Applied Ergonomics 77 (2019) 92–99

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

The effects of flight complexity on gaze entropy: An experimental study with fighter pilots

Carolina Diaz-Piedra[a,b,*], Hector Rieiro[a], Alberto Cherino[c], Luis J. Fuentes[d], Andres Catena[a], Leandro L. Di Stasi[a,e,*]

[a] Mind, Brain, and Behavior Research Center-CIMCYC, University of Granada, Campus de Cartuja s/n, 18071, Granada, Spain
[b] College of Health Solutions, Arizona State University, 500 N. 3rd St, 85004, Phoenix, AZ, USA
[c] First Attack Helicopter Battalion I – BHELA I (Spanish Army Airmobile Force), Almagro, Ciudad Real, Spain
[d] Department of Basic Psychology and Methodology, University of Murcia, 30100, Murcia, Spain
[e] Joint Center University of Granada - Spanish Army Training and Doctrine Command, C/ Gran Via de Colon, 48, 18071, Granada, Spain

ARTICLE INFO

Keywords:
Brain activity
Cognition
EEG
Eye movements
Eye tracking
Flight

ABSTRACT

We studied the effects of task load variations as a function of flight complexity on combat pilots' gaze behavior (i.e., entropy) while solving in-flight emergencies. The second company of the Spanish Army Attack Helicopter Battalion ($n = 15$) performed three sets of standardized flight exercises with different levels of complexity (low [recognition flights], medium and high [emergency flights]). Throughout the flight exercises we recorded pilots' gaze entropy, as well as pilots' performance (assessed by an expert flight instructor) and subjective ratings of task load (assessed by the NASA-Task Load Index). Furthermore, we used pilots' electroencephalographic (EEG) activity as a reference physiological index for task load variations. We found that pilots' gaze entropy decreased ~2% (i.e., visual scanning became less erratic) while solving the emergency flight exercises, showing a significant decreasing trend with increasing complexity ($p < .05$). This is in consonance with the ~12% increase in the frontal theta band of their EEG spectra during said exercises. Pilots' errors and subjective ratings of task load increased as flight complexity increased (p-values $< .05$). Gaze data suggest that pilots used non-deterministic visual patterns when the aircraft was in an error-free state (low complexity), and changed their scanning behavior, becoming more deterministic, once emergencies occurred (medium/high complexity). Overall, our findings indicate that gaze entropy can serve as a sensitive index of task load in aviation settings.

1. Introduction

The interaction with any safety-critical system, as during flying an aircraft, is a highly demanding and risky task that requires continuous monitoring of the system and environmental parameters. Often, when flight demands exceed the pilot's mental resources (overload), the pilot's performance is affected. That is, the deleterious effects of pilot's overload can compromise the system efficiency and safety. The evolution of avionics technologies, the development of highly standardized procedures, and the implementation of long, thorough flight training programs have helped tremendously to reduce pilot's task load and to improve flight safety. However, even in well-designed systems, unexpected events may still happen, which can increase operators' task load.

Pilot's task load levels can be inferred through several assessment methods, but the need for more sensitive, practical tools has been widely recognized for years. For example, the NASA-Task Load Index (Hart, 2006; Hart and Staveland, 1988), a well-known and widely spread instrument, has several limitations and problems similar to those imposed by self-reported measures (e.g., social desirability bias, halo and leniency effects, non-conscious activation; Podsakoff et al., 2003). Such shortcomings might limit its applicability to the aviation domain (see Table 1). Central psychophysiological indices, such as eye movement metrics, can give an insight into pilot's task load fluctuations over time in an objective, unbiased fashion that does not interfere with performance in real-life situations (Marinescu et al., 2017). Therefore, the possibility of continuously monitoring the pilot's (cognitive) state offers new opportunities to increase flight safety (see Peißl et al., 2018 for a recent review).

---

[*] Corresponding authors. Mind, Brain, and Behavior Research Center-CIMCYC, University of Granada, Campus de Cartuja s/n, 18071 Granada, Spain.
E-mail addresses: dipie@ugr.es (C. Diaz-Piedra), distasi@ugr.es (L.L. Di Stasi).

https://doi.org/10.1016/j.apergo.2019.01.012
Received 8 May 2018; Received in revised form 3 December 2018; Accepted 27 January 2019

Table 1
Summary of major advantages and disadvantages of self-reports and different eye trackers to track pilot's mental workload inside an aircraft/flight simulator. The table is, in part, based on the works of Kramer (1991), and Luximon and Goonetilleke (2001).

			Eye tracking devices					
	Self-reports (i.e., NASA-TLX)		Wireless eyeglasses-frame eye tracker 30–120 Hz		Remote/embedded eye tracker > 250 Hz		EOG-based eye tracker > 256 Hz	
	Pros	Cons	Pros	Cons	Pros	Cons	Pros	Cons
Price	✓			✓		✓		✓
Required training	✓		✓			✓		✓
Practicality Test preparation and administration[a]	✓		✓			✓		✓
Operator acceptance	✓		✓		✓			✓
Face validity The extent to which a test appears to measure what it is intended to measure (Johnson, 2013).	✓			✓		✓		✓
Generality of application The degree to which a technique has been successfully employed in laboratory, simulator, and operational environments.	✓		✓		✓		✓	
Diagnosticity The capability to discriminate among types of demanded resources.	✓		✓		✓		✓	
Instrument bulkiness	✓		✓			✓		✓
Freedom of movements	✓		✓			✓		✓
Intrusiveness The capability to measure mental workload without interfering with performance.		✓	✓		✓			✓
Continuous measurement		✓	✓		✓		✓	
Unbiased data		✓	✓		✓		✓	
Data quality/artifacts[a]	–	–		✓	✓			✓

Note. EOG = Electrooculography; NASA-TLX = NASA Task Load Index.
[a] See Supplementary material.

1.1. *Eye movement metrics to study the pilot's task load*

Since pioneering studies in the 1950s (e.g., Fitts et al., 1950; Tiffin and Bromer, 1943), eye movement metrics have represented a powerful tool for researchers to study pilot's task load variations (Di Stasi and Diaz-Piedra, 2019; Diaz-Piedra et al., 2016). Pupil dilation and blink rate represent two of the most popular indices adopted in operators' task load studies (Heard et al., 2018). Unfortunately, in real aviation settings, both indices present several shortcomings due to environmental factors, as changes in luminosity and humidity levels (Peysakhovich et al., 2017; Thomas et al., 2015). Gaze-behavior indices (e.g., saccades and fixations) are less sensitive to environmental factors and, therefore, represent a better candidate to monitor pilot's task load variations (Glaholt, 2014).

Pilot's eyes are constantly making voluntary rapid movements (i.e., saccades) to sample ambient regions and to extract relevant information (i.e., fixations). The sequential periods of rapid saccades and steady fixations define the pilot's visual scanning. That is, during each fixation, in addition to extracting information, the pilot has to make a decision as to where to look next. The spatial/temporal randomness of the pilot's visual scanning can be quantified by gaze entropy (Harris et al., 1986, for a recent review on gaze entropy, see Shiferaw et al., 2019). Gaze entropy-based metrics provide a good indication of the dispersion of the gaze over the visual field. Additionally, compared to other sensitive task load indices (e.g., saccadic velocity, see Glaholt, 2014 for a review), gaze entropy relies on less sensitive detection methods and less sophisticated eye tracking systems (a sample rate of ~30 Hz is enough to record gaze entropy). Consequently, gaze entropy-based metrics seem a more suitable and robust alternative task load index for real aviation settings (Ruigrok and Hoekstra, 2007; Shiferaw et al., 2018).

Studies focused on pilot's gaze entropy as a task load index have found contradictory results, however. For example, Di Nocera and colleagues observed changes in gaze entropy over different simulated flight phases: highly demanding flight procedures (simulated takeoff and landing) were associated with higher dispersion of the eye fixation (higher entropy), whereas less demanding phases (climb, descend, and cruise phases) were associated with lower dispersion (Di Nocera, Camilli and Terenzi, 2007). Two more recent studies that included emergency operational procedures also found that gaze entropy increased after the pilots discovered a cockpit instrument failure, which likely increased task load levels (van de Merwe et al., 2012; van Dijk, van de Merwe and Zon, 2011). These recent findings seem counterintuitive in the light of the original research that found opposite trends when pilots flew under different task loads (dual-task paradigm): the gaze entropy rate decreased with increased flight complexity (Harris et al., 1982; Tole et al., 1982). (Note that the gaze entropy rate, like the gaze entropy, indicates the randomness observed in the visual scanning [Harris et al., 1986]). Differences in entropy estimation procedures

Fig. 1. Experimental design and virtual scenario. A) The diagram illustrates an example of the emergency training session performed by one pilot (on the left) and the experimental set-up (on the right). It always started and ended with a recognition flight. In between, pilots underwent six flight exercises where a failure was intentionally introduced to the pilot's instrumentation. For example, in this case pilot #4 started and finished the session with the low complexity flights (in black), and also performed the medium (in orange) and the high complexity flights (in red). During the entire emergency training session, pilots wore the SomnoWacth EEG + 6 and the Tobii Glasses 2.0. After each flight, pilots filled in the NASA-Task Load Index. All flight exercises departed from the same simulated heliport base (*Tigerland*, *ad hoc* created virtual scenario). B) A snapshot showing the God's-eye view from a top viewpoint. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

might have generated such discrepancies. Most important is that differences in experimental procedures (e.g., Did the pilot flight error-free flights or emergency flights? Did the pilot flight under instrument or visual flight rules? Did the pilot employ standard procedures to solve a malfunction?) would lead to different results. For example, van de Merwe and colleagues (van de Merwe et al., 2012; van Dijk, van de Merwe and Zon, 2011) explained that their manipulation might not represent mental workload, but "simply increased visual scanning activities around the cockpit to find the problem". Thus, overall, the question of how task load modulates the pilot's visual scanning (i.e., gaze entropy) remains open.

Here, we aimed to determine the effects of task load variations (as a function of flight complexity) on the gaze entropy of combat pilots solving in-flight emergencies. Pilots underwent three sets of standardized flight exercises with different task complexity (low, medium, and high) in a high-fidelity helicopter simulator. As pilots were flying under visual flight rules and they used standard procedures (checklists) to solve emergencies, we expected that gaze entropy would decrease during the most complex flights.

To externally validate our results, we recorded the pilot's electroencephalographic (EEG) activity. EEG is one of the most informative methods to monitor task load in real time (Tracey and Flower, 2014). In particular, several studies have found that EEG theta power, a well-known task load index, increases as the task demands increase (for a review, see Borghini et al., 2014). Increased task demands include, for example, increased memory load (Gevins and Smith, 2003), retrieval attempts (Klimesh et al., 2005), and time pressure (Slobounov et al., 2000). We also collected expert assessments of performance, as well as subjective ratings of task load. Therefore, we expected that EEG theta power, number of errors, and perceived task load would increase during the most difficult flights.

2. Methods

2.1. Ethical approval

We conducted the study in conformity with the Code of Ethics of the World Medical Association (WMA, 2008) and under the guidelines of the University of Granada's Institutional Review Board (IRB approval #866). Each pilot signed the informed consent prior to the start of the study.

2.2. Participants

All members ($n = 15$) of the second company of the Attack Helicopter Battalion (Spanish Army Airmobile Force, Almagro, Ciudad Real, Spain) volunteered to participate in the study. Inclusion criteria were (1) normal or corrected to normal vision and (2) flight status at the time, indicating recent good health. Pilots were excluded if they had low levels of arousal and/or have been consuming alcohol before the flight session. One pilot was excluded after the arousal assessment (see *Procedure* section). The final sample included 14 men, who averaged (± standard deviation, SD) 1864.1 ± 878.5 flight hours in all aircraft types, and 768.8 ± 577.9 in the Airbus Attack Helicopter Tiger. Mean age and body mass index (± SD) were 39.0 ± 5.6 years, and 24.9 ± 1.6 kg/m^2. The pilots studied were lieutenants (n = 4), sergeant majors (n = 3), sergeant and master sergeant (n = 2, each), capitan, major, and second lieutenant (n = 1, each). Participants slept an average of (± SD) 6.25 ± 0.82 h the previous night. Most of them (64.3%) had consumed coffee (~95 mg of caffeine/1 cup of coffee) before the flight session.

2.3. Experimental design

The experiment followed a repeated measures design. Pilots underwent a full emergency training session consisting of three sets (~20 min long each) of flight exercises of different complexity (low, medium, and high, see next section and Fig. 1). Thus, *flight complexity* served as the within-subjects factor. Gaze entropy, EEG power spectra, performance, and the pilot's subjective ratings of task load were the dependent variables.

2.4. Apparatus and simulated scenario

2.4.1. Helicopter simulator and flight exercises

Pilots flew the high-fidelity fixed-base Armed Reconnaissance Helicopter Tiger Full Flight & Mission Simulator (Indra S.A., Alcobendas, Spain) (henceforth, the Tiger simulator, see Fig. 1A). It simulates all aspects of the Tiger's operation and environment, and is used to train pilots and other flight deck crew.

To mitigate the influence of fatigue on the simulated emergency training, the flight session lasted ~60 min (Di Stasi, McCamy, et al., 2016) and consisted of eight flight exercises with visual meteorological conditions (i.e., weather conditions that are clear enough to allow the pilot to see where the aircraft is going). The flight exercises and the simulated environment (*Tigerland*, see Fig. 1B) were similar to those pilots have to perform in their everyday military training. Each flight exercise started with a rapid checklist procedure and ended after the landing. Two out of eight flights (recognition flights, the *1*st and the *8*th flights) represented the Low complexity flights, as they did not present any emergency situation. Although there were just two low complexity flights, these exercises were the longest flights. Six out of eight flights did present emergency situations. Among these six flight exercises, three represented the Medium complexity flights (flying with a multifunction display failure simultaneously with metal fragments in the engines lubricating oil; flying with one engine inoperative circuit; and flying simultaneously with one engine inoperative circuit and with an engine on fire). Finally, the remaining three flight exercises represented the High complexity flights (performing an auto-rotation landing; flying simultaneously with one engine inoperative circuit and suffering a hydraulic failure while performing a crosswind landing; and landing while suffering a tail rotor failure). The order of appearance of the flights with emergencies (that is, from the *2*nd to the *6*th) varied among pilots. For technical reasons, the *7*th flight was always a tail rotor failure (this exercise increases the likelihood of the simulator failure). A flight instructor (the same for all the pilots, a master sergeant with more than 3000 h of flights in all aircraft types) assessed pilot's performance as the percentage of correct required maneuvers that were executed during each flight exercise.

2.4.2. Eye movement recordings

We sampled eye movements binocularly at 30 Hz, using the Tobii Glasses 2.0 (Tobii AB, Sweden), a wearable eye tracking system. The device consists of an eye tracking unit mounted onto an eyeglasses frame and a small recording unit. The eye tracker is connected to the recording unit via a HDMI cable. Recordings are stored in a SD memory card. See recent works by Diaz-Piedra and colleagues (Diaz-Piedra et al., 2017) for a detailed description of the system.

2.4.3. Electroencephalographic recording

We recorded the pilot's EEG activity using the SOMNOwatch + EEG-6 (Somnomedics, Germany), which consists of two small thin boxes (SOMNOwatch and EEG headbox with ten electrodes) that were kept into the flight suit (pocket on the upper left sleeve) and attached to the nametag Velcro patch, respectively. This kind of device has been used before to perform real flight recordings (Di Stasi et al., 2015), and it is therefore robust to movements and noise (i.e., artifacts from electrode movement that lead to changes in contact impedance or even the generation of a triboelectric response on the wires). The device samples data at 256 Hz applying a band pass filter (0.1 Hz - 80 Hz). Impedance was kept below 5 kΩ for all electrodes. We used a monopolar montage with gold cup electrodes (Natus Neurology Incorporated - Grass Products Warwick, US) at five active scalp sites: F3, F4, C3, C4, and Cz placed according to the international 10/20 system (Klem et al., 1999), and using the mastoids (A1 and A2) as references. Ground was placed at FpZ. The channel Cz was recorded by default as an internal device requirement (internal reference). We recorded vertical and horizontal eye movements placing an electrode ~1 cm out from the outer canthus of the right eye and another ~1 cm below the left eye. The device collects internally the raw EEG data. We used the DOMINO Light software (version 14.0, Somnomedics, Germany) to export raw signals to EDF + files. Then, we imported the EDF + files, preprocessed and analyzed them using Matlab (Mathworks Inc., USA, see Electroencephalographic analyses: Band power spectra section).

2.5. Questionnaires

2.5.1. Stanford sleepiness scale (SSS)

The SSS (Hoddes et al., 1973) provides a global measure of alertness, ranging between 1 and 7. It contains seven statements ranging from “Feeling active, vital, alert, or wide awake” (score 1) to “No longer fighting sleep, sleep onset soon, having dream-like thoughts” (score 7).

2.5.2. NASA-task load index (NASA-TLX)

The NASA-TLX (Hart, 2006; Hart and Staveland, 1988) is an indicator of the degree of task load that pilots experienced while performing the flight exercises. The NASA-TLX is a scale with six bipolar dimensions: mental demand (MD); physical demand (PD); temporal demand (TD); own performance; effort; and frustration. The first three dimensions (MD, PD, TD) reflect task-related factors such as task complexity. NASA-TLX values range between 0 and 100, with higher values indicating higher task load.

2.6. Procedure

Pilots attended the Helicopter Training Center (CEFAMET) situated in the “Coronel Sánchez Bilbao” base (Almagro, Ciudad Real, Spain) for the experiment. CEFAMET facility houses several helicopter flight training simulator platforms, which provides general skills and tactical training for pilots. The training includes instructions on the use of the mission communications, navigation, and weapon systems. Here, as part of a broader assessment of physical and cognitive performance, pilots flew the Tiger simulator. All pilots were naïve to the aim of the experiment. In the context of assessing day-to-day military duties, we allowed pilots to continue their routine schedules. The flight instructor gave to each pilot a pre-flight briefing about the simulated emergency training session. After that, we collected sociodemographic and health data, as well as flight hours, and we placed the EEG electrodes. Before electrodes were placed, the pertinent areas of skin in the scalp and areas around the eyes were cleaned up with a slightly abrasive paste. Gold electrodes were filled with conductive paste and attached with collodion. We drove the pilot to the simulator after finishing the set-up of the electroencephalograph. There, the eye tracking system was set up and we performed the calibration procedure consisting in a short fixation target. The calibration procedure was repeated between flight exercises as needed. Right before the *1*st flight exercise, pilots filled in the SSS for screening purposes. Participants who scored more than 3 (Morales et al., 2017) were excluded from further testing ($n = 1$).

The simulation lasted around 1 h (0:54.3 ± 0:07.1 h). Pilots started the simulation with a recognition flight exercise (low complexity). Then, they performed six flight exercises that required emergency maneuvers (medium and high levels of complexity). Pilots finished the simulation with a final recognition flight similar to the first one. After each flight exercise, pilots filled in the NASA-TLX. Lastly, the instructor gave the pilot the pertinent debriefing, before removing the electrodes. All participants had the order to not share the contents of the flight session with their colleagues.

2.7. Electroencephalographic analyses

We analyzed EEG data using the MatLab EEGLab software package (Mathworks Inc., USA). We filtered the data using a Butterworth filter with the -3 dB bandpass corresponding to the interval [0.5 Hz-32 Hz]. We corrected for eye artifacts using a regression procedure (Gratton et al., 1983) to subtract the signal recorded with the two electrodes placed around the eyes from each data electrode.

We split the continuous EEG data in periods of variable length corresponding to each of the 8 flight exercises performed, and then split the data for each flight in non-overlapping 2-s segments. To reduce the influence of artifacts, both physiological and non-physiological, in the analysis, we discarded data segments containing voltage values outside the [-100 μV, 100 μV] interval. We estimated the power spectra for each flight exercise (low, medium, and high complexity) by averaging and normalizing the Fourier transforms of the data contained in the valid segments, weighed using a Hamming window (Bartlett's method). The spectra was divided in four power bands (δ, delta, 0.5–4 Hz; θ, theta, 4–8 Hz; α, alpha, 8–13 Hz; β, beta, 13–30 Hz), and we obtained average values of power for each frequency band, channel, and flight exercise (Di Stasi et al., 2015). Due to the reduced number of participants and, in order to reduce degrees of freedom of the error term, the power values for the frontal (F3, F4) and central (C3, C4) brain regions were averaged (channels were highly correlated, $r = 0.91$, and $r = 0.96$, respectively).

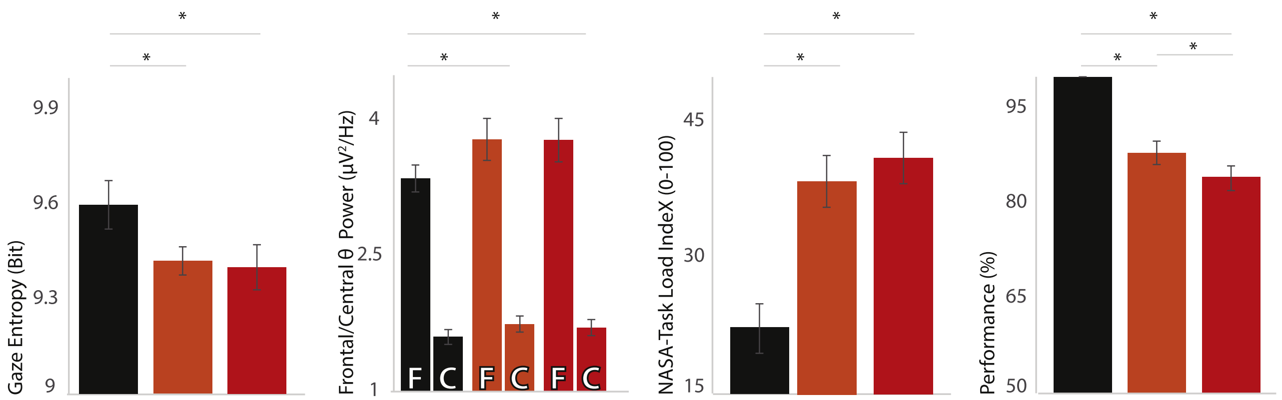

Fig. 2. The effects of flight complexity on psychophysiological indices (gaze entropy and electroencephalographic frontal [F] and central [C] theta power), subjective ratings of task load, and performance. Data from the low complexity flights indicated in black, from the medium complexity flights in orange, and from the high complexity flights, in red. For NASA-Task Load Index, higher scores indicate higher perceived levels of task load. Error bars represent the SEM across participants. The asterisk represents corrected *p*-values < .05. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

EEG data from one pilot were lost due to a failure of the recording system.

2.8. Gaze entropy analyses

We detected and analyzed gaze parameters as in Di Stasi and colleagues (Di Stasi, Diaz-Piedra et al., 2016). Briefly, we identified blink periods as portions of the raw data where eye information was missing for 100 ms or more, and removed these segments from the analysis. We further removed the 200 ms before and after each blink or semi-blink to eliminate the initial and final parts during which the pupil was still partially occluded. To measure (gaze) entropy, we used Shannon's entropy formula (Shannon, 1948) defined as:

$$H_g(X) = -\sum p(x, y)\cdot\log_2 p(x, y)$$

where p(x,y) is the probability of the pilot's gaze falling in the (x,y) position of the visual field for a given sample, estimated from the full recording. This gives a measure of the average uncertainty over the position of gaze on an instant in time during the flight exercise, or, equivalently, the information provided by a single observation, measured in bits. Therefore, it is a measure of gaze dispersion. In order to calculate the gaze entropy, we discretized the visual field (81 × 48 degrees of visual angle [dva]; i.e. the total visual field allowed by the eye tracker device) in 3888 bins of 1 × 1 dva, and calculated the probabilities of the gaze falling in each of these bins on any given time sample. The gaze entropy values for each flight exercise – within the same flight complexity (low, medium, and high) – were averaged. Gaze data from one pilot were lost due to a failure of the recording system.

2.9. Statistical analyses

We performed separate repeated measures analysis of variance (ANOVA) for each dependent variable (i.e., gaze entropy, subjective ratings of task load, as well as the pilot's flight performance) with *flight complexity* as the within-subjects factor (three levels: low complexity, medium complexity, and high complexity). Thus, data were analyzed using a within-subjects design (i.e., comparing each pilot to himself across the three levels of flight complexity), and therefore the variability between pilots was part of the error term. For the EEG data, we performed a 3 (*flight complexity*) x 2 (*channel*) x 4 (*frequency band*) repeated-measures ANOVA. We used the Greenhouse-Geisser adjustment to correct for the violation of the sphericity assumption, thus all *p*-values are reported with this correction. We also performed separate trend analyses for each dependent variable. We used the Bonferroni correction for multiple comparisons. We used partial η^2 (η_p^2, calculated for a repeated measures design) to estimate the effect size. Significance levels were always set at $p < .05$.

3. Results

We aimed to determine the effects of task load (by means of flight complexity) on the pilot's central psychophysiological responses (brain activity and eye movements), performance, and the pilot's subjective ratings of task load. Pilots performed three sets (of ~20 min each) of standardized simulated flight exercises while we recorded their EEG signals and eye movements.

3.1. Effects of flight complexity on pilot's gaze entropy

Flight complexity affected the pilot's gaze entropy, $F(2,24) = 5.93$, $p = .008$, $\eta_p^2 = 0.33$; showing a significant linear trend $F(1,12) = 9.11$, $p = .011$, $\eta_p^2 = 0.43$. Gaze entropy was statistically lower in the emergency flight exercises (medium [mean bits ± SD, 9.42 ± 0.16] and high [mean bits ± SD, 9.40 ± 0.26] complexity) than the recognition flights (low complexity [mean bits ± SD, 9.59 ± 0.27]); corrected *p*-values < .05, indicating that exploration pattern become more stereotyped (i.e., less random) during the more complex flights (see Fig. 2; Table 2).

3.2. Effectiveness of the flight complexity manipulation: EEG, performance, and subjective data

To examine the effectiveness of the flight complexity manipulation, we analyzed the pilot's brain activity (EEG power spectra) as well as performance (correctness of executed flight procedures), and the pilot's subjective ratings of task load (i.e., NASA-TLX scores).

As expected, we observed main effects of both *channel* and *frequency band* on the pilots' EEG power spectra; $F(1,12) = 130.12$, $p < .001$, $\eta_p^2 = 0.92$; $F(3,36) = 192.25$, $p < .001$, $\eta_p^2 = 0.94$; respectively. *Flight complexity* also modulated the EEG power spectra across all EEG frequency bands; $F(2,24) = 5.93$, $p = .008$, $\eta_p^2 = 0.33$; showing a significant linear trend $F(1,12) = 6.03$, $p = .030$, $\eta_p^2 = 0.34$ (mean ± SD for low, medium, and high complexity flights: 5.70 ± 1.05, 6.21 ± 1.21, 6.10 ± 1.18, respectively). The interaction of *flight complexity*, *channel*, and *frequency band* was significant; $F(6,72) = 4.072$, $p = .043$, $\eta_p^2 = 0.25$. Simple effects analysis of this interaction showed that the power spectra in theta band at the frontal and the central channels were lower during the recognition flights (low complexity) compared to the emergency flights (medium and high complexity) (mean ± SD for frontal channels: 3.36 ± 0.54 vs.

Table 2
The effects of flight complexity on psychophysiological indices, performance, and subjective ratings of task load. Electroencephalographic delta, theta, alpha, and beta power measured at frontal (F3/F4) and central (C3/C4) locations, gaze entropy, correctness of the flight procedures, and perceived task load measured by the NASA-Task Load Index (NASA-TLX).

		Flight complexity		
		Low	Medium	High
		M (SD)		
Delta (μV^2/Hz)	Frontal	26.163 (6.475)	28.464 (7.123)	28.027 (7.211)
	Central	9.774 (2.164)	10.457 (2.754)	10.305 (2.369)
Theta (μV^2/Hz)	Frontal*	3.360 (0.536)	3.790 (0.836)	3.783 (0.866)
	Central*	1.605 (0.293)	1.747 (0.318)	1.709 (0.318)
Alpha (μV^2/Hz)	Frontal	1.423 (0.391)	1.453 (0.241)	1.393 (0.221)
	Central	1.080 (0.342)	1.103 (0.317)	1.068 (0.281)
Beta (μV^2/Hz)	Frontal	1.108 (0.467)	1.283 (0.755)	1.203 (0.692)
	Central	1.106 (0.848)	1.398 (1.496)	1.286 (1.292)
Gaze entropy (bits)*		9.60 (0.276)	9.42 (0.160)	9.40 (0.257)
Performance (%)*		100 (0)	87.98 (6.87)	84.18 (7.26)
NASA-TLX *		21.77 (10.24)	38.57 (10.70)	41.16 (10.64)

Note: M = Mean; SD = Standard deviation.
* corrected *p*-values < .05.

3.79 ± 0.83 and 3.78 ± 0.87; central channels: 1.60 ± 0.29 vs. 1.74 ± 0.32 and 1.71 ± 0.32, corrected *p*-values < .05) (see Fig. 2; Table 2). In both cases, there were no differences between medium and high complexity flights in theta EEG power. No significant correlations have been found between EEG data and gaze entropy (data not shown).

Coherently, pilots' subjective ratings of task load (as expressed by the NASA-TLX scores) differed depending on *flight complexity*, $F(2,26) = 43.68$, $p < .001$, $\eta_p^2 = 0.77$; showing a significant linear trend $F(1,13) = 61.86$, $p < .001$, $\eta_p^2 = 0.83$. Pilots reported higher levels of perceived task load after performing the emergency flights (mean NASA-TLX scores ± SD for medium: 38.57 ± 10.70 and high complexity flights: 41.16 ± 10.63) compared to the recognition flights (mean NASA-TLX scores ± SD for low complexity flights: 21.77 ± 10.24); corrected *p*-values < .05. There were no differences between medium and high complexity flights in NASA-TLX scores.

Finally, pilots' performance, as measured by the flight instructor, lowered as *flight complexity* increased, $F(2, 26) = 45.84$, $p < .001$, $\eta_p^2 = 0.78$; showing a significant linear trend $F(1,13) = 66.48$, $p < .001$, $\eta_p^2 = 0.84$. Pilot's performance during the recognition flights was significantly higher (100% ± 0.0 of required maneuvers performed) compared to the emergency flights (for the medium and high complexity flights pilots performed 87.99% ± 6.87 and 84.18% ± 7.23, respectively, of the required maneuvers); corrected *p*-values < .05 (see Fig. 2; Table 2).

Overall, these results indicated the correct manipulation of *flight complexity*: medium/high complexity flight exercises (from the 2nd to the 7th flight) led to higher power spectra in theta band at the frontal and the central channels, subjective perception of higher task load, and less correct procedures than the low complexity flights (the 1st and the 8th flight).

4. Discussion

Although environmental and structural factors can affect flight safety, the pilot's overload seems likely to remain the single greatest functional limitation to reach a fully reliable and error-free system. Currently, monitoring the pilots' task load is guesswork and there are few real-time cues (e.g., flight data and pilot's actions) available for the aircrew to assess if the pilot is safely interacting with the aircraft (Dehais et al., 2008). Thus, sensitive methods to monitor pilot's task load and therefore to improve flight safety are needed. Previous studies have shown that both eye movements-based and EEG-based indices, recorded separately and together, represent good metrics to track pilots' workload in real and simulated aviation settings (see Charles and Nixon, 2019, for a recent review). Here, we aimed to find conclusive evidence about the effects of task load (as a function of flight complexity) on pilot's gaze entropy while solving several in-flight emergencies. We examined, for the first time, how gaze data (randomness of visual scanning) changed together with EEG theta power, a well-known index of pilot's task load in real and simulated flights (Borghini et al., 2014; Di Stasi et al., 2015). Our combined results indicate that highly task demanding situations (in-flight emergencies) reduced pilot's gaze entropy. These results confirm the original findings obtained, in a similar experimental setting, while pilots were flying using instrument reading procedures (Harris et al., 1982; Tale et al., 1982), and reconcile previous disparate results.

4.1. The effects of flight complexity on EEG, performance, and subjective data

We analyzed the pilot's EEG activity, performance, and subjective ratings of task load to externally validate pilot's gaze entropy changes associated with flight complexity (i.e., task load). Overall, these indices provide unambiguous evidence about our successful manipulation of task complexity: when solving in flight-emergencies, pilots experienced higher level of task load at physiological and subjective levels, as well as lowered their performance.

EEG power reflects the amount of neurons that discharge at the same time (Klimesch, 1999). This discharge generates oscillatory activities that are task dependent; that is, oscillations occur more frequently during high than low demanding tasks (Astolfi et al., 2011). Here, we found higher EEG frontal and central theta power while pilots were performing the in-flight emergencies (medium and high flight complexity), as compared to recognition flights (low flight complexity). These findings confirm previous results on the relationship between EEG theta power and cognitive effort (Shaw et al., 2018), as well as between task load and EEG theta power in pilots (Borghini et al., 2014; Di Stasi et al., 2015; Smith and Gevins, 2005). In addition, this effect might be related to the specific features of the most demanding flights: a higher visual memory load was required when pilots were retrieving information to solve the in-flight emergencies (Klimesch, 1999; Onton et al., 2005). Furthermore, variations in arousal levels also affect EEG theta power (Eoh et al., 2005). Because task complexity modulates arousal (Yerkes and Dodson, 1908), the occurrence of the in-flight emergencies might have modulated pilot's arousal levels, which in turn could influence EEG power signals (Di Stasi et al., 2015).

Consistently, pilot performance was lower (the number of errors increased) and perceived levels of task complexity were higher for the in-flight emergencies exercises. Performance degradation and subjective results are in line with earlier studies using similar experimental procedures (e.g., Tsang and Wilson, 1997).

4.2. The effect of task complexity on pilot's gaze entropy

We observed a significant decrease in pilot's gaze entropy when pilots were solving the in-flight emergencies, compared to the recognition flights (low complexity). That is, pilots followed a more systematic visual scanning during the most complex scenarios than during the less complex flights, confirming the original studies (e.g., Harris et al., 1986). This suggests the possibility that pilots might use nondeterministic visual patterns when the aircraft is in an error-free state (recognition flight) and changed their scanning behavior once emergencies are detected. These results might also reflect the well-documented "attentional tunneling" phenomenon (perceptual narrowing or weapon focus) (Christianson, 1992; Easterbrook, 1959; Loftus et al., 1987; Mackworth, 1965) that generally occurs when the focus of attention narrows with high task demands, memory load, or stress (Dirkin, 1983; Hockey, 1997). Furthermore, our findings confirm previous results observed in non-aviation scenarios. For example, several driving studies have also found that spatial gaze variability lowers as the task complexity increases (dual-task paradigm, e.g., Recarte and Nunes, 2003; Schieber and Gilland, 2008). For several reasons, piloting a rotorcraft is, arguably, more complex and dangerous than driving a car (Chuang et al., 2013). Therefore, although it may not be a straightforward comparison, it also supports our complexity-based explanation.

One might wonder if the present results are in contradiction with those recently observed in healthcare settings (Kataoka et al., 2011; Di Stasi, Diaz-Piedra et al., 2016; Di Stasi et al., 2017; Diaz-Piedra et al., 2017), that have found an opposite trend for gaze entropy with increased levels of operator's task load. This possibility seems unlikely in light of one of the main characteristic in the aviation system: the highly standardized procedures required to pilots to solve emergencies. That is, the level of standardization in aviation procedures is extremely high, and somewhat lower (or inexistent) in healthcare operational settings (Kapur et al., 2015). Pilots are required to carefully follow written standard operating procedures at any stage of the flight and at any operation (normal, abnormal, and emergency operations). Doing so, they acquire a full understanding of the aircraft in order to conduct safe and efficient flights (Giles, 2013). This standardization is crucial for flight safety and affects directly pilots' visual scanning strategies (e.g., Haslbeck and Zhang, 2017). For example, one of the most recent studies analyzing gaze entropy in flight exercises recruited participants with no previous experience of real/simulated flight (Allsop and Gray, 2014). Thus, those results (and others from similar studies) might not be generalizable to studies with experienced pilots. In our study, all participants were members of the same elite military battalion (highly-trained pilots).

Finally, previous studies have investigated gaze entropy as pilot's task load index in error-free flight phases (Di Nocera, Camilli and Terenzi, 2007). Here, with the exception of the recognition flights, all the exercises included the response to certain emergencies. The medium complexity scenarios required following a correct sequence of actions based on emergency checklists, whereas the high complexity situations required immediate actions, normally carried out from memory. In both cases, once the emergency occurred, pilots have to perform specific well-learnt steps to solve it. This might confirm the less random (i.e. more deterministic) visual scanning strategy. Thus, the most parsimonious explanation for the decrease in gaze entropy observed here might be explained by an increase of task load levels.

Pilot's gaze entropy as a workload index satisfies several neuroergonomics criteria to establish an ideal measure of workload (see Table 1), such as its non-invasiveness or its convenience. It does not fully satisfy the sensitivity criterion, however (Luximon and Goonetilleke, 2001), as pilot's gaze entropy was not able to discriminate between the three levels of flight complexity (low, medium, and high). It differentiated only between the error-free state (low complexity: recognition flights) and error state flights (medium and high complexity: in-flight emergencies). It is plausible that the way we classified the six in-flight emergency exercises into two different subsets of states (medium and high complexity flights exercises) was not sufficiently accurate. We created the states/subsets of states based on the expert judgment of the chief flight instructor of the squadron (a highly experienced fighter pilot with more than 2000 flight hours on the Tiger helicopter) about the cognitive demands and the activities that the pilot should perform in each of the eight flight exercises, and the overall risk of critical outcomes (i.e., crashes). Disregarding the multitasking nature of flying a helicopter (recognition flights), the in-flight emergency exercises included the concurrent execution of overlapping maneuvers/procedures in a situation with high chances of crashing. Thus, it is plausible that the medium and the high flight complexity exercises demanded resources not different enough to induce evident changes in the pilot's workload levels. Pilots' EEG-power activity, performance, and subjective ratings of task load corroborate this explanation. Future studies should disentangle this issue.

Moreover, the methodological challenges imposed by the respective devices limited real synchronization. Further research is needed to address this issue. It would be utterly relevant, for instance, to address the basic relationship, as well as temporal contingency, between different eye movements indices (e.g., fixation duration, saccadic velocity) and electrophysiological indices (e.g., component amplitude in event-related potentials) of task complexity in realistic scenarios.

To conclude, our findings could help to bridge the gap between flight safety and neuroscience, by offering valid and conclusive evidence on the sensitivity of pilot's gaze entropy to monitor pilot's task load while performing ecological and complex tasks. The continuous monitoring not only would allow detecting overload (and underload) situations, but providing online feedback to the system (either automated equipment or the crew) which would be able to take countermeasures and prevent fatal errors. Furthermore, our findings can be applied to other safety critical-systems (e.g., nuclear platform operators) where vision is the dominant sensory system supporting the operator's functions, and the management of emergency/abnormal operations are mainly guided by standardized emergency operating procedures (Bhavsar et al., 2017; Gracia and Martínez-Córcoles, 2018).

Declarations of interest

None.

Acknowledgements

This study was funded by the CEMIX UGR-MADOC/Santander Bank (Project PINS 2014-16 to CDP & LLDS) and partially supported by a Spanish Ministry of Economy and Competitiveness grant (DEP2013-48211-R). The funding organizations had no role in the design or conduct of this research.

Research by LLDS is supported by the Ramón y Cajal fellowship program (RYC-2015-17483). We thank Tobii Group for the Eye Track Awards Stipend awarded to LLDS. We thank Dr. Jesus Vera (University of Granada) and the staff from Indra Sistemas, S.A., for their help during the data collection, and Jose M. Morales, MSc (University of Granada) for his help in data processing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apergo.2019.01.012.

References

Allsop, J., Gray, R., 2014. Flying under pressure: effects of anxiety on attention and gaze behavior in aviation. J. Appl. Res. Mem. Cogn. 3, 63–71.

Astolfi, L., Toppi, J., Borghini, G., Vecchiato, G., Isabella, R., De Vico Fallani, F., Babiloni, F., 2011. Study of the functional hyperconnectivity between couples of pilots during flight simulation: an EEG hyperscanning study. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. 2338–2341 2011.

Bhavsar, P., Srinivasan, B., Srinivasan, R., 2017. Quantifying situation awareness of control room operators using eye-gaze behavior. Comput. Chem. Eng. 106, 191–201.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F., 2014. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci. Biobehav. Rev. 44, 58–75.

Charles, R.L., Nixon, J., 2019. Measuring mental workload using physiological measures: a systematic review. Appl. Ergon. 74, 221–232.

Christianson, S.Å., 1992. Emotional stress and eyewitness memory: a critical review. Psychol. Bull. 112, 284–309.

Chuang, L.L., Nieuwenhuizen, F.M., Bülthoff, H.H., 2013, July. A fixed-based flight simulator study: the interdependence of flight control performance and gaze efficiency. In: Harris, D. (Ed.), International Conference on Engineering Psychology and Cognitive Ergonomics. Springer International Publishing, Cham, Switzerland, pp. 95–104.

Dehais, F., Causse, M., Pastor, J., 2008, March. Embedded eye tracker in a real aircraft: new perspectives on pilot/aircraft interaction monitoring. In: Zellweger, A., Donohue, G., Duong, V. (Eds.), Proceedings from the 3rd International Conference on Research in Air Transportation. Federal Aviation Administration, Fairfax, USA, pp.

303–310.

Di Nocera, F., Camilli, M., Terenzi, M., 2007. A random glance at the flight deck: pilots' scanning strategies and the real-time assessment of mental workload. J. Cogn. Eng. Decis. Making 1, 271–285.

Di Stasi, L.L., Diaz-Piedra, C., 2019. Eye movements research in aviation: past, present, and future. In: Vidulich, M.A., Tsang, P.S. (Eds.), Improving Aviation Performance through Applying Engineering Psychology, Advances in Aviation Psychology (Vol. 3). CRC Taylor and Francis Group, Boca Raton, FL, US, pp. 143–162.

Di Stasi, L.L., Diaz-Piedra, C., Rieiro, H., Sanchez Carrión, J.M., Berrido, M.M., Olivares, G., Catena, A., 2016a. Gaze entropy reflects surgical task load. Surg. Endosc. 30, 5034–5043.

Di Stasi, L.L., Diaz-Piedra, C., Ruiz-Rabelo, J.F., Rieiro, H., Carrion, J.M.S., Catena, A., 2017. Quantifying the cognitive cost of laparo-endoscopic single-site surgeries: gaze-based indices. Appl. Ergon. 65, 168–174.

Di Stasi, L.L., Diaz-Piedra, C., Suárez, J., McCamy, M.B., Martinez-Conde, S., Roca-Dorda, J., Catena, A., 2015. Task complexity modulates pilot electroencephalographic activity during real flights. Psychophysiology 52, 951–956.

Di Stasi, L.L., McCamy, M.B., Martinez-Conde, S., Gayles, E., Hoare, C., Foster, M., Macknik, S.L., 2016b. Effects of long and short simulated flights on the saccadic eye movement velocity of aviators. Physiol. Behav. 153, 91–96.

Diaz-Piedra, C., Rieiro, H., Suarez, J., Rios-Tejada, F., Catena, A., Di Stasi, L.L., 2016. Fatigue in the military: towards a fatigue detection test based on the saccadic velocity. Physiol. Meas. 37, 62–75.

Diaz-Piedra, C., Sanchez-Carrion, J.M., Rieiro, H., Di Stasi, L.L., 2017. Gaze-based technology as a tool for surgical skills assessment and training in urology. Urology 107, 26–30.

Dirkin, G.R., 1983. Cognitive tunneling: use of visual information under stress. Percept. Mot. Skills 56, 191–198.

Easterbrook, J.A., 1959. The effect of emotion on cue utilization and the organization of behavior. Psychol. Rev. 66, 183–201.

Eoh, H.J., Chung, M.K., Kim, S.-H., 2005. Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. Int. J. Ind. Ergon. 35, 307–320.

Fitts, P.M., Jones, R.E., Milton, J.L., 1950. Eye movements of aircraft pilots during instrument-landing approaches. Aero. Eng. Rev. 9, 24–29.

Gevins, A., Smith, M.E., 2003. Neurophysiological measures of cognitive workload during human-computer interaction. Theor. Issues Ergon. Sci. 4, 113–131.

Giles, C.N., 2013. Modern airline pilots' quandary: standard operating procedures—to comply or not to comply. J. Aviat. Technol. Eng. 2, 2–12.

Glaholt, M., 2014. Eye Tracking in the Cockpit: a Review of the Relationships between Eye Movements and the Aviators Cognitive State (Report No. R153). Defence Research and Development Canada, Toronto, Ontario, Canada Retrieved from. http://www.dtic.mil/docs/citations/AD1000097.

Gracia, F.J., Martínez-Córcoles, M., 2018. Understanding risky behaviours in nuclear facilities: the impact of role stressors. Saf. Sci. 104, 135–143.

Gratton, G., Coles, M.G., Donchin, E., 1983. A new method for off-line removal of ocular artifact. Electroencephalogr. Clin. Neurophysiol. 55, 468–484.

Harris, R.L., Glover, B.J., Spady, A.A., 1986. Analytical Techniques of Pilot Scanning Behavior and Their Application (Report No. NASA TP-2525). NASA Langley Research Center, Hampton, VA, US Retrieved from. https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19860018448.pdf.

Harris, R.L., Tole, J.R., Ephrath, A.R., Stephens, A.T., 1982. How a new instrument affects pilots' mental workload. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 26, 1010–1013 Los Angeles, CA, US: Sage Publications.

Hart, S.G., 2006. NASA-task load index (NASA-TLX); 20 years later. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 50, 904–908.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv. Psychol. 52, 139–183.

Haslbeck, A., Zhang, B., 2017. I spy with my little eye: analysis of airline pilots' gaze patterns in a manual instrument flight scenario. Appl. Ergon. 63, 62–71.

Heard, J., Harriott, C.E., Adams, J.A., 2018. A survey of workload assessment algorithms. IEEE Trans. Hum.-Mach. Syst. https://doi.org/10.1109/THMS.2017.2782483.

Hockey, G.R.J., 1997. Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. Biol. Psychol. 45, 73–93.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., Dement, W.C., 1973. Quantification of sleepiness: a new approach. Psychophysiology 10, 431–436.

Johnson, E., 2013. Face validity. In: Volkmar, F.R. (Ed.), Encyclopedia of Autism Spectrum Disorders. Springer, New York, NY, US, pp. 46.

Kapur, N., Parand, A., Soukup, T., Reader, T., Sevdalis, N., 2015. Aviation and healthcare: a comparative review with implications for patient safety. JRSM Open 7, 2054270415616548.

Kataoka, J., Sasaki, M., Kanda, K., 2011. Effects of mental workload on nurses' visual behaviors during infusion pump operation. Jpn. J. Nurs. Sci. 8, 47–56.

Klem, G.H., LuÈders, H.O., Jasper, H.H., Elger, C., 1999. The ten-twenty electrode system of the International Federation. Electroencephalogr. Clin. Neurophysiol. 52, 3–6.

Klimesch, W., 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Rev. 29, 169–195.

Klimesch, W., Schack, B., Sauseng, P., 2005. The functional significance of theta and upper alpha oscillations. Exp. Psychol. 52, 99–108.

Kramer, A.F., 1991. Physiological Metrics of Mental Workload: A Review of Recent Progress. (Report No. NPRDC-TN-90-23). Office of Naval Technology, Arlington, VA, US.

Loftus, E.F., Loftus, G.R., Messo, J., 1987. Some facts about "weapon focus". Law Hum. Behav. 11, 55–62.

Luximon, A., Goonetilleke, R.S., 2001. Simplified subjective workload assessment technique. Ergonomics 44, 229–243.

Mackworth, N.H., 1965. Visual noise causes tunnel vision. Psychonomic Sci. 3, 67–68.

Marinescu, A.C., Sharples, S., Ritchie, A.C., Sánchez López, T., McDowell, M., Morvan, H.P., 2017. Physiological parameter response to variation of mental workload. Hum. Factors 60, 31–56.

Morales, J.M., Diaz-Piedra, C., Rieiro, H., Roca-González, J., Romero, S., Catena, A., Di Stasi, L.L., 2017. Monitoring driver fatigue using a single-channel electroencephalographic device: a validation study by gaze-based, driving performance, and subjective data. Accid. Anal. Prev. 109, 62–69.

Onton, J., Delorme, A., Makeig, S., 2005. Frontal midline EEG dynamics during working memory. Neuroimage 27, 341–356.

Peißl, S., Wickens, C.D., Baruah, R., 2018. Eye-tracking measures in aviation: a selective literature review. Int. J. Aerospace Psychol. https://doi.org/10.1080/24721840.2018. 1514978.

Peysakhovich, V., Vachon, F., Dehais, F., 2017. The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load. Int. J. Psychophysiol. 112, 40–45.

Podsakoff, P.M., MacKenzie, S.B., Lee, J.Y., Podsakoff, N.P., 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. J. Appl. Psychol. 88, 879–903.

Recarte, M.A., Nunes, L.M., 2003. Mental workload while driving: effects on visual search, discrimination, and decision making. J. Exp. Psychol. Appl. 9, 119–137.

Ruigrok, R.C., Hoekstra, J.M., 2007. Human factors evaluations of Free Flight: issues solved and issues remaining. Appl. Ergon. 38, 437–455.

Schieber, F., Gilland, J., 2008. Visual entropy metric reveals differences in drivers' eye gaze complexity across variations in age and subsidiary task load. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 52, 1883–1887.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423.

Shaw, E.P., Rietschel, J.C., Hendershot, B.D., Pruziner, A.L., Miller, M.W., Hatfield, B.D., Gentili, R.J., 2018. Measurement of attentional reserve and mental effort for cognitive workload assessment under various task demands during dual-task walking. Biol. Psychol. 134, 39–51.

Shiferaw, B., Downey, L., Crewther, D., 2019. A review of gaze entropy as a measure of visual scanning efficiency. Neurosci. Biobehav. Rev. 96, 353–366.

Shiferaw, B.A., Downey, L.A., Westlake, J., Stevens, B., Rajaratnam, S.M., Berlowitz, D.J., Howard, M.E., 2018. Stationary gaze entropy predicts lane departure events in sleep-deprived drivers. Sci. Rep. 8, 2220.

Slobounov, S.M., Fukada, K., Simon, R., Rearick, M., Ray, W., 2000. Neurophysiological and behavioral indices of time pressure effects on visuomotor task performance. Cognit. Brain Res. 9, 287–298.

Smith, M.E., Gevins, A., 2005. Neurophysiologic monitoring of mental workload and fatigue during operation of a flight simulator. In: Caldwell, J.A., Wesensten, N. (Eds.), Biomonitoring for Physiological and Cognitive Performance during Military Operations. SPIE, Bellingham, Washington, USA, pp. 116–127.

Thomas, L.C., Gast, C., Grube, R., Craig, K., 2015. Fatigue detection in commercial flight operations: results using physiological measures. Procedia Manuf. 3, 2357–2364.

Tiffin, J., Bromer, J., 1943. Analysis of Eye Fixations and Patterns for Eye Movement in Landing a Piper Cub 5-3 Airplane (Report No. 14). CAA Division of Research, Washington, DC, US.

Tole, J.R., Stephens, A.T., Vivaudou, M., Harris, R.L., Ephrath, A.R., 1982, October. Entropy, instrument scan, and pilot workload. In: Proceedings of the International Conference on Systems, Man & Cybernetics. IEEE, New York, NY, USA, pp. 588–592.

Tracey, I., Flower, R., 2014. The warrior in the machine: neuroscience goes to war. Nat. Rev. Neurosci. 15, 825–834.

Tsang, P., Wilson, G., 1997. Mental workload. In: Salvendy, G. (Ed.), Handbook of Human Factors and Ergonomics, second ed. Wiley, New York.

van de Merwe, K., van Dijk, H., Zon, R., 2012. Eye movements as an indicator of situation awareness in a flight simulator experiment. Int. J. Aviat. Psychol. 22, 78–95.

van Dijk, H., van de Merwe, K., Zon, R., 2011. A coherent impression of the pilots' situation awareness: studying relevant human factors tools. Int. J. Aviat. Psychol. 21, 343–356.

WMA, 2008. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects. Retrieved August 21, 2012, from http://www.wma.net/en/30publications/10policies/b3/.

Yerkes, R.M., Dodson, J.D., 1908. The relation of strength of stimulus to rapidity of habit-formation. J. Comp. Neurol. Psychol. 18, 459–482.