2 = 0.194$
		LPF Left	$F_{(2,40)}$ =4.154, $p < 0.05$, $\eta^2 = 0.172$

TABLE 8 | Multivariate tests.

Interacting factors	Levels of factors	Metrics	Multivariate test (Pillai's trace)
Light vs. Task Difficulty	Dark room (Auditory N-back) vs. dynamic light room (Auditory N-back)	STDP Left	$F_{(2,19)} = 18.634$, $p < 0.05$, $\eta^2 = 0.662$
		STDP Right	$F_{(2,19)} = 8.398$, $p < 0.05$, $\eta^2 = 0.469$
		L1NS Left	$F_{(2,19)} = 20.084$, $p < 0.05$, $\eta^2 = 0.679$
		L1NS Right	$F_{(2,19)} = 7.375$, $p < 0.05$, $\eta^2 = 0.437$
		LPF Left	$F_{(2,19)} = 17.182$, $p < 0.05$, $\eta^2 = 0.644$
		LPF Right	$F_{(2,19)} = 6.747$, $p < 0.05$, $\eta^2 = 0.415$
	Dark room (Auditory Arithmetic) vs. dynamic light room (Auditory Arithmetic)	STDP Left	$F_{(2,19)} = 4.073$, $p < 0.05$, $\eta^2 = 0.3$
		L1NS Left	$F_{(2,19)} = 4.813$, $p < 0.05$, $\eta^2 = 0.336$
Task Type vs. Task Difficulty	Auditory Arithmetic (Dynamic light room) vs. Auditory N-back (Dynamic light room)	STDP Left	$F_{(2,19)} = 21.436$, $p < 0.05$, $\eta^2 = 0.693$
		STDP Right	$F_{(2,18)} = 11.517$, $p < 0.05$, $\eta^2 = 0.561$
		L1NS Left	$F_{(2,19)} = 20.509$, $p < 0.05$, $\eta^2 = 0.683$
		L1NS Right	$F_{(2,18)} = 12.273$, $p < 0.05$, $\eta^2 = 0.577$
		LPF Left	$F_{(2,18)} = 15.759$, $p < 0.05$, $\eta^2 = 0.636$
		LPF Right	$F_{(2,18)} = 10.168$, $p < 0.05$, $\eta^2 = 0.53$
Presentation vs. Task Difficulty	Auditory N-back (Dark room) vs. Visual N-back	STDP Left	$F_{(2,19)} = 10.415$, $p < 0.05$, $\eta^2 = 0.523$
		L1NS Left	$F_{(2,19)} = 11.691$, $p < 0.05$, $\eta^2 = 0.552$
		LPF Left	$F_{(2,19)} = 10.732$, $p < 0.05$, $\eta^2 = 0.53$
	Auditory N-back (Dynamic light room) vs. Visual N-back	STDP Left	$F_{(2,19)} = 3.717$, $p < 0.05$, $\eta^2 = 0.281$
		STDP Right	$F_{(2,19)} = 4.442$, $p < 0.05$, $\eta^2 = 0.319$
		L1NS Left	$F_{(2,19)} = 3.954$, $p < 0.05$, $\eta^2 = 0.294$
		L1NS Right	$F_{(2,19)} = 3.781$, $p < 0.05$, $\eta^2 = 0.285$
		LPF Left	$F_{(2,19)} = 3.872$, $p < 0.05$, $\eta^2 = 0.29$
		LPF Right	$F_{(2,19)} = 3.295$, $p < 0.05$, $\eta^2 = 0.258$

effect sizes in L1NS left eye for Visual N-Back, LPF left eye for Auditory N-Back Darkroom, L1NS left eye for Auditory N-Back Dynamic lightroom, LPF left eye for Auditory Arithmetic Dark room, and LPF left eye for Auditory Arithmetic Dynamic lightroom. Though STDP left and L1NS left were able to significantly distinguish between task difficulties in all conditions except dark room arithmetic, LPF left and LPF right were able to significantly distinguish between task difficulties in the darkroom arithmetic test. This infers that each metric performed significantly in each test condition. We also observed that the trend of increase in metric values with respect to increase in task difficulty is the same for changes in light conditions for visual and auditory presentations. Though we found an interaction effect between task difficulty and lighting conditions for pupil-based metrics, *t*-test results showed that our pupil-based metrics significantly distinguished between task difficulties in different lighting conditions. Similarly, a set of pupil-based metrics could substantially distinguish between task difficulties in different task types and presentation conditions despite significant interaction between the factors.

FLIGHT SIMULATOR STUDY

Once the ocular parameters' robustness was evaluated with standard methods, we applied the same for aviation specific tasks. We conducted user studies in the high-fidelity NALSim flight simulator that is based on the Learjet aircraft model (Kamali et al., 2014). The purpose of the user study was to check the usability of EEG- and ocular parameter-based cognitive workload estimators and to investigate the effect of secondary tasks on cognitive workload.

All simulations started with similar initial conditions such as landing gear down, on ground, and a trim speed of 120 knots. The airport altitude was 89 0m above mean sea level. The baseline task was to conduct a controlled take-off, followed by a climb phase and a wings level flight (**Figure 3**).

Participants

Biswas and JeevithaShree (2018) have suggested in their study that it is advisable to first test any new technology using participants with little or no knowledge about the platform. Accordingly, we chose students from our university for the flight simulator study. We conducted 36 simulations with 12 participants for three different task difficulty conditions C1, C2, and C3 mentioned in **Table 9**. Participants were aged between 22 and 40 years. The male-to-female ratio was 3:9. As the participants were non-pilots and new to the flight simulator environment, their cognitive workload variations were expected to be higher than that of experienced pilots (Antonenko et al., 2010). Hence, this sampling strategy satisfied our aim to relate different physiological parameters at varying levels of cognitive workload.

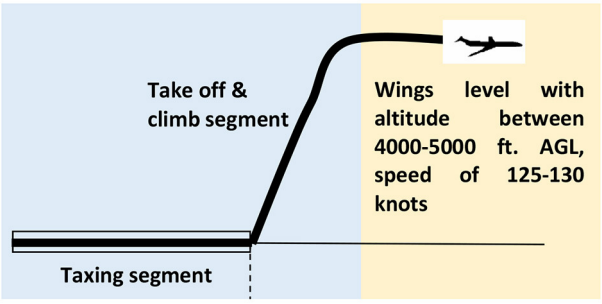

FIGURE 3 | Test scenario.

Design

We conducted 36 simulations with 12 participants for three different task difficulty conditions mentioned in **Table 9**.

The secondary task in C3 necessitated the participants to select a randomly positioned button in the secondary display based on an aural warning. While introducing the secondary task in C3, participants were instructed to prioritize their primary task, which was maintaining altitude within limits of ±1,000 ft.

Materials

Simulation Setup

Simulation studies were conducted using the NALSim flight simulator at I3D lab, Indian Institute of Science, Bangalore, India. NALSim is a cost-effective ground-based variable stability flight simulator developed for Indian aircraft design programs. NALSim architecture is being used by a premier flight test pilot school of Indian Airforce for pilot training on aircraft handling qualities. The advantage of this simulator is that it is designed to provide a platform for researchers and aerospace students to understand aircraft dynamics, conduct aircraft configuration design studies, and handle quality studies and PVI studies.

The pilot's view in the simulator consists of out-of-the-window visual scenery and a head down display (**Figure 4A**).

Eye Gaze Tracker

Tobii pro wearable eye tracker (**Figure 4B**) was used in the study for recording eye gaze parameters (Eye tracking for research., 2018). This system measures three-dimensional gaze direction with an accuracy of 0.4° of visual angle. The sampling frequency of the device is 100 Hz.

TABLE 9 | Task scenarios for Flight simulator study.

C1	C2	C3
Take off: Initial checks, apply full throttle, take off at 130 knots speed. Climb and continue with a level flight for 4 min.	Similar to C1 till level segment. Maintain altitude between 4,000 and 5,000 ft above MSL for 4 min.	A secondary task was added in addition to C2 condition. The secondary task was defined as pointing and selection in an adjacent secondary head down touchscreen display.

EEG Headset

A study by Grummett et al. (2015) compares few inexpensive and wireless EEG systems for such experimental studies. The study supports the suitability of Emotiv headset for alpha responses and visual steady-state responses (VSSR). As the experiments designed in this study are related to VSSR, we have used portable, low-cost Emotiv Insight 5 channel EEG headset (**Figure 4C**) under controlled conditions for recording EEG data (Insight User Manual., 2018). EEG signals have a useful bandwidth in the range of the different frequency bands as given in **Table 10**.

The electrode positions that were studied are AF3 (left frontal), AF4 (right frontal), T7 (left temporal), T8 (right temporal), and Pz (central parietal).

Procedure

We instructed participants to wear the EEG headset and the Tobii-pro eye-tracking glasses for the user trials. The contacts of the EEG headset were checked, and the eye tracker was calibrated prior to each simulation. All the participants were instructed about the procedure of the experiment and given 15 min of practice time to get acquainted with flying. It was ensured that participants were capable of performing wings level flight with constant altitude before starting the actual test scenario.

Results

EEG, eye gaze, aircraft performance parameters, and inceptor control data were analyzed to infer the demand on pilot's cognitive workload. This section details the analysis results. We removed the outliers in the EEG and the gaze data using outer fencing. Normality in the data was checked using the Anderson–Darling test. As we found that both EEG and gaze data were not normally distributed, we used non-parametric tests such as Friedman test and Wilcoxon pairwise signed-rank test to analyze the significance of difference in cognitive workload.

EEG Signal Analysis

Figure 5 shows the median of power in each frequency band for C1, C2, and C3. We found that the median EEG signal power level increased from C1 through C3 in all the frequency bands. However, we did not find a significant increase in median power for the gamma frequency band ($p > 0.5$). Hence, we have not considered the gamma band for further discussions. The results from the statistical tests are summarized in **Table 11**.

We found that EEG signal power in the LB and theta band showed a significant increase between C2–C3 and C1–C3 ($p < 0.05$). Accordingly, C3 has a relatively higher cognitive workload.

Ocular Parameter Analysis

The following ocular parameters were selected for the flight 3 simulator study:

a. **Gaze Fixation:** We used the fixation classification algorithm introduced by Tobii, called the Tobii I-VT filter (Olsen, 2012), for extracting fixations. According to Olsen, I-VT filter classifies eye movements based on the velocity of the eye's

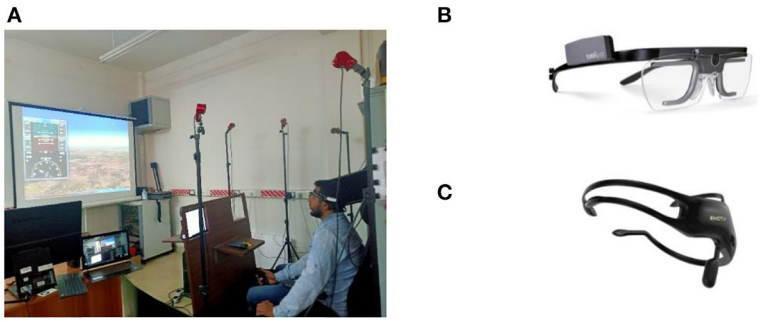

FIGURE 4 | Apparatus. **(A)** Flight simulator setup. **(B)** Eye gaze tracker from Tobii [courtesy (14)]. **(C)** EEG headset from Emotiv [courtesy (15)].

TABLE 10 | EEG signal bandwidths.

Brain frequency bands	Frequency range (Hz)	State of mind	Changes in band power with increasing task demand
Theta band	4–8	Sleepy, drowsy, meditative, and dreaming	Increases, associated with fatigue (Borghini et al., 2012) and information retrieval
Alpha band	8–13	Relaxed, calm, lucid state of mind	Increases with aural secondary task indicating increased information processing and reduced concentration ability (Schrauf et al., 2011)
Low beta (LB) band	13–21	Alert, active concentration, busy, and anxious state of mind	Increases (Pavlov and Kotchoubey, 2017)
High beta (HB) band	21–30	Focus, quick thinking, working	Increases (Pavlov and Kotchoubey, 2017)
Gamma	31–100	Optimal frequency for thinking, active thought, peak focus	Increases

directional shifts. Gaze is classified as a saccade if the velocity is above a particular value of threshold (default−30°/s), otherwise, it is classified as a fixation.

We computed fixation rate as the total time of fixations divided by simulation duration. **Figure 6** shows the mean fixation rate for the three test scenarios. Friedman's statistics did not show any significant difference in the fixation rates between the test conditions. However, we observed an increasing trend of average fixation rate from C2 to C3 ($Z = 1.88$, $p = 0.058$).

b. **Distribution patterns of fixation:** We used NNI in this study as an indirect cognitive workload estimator.

We computed NNI as the ratio of the nearest-neighbor distance (d_{NN}) to the mean random distance (d_{MRD}).

$$NNI = \frac{d_{NN}}{d_{MRD}}$$

Where $d_{NN} = \sum_{i=1}^{N} min[\frac{\sqrt{(fx_i - fx_j)^2 - (fy_i - fy_j)^2)}}{N}]$ and $d_{MRD} = 0.5\sqrt{Area\ of\ interest/N}$

Area of interest is computed as the rectangular area that the x and y gaze coordinates cover. fx and fy are the x and y eye coordinates, respectively, and i and j are the successive time instances in x and y. N is the total number of data points.

The mean of NNI scores for the three test conditions are shown in **Figure 7**.

We found that distribution of eye fixations significantly differs with different test conditions [$\chi^2(2) = 9.50$, $p < 0.01$]. Furthermore, pairwise comparison using signed-rank test showed that the eye fixations were more randomly distributed in space for C3 ($p < 0.01$ for C1–C3 and $p < 0.1$ for C2–C3).

c. **Pupil dilation dynamics:** We formulated the following metrics to extract features from the frequency spectrum of pupil dilation data. The three-frequency domain-based pupil dilation metrics discussed in section User Study on Psychometric Tests are L1NS, STDP, and LPF. In section Discussion, we found that L1NS shows the ability to significantly distinguish between task difficulties in all conditions. Hence, we used L1NS in the flight simulator study.

i. L1 Norm of Spectrum (L1NS): Frequency domain-based L1NS on pupil dilation was computed based on the algorithm proposed by Prabhakar and Biswas (2018). Single-sided spectrum of the left and right pupil dilation time-series data

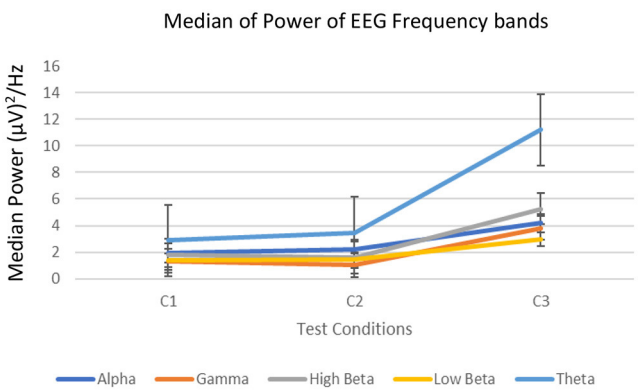

FIGURE 5 | Median power of different EEG bands.

TABLE 11 | Statistical test results indicating changes in cognitive workload.

Friedman Test			
Power levels in EEG Frequency bands			
Alpha band		$\chi^2(2) = 8.0, p < 0.05$	
LB band		$\chi^2(2) = 8.166, p < 0.05$	
HB band		$\chi^2(2) = 10.5, p < 0.01$	
Theta band		$\chi^2(2) = 10.66, p < 0.01$	
Wilcoxon signed-rank test			
	C1–C2	C1–C3	C2–C3
Alpha band	–	$Z = 13, p < 0.05$	$Z = 16, p = 0.07$
LB band	–	$Z = 15, p < 0.05$	$Z = 13, p < 0.05$
HB band	–	$Z = 13, p < 0.05$	$Z = 10, p < 0.05$
Theta band	–	$Z = 1, p < 0.01$	$Z = 17, p = 0.07$

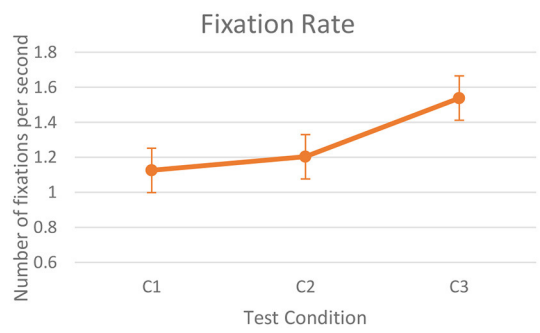

FIGURE 6 | Fixation rate.

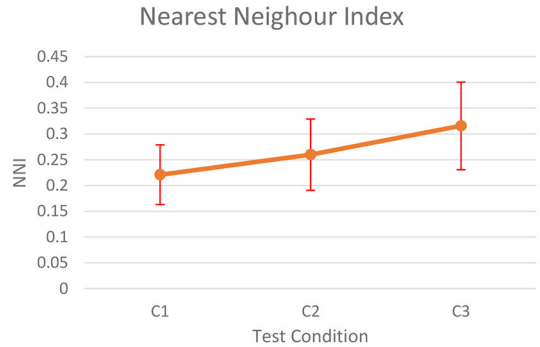

FIGURE 7 | Nearest Neighborhood Index.

(Y_k) was computed using Fast Fourier transform given as:

$$\widetilde{Y}_j = \sum_{k=1}^{N} Y_k e^{\left(\frac{-2\pi i}{N}\right)(k-1)(j-1)}$$

Frequency components from 1 to 5 Hz were summed to compute L1NS. **Figure 8** shows the comparison plots for the left and right pupil diameters.

We found that rate of change of pupil diameter was significantly different for right pupil for [$\chi^2(2) = 6.17$, $p < 0.05$, $\eta^2 = 0.2569$]. Pairwise comparison showed that C3 had maximum changes in pupil diameters ($p < 0.01$ for C2–C3 and $p < 0.1$ for C1–C3).

ii. Median of SI velocity in °/second: As defined by Abadi and Gowen (2004), horizontal eye movements within 0.4° in the X axis where eye-gaze returns to the same position between 60 and 870 ms are known as SI (Prabhakar and Biswas, 2018). We used the algorithm described in Biswas and Langdon (2015) to compute median SI velocity.

Figure 9 shows the median SI velocity for the three test conditions. Friedman test did not show any significant change in the rate of change of median SI velocity for the three conditions.

Table 12 consolidates the inferences of all the gaze measurement methodologies discussed in this section.

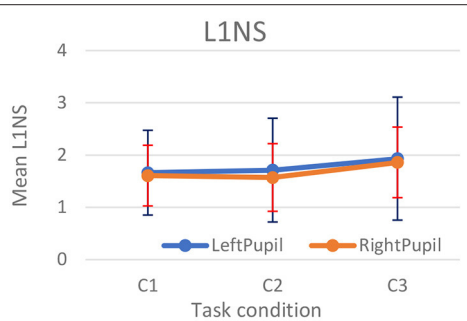

FIGURE 8 | L1NS for left and right pupil diameters.

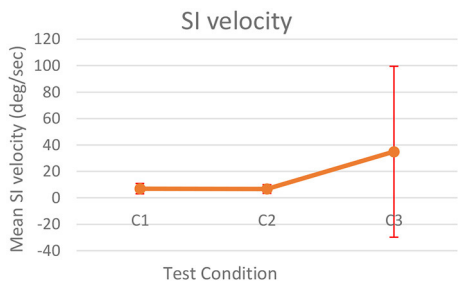

FIGURE 9 | Median of SI velocity.

Flying Performance Analysis

We used DC and aggressiveness of participants' inceptor control columns as an indicator of cognitive workload experienced by the pilot. DC indicates the percentage of time a participant controls his/her input on the inceptor. DC is computed as follows:

$$DC = 100\%^* \frac{1}{t_n - t_2} \sum_{i=2}^{n} x_i$$

$$\text{Here, } x_i = \begin{cases} 0 & \text{for } \frac{\delta_i - \delta_{i-1}}{t_i - t_{i-1}} < \text{noise threshold and } |\delta_i| < \delta_{max} \\ 1 & \text{otherwise} \end{cases}$$

Aggressiveness is the rate of change of inceptor control movements. The formula is as follows:

$$Aggressiveness = \sqrt{\frac{1}{n-1} \sum_{i=2}^{n} \left(\frac{\delta_i - \delta_{i-1}}{t_i - t_{i-1}}\right)^2}$$

t is the simulation time, n is the number of data points, δ_i is the inceptor deflection in mm, and δ_{max} is the maximum stick deflection.

Plotting aggressiveness vs. DC is known as the PIW plot. Higher aggressiveness relates to more random control commands and higher DC infers that more time is required to control. We used the PIW plot to infer the variations in workload.

Figure 10 shows the PIW plot of mean values of both the parameters for C1, C2, and C3. We observed that participants' aggressiveness levels were similar for both C2 and C3. However, participants had to spend more time controlling the inceptor in order to maintain level flight in case of C3 [$F_{(4,26)} = 2.72$, $p < 0.1$, $\eta^2 = 0.247$]. Furthermore, pairwise comparison using Tukey Kramer test showed that C3 had statistically higher DC than C2 ($p < 0.1$).

Participants' inceptor control strategy, together with his/her flying performance, indicates his/her cognitive workload (Hebbar and Pashilkar, 2017). Accordingly, we computed

TABLE 12 | Summary of analysis of ocular parameters.

Friedman Test			
Fixation rate			–
NNI			χ^2 (2) = 9.5, $p < 0.01$
L1NS			χ^2 (2) = 6.17, $p < 0.05$
SI			–
Wilcoxon signed-rank test			
	C1–C2	**C1–C3**	**C2–C3**
Fixation rate	–	–	$Z = 1.88$, $p = 0.058$
NNI	–	$Z = 2.74$, $p < 0.01$	$Z = 1.65$, $p < 0.1$
L1NS	–	$Z = 1.8$, $p = 0.07$	$Z = 2.5$, $p < 0.01$
SI	–	–	–

flight performance in terms of RMSE in altitude and airspeed deviations (**Figure 11**). We found that errors increased significantly with additional demand of secondary task (for airspeed, $p < 0.02$).

Conformance Among Parameters

Table 13 consolidates the results of cognitive workload estimation metrics discussed. A level of significance < 0.1 is indicated in green.

In the case of C3, we found comparative similarity ($p < 0.05$) between LB and HB bands of EEG power, NNI and L1NS of ocular parameters, and the DC metric from flight parameters. Increased task difficulty was observed by the parameters mentioned above with secondary task (C3). However, ocular parameters such as SI velocity and fixation rate did not show a significant increase. This corroborates with the results from the psychometric study.

Subsequently, we carried out Spearman's pairwise rank correlation analysis between the significant parameters given in **Table 13** (**Figure 12**). Spearman's rank correlation coefficient (ρ) is computed as

$$\rho = 1 - \frac{6\sum_i^n d_i}{n^3 - n}$$

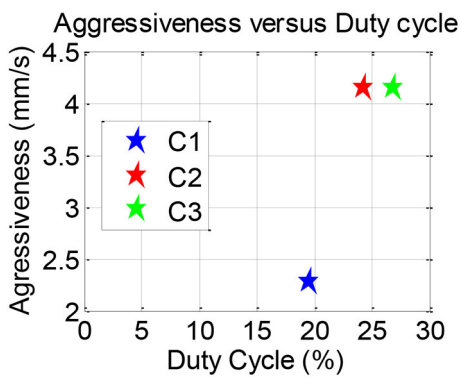

FIGURE 10 | PIW plots.

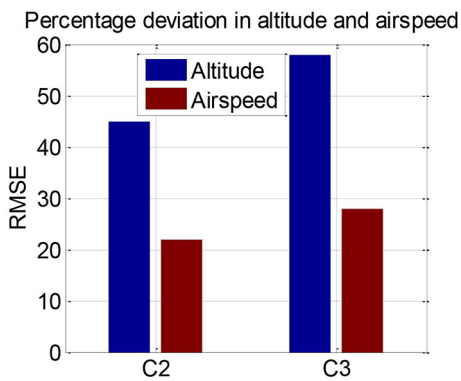

FIGURE 11 | Percentage RMSE in altitude and airspeed.

Here, d_i is the difference between the ranks of each observation and n is the number of observations.

Firstly, we compared the correlation between the EEG parameters. We observed a consistent positive trend between LB, HB, alpha, and theta frequency band powers ($0.6 < \rho < 0.9$, $p < 0.001$). LB and theta bands showed a very strong association ($\rho > 0.8$, $p < 0.001$). When comparing the correlation between ocular parameters, we found statistically significant ($\rho > 0.7$, $p < 0.001$) positive correlation between NNI and L1NS.

Furthermore, we repeated the correlation analysis between EEG, ocular parameters, and the DC metric. **Figure 12** shows the correlation plots for the significant parameters. We observed positive correlation between EEG and L1NS (LB and L1NS: $\rho = 0.3791$, $p < 0.005$; and theta and L1NS: $\rho = 0.38$, $p < 0.005$) and between EEG and NNI (LB and NNI: $\rho = 0.4038$, $p < 0.05$; theta and NNI: $\rho = 0.4283$, $p < 0.05$). We detected positive correlation between DC and EEG (DC and theta: $\rho = 0.3652$, $p < 0.05$ and DC and LB: $\rho = 0.3338$, $p < 0.05$), and DC and ocular parameters (DC and L1NS: $\rho = 0.4627$, $p < 0.005$; DC and NNI: $\rho = 0.3251$, $p < 0.05$).

Discussions

Flight simulator studies were designed with three task conditions. The tasks were intended to increase the difficulty levels from C1 to C3. The recorded spectral power in beta and theta bands show a progressive increase from C1 to C3 and C2 to C3 (Table 14). These results have physiological significance in terms of human information processing. As discussed in **Table 10**, theta activity is associated with information retrieval. Beta band represents fast activity and is an indicative of increased thinking and focus levels. Hence, EEG data suggest that introducing a secondary task in C3 increased load on the participant's working memory. Results from ocular parameters also suggest a similar trend. NNI and L1NS showed a significant increase in C3. Hence, gaze fixations were more random and variations in pupil diameter were more predominant during C3. Data from the participant's flying performance and his/her control strategy [$F_{(4,26)} = 2.72$, $p < 0.0839$, $\eta^2 = 0.247$] also indicated increased cognitive workload with the inclusion of the secondary task.

Correlation analysis between EEG, ocular, and flying performance data for all the participants indicated positive correlation among all the parameters. Among the EEG frequency band power, LB and theta bands were found to be highly correlated. We also found that EEG theta and LB power, NNI, L1NS, and DC have a statistically significant positive correlation.

GENERAL DISCUSSIONS

We presented two studies on validating physiological measures to estimate pilot's cognitive workload in demanding scenarios. In the first study, different task difficulty levels were achieved through proven psychometric tasks such as N-Back and arithmetic tasks. As discussed in section Discussion, the test results confirmed our ocular parameters' robustness in estimating cognitive workload for varying task difficulties and varying illumination conditions. We found that L1NS,

TABLE 13 | Comparison between parameters.

	Alpha band power	LB band power	HB band power	Theta band power	SI velocity	NNI	Fixation rate	L1NS	DC
C1–C2	$p > 0.1$	$p > 0.1$	$p > 0.1$	$p > 0.1$	$p > 0.1$	$p > 0.1$	$p > 0.1$	$p > 0.1$	$p > 0.1$
C1–C3	$p < 0.05$	$p = 0.05$	$p < 0.05$	$p < 0.01$	$p > 0.1$	$p < 0.01$	$p > 0.1$	$p = 0.07$	$p > 0.1$
C2–C3	$p = 0.07$	$p < 0.05$	$p < 0.05$	$p = 0.07$	$p > 0.1$	$p < 0.1$	$p = 0.058$	$p < 0.01$	$p < 0.05$

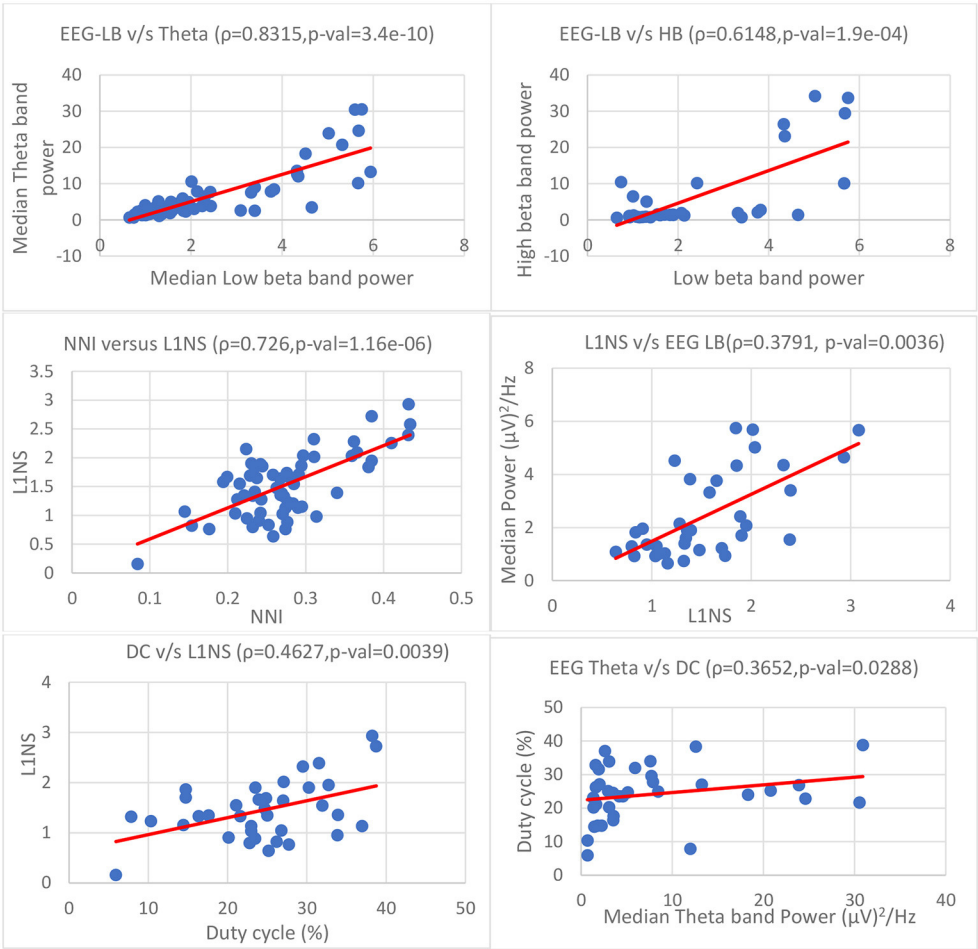

FIGURE 12 | Correlation between EEG, ocular, and flying performance parameters (ρ is the pairwise linear correlation coefficient; p-val is the level of significance). (From top left) EEG: LB vs. theta, LB vs. HB; Ocular parameters: NNI vs. L1NS, L1NS vs. EEG LB; Flying performance: DC vs. L1NS, DC vs. theta.

STDP, and LPF of pupil diameter were able to distinguish between different cognitive states corresponding to task difficulties irrespective of changes in lighting conditions. The results also proved that increase in task difficulty causes a decrease in performance. In our flight simulator study, manipulation of task difficulty was achieved in a controlled test environment with realistic flight scenarios. We used NALSim simulator and designed test scenarios that were representative of the real flight conditions. We used EEG, ocular, and flying performance parameters to estimate pilot's cognitive workload to address our third objective. We used standard statistical hypothesis methods to report the comparative results. Results from the user studies concluded that, in general,

- Distribution pattern of gaze fixations was more random with increase in task difficulty. This was proven by the NNI parameter in the flight simulator study.
- Pupil dilation-based L1NS metric showed significant increase in N-back and arithmetic tasks and aircraft flight task.
- In the case of EEG data, low beta and theta band powers were consistently more sensitive to task difficulty. Test condition

with secondary task showed the highest cognitive workload among all scenarios.

We used participants' inceptor control strategy and their flying performance as another indicator for comparison with the physiological parameters. C3 showed higher DCs among the three test cases. The higher DC suggests that participants in the C3 test condition had to use the inceptor controls more rigorously than for other task conditions. Our final objective was to find a correlation between the multiple observations. We found that low beta and theta EEG band power, the gaze base ocular parameter NNI, the pupil dilation-based ocular parameter L1NS, and the performance-based parameter DC are indicators of cognitive workload variations and have positive correlation ($p < 0.05$) among themselves.

The primary aim of this study was to identify and correlate the different physiological and performance-based metrics as an indicative measure of pilot's cognitive workload. However, it is known that for the same task, novice pilots experience higher cognitive workload than experienced pilots (Antonenko et al., 2010). Hence, the results of the study are limited to understanding the correlation between the measures and not to compare the cognitive workload of the pilots. Future research would focus on validating the metrics discussed in this study with pilot evaluations for the entire flight envelope. Additionally, based on the available database, we plan to use machine learning techniques to classify pilots' cognitive status in real time.

These findings can also be extended to automotive domain where the drivers are always engaged in tasks that demand their attention and increase their cognitive load while driving. The estimated cognitive load from the proposed technique can provide necessary information to the car for making smart decisions when the driver undergoes increase in cognitive load.

Furthermore, cognitive workload estimation principles may turn out to be highly relevant for design optimization of any new product. Innovation starts with user's need, which is then fulfilled by creating new solutions or improving existing solutions. However, the challenge lies in identifying the real need of the users. Design thinking has been very successful in adopting a human-centered approach in identifying the need of the users in society. Hence, estimating a user's cognitive status is extremely critical to understand the underlying factors that govern responses of human mind and human actions. An accurate understanding of the cognitive processes can create an efficient design that can create a superior user experience.

CONCLUSIONS

This paper discussed the application of non-invasive physiological measures along with task performance-based metrics to estimate pilot's cognitive workload. Initially, we conducted studies to estimate ocular parameters' ability to distinguish between variations in cognitive workload corresponding to differences in task difficulties. We also evaluated the robustness of our metrics in different ambient light conditions. In the second study, three different workload estimation methodologies were validated and compared. Participants were assigned different dimensions of task levels, such as primary and secondary tasks and maintaining one or many flight parameters. It was observed that the introduction of the secondary task (condition C3) along with flying caused a significant increase in cognitive workload. Degradation in performance due to such secondary tasks can be estimated from the proposed metrics. Thus, results discussed in this study propose a methodology for estimating pilot's cognitive workload based on his/her physiological measures such as EEG, ocular parameters, and the pilot's flying performance.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study. Written informed consent was also obtained from the participants for the publication of any potentially identifiable images and data used in this study.

AUTHOR CONTRIBUTIONS

PH has contributed in the formulation of the test cases, conducting the user studies, and analyzing the ocular parameters of the flight simulator experiments. KB has contributed in the conducting the user studies data extraction and analyzing the EEG data of the flight simulator experiments. GP contribution is the conduct of the psychometric test and analyzing the results. PB contribution is overall guidance, discussion of the results, and reasoning and conclusions. AP has provided overall guidance and support. All authors contributed to the article and approved the submitted version.

REFERENCES

Abadi, R. V., and Gowen, E. (2004). Characteristics of saccadic intrusions. *Vis. Res.* 44, 2675–2690. doi: 10.1016/j.visres.2004.05.009

Antonenko, P., Paas, F., Grabner, R., and Gog, T. V. (2010). Using electroencephalography to measure cognitive load. *Educ. Psychol. Rev.* 22, 425–438. doi: 10.1007/s10648-010-9130-y

Babu, M. D., JeevithaShree, D. V., Prabhakar, G., Saluja, K. P. S., Pashilkar, A., and Biswas, P. (2019). Estimating pilots cognitive load from ocular parameters through simulation and inflight studies. *J. Eye Mov. Res.* 12:3. doi: 10.16910/jemr.12.3.3

Beatty, J., and Lucero-Wagoner, B. (2000). "The pupillar system," in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson, (Cambridge: Cambridge University Press), 142–162.

Biella, M., Wies, M., Charles, R., Maille, N., Berberian, B., and Nixon, J. (2017). "How eye tracking data can enhance human performance in tomorrow's cockpit. *Results from a flight simulation study in FUTURE SKY SAFETY," in Proceedings of Joint AIAA and Royal Aeronautical Society Fall Conference on Modeling and Simulation for ATM* (London), 13–15.

Biswas, P., and JeevithaShree, D. V. (2018). "Eye Gaze Controlled MFD for Military Aviation," in *23rd International Conference on Intelligent User Interfaces* (ACM), 79–89. doi: 10.1145/3172944.3172973

Biswas, P., and Langdon, P. (2015). Multimodal intelligent eye-gaze tracking system. *Int. J. Hum. Comp. Interact.* 31, 277–294. doi: 10.1080/10447318.2014.1001301

Bjäreholt, E. A Simple N-Back game written in Python, compatible with Android. (2014). Available online at: https://github.com/ErikBjare/N-Back/ (accessed November 18, 2019).

Bodala, I. P., Li, J., Thakor N. V. and Nashash H. A. (2016). EEG and eye tracking demonstrate vigilance enhancement with challenge integration. *Front. Hum. Neurosci.* 10:273. doi: 10.3389/fnhum.2016.00273

Boeing (2018). *Statistical Summary of Commercial Jet Airplane Accidents-Worldwide Operations—1959-2018, 50th Edn*. Washington, DC: Aviation Safety, Boeing Commercial Airplanes Seattle.

Borghini, G., Vecchiato G., Toppi J., Astolfi L., Maglione A., Isabella R., et. al. (2012). "*Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices," in Proceedings of 34th Annual International Conference of the IEEE EMBS* (SanDiego, CA). doi: 10.1109/EMBC.2012.6347469

Borys, M., Tokovarov, M., Wawrzyk, M., Wesolowska, K., Plechawska, M., Dmytruk, R., et al. (2017). "An analysis of eye tracking and electroencephalography data for cognitive load measurement during arithmetic tasks," in *Proceedings of 10th International Symposium on Advanced Topics in Electrical Engineering* (Bucharest). doi: 10.1109/ATEE.2017.7905130

Causse, M., Fabre, E., Giraudet, L., Gonzalez, M., and Peysakhovich, V. (2015). "EEG/ERP as a measure of mental workload in a simple piloting task," in *Proceedings of 6th International Conference on Applied Human Factors and Ergonomics and Affiliated Conferences, Procedia Manufacturing* (Las Vegas, NV), 5230–5236. doi: 10.1016/j.promfg.2015.07.594

Cheng, S., and Hsu, H. (2011). "*Mental fatigue measurement using EEG*," in Risk Management Trends, ed G. Nota (InTech). doi: 10.5772/16376

Coulacoglou, C., and Saklofske, D. H. (2017). Psychometrics and psychological assessment: principles and applications. *Acad. Press* 91–130. doi: 10.1016/B978-0-12-802219-1.00005-5

Demberg, V., and Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PLoS ONE* 11:e0146194. doi: 10.1371/journal.pone.0146194

Di Nocera, F., Camilli, M., and Terenzi, M. (2007). *A* random *glance at the flight deck: Pilot's scanning strategies and the real-time assessment of mental* workload. *J. Cogn. Eng. Decis. Making* 1, 271–285. doi: 10.1518/155534307X255627

Ebbatson, M., Huddlestone, J., Harris, D., and Sears, R. (2007). The application of frequency analysis based performance measures as an adjunct to flight path derived measures of pilot performance. *Hum. Factors Aerospa. Saf.* 6, 383–394.

Eye tracking for research. (2018). Available online at: https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/ (accessed January 22, 2020).

Friedman, N., Feketa, T., Gal, G., and Shriki, O. (2019). EEG based Prediction of cognitive load in intelligence tests. *Front. Hum. Neurosci.* 13:191 doi: 10.3389/fnhum.2019.00191

Gentili, R. J., Rietschel, J. C., Jaquess, K. J., Lo, L. C., Prevost, C. M., Miller, M. W., et al. (2014). "Brain biomarkers-based assessment of cognitive workload in pilots under various task demands," in *Proceedings of 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Chicago, IL), 5860–5863. doi: 10.1109/EMBC.2014.6944961

Gevins, A., Smith, M. E., McEvoy L., and Yu, D. (1997). High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cereb Cortex*. 7, 374–85. doi: 10.1093/cercor/7.4.374

Granholm, E., Asarnow, R. F., Sarkin, A. J., and Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology* 33, 457–461. doi: 10.1111/j.1469-8986.1996.tb01071.x

Grummett, T. S., Leibbrandt, R. E., Lewis, T. W., DeLosAngeles, D., Powers, D. M. W., Willoughby, J. O., et al. (2015). Measurement of neural signals from inexpensive, wireless and dry EEG systems. *Physiol. Meas.* 36:1469. doi: 10.1088/0967-3334/36/7/1469

Hanson, C., Schaefer, J., Burken, J. J., Larson, D., and Johnson, M. (2014). *Complexity and Pilot Workload Metrics for the Evaluation of Adaptive Flight Controls on a Full-Scale Piloted Aircraft*. Edwards, CA: NASA Dryden Flight Research Center.

Harper, R.P. Jr., and Cooper, G. E. (1986). Handling Qualities and Pilot Evaluation. *AIAA J. Guidance. Cont. Dyn*. 9, 515–529. doi: 10.2514/3.20142

Hebbar, P.A., and Pashilkar, A.A. (2017). Pilot performance evaluation of simulated flight approach and landing maneuvers using quantitative assessment tools. *Sādhanā* 42, 405–415. doi: 10.1007/s12046-017-0613-0

Hess, E. H. (1975). *The Tell-Tale eye: How Your Eyes Reveal Hidden Thoughts and Emotions.* New York, NY: Van Nostrand Reinhold Co.

Hyönä, J., Radach, R., and Deubel, H. (2003). "Chapter 23 - eye tracking in usability evaluation: a practitioner's guide." in *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, eds R. Radach, J. Hyona, and H. Deubel (Amsterdam: Elsevier Science), 493–517.

Insight User Manual. (2018). Available online at: https://emotiv.gitbook.io/insight-manual/ (accessed January 22, 2020).

Kamali, C., Hebbar, P. A., Vijeesh, T., and Moulidharan, S. (2014). Real time desktop flying qualities evaluation simulator. *Defen. Sci. J.* 64, 27–32. doi: 10.14429/dsj.64.4961

Kramer, A. (1991). "Physiological metrics of mental workload: a review of recent progress," in *Multiple Task Performance*, ed D. Damos (Taylor and Francis), 279–328.

Lee, K. (2010). *Effect of Flight Factors on Pilot Performance, Workload and Stress at Final Approach to Landing Phase of Flight* (dissertation). Orlando, Florida, University of Central Florida.

Li, W. C., Lin, J. J., Braithwaite, G., and Greaves, M. (2016). "The development of eye tracking in aviation (ETA) technique to investigate pilot's cognitive processes of attention and decision-making," in *Proceedings of 32nd conference of the European Association for Aviation Psychology* (Cascais).

Marshall, S. (2007). Identifying cognitive state from eye metrics. *Aviat. Space Environ. Med.* 78(Suppl. 1), B165–B175.

Mohanavelu, K., Poonguzhali, S., Ravi, D., Singh, P. K., Mahajabin, M., Ramachandran, K., et al. (2020). Cognitive workload analysis of fighter aircraft pilots in flight simulator environment. *Defen. Sci. J.* 70, 131–139. doi: 10.14429/dsj.70.14539

Olsen, A. (2012). *The Tobii I-VT Fixation Filter*. Tobii Technology. Retrieved from: https://www.tobiipro.com/siteassets/tobii-pro/learn-andsupport/analyze/how-do-we-classify-eye-movements/tobiipro-i-vt-fixation-filter.pdf

Othman, N., and Romli, F. I. (2015). "*Mental workload evaluation of aircraft operators using pupil dilation and NASA –task load index," in Proceedings of International Symposium on Advancement in Ergonomics and Safety,* Arau.

Palinko, O., Kun, A. L., and Shyrokov, A. (2010). "*Estimating cognitive load using remote eye tracking in a driving simulator," in Proceedings of the 2010 Symposium on Eye-Tracking Research and Applications, ETRA 2010* (Austin, TX). doi: 10.1145/1743666.1743701

Pavlov, Y.G and Kotchoubey, B. (2017). EEG correlates of working memory performance in females. *BMC Neurosci. 18*:26. doi: 10.1186/s12868-017-0344-5

Petkar, H., Dande, S., Yadav, R., Zeng, Y., and Nguyen, T.A. (2009). "*A pilot study to assess designer's mental stress using eye gaze system and electroencephalogram," in Proceedings of the ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, IDEC/CIE (San* Diego, CA). doi: 10.1115/DETC2009-86542

Prabhakar, G., and Biswas, P. (2018). Detecting drivers' cognitive load from saccadic intrusion. *Transport. Res. F Traff. Psychol. Behav*. 54:17. doi: 10.1016/j.trf.2018.01.017

Prabhakar, G., Mukhopadhyay, A., Murthy, L., Modiksha, M. A. D. A. N., and Biswas, P. (2020). Cognitive load estimation using ocular parameters in automotive. *Transport. Eng. 2*:100008. doi: 10.1016/j.treng.2020.100008

Reising, J. M., Ligget, K. K., Solz, T. J., and Hartsock, D. C. (1995). "A comparison of two head up display formats used to fly curved instrument approaches," in *Proceedings of the Human Factors and Ergonomics Society, 39th Annual Meeting* (San Diego, CA), 1–4. doi: 10.1177/154193129503900102

Rudi, D., Kiefer, P., and Raubal, M. (2020). The instructor assistant system (iASSYST) - utilizing eye tracking for commercial aviation training purposes. *Ergonomics* 63, 61–79, doi: 10.1080/00140139.2019.1685132

Ryffel, C. P., Muehlethaler C. M., Huber, S. M., and Elfering, A. (2019). Eye tracking as a debriefing tool in upset prevention and recovery training (UPRT) for general aviation pilots. *Ergonomics* 62, 319–329. doi: 10.1080/00140139.2018.1501093

Scharinger, C., Kammerer, Y., and Gerjets, P. (2015). Pupil dilation and EEG alpha frequency band power reveal load on executive functions for link-selection processes during text reading. *PLoS ONE* 10:e0130608. doi: 10.1371/journal.pone.0130608

Schrauf, M., Sonnleitner, A., and Kincses, W. E. (2011). "*EEG alpha spindles as indicators for prolonged brake reaction time during auditory secondary task in a real road driving study*," in *Proceedings of Human Factors and Ergonomics Society Annual Meeting* (Las Vegas, NV). doi: 10.1177/1071181311551045

Sharma, S., Baijal, R., and Sinha, A. (2012). Mental workload assessment during different simulated instrument meteorological conditions, in clouds and during dark night. *Indian J. Aerosp. Med.* 56, 11–20.

Shepherd, M. J., MacDonald, A., Gray, W. R., and Cobb R. G. (2009). "Limited simulator aircraft handling qualities evaluation of an adaptive controller," in *Proceedings of IEEE Aerospace Conference, Big Sky, Montana, USA* (Big Sky, MT), 1–12. doi: 10.1109/AERO.2009.4839613

Smith, J. K., and Caldwell, J. A. (2004). *Methodology for Evaluating the Simulator Flight Performance of Pilots*. *Brooks City Base, TX: Air Force Research Laboratory.*

Thomas, P., Biswas P., and Langdon P. (2015). "State-of-the-Art and Future Concepts for Interaction in Aircraft Cockpits," in *Universal Access in Human-Computer Interaction. Access to Interaction. UAHCI 2015. Lecture Notes in Computer Science, Vol. 9176*. eds M. Antona, C. Stephanidis (Cham: Springer) 538–549. doi: 10.1007/978-3-319-20681-3_51

Tokuda, S., Obinata, G., Palmer, E., and Chaparo, A. (2011). "Estimation of mental workload using saccadic eye movements in a free-viewing task," in *Proceedings of Conference of the IEEE Engineering in Medicine and Biology Society*, 4523–4529. doi: 10.1109/IEMBS.2011.6091121

Trejo, L. J., Kubitz, K., Rosipal, R., Kochavi, R. L., and Montgomery, L. D. (2015). EEG-based estimation and classification of mental fatigue. *Psychology* 6, 572–589. doi: 10.4236/psych.2015.65055

Vrzakova, H., and Bednarik, R. (2012). "Hard lessons learned: Mobile eye-tracking in cockpits," in *Proceedings of the 4th workshop on Eye Gaze in Intelligent Human Machine Interactions* (New York, NY), 1–6. doi: 10.1145/2401836.2401843

Wanyan, X., Zhuang, D.,Wei, H, and Song, J. (2011). Pilot attention allocation model based on fuzzy theory. *Comput. Math. Appl.* 62, 2727–2735. doi: 10.1016/j.camwa.2011.06.061

Wei, Z., Damin, Z., Xiaoru, W., Chen, L., and Huan, Z. (2014). A model for discrimination and prediction of mental workload of aircraft cockpit display interface. *Chin. J. Aeronaut.* 27, 1070–1077. doi: 10.1016/j.cja.2014.09.002

Xu, J., Wang, Y., Chen, F., Choi, H., Li, G., Chen, S., et al. (2011). "Pupillary response based cognitive workload index under luminance and emotional changes," in *Proceedings of Annual Conference Extended Abstracts on Human Factors in Computing Systems* (Vancouver, BC), 1627–1632. doi: 10.1145/1979742.1979819

Yerkes, R.M., and Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Comp. Neurol. Psychol.* 18, 459–482. doi: 10.1002/cne.920180503

Zongmin, W., Damin, Z., Xiaoru, W., Chen, L., and Huan, Z. (2014). A model for discrimination and prediction of mental workload of aircraft cockpit display interface. *Chin. J. Aeronaut.* 27, 1070–1077.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.