

Neuroscientific tools in the cockpit: towards a meaningful decision support system for fatigue risk management

Maria C. Papanikou^{1,*}, *Christos A. Frantzidis*^{2,3}, *Anna Nikolaidou*^{2,4}, *Christina Plomariti*²,
*Maria Karagianni*², *Vasilis Nigdelis*², *Aliki Karkala*^{3,4}, *Christiane Nday*², *Giorgos Ntakakis*²,
Akis Krachtis^{2,4}, *Panagiotis D. Bamidis*^{2,3}, and *Chrysoula Kourtidou -Papadeli*^{2,3,4}

¹Amsterdam University of Applied Sciences, Faculty of Technology, Aviation Academy, The Netherlands

²Aristotle University of Thessaloniki, Medical School, Laboratory of Medical Physics, Greece

³Greek Aerospace Medical Association-Space Research, GASMA-SR, Greece

⁴AeroMedical Center of Northern Greece, Greece

Abstract. Fatigued pilots are prone to experience cognitive disorders that degrade their performance and adherence to high safety standards. In light of the current challenging context in aviation, we report the early phase of our ongoing project on the re-evaluation of human factors research for flight crew. Our motivation stems from the need for aviation organisations to develop decision support systems for operational aviation settings, able to feed-in in the organisations' fatigue risk management efforts. Key criteria to this end are the need for the least possible intrusiveness and the added information value for a safety system. Departing from the problems in compliance-focused fatigue risk management and the intrusive nature of clinical studies, we report a neuroscientific methodology able to yield markers that can be easily integrated in a decision support system at the operational level. Reporting the preliminary phase of our live project, we evaluate the tools suitable for the development of a system that tracks subtle pilot states, such as drowsiness and micro-sleep episodes.

1 Introduction

Fatigue, stress and other conditions (e.g. mild stroke) can lead to micro-sleep episodes and subtle pilot incapacitation, meaning the progressive deterioration of a pilot's state, which "escapes the normal pilot medical screening" [1,2]. The deterministic approach to cockpit technology, the innovation research exploring single-pilot operations, as well as the current crisis in the aviation industry, form a context in commotion. In the covid-19 crisis, there is an unravelling discussion on replacing passenger flights with cargo operations, implicating longer flights, and, combined with night work, a more demanding and disruptive work environment. In addition, many pilots have lost their jobs, inducing greater distress,

* Corresponding author: m.papanikou@hva.nl

subjective affective states, manifested also in symptoms in the central and peripheral nervous system of the “survivors” [3]. These conditions are likely to impact sleep quality in terms of inability to fall asleep (prolonged sleep onset), frequent awakenings and non-invigorating sleep associated with attenuation of sleep depth (shorter slow wave activity duration). It is therefore argued that, a lack of human performance data system, allows lurking fatigue risks such as drowsiness to go undetected, and alertness levels to decrease during the critical flight phases (e.g. approach and landing). In addition, latest research highlights the need for airline level interventions [4] addressing pilot fatigue detection and including mental health. Previous studies, however, mainly propose models unable to capture complex human factors, due to isolated data collection, intrusiveness of detection and cost limitations. Aiming to enhance the arsenal of the neuroscientific tools, we propose a neuroergonomics framework, which consists of a detailed polysomnographic (PSG) sleep study accompanied with an elaborate neurocognitive and neurophysiological assessment. The proposed framework aims to elicit risk factors that are associated with fatigue patterns and to provide the appropriate mitigation approaches through non-pharmacological interventions. Developments regarding machine learning recommendation systems and potential intervention approaches are also described.

2 Fatigue risk management

At the operational level, in an effort to proactively manage aviation safety, safety management systems (SMS) have been set up to drive decisions based on data. Fatigue (risk) management involves the development and execution of a data-driven Fatigue Risk Management System (FRMS). Due to concrete limitations towards a truly data driven support system, fatigue is addressed as an isolated rest problem, and crew scheduling follows aeromedical requirements, the Flight Time Limitations (FTLs) to proactively manage fatigue. Aeromedical regulations and the selection and assessment of pilots do not predict future performance and health conditions but rather evaluate the condition of pilots at the time of assessment. Fatigue, however, may also induce boredom or even sleepiness which further deteriorates the pilots’ psychomotor performance. Specifically, cockpit design and automation are linked to pilot boredom, and concerns of complacency and distraction are raised in the aviation community [5]. In addition, recent studies [4] note that the pilots are concerned about losing their licence, with 70% of the pilots being reluctant to report fatigue and wellbeing problems. Fatigue risk management hence requires the appropriate decision support systems (DSS) incorporating pilot characteristics, yet with operationalisable solutions.

In light of the latter gaps and the aforementioned focus on compliance, detection and pilot referral is lacking development at an operational level and fatigue management is based on regulatory guidance (i.e. FTLs). As a result, pilot fatigue interventions traditionally focus on self-management techniques. For example, studies [6] report that the top three self-management mitigation measures for fatigue include: a) more sleep, b) reducing workload, and c) being self-aware and making healthy lifestyle choices. Accordingly, fatigue studies are based on FTLs and fatigue risk prediction for better crew planning using mathematical models [7], aiding compliance in operational planning. From these and similar approaches tools were developed, such as the Boeing Alertness Model (BAM) that focuses on capturing sleep inputs in order to aid planning and scheduling. The limitations of BAM are highlighted by the Flight Safety Foundation, where they note that the model lacks validation within an airline environment. Other efforts to monitor fatigue include applications such as the Jeppesen Crew Alert, aiming to complement FRMS. The use of such tools is intended for pilots, who can insert personalised settings in order to record sleep patterns, and to generate and share fatigue reports. An advantage of this approach is that it helps pilots to be more

aware of their alertness levels. Such tools extend to assessing controlled rest. For example, an actigraphy watch is used for the detection of sleep disorders. Latest applications in research show the positive effects of the latter’s use for controlled rest [8], yet, despite individual pilots reporting the benefits of its use, applications at an airline level are not reported.

Neuroscience offers more insight into fatigue management by deploying sleep studies, which, however, are not utilised for operational improvements but mainly highlight the importance of fatigue and criticise the FTLs, engaging in policy making discussions. Eye-tracking studies are also popular in fatigue detection [9], as well as the use of additional tools, such as the Psychomotor Vigilance Task (PVT), to assess reaction times and attention. Latest applications of the PVT include short test flights (approx. 30 minutes each) in simulated environments [10]. Importantly, however, a recent study of pilot cognitive workload using headbands in real – time operations showed differences from a simulated environment [11]. Significant differences between simulated and real-time environments are reported, where pilots commit errors they did not in the simulator [11]. Specifically, the study reports that in high workload phases of flight, errors were committed in the simulated environment were less than in the real-time environment. Hence, although, the majority of studies report lurking risks and the benefits of risk managing fatigue beyond the FTLs, there is no operational-friendly system that is able to create, simple yet informative, data banks the FRMS requires. The papers on the value of biomedical engineering and data mining far exceed the acceptance of their proposed solutions for flight crew performance.

In addition, sampling limitations are noted as previous neuroscience-focused studies included student [12] or inactive pilots [13]. Similarly, pilot characteristics and performance, and their challenges in complex environments, do not go beyond psychological profiling for selection purposes. In light, however, of studies documenting the pilot’s professional changes [14, 15], we propose that the inclusion of professional characteristics when studying subjects can help develop a better decision system. Decrease of training, safety practices led by compliance and cost-cutting needs, dependency on employers, the imbalance of investments between the human and the machine (human out of the loop), lack of reporting, low status and discretion were found as the long-term changes induced to flight crew as the industry grew and changed [14]. Airmanship 2.0 [15] deals with the skills required to operate an aircraft, extending from manual skills and the level of automation (LoA) to macro-cognitive factors and key areas of concern. In particular, pilots are found re-skilled in terms of gaining new flying skills due to automation and de-skilled in terms of manual flying. Key areas of concern include fatigue management, failure diagnostics, emotional self-control, complexity oversight and knowledge of manuals. For the purpose of our project, we distinguish between *Automation* and *Employment*. The latter encompasses elements that can amplify the areas of concern under Airmanship 2.0. Based on the latter studies, Table 1 summarises the framework that underpins our live study. We distinguish between the Old and New generation of pilots and, automation and employment characteristics. Age is included due to its significant relation with cognitive abilities and with knowledge, meaning the expertise, acquired skills and procedural knowledge [16].

Table 1. Airmanship 2.0 adapted from [14,15]

	Airmanship 2.0	
	Automation	Employment
OldGen 46-65	De-skilled (Aviate-navigate- communicate-manage)	Reduced training-decay of knowledge- less discretion

	Airmanship 2.0	
	Automation	Employment
NewGen 20-45	Re-skilled (Manage-communicate- navigate-aviate)	Debt of training-low income- employment dependency

We return to our underpinning framework in light of the preliminary findings reported in this paper. Before engaging with the data, we describe the deployed methodology for this early phase of our project using three cases, one from the OldGen typology and two from the NewGen typology.

3 Methodology

In this section we describe the underlying data acquisition as well as the analysis employed for identifying potential, fatigue risk factors. The methodological objectives support our key aim, namely developing a DSS that is enabled by contextual characteristics. We deploy three cases in order to explain our methodology. Each case underwent a different set of experiments and assessment in order to explore intrusiveness levels and operational value. The latter aimed in a) determining the value of tools for an operational setting, and b) exploring the value of pilot profiling in the interface with the tools. Below we explain the neuroscientific tools for the experiments, and we describe the cases that represent the typologies of OldGen and NewGen pilots.

3.1. Polysomnographic data acquisition

A Nihon Kohden electroencephalographic (EEG) device equipped with 32 Ag/AgCl electrodes was used. The EEG data were collected from 19 electrodes located according to the 10-20 International System. A ground electrode was placed on the prefrontal midline (Fpz) positions and two reference electrodes were placed on the left and right mastoids. The electrooculogrammic (EOG) activity gathered both vertical (blink) and horizontal movements by placing bipolar electrodes above and below the right eye and on the outer canthi of both eyes. Electromyographic (EMG) activity was recorded by a bipolar electrode placed on the left and right chin, whereas electrocardiographic (ECG) timeseries was recorded by a bipolar electrode placed above the chest.

Pre-processing pipeline

The entire night PSG data were initially divided into 10 segments of equal size in order to facilitate more efficient computations. From each segment a baseline removal took place initially by subtracting from each signal its mean value. This resulted in eliminating amplitude linear shifts (DC bias). Regarding the EEG data a common average re-reference procedure was performed. Then, digital filtering was performed through Butterworth filters of 2nd order as follows:

1. High-pass filter (cut-off frequency 0.5 Hz) for further removing linear trends
2. Low-pass filter (cut-off frequency 50 Hz) for rejecting unnecessary high frequency content
3. Band stop (notch) filter among 47-53 Hz for removing industrial noise centered at 50 Hz
4. Band stop (notch) filter among 97-103 Hz for removing industrial noise harmonics centered at 100 Hz

5. Band stop (notch) filter among 147-153 Hz for removing industrial noise harmonics centered at 150 Hz

In the case of the ECG signal the cut-off frequency of the low-pass filter was set at 20 Hz and at 15 Hz for the EOG. In the case of the EMG signal the cut-off frequency of the high-pass filter was set at 10 Hz and for the low-pass at 70 Hz. Then, the algorithm of the Independent Component Analysis (ICA) was ran for each segment through the EEGLAB user interface under Matlab environment. Visual inspection of the ICA components by two experienced neuroscientists resulted in the rejection of artefactual sources due to blinks, muscle movements, bad electrode placement, ECG modulation. Finally, the data were epoched into continuous, non-overlapped epochs of 30 second duration and two experienced sleep experts were manually scored according to the guidelines of the American Association of Sleep Medicine (AASM) [17].

3.2. Cognitive screening

Neurocognitive assessment was performed through a single session of the CogScreen aeromedical edition (Cogscreen AM). The examination was performed in the premises of the AeroMedical Center and was performed by two experienced medical doctors. It is a computerized cognitive battery able to detect even subtle alterations in cognitive functioning. Although CogScreen does not focus on aviation knowledge or flying skills, it could provide quantified estimations of the underlying perceptual cognitive, and information processing abilities associated with flying. An initial rough estimate is provided by the Logistic Regression Probability Variation (LRPV) score of brain dysfunction. This ranges from 0 to 1 and the higher the score the more likely a cognitive degradation is. However, it is highly correlated with physiological aging and its value represents only the probability of impairment existence and not its impact. Further information is provided by a base rate analysis, which describes the number of tasks that the participant scored lower than the 5th percentile and 15th percentile when compared to a selected normative comparison group. This group mainly consists of astronauts, US military and civil aviation pilots. However, it is possible for the examiner to establish her/his own database of pilots [18]. There are also tasks associated with aviation factors such as deductive reasoning, motor coordination under speeded conditions, visual learning and recall, visual scanning, perceptual speed, and working memory and visual/psychomotor tracking accuracy.

3.3. Cases

In this report, we include three cases of male Caucasian right-handed active pilots. The pilots were assessed between flights following their roster. The sleep studies took place in the night between their flights and the other tools were utilised in their day off. The participants did not report any medical or psychiatric history until now. The participants did not suffer from any neurologic or psychiatric disorder or human immunodeficiency virus (HIV). More specifically, the participants' medical history did not contain any instance of hypertension, kidney disease, diabetes, liver, heart, respiratory disease or cancer diagnosis. Neurological examinations did not reveal any head injury, stroke, brain tumor or seizure. Psychiatric examinations did not reveal any anxiety, depression, learning disabilities, mood or attention deficit disorder. Furthermore, the participants were systemically sustained from alcohol or any drug substance consumption. Different characteristics were taken into account in their selection, such as aircraft type, rank, airline business model/operations, and career progression.

OldGen

The first profile was of an airline captain (Case 1). Case 1 was aged 53 years old, with 18 years of education (basic education plus six academic years). He was an active captain for a large full-cost carrier with an ATP license type and 18000 total flight hours. Following the sleep study and the Cogscreen AM session, the participant was interviewed, in order to assess his perception of the current challenges, and to assess acceptability, intrusiveness feelings, and to compare feelings and perceptions with the experiments' results. The interview lasted 1 hour and 28 minutes. The semi-structured interview protocol followed previously validated factors in order to develop the coding framework for analysis based on a) mental and temporal demands, b) perceived performance, and c) frustration fatigue. Physical workload was excluded, whereas temporal workload was analysed based on time pressure and pace of tasks, due to the different phases of flight.

NewGen

The second profile was of a first officer (Case 2). Case 2 was 35 years old, with 16 years of education (basic education plus four academic years). He was an active ATP licensed first officer for a small full-cost airline operating a Bae 146 with 4000 total flight hours. Case 2's state was assessed only through the sleep diagnostic examination. Finally, the third profile was of a male first officer (Case 3). Case 3, aged 32 years old, with the same level of education as Case 2. He was an active ATP licensed first officer for a small low-cost carrier operating an A320 with a total of 2000 flight hours and was at the time preparing to transition to the captain's role. Case 3 was interviewed following the same protocol as with Case 1. The interview lasted 2 hours and 15 minutes. Case 3 was excluded from the sleep study but was included in the Cogscreen AM session.

3.4. Neurophysiological feature extraction for drowsiness estimation

The estimation employs features derived from the ECG, EMG and EOG in order to discriminate active awake from drowsiness. The ECG analysis involves time-frequency analysis of the heart rate variability (HRV) through the Lomb-Scargle periodogram (LSP) [13]. Firstly, the ECG signal was normalized and then R peaks were detected, forming thus the R-R timeseries. Since, the R-R timeseries does not follow a uniform sampling like the other biosignals we extracted its spectral component employing more advanced mathematical tools than the Fourier transform. More specifically, the LSP method resulted in the calculation of the following spectral features: The frequency with the maximum HRV power and the relative energy ratios of the ultra-low Frequency (≤ 0.0033 Hz), very low frequency (0.0033 – 0.04 Hz), low frequency (0.04 – 0.15 Hz), high frequency (0.15-0.4 Hz), very high frequency (≥ 0.4 Hz) and the range in which the frequency with the maximum power was detected. The EOG activity was estimated as 1) the number of blinks per minute (blink rate), 2) mean blink amplitude, 3) the mean rise time and 4) blink duration. The EMG spectral features were estimated as the relative energy contribution of very low (≤ 4 Hz), low (4-8 Hz), medium (8-12 Hz), high (12-24 Hz), very high (24-48 Hz) and ultra-high (48-96 Hz) of the chin EMG activity. The Orthogonal Discrete Wavelet Transform (ODWT) was employed for estimating the energy contributions with optimal time-frequency resolution [14]. The analysis involved 18 instances of active wake and 23 instances of drowsiness. The identification of each instance was assessed by two sleep experts. Each instance duration was 1 minute (60 seconds). In summary, the analysis involved 41 minutes of PSG recordings.

4 Findings

4.1. Case 1

The coding framework (cf. 3.3) was applied on the interview data. As shown in Table 1, the participant shared concerns about airmanship, characterising the cockpit “so virtual [that] it is like a video game”, without having the knowledge of what lays “behind the system”. Beyond flying skills, the decision making, and other macro-cognitive and resource management factors are affected. The participant’s perceived mental workload was high in terms of demands in effort and tasks that do not only include the cockpit ones but 24/7 engagement with the company, online trainings, and dual roles. In detail, the participant reported an “extremely high” workload, and working in an environment where “everybody is tired” and employees are “24/7 [...] company-minded”. Temporal demands were discussed in different flight phases with an emphasis on alertness in critical phases. Furthermore, frustration fatigue was discussed through the profession’s changes affecting their knowledge, skills, and abilities (KSA). Employment concerns amplified frustration fatigue, involving emotions of stress, anger, and cynicism. In this context, the participant stressed the importance of experience to mediate lurking risks. The participant also noted the increase in his flight hours and his reluctance to report in the beginning of his career, whilst new pilots were characterised as ones lacking stimuli. The participant reported the intention to accept a wearable device in the cockpit.

Table 2. Perceptions Case 1.

Mental demands	Temporal demands	Perceived performance	Frustration
Effort: high	Time pressure: extremely high	Accomplished tasks: Manual	Profession: deterioration of KSA
Tasks: continue outside the cockpit.	Pace of tasks: irregular, lack of stimuli at cruising altitude.	Least satisfaction: Decision making Information processing Reaction time Team leadership	Employment: security, aggressive environment.

Sleep macro-architecture

The participant slept for 244.5 minutes (Figure 1). The sleep latency recorded as the duration from the first epoch with eyes closed and drowsiness until the first sleep (NREM1) epoch and was at 23.5 minutes. The duration of each sleep stage was 58.5 minutes for N1, 148.5 minutes for N2, 18 minutes for N3, 19.5 minutes for REM. Latency periods were 37 minutes for N2, 138 for N3 and 252 for REM. The slow wave activity (SWA) covered the 7.36% of the total sleep time, while the REM period covered the 7.98% of the total sleep time. Moreover, the participant woke up 11 times and the total awakening period was 77.5 minutes. So, its sleep efficiency measured as the total sleep time divided by the entire PSG duration was estimated at the 65.2%. Finally, the participant experienced 44 arousal episodes. There were also 55 spindles and 105 K-complexes.

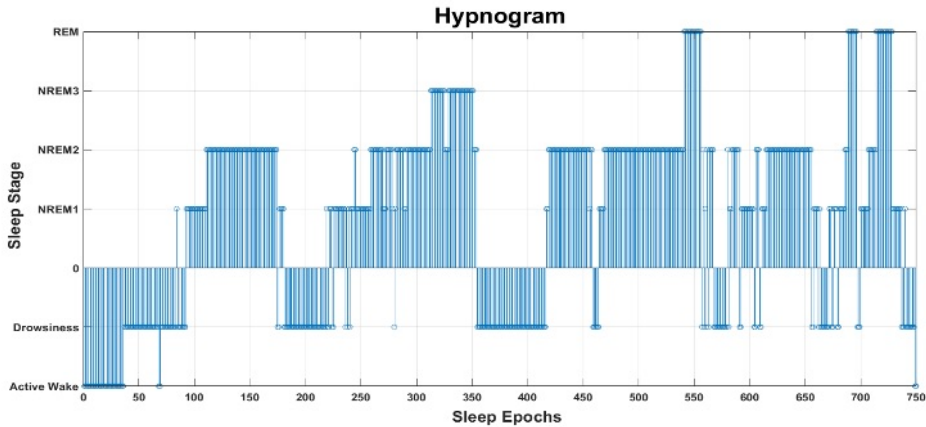


Figure 1. Hypnogram of OldGenCap representing the sequence of the sleep stages across the entire sleep recording through polysomnography

Neurocognitive Assessment

The LRPV score of Case 1 was 1. The aviation factor scores are displayed in Table 3.

Table 3. Case1 performance on cognitive domains associated with aviation characteristics

Factors	Description	z-score	T-score
Attribute Identification	Deductive reasoning	-4.14	8.56
Motor Coordination	Motor coordination under speeded conditions	-0.57	55.70
Visual Association Memory	Visual learning & recall	-2.08	29.19
Speed/Working Memory	Visual scanning, perceptual speed & working memory	-0.32	46.84
Tracking	Visual/psychomotor tracking accuracy	0.10	48.97

4.2. Case 2

Sleep macro-architecture

Case 2 slept for 471.5 minutes (Figure 2). The sleep latency was at 24 minutes. The duration of each sleep stage was 172 minutes for N1, 212.5 minutes for N2, 77 minutes for N3, 10 minutes for REM. Latency periods were 30 minutes for N2, 44 for N3 and 88.5 for REM. The slow wave activity (SWA) covered the 16.33% of the total sleep time, while the REM period covered the 2.12% of the total sleep time. Moreover, the participant woke up 18 times and the total awakening period was 53 minutes. So, its sleep was estimated at the 82 %. Finally, the participant experienced 18 arousal episodes. There were also 27 spindles and 111 K-complexes.

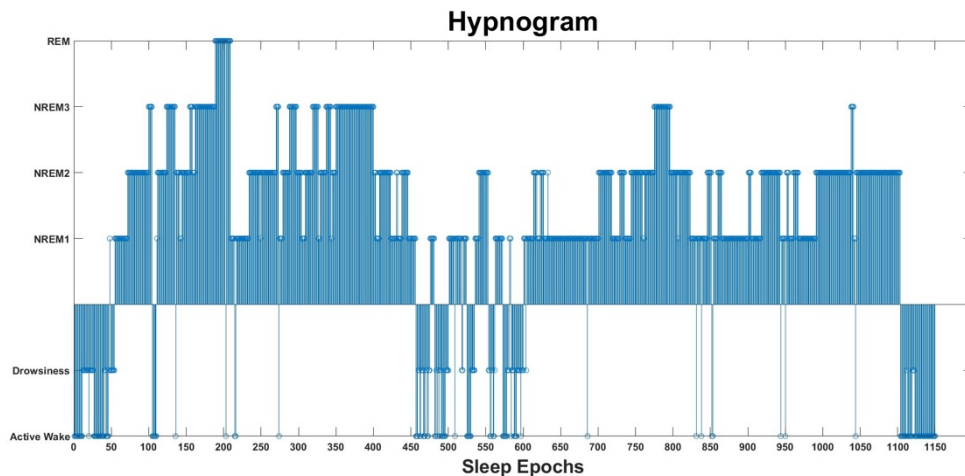


Figure. 2. Hypnogram of Case 2 representing the sequence of the sleep stages across the entire sleep recording through polysomnography

4.3. Case 3

Neurocognitive Assessment

The LRPV score of Case 3 was 1. The aviation factor scores are displayed in Table 4 below.

Table 4. Case 3 performance on cognitive domains associated with aviation characteristics.

Factors	Description	z-score	T-score
Attribute Identification	Deductive reasoning	-2.33	26.70
Motor Coordination	Motor coordination under speeded conditions	-2.03	70.28
Visual Association Memory	Visual learning & recall	-0.22	47.85
Speed/Working Memory	Visual scanning, perceptual speed & working memory	-1.57	34.30
Tracking	Visual/psychomotor tracking accuracy	-1.16	61.58

Case 3 interview data are presented in Table 5. Case 3 was frustrated and discussed in depth his concerns about the use of FTLs in small airlines, where rest according to regulations is followed to risk manage fatigue. Case 3 focused on the operational needs of airlines indicating that “what can be done in the minimum will be done in the minimum” and showing the interviewer a heavy night shift roster. Case 3 reports signs of accumulative fatigue, stating that he is flying the maximum hours possible and then returning to work tired “even after days off”. “Bad management of flight crew” and that FRMS “is not in place”, centered the discussion showing lack of peer or company support. Daily incidents in the airline’s operations were characterised as “safe failures”, but several AOGs (aircraft on ground) put more pressure on the operations. Regarding the temporal demands, Case 3 noted that there is a decrease in workload but that it is critically high in take-off and approach. Job insecurity (frustration fatigue) and high mental demands were noted as by Case 1. However, tasks are perceived as routine, and performance is perceived as appropriate for operating the aircraft. As in Case 1, Case 3 reported the intention to accept a wearable device in the cockpit.

Table 5. Perceptions Case 3.

Mental demands	Temporal demands	Perceived performance	Frustration
Effort: high	Time pressure: high	Accomplished tasks: Operating the aircraft	Profession: skills vs. knowledge of systems
Tasks: routine	Pace of tasks: high demand in critical phases of flight.	Least satisfaction: Developmental training	Employment: insecurity

4.4. Neurophysiological drowsiness estimation

The Matlab software was used to develop a drowsiness classifier based on the aforementioned features. We explored decision trees (fitted binary classification decision trees) as potential classifiers in order to result in a rule-based algorithm that would be easily interpreted in aviation and industrial settings. More specifically, we used active wake and drowsiness data from the two participants to train and test the classifier. There were 32 active wake and 33 drowsiness data epochs. The train set consists of 44 epochs and the test set of 21 epochs. The decision tree in Figure 3 was developed based on the train data and then its performance was validated in the test data, which were not previously used in the training phase. The root (most important feature) of the decision tree was the low frequency of the HRV timeseries. Then EMG low frequency and blink rate (number of blinks per minute) were demonstrated to further discriminate among the two classes. Finally, the blink amplitude was also an important factor. The resulted tree correctly classified 17/21 of the test instances. So, its classification performance was at 80.95%. A more detailed description of the method's classification accuracy is the following: a) True Positive 9/10 (90%), b) True Negative: 8/11 (72.73%), c) False Positive: 3/11 (27.27%), d) False Negative 1/10 (10%).

Neurophysiological Drowsiness Detector

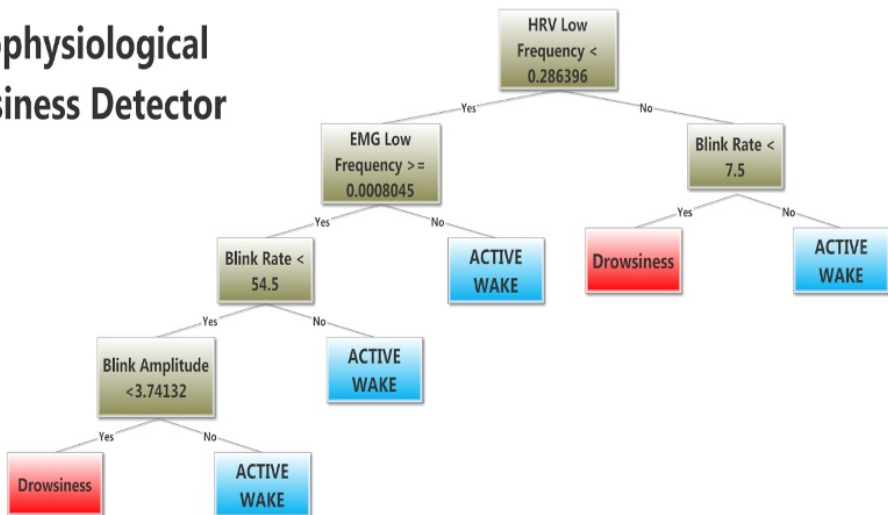


Figure 3. Visualization of the rule-based algorithm (binary decision tree) for drowsiness detection based on the participants' heart-rate variability, chin electromyographic and eye blink activity.

5 Discussion & recommendations

The Airmanship 2.0. based framework is supported by all cases in the preliminary findings, and the typologies aided analysis and interpretation. In accordance with the literature on airline fatigue risk management, we explored the options that best fit the purpose of our methodology and the objectives of our project. These options satisfy the criteria of minimal intrusiveness and of operationalisability of tools for DSS development. Based on these criteria, Cogscreen AM was included in our deployed tools because of its low cost, ease of deployment and validity. In the Cogscreen sessions with Cases 1 and 3 inadequate scores indicate a degradation of cognitive abilities. Case 1 had over 30 years of professional experience in military and civil aviation, including the employment in different airlines, and the operation of various aircraft types with different approaches to the role of technology in the cockpit. Case 3 had 9 years of experience and mainly in the flight training industry before transiting to airlines performing better in the session but still producing low scores. Although Cogscreen AM was initially designed by the FAA based on military pilots, applications took place in samples of commercial pilots reporting a clear prediction of cockpit performance [18]. In cases where Cogscreen AM was applied in older groups of aviation employees the scores were lower as in Case 1. Our preliminary results, however, do not support the mediating effect of training and experience reported in other aviation professional groups such as air traffic controllers [19], showing a need for greater analysis on LoA. A re-evaluation of the assessment tool that takes into consideration the local population characteristics, including the attitudes of pilots and their airmanship, could be further explored to address such questions. In addition, as explained earlier (cf. Section 3.3), aeromedical examiners could create their own data banks. Similar concerns and propositions are reported where a multi-modal deep learning network can aid the classification of mental modes of pilots [20]. Nevertheless, in combination with the interview data, the Cogscreen results gain importance because these are in contrast with the perceived performance of Case 1 and Case 3. Both cases show confidence, yet in different systems; Case 1 in experience and Case 3 in automation. The latter supports the transition between generations from *aviate-navigate-communicate-manage* to *manage-communicate-navigate-aviate*.

Moreover, the sleep analysis of Cases 1 and 2 showed that there is a possibility of a sleep disorder. In Case 1, drowsiness lasted much longer than normal (almost up to 100 epochs), including 30 seconds outliers going from drowsiness to active wake. The lack of rest is evident (cf. Figures 1, 2), as N1 should have been within a range of 6-10 minutes and for Case 1 it was of 58.5 minutes, with lapses between N1 and drowsiness. Similarly, the participant went into N2 for less than the expected normal range of 1 to 2 hours. Importantly, the short periods of stages N3 and REM (i.e. N3 18 minutes and REM 19 minutes) indicate no rest and the fall into REM due to sleep deprivation. Similarly, Case 2 also shows signs of a sleep disorder. Although Case 2 slept longer than Case 1, short periods of REM are also observed, showing that the pilot did not rest. Because, however, we did not perform pathological examinations the possibility of this being an effect of apnea could not be ruled out. It is recommended that other studies examine apnea due to its effects. It is noteworthy to remind readers the Go! 2009 incident, when both pilots fell asleep during cruising altitude and the NTSB reported obstructive sleep apnea as the cause of the incident. Our preliminary report hence finds support in cases and concerns of pilot drowsiness [20]. In order, however, to be able to deploy the value of polysomnography from a research to an operational setting in a meaningful manner for fatigue risk management, a less time-consuming instrumentation is required. Decision trees are commonly used in aviation operations and have many similarities with other failure analysis trees applied in the industry, able to possibly gain greater acceptance by safety professionals. Furthermore, such data can aid peer pilot support

programmes, extending the DSS for the enabling of awareness and self-declaration, and, if need be, referral.

However, although the classification of the proposed decision tree is scientifically acceptable (above 80%), its employment in aviation condition needs its further improvement. The latter could be achieved by including data from more participants (≥ 25), which will allow the algorithm's generalization. Then, more sophisticated machine learning algorithms (Support Vector Machines, neural networks) could be also employed to further improve the classification accuracy [21, 22]. In case of also employing brain activity data (e.g. wearable EEG devices), the number of estimated features would result in a big data perspective which may facilitate the use of deep learning algorithms [23]. The above steps may serve for a future roadmap towards the reliable integration of deep learning applications as onboard infrastructure for drowsiness estimation. The features used in the decision tree are in accordance with previous neuroscientific evidence. More specifically, the low frequency of the HRV timeseries is associated with the sympathetic branch of the autonomic nervous system. Drowsiness and sleep onset are closely related with the inhibition of sympathetic activity [24]. According to the AASM sleep scoring rules, drowsiness is mainly characterized by the onset of the alpha rhythm [25], which is associated by attenuated chin EMG and blink activity. Since, the features involved in the classification was only selected by the tree algorithm and not the developers, this is an evidence that the proposed methodology is of scientific and clinical validity. At operational levels, it would be easily implemented through microphones or cameras, whereas wearable devices with a bipolar electrode could provide more reliable, spectral EMG features.

6 Conclusion

Our paper marks a call for the selection of meaningful data and their sustainable use in air transport organisations. In the dynamic aviation environment, changes affect the pilots' willingness to report fatigue, their nature of work and their lifestyle. In this context, a pilot's capacity for information processing is radically decreased and their attention level diminishes, which may result in erroneous responses to crucial flight tasks. Following the Airmanship 2.0. dimensions, we reported the preliminary findings of a live project. The proposed systematic data collection method aims to promote evidence-based decision making and to address complex and least quantifiable pilot performance parameters, such as fatigue and sleep disorders. Our methodology employed both bio-signal processing and machine learning techniques to provide an objective estimator of drowsiness level. However, a greater amount of data is required in order to provide definitive results that could provide evidence-based feedback for policy making activities in the field of pilot fatigue. More advanced recording modalities such as EEG activity could be also employed to quantify the fatigue level prior entering a drowsiness state. However, we should bear in mind that a compromise among classification accuracy and unobtrusiveness should be made in order to result in applicable scientific solutions easily accepted by the aviation community. Further research should explore the relationship between boredom and drowsiness in the cruising altitude of flight with high LoA, the characteristics of pilots today and possible applications of such tools in peer pilot support programmes.

References

1. Besco, R.O. (1990). Subtle incapacitation of pilots: How to tell if your captain has died. Accident Prevention.

2. ECA (2012). Pilot fatigue. Available online at https://www.eurocockpit.be/sites/default/files/eca_barometer_on_pilot_fatigue_12_11_07_f.pdf
3. Cropanzano, R. S., Ambrose, M. L., & Van Wagoner, P. (2019). 11 Organizational Justice and Workplace Emotion. *Social Psychology and Justice*.
4. Cahill, J., Cullen, P. and Gaynor, K. (2019). Pilot Wellbeing & Work-Related Stress (Wrs). 27th International Symposium on Aviation Psychology.
5. Gawron, V. (2019). Automation in aviation – Guidelines. Mitre Technical Report MTR190018. McLean, VA: Mitre Corporation.
6. Levin, E., MR, M., COIMBRA, F., Keller, J. and Teo, A. (2019). Fatigue in Collegiate Aviation." *International Journal of Aviation, Aeronautics, and Aerospace* 6.4.
7. HELLERSTRÖM, D., Eriksson, E., ROMIG, E. and KLEMETS, T. (2010). Flight time limitations and fatigue risk management: comparison of three regulatory approaches", Boeing Flight Safety Foundation.
8. Hilditch, C.J., and Flynn-Evans, E.E. (2019). Controlled Rest: Profile of Use, Challenges, and Best Practices.
9. Rostamina, S., Mayberry, A., Ganesan, D., Marlin, B. and Gummeson, J. (2017). iLid: Low-power Sensing of Fatigue and Drowsiness Measures on a Computational Eyeglass. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1(2). pp. 1-26.
10. O'Hagan, A.D., Issartel, J., Wall, A., Dunne, F., Boylan, P., Groeneweg, J., Herring, M., Campbell, M. and Warrington, G. (2019). Flying on empty: effects of sleep deprivation on pilot performance. *Biological Rhythm Research*, pp. 1-22.
11. Gateau, T., Ayaz, H. and Dehais, F. (2018). In silico vs. over the clouds: on-the-fly mental state estimation of aircraft pilots, using a functional near infrared spectroscopy-based passive-BCI. *Frontiers in human neuroscience* 12.
12. Cordeiro, R. and Carvalho J. (2019) Positive mental health in the workplace: a new global and preventive approach for flight staff. *Авіаційна та екстремальна психологія у контексті технологічних*. 3.
13. CAA, NLR, & NIN. (2018). Independent research on pilot fatigue measurement by the Netherlands Aerospace Center. Available at http://publicapps.caa.co.uk/docs/33/Pilot%20Fatigue%20Measurement%20Research_Final%20Report_v0.9.1clean.pdf.
14. Papanikou, M. (2017). The philosophy of safety: change and the re-organisation of the aviation industry. Ph.D. Thesis, University of Greenwich, UK.
15. Mohrmann, F., & Stoop, J. (2019). Airmanship 2.0: Innovating aviation human factors forensics to necessarily proactive role. *International Society of Aviation Safety Investigators (ISASI). Annual Seminar*.
16. Warr, P. (2001). Age and work behaviour: Physical attributes, cognitive abilities, knowledge, personality traits and motives. *International review of industrial and organizational psychology*, 16, 1-36.
17. Moser, D., Anderer, P., Gruber, G., Parapatics, S., Loretz, E., Boeck, M., ... & Saletu, B. (2009). Sleep classification according to AASM and Rechtschaffen & Kales: effects on sleep scoring parameters. *Sleep*, 32(2), 139-149.
18. Taylor, J., O'Hara R., Mumenthaler, M. and Yesavage, J. (2000). Relationship of CogScreen-AE to flight simulator performance and pilot age. *Aviat Space Environ Med* 7. pp. 373-80.
19. Taylor, J., O'Hara R., Mumenthaler, M., Rosen, A. & Yesavage, J. (2005). Cognitive Ability, Expertise, and Age Differences in Following Air-Traffic Control Instructions. *Psychology and Aging*. 20(1). pp. 117-133.

20. Corbett, M.A. (2009). Science & Technology Watch: A Drowsiness Detection System for Pilots: Optalert®. *Aviation, space, and environmental medicine*, 80(2). pp.149-149.
21. Frantzidis, C. A., Bratsas, C., Papadelis, C. L., Konstantinidis, E., Pappas, C., & Bamidis, P. D. (2010). Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli. *IEEE Transactions on Information Technology in Biomedicine*, 14(3), 589-597.
22. Spyrou, I. M., Frantzidis, C., Bratsas, C., Antoniou, I., & Bamidis, P. D. (2016). Geriatric depression symptoms coexisting with cognitive decline: a comparison of classification methodologies. *Biomedical Signal Processing and Control*, 25, 118-129.
23. Chriskos, P., Frantzidis, C. A., Gkivogkli, P. T., Bamidis, P. D., & Kourtidou-Papadeli, C. (2019). Automatic sleep staging employing convolutional neural networks and cortical connectivity images. *IEEE transactions on neural networks and learning systems*.
24. Reyes del Paso, G. A., Langewitz, W., Mulder, L. J., Van Roon, A., & Duschek, S. (2013). The utility of low frequency heart rate variability as an index of sympathetic cardiac tone: a review with emphasis on a reanalysis of previous studies. *Psychophysiology*, 50(5), 477-487.
25. Moser, D., Anderer, P., Gruber, G., Parapatics, S., Loretz, E., Boeck, M., ... & Saletu, B. (2009). Sleep classification according to AASM and Rechtschaffen & Kales: effects on sleep scoring parameters. *Sleep*, 32(2), 139-149.