A Multimodal Sensor Suite for Real-Time Prediction of Pilot Workload in Operational Settings

Kaunil Dhruv
Institute of Cognitive Science
University of Colorado, Boulder
kaunil.dhruv@colorado.edu

Trevor Grant
Institute of Cognitive Science
University of Colorado, Boulder
trevor.grant@colorado.edu

Kevin Durkee
Aptima
Dayton, Ohio
kdurkee@aptima.com

Leanne Hirshfield
Institute of Cognitive Science
University of Colorado, Boulder
leanne.hirshfield@colorado.edu

ABSTRACT

In this study we explore neuro-physiological correlates of mental workload and task difficulty during a simulated aircraft piloting task by simultaneously collecting Electrocardiogram (ECG), galvanic skin response (GSR), 2-channel functional near-infrared specroscope (fNIRS), respiration and eye pupil size data of 10 different participants. Using this data we investigate the significance of each of these modalities in predicting workload and finally we propose a hybrid workload prediction system by combining these modalities using an SVM classifier.

KEYWORDS

BCI, Mental Workload, fNIRS

ACM Reference Format:
Kaunil Dhruv, Trevor Grant, Kevin Durkee, and Leanne Hirshfield. 2018. A Multimodal Sensor Suite for Real-Time Prediction of Pilot Workload in Operational Settings . In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Over recent years, research involving the use of neuro physiological sensor streams to quantitatively measure and predict workload experienced by an individual has gained momentum with the complexity of its applications ranging from driving cars [1, 3] to playing music [30] and web surfing [18]. Such systems often pair a neuro-imaging modality viz. fNIRS and electroencephalogram (EEG) with physiological sensors such as ECG, electroculogram (EOG), respiration and GSR. For example, *Sangte Ahn et al.* and *Gianluca Borgini et al.* explored the combination of ECG, EOG, EEG and/or fNIRS modalities to investigate the effect of sleep deprivation on a subjects' performance in a simulated car driving study.
Data collected from these modalities are then fused to build a classifier trained to discretely predict workload using machine learning techniques [17].
While these studies effectively correlate performance in simulated tasks with workload, their extension to a practical setting is limited by the footprint and portability of these sensors. With this challenge in mind, our research makes the following contributions: We demonstrate the utility of merging multi-modal sensor data for the measurement of workload, using specifically selected wearable sensors that are comfortable, wireless, and practical for use in operational settings. These sensors include fNIRS, ECG, EDA, respiration, and eye-tracking sensors. We do this by evaluating the sensitivity of each data stream to our workload manipulations, and showing model performance achieved by merging data streams from multiple sensors.

Background and Literature Survey

Despite the fact that aircraft pilots face less traffic compared to their on-ground counterparts, piloting is a complex activity that requires both skill and technical expertise [21]1. Every decision that an aircraft pilot makes requires the consideration of a multitude of variables and uncertainties all while maintaining clear communication with ground control and their crew members.
From an HCI perspective, piloting is a cognitively resource intensive task exercising working memory to satisfy task

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Woodstock '18, June 03–05, 2018, Woodstock, NY
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

demands[29]. Delineating the cognitive activities and their effects on the parasympathetic and physiological systems[13] can help us shed some light on the mental workload that a pilot experiences during challenging situations. Potentially allowing us opportunities to enhance their skills set and training routines[16].
Numerous studies have used neuro-imaging modalities such as EEG [7, 8, 33] and fNIRS[9, 31] paired with other physiological sensors to evaluate neuro-physiological correlates of pilots with their task performance[14] in a simulated environment. These physiological sensors typically include ECG, EDA, GSR and EOG[20, 32].

Subjective Assessment of Mental workload

The construct of mental workload in human factors research albeit a vague one is centered around the definition proposed by *Kantowitz et al.* who define mental workload as a ratio of task demands to the inherent cognitive capacity of a subject. As task demands increase the performance of a subject during this task decreases due to their limited cognitive capacity [10, 22, 23]. Therefore the cognitive effort invested by a subject while executing the task provides us a direct insight into the cognitive demands of a task.
Techniques for measuring cognitive effort and thus, the workload can be divided into subjective ratings, performance based and physiological measures [27].
Subjective ratings collected using questionares such as NASA-TLX (post task completion), SPAM (recording while a task is in progress) or SAGAT (pausing the task) are strong indicators of the workload that a subject experiences. However, in many cases subjects have to be interrupted while working on the primary task [26] and their own assessments suffer from an inherent expectation bias about the task [28].
Due to the limited cognitive resources available to subjects under heavy workloads, the parasympathetic nervous system conserves energy and slows down the metabolism to perform adequately at the primary task [13]. In their study *Hancock et al.* show the increased cognitive processing and energy demands required to match the resource demands of a high difficulty task are reflected by the aforementioned physiological variables[11]. While this is indicative superior modelling capacities of multi-modal sensors it is our study investigates it is counter productive to our study.
In order to verify that the workload manipulations are being mentally perceived by subjects with different difficulty, we decided to use joystick deflections from the neutral position as the behavioural measure[4].

2 EXPERIMENT

Participants

10 participants (9 male and 1 female) from the University of Colorado, Boulder participated in this the study. They ranged in age from 23 to 42 years (M = 26.4 yrs, SD = 8.6) and gave informed consent under the guidelines and restrictions of the university's institutional review board. As the participants were fairly new to the X-Plane simulator and were allowed to practice until they were familiar with the simulator plugin.

Testbed Environment

Overview. The experiment setup consisted of a computer equipped with an X-Plane 11 Flight Simulator installation. In order to mimic a typical in-flight cockpit of an FA-18F, the simulator was configured to simulate the same aircraft to be maneuvered with a Thrust-Master Joystick and Rudders as shown in Figure 1. Using pilot testing and prior literature **cite** as guidance, we created two scenarios of high and low difficulty levels within the X-Plane environment. For the low difficulty level, participants were instructed to fly the plane while maintaining an altitude of ±5000 feet from 10,000 feet while the weather conditions were clear and sunny, with no wind. In the high difficulty level condition the participants had to keep their altitude within the more restrictive ±500 feet from 10,000 feet, while operating under extreme weather conditions that included high wind and levels of rain. After the participants were comfortable with X-Plane plugin's workflow and the test-bed environment, the data acquisition began during which they were left alone to their own devices providing them a distraction free zone.

Figure 1: Testbed Environment with desk mounted Tobii Eye Tracker (highlighted in yellow)

TOME - Data Collection Infrastructure

One of the key contributions of our study is the use of psychophysiology modalities in a non-invasive and distraction free environment.

Nicknamed TOME [Figure 2]- The Tools for Objective Measurement and Evaluation system is a diagnostic tool-set for

Table 1: Weather changes and altitude constraints for each task

Parameter	Low Difficulty	High Difficulty
Altitude	10000 ±5000	10000. ±500
Wind Speed	55 kts.	430 kts.
Sky	Clear	Cloudy
Rain Speed	16mph	110mph

Figure 2: A screenshot of the TOME Dashboard

cost-effectively supporting test and evaluation practitioners and augmenting system acquisition decisions through advanced workload measurement and performance assessment strategies. TOME applies a hybrid modeling architecture that capitalizes on the strengths of various computational modeling and machine learning techniques, while mitigating over reliance on any single approach. The primary component of the TOME system is the real-time workload assessment toolkit, which consists of several components intended to measure, harness and process multi-modal sensor data, and using these data to infer cognitive workload. The Body Area Network (BAN) transmitter is a service application that currently runs on Android mobile devices. The BAN transmitter's main responsibility is connecting to the various sensors a person is wearing and transmitting that information to the TOME server. The BAN has multiple ways it interfaces with the system, including data streaming through a message broker, making requests through HTTP web services provided by the TOME server, and interfacing with sensors through various communication methods, such as Bluetooth Low Energy (BLE). The BAN is designed to serve as a gateway between the TOME server and sensors. This prevents the server from having to manage sensor connections directly, making the system more scalable and manageable. Additionally, because the BAN can remotely connect to the server (via WiFi or cellular network), users do not need to stay within a certain physical range of the server, thus carrying the potential to make them more mobile. Another benefit of the BAN transmitter is that it runs as a background service. This ensures the application is always running and maintaining an active connection with the server. It also limits user interaction with the system, which allows users to focus on their tasking without distraction. The TOME server performs a variety of functions, including centralized real-time data processing, management of user states, management of experimental test conditions, and persistent data storage by utilizing cloud computing resources. In short, the TOME server manages all data in the system, executes algorithms, and hosts web services that interact with the system. Also running on the TOME server is a web server that provides a front-end user interface for the entire system. These applications include several different pages for viewing data, entering forms, and performing administrative functions. This also provides a convenient mechanism for exporting all collected data into a comma separated value format that can be ingested by virtually any commercial statistical software package, including both unprocessed sensor data and post-processed algorithm derived measures. Because the displays are web-based, they are compatible with a wide variety of devices, specifically any device that can run a web browsing application. The TOME backend server includes a processing module that executes algorithms to generate alerts and derived features within the system, including inferred user states such as cognitive workload. The TOME project includes an API for algorithm development. This API includes several interfaces and abstract classes to help developers create new algorithms that can easily be used by the system. It also comes with utility methods to help evaluate algorithms in bulk. The system currently supports two main types of algorithms: 1) Data Algorithms: Algorithms that receive data messages to generate new features within the system, often referred to as "derived" features; and 2) Alert Algorithms: Algorithms that receive data messages to generate alerts within the system. Both types of algorithms fundamentally work the same way; they mostly differ with respect to the type of information they generate.

Sensor Setup

Heart Rate, fNIRS, Respiration Rate, Galvanic Skin Response (GSR) and Eye Tracking data were simultaneously recorded as the participant completed the experiment. Special measures were undertaken in order to maintain uniformity in sensor placement, invariant to participants' physical attributes.

Each participant wore a compression t-shirt embedded with electrodes around its chest cavity which connects to a Polar H10 heart rate monitor placed on their necks. Respiration rate was collected using a Zephyr Bio-harness placed at the bottom of their sternum. GSR was recorded using a Polar M4500 smartwatch worn by the participants' on their right wrist. Lastly a 2 channel fNIRS device, the Plux Pioneer was placed on the participants forehead at the mid-point of their

Figure 3: Participant equipped with the 4 sensors - 1. Plux 2 Channel fNIRS, 2. Zephyr Bioharness (Respiration), 3. Polar HR10 (Heart Rate) and 4. Smartwatch (GSR)

respective FpZ locations using the measured using the 10-20 system [24].

Protocol

The custom X-Plane plugin allowed the researchers create custom experimental designs for each participants, with custom difficulty level settings. The weather parameters and their respective values for low and high difficulty tasks are tabulated in Table 2. While navigating through the skies, participants were also instructed to maintain flight altitude at a minimum of 10,000 ft. with an error margin allowance as per the task difficulty. A set of instructions pertaining to the altitude of their flights would appear on the center-left side of the simulator screen which the participants were instructed to follow [Figure 4].

Figure 4: X-Plane 11 GUI Showing a blocking window for Rest

A task representing each condition (low and high) consisted of flying for 60 seconds while maintaining the target altitude range amongst the weather conditions created by the plug-in. After each 60 second long task, the X-Plane simulator would pause the screen mid-flight and participants would rest for 45 seconds to allow their brain's metabolic activity to return to baseline. The protocol included a randomized block, with each block containing one low and one high difficulty level condition. The block design had 3 blocks. Thus, each condition (low and high) was experienced three times for each participant.

3 DATA ANALYSIS

The following sections delineate four processes involved in building a typical BCI system - pre-processing, feature extraction, feature selection and classification.

Data Pre-Processing

Since the signalling medium is not uniform between the 5 sensors (optical density of fNIRS vs electric potential of heart rate monitor), data acquired from each sensor goes through different pre-processing routines.

Functional Near Infra-Red Spectroscope: Raw data acquired (ADC) from fNIRS device is first converted into absorption coefficient $I_R(\mu_A)$ using the transfer function outlined below:

$$I_R = \frac{c * ADC}{2^n} \tag{1}$$

Here, c is a proportionality constant depending on bit precision (n) of ADC values configured within the OpenSignals Software. We used values 0.15 and 16 bits respectively[5]. Absorption coefficient (A) is then estimated using Modified Beer Lambert Law. Later Optical Density (OD) is calculated using the Differential Path-length algorithm [2]. For our device we used DPF values most commonly cited in the literature - 6.51 and 5.86 for wavelengths at 690 and 830 nm respectively. Due to the intricate optical properties and high frequency nature of fNIRS sensors, the OD values acquired are sensitive to to the displacement of opt-odes from their locations resulting from head movements, micro movements resulting from cardiac pulses (Mayer Waves) or respiratory activities of the subjects [6]. Such a displacement of opt-odes can manifest as sudden spikes in the acquired optical densities. These motion artifacts are band-pass filtered from the μ_A values between 0.1 and 0.5Hz. Finally ΔHbO and ΔHbR values are estimated from the two wavelengths using the coefficients described in the work of Franceschini et al[12]. In order to classify MWL from ΔHbO and ΔHbR, mean concentration change and slope of a linear regressor fitted to each 1 second window for the 2 channels were then used as features for the classifier.

Heart Rate Monitor: To classify MWL from heart rate, raw ECG data was band-pass filtered between 0.1 and 30Hz followed by detrending to remove baseline shift **cite**. After extracting the QRS complex from the filtered signal, emergence of an R peak in the waves indicated a heart beat **cite** and hence, time difference between two consequent R peaks and their frequency (RR Interval) were then used as features to condense ECG data.

From Respiration Monitor, GSR and Eye Tracker. For extracting features from Tobii eye tracker, GSR in smartwatch and Respiration monitor we first removed outliers by calculating z-score and identifying the data points > 2 units from the mean as outliers. The results of performing outlier detection are shown in Figure 5. Later, mean and standard deviations over a 1 second window from the above features were used as features for predicting MWL during the piloting task.

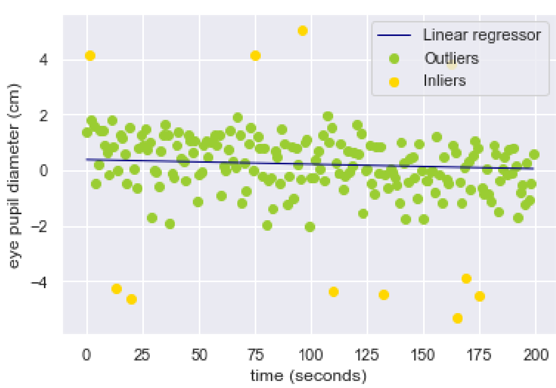

Figure 5: Results of outlier detection left pupil diameter from TOBII Eye Tracker

Feature selection

After pre-processing the data we extracted a total of 17 features [Table 2] from the set of 5 sensors outlined above. The mental workload that a subject experiences is then classified using these features. However, neuro-physiological processes respond differently to external stimulus provided and as a result their contribution to the classifier varies. Selection of an optimal feature set is critical to creating an efficient classification pipeline. While more sophisticated methods do exist in the literature for identifying an optimal feature set [25], we found Pearson's correlation coefficient to provide the best trade off between computational complexity and low dimensionality of the data output from each of the sensors in our study[19, 34]. The correlation matrix for data-sets formed by low and high workload are shown in Figures 6 and 7 respectively. We identified 5 features which exhibited a good ($p > 0.5$) correlation between neuro-physiological changes high and low stimuli.

Index	Source	Feature
1	Polar HR10	RR interval
2		Heart Rate
3	Bioharness	Respiration rate SD
4		**Respiration rate mean**
5	Tobii 4C	**Left pupil diameter**
6		**Right pupil diameter**
7	Polar M4500	Pulse Rate
8		Activity (GSR)
9	fNIRS	Channel 1 ΔHbO mean
11		Channel 1 ΔHbR mean
12		**Channel 1 ΔHbO slope**
13		Channel 1 ΔHbR slope
14		Channel 2 ΔHbO mean
15		Channel 2 ΔHbR mean
16		**Channel 2 ΔHbO slope**
17		Channel 2 ΔHbR slope

Table 2: Features extracted from the 5 sensor streams. Optimal set identified using p > 0.5 and p<-0.5 highlighted in bold.

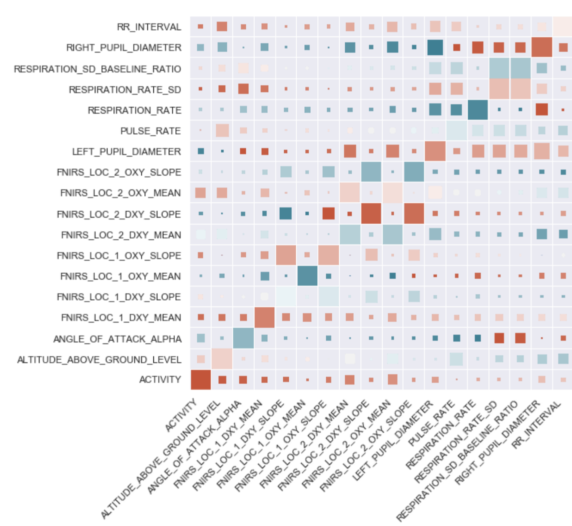

Figure 6: Correlation Matrix for low difficulty data-set

Classification

Using feature set extracted by the above procedure, we trained a supervised classifier to classify the mental workload experienced by a participant as an effect of the changes in the

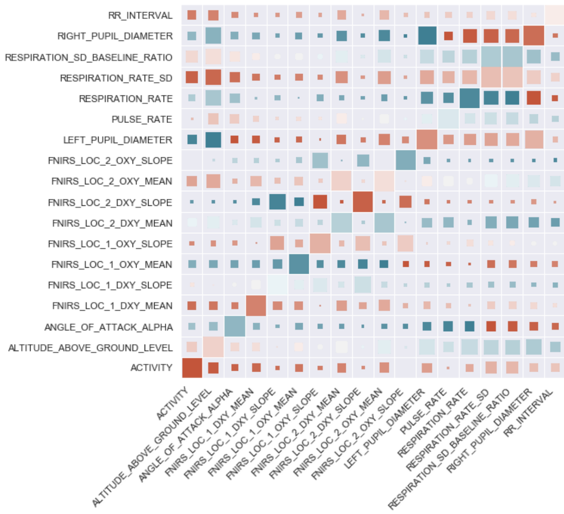

Figure 7: Correlation Matrix for high difficulty data-set

difficulty levels of the simulator. For this study we implemented three base classification algorithms based on prior literature [25] namely, support vector machines (SVM), linear discriminant analysis (LDA) and naive bayes (NB). Each of the above classifiers were then evaluated using leave-one-participant-out-cross-validation paradigm [15]. As per this paradigm we trained each classifier 10 times such that at each cross-validation stage, the classifier is trained on the data-set containing features from 9 randomly sampled participants and later tested on the features from the 10^{th} participant. The performance of three based classifiers is then evaluated by averaging the f1-score and accuracy of the classifiers from the 10 validation stage.

4 RESULTS AND INTERPRETATION

Using the first order derivative of joystick deflections (ΔJ) from the neutral axis we were able to perform workload manipulation check on our participant's data. Figure 8 shows the average of deflections for the first High and Low difficulty task encountered by the participants respectively. As per the figure, the ΔJ for high difficulty tasks are significantly far from the neutral axis ($y = 0$).

Results of cross-validation for the three classifiers are tabulated in Table 3.

Since SVM performs significantly better than LDA and NB, we will carryout further analysis using SVM as the base classifier. We also investigated the best combination of the three modalities that passed Pearson's test by training an SVM classifier using the 10 fold CV paradigm described in section 4.3. Comparing the results tabulated in Table 4 and 3 we can see that fusing the 3 modalities together results in the best performing classifier compared to a subset of modalities alone.

Base Classifier	F1-Score	Accuracy
Support Vector Machine	**0.78**	**0.79**
Linear Discriminant Analysis	0.73	0.71
Naive Bayes	0.69	0.66

Table 3: Averaged scores for the base classifiers using 10-fold leave one participant out cross validation.

Modality	F1-Score	Accuracy
fNIRS	0.71	0.698
Tobii 4C	0.76	0.78
Bioharness	0.75	0.77
Bioharness + Tobii 4C	0.77	0.78
Bioharness + fNIRS	0.69	0.69
fNIRS + Tobii 4C	0.76	0.75
Bioharness + fNIRS + Tobii 4C	0.78	0.79

Table 4: Performance scores of a combination of fNIRS, Eye tracking and Respiration modalities.

To investigate the performance of individual modalities we also carried out a one way ANVOA test on the three groups of sensor streams. As seen in Table 5 features from the 3 devices show a significant variance with respect to the change in task difficulty (p < 0.005)

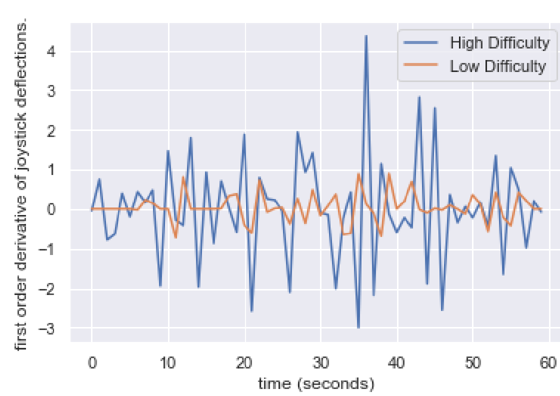

Figure 8: Average Joystick deflections across 10 participants for the two difficulty levels.

Modality	p-value
fNIRS	$6 * 10^-4$
Zephyr Bioharness	$8 * 10^-13$
Tobii 4C	$5 * 10^-21$

Table 5: p-values calculated using 3 way ANOVA for the two difficulty levels.

5 CONCLUSION AND FUTURE WORK

Through this study we demonstrate the correlation between MWL and the neuro-physiological responses exhibited by the subjects undergoing a simulated aircraft piloting task.

REFERENCES

[1] Sangtae Ahn, Thien Nguyen, Hyojung Jang, Jae G. Kim, and Sung C. Jun. 2016. Exploring Neuro-Physiological Correlates of Drivers' Mental Fatigue Caused by Sleep Deprivation Using Simultaneous EEG, ECG, and fNIRS Data. *Frontiers in Human Neuroscience* 10 (2016), 219. https://doi.org/10.3389/fnhum.2016.00219

[2] Arjen Amelink, Henricus JCM Sterenborg, Martin PL Bard, and Sjaak A Burgers. 2004. In vivo measurement of the local optical properties of tissue by use of differential path-length spectroscopy. *Optics letters* 29, 10 (2004), 1087–1089.

[3] Gianluca Borghini, Laura Astolfi, Giovanni Vecchiato, Donatella Mattia, and Fabio Babiloni. 2014. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience Biobehavioral Reviews* 44 (2014), 58 – 75. https://doi.org/10.1016/j.neubiorev.2012.10.003 Applied Neuroscience: Models, methods, theories, reviews. A Society of Applied Neuroscience (SAN) special issue.

[4] Jan Boril, Miroslav Jirgl, and Rudolf Jalovecky. 2016. Use of flight simulators in analyzing pilot behavior. In *IFIP International Conference on Artificial Intelligence Applications and Innovations*. Springer, 255–263.

[5] Bethany K Bracken, Noa Palmon, Seth Elkin-Frankston, and Filipe Silva. 2018. PORTABLE, DURABLE, RUGGED, FUNCTIONAL NEAR-INFRARED SPECTROSCOPY (fNIRS) SENSOR. US Patent App. 15/673,934.

[6] Sabrina Brigadoi, Lisa Ceccherini, Simone Cutini, Fabio Scarpa, Pietro Scatturin, Juliette Selb, Louis Gagnon, David A Boas, and Robert J Cooper. 2014. Motion artifacts in functional near-infrared spectroscopy: a comparison of motion correction techniques applied to real cognitive data. *Neuroimage* 85 (2014), 181–191.

[7] Jr JA Caldwell and JA Lewis. 1995. The feasibility of collecting in-flight EEG data from helicopter pilots. *Aviation, space, and environmental medicine* 66, 9 (1995), 883–889.

[8] Daniel E Callan, Gautier Durantin, and Cengiz Terzibas. 2015. Classification of single-trial auditory events using dry-wireless EEG during real and motion simulated flight. *Frontiers in systems neuroscience* 9 (2015), 11.

[9] Mickael Causse, Zarrin Chua, Vsevolod Peysakhovich, Natalia Del Campo, and Nadine Matton. 2017. Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. *Scientific reports* 7, 1 (2017), 1–15.

[10] Mickaël Causse, Frédéric Dehais, and Josette Pastor. 2011. Executive functions and pilot characteristics predict flight simulator performance in general aviation pilots. *The International Journal of Aviation Psychology* 21, 3 (2011), 217–234.

[11] Rebecca L Charles and Jim Nixon. 2019. Measuring mental workload using physiological measures: a systematic review. *Applied ergonomics* 74 (2019), 221–232.

[12] Maria Angela Franceschini, Vlad Toronov, Mattia E Filiaci, Enrico Gratton, and Sergio Fantini. 2000. On-line optical imaging of the human brain with 160-ms temporal resolution. *Optics express* 6, 3 (2000), 49–57.

[13] Peter A Hancock, Najmedin Meshkati, and MM Robertson. 1985. Physiological reflections of mental workload. *Aviation, space, and environmental medicine* (1985).

[14] Thomas C Hankins and Glenn F Wilson. 1998. A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, space, and environmental medicine* 69, 4 (1998), 360–367.

[15] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction.* Springer Science & Business Media.

[16] Robert T. Hays, John W. Jacobs, Carolyn Prince, and Eduardo Salas. 1992. Flight Simulator Training Effectiveness: A Meta-Analysis. *Military Psychology* 4, 2 (1992), 63–74. https://doi.org/10.1207/s15327876mp0402_1 arXiv:https://doi.org/10.1207/s15327876mp0402$_1$

[17] Keum-Shik Hong, M. Jawad Khan, and Melissa J. Hong. 2018. Feature Extraction and Classification Methods for Hybrid fNIRS-EEG Brain-Computer Interfaces. *Frontiers in Human Neuroscience* 12 (2018), 246. https://doi.org/10.3389/fnhum.2018.00246

[18] Angel Jimenez-Molina, Cristian Retamal, and Hernan Lira. 2018. Using Psycho-Physiological Sensors to assess Mental Workload in Web Browsing. *Sensors* 18 (01 2018). https://doi.org/10.3390/s18020458

[19] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. 2011. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing* 3, 1 (2011), 18–31.

[20] Ioana Koglbauer, K Wolfgang Kallus, Reinhard Braunstingl, and Wolfram Boucsein. 2011. Recovery training in simulator improves performance and psychophysiological state of pilots during simulated and real visual flight rules flight. *The international journal of aviation psychology* 21, 4 (2011), 307–324.

[21] Loukia D Loukopoulos, R Key Dismukes, and Immanuel Barshi. [n.d.]. Concurrent task demands in the cockpit: Challenges and vulnerabilities in routine flight operations.

[22] George A Miller. 1994. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review* 101, 2 (1994), 343.

[23] Charles H. Morris and Ying K. Leung. 2006. Pilot mental workload: how well do pilots really perform? *Ergonomics* 49, 15 (2006), 1581–1596. https://doi.org/10.1080/00140130600857987 arXiv:https://doi.org/10.1080/00140130600857987 PMID: 17090505.

[24] Takayuki Nakahachi, Ryouhei Ishii, Masao Iwase, Leonides Canuet, Hidetoshi Takahashi, Ryu Kurimoto, Koji Ikezawa, Michiyo Azechi, Osami Kajimoto, and Masatoshi Takeda. 2010. Frontal cortex activation associated with speeded processing of visuospatial working memory revealed by multichannel near-infrared spectroscopy during Advanced Trail Making Test performance. *Behavioural brain research* 215, 1 (2010), 21–27.

[25] Noman Naseer and Keum-Shik Hong. 2015. fNIRS-based brain-computer interfaces: a review. *Frontiers in Human Neuroscience* 9 (2015), 3. https://doi.org/10.3389/fnhum.2015.00003

[26] Calvin KL Or and Vincent G Duffy. 2007. Development of a facial skin temperature-based methodology for non-intrusive mental workload

measurement. *Occupational Ergonomics* 7, 2 (2007), 83–94.

[27] Gary B Reid and Thomas E Nygren. 1988. The subjective workload assessment technique: A scaling procedure for measuring mental workload. In *Advances in psychology*. Vol. 52. Elsevier, 185–218.

[28] Eike A Schmidt, Michael Schrauf, Michael Simon, Martin Fritzsche, Axel Buchner, and Wilhelm E Kincses. 2009. Drivers' misjudgement of vigilance state during prolonged monotonous daytime driving. *Accident Analysis & Prevention* 41, 5 (2009), 1087–1093.

[29] Young Woo Sohn and Stephanie M. Doane. 2004. Memory Processes of Flight Situation Awareness: Interactive Roles of Working Memory Capacity, Long-Term Working Memory, and Expertise. *Human Factors* 46, 3 (2004), 461–475. https://doi.org/10.1518/hfes.46.3.461.50392 arXiv:https://doi.org/10.1518/hfes.46.3.461.50392 PMID: 15573546.

[30] Patricia Vanzella, Joana B. Balardin, Rogério A. Furucho, Guilherme Augusto Zimeo Morais, Thenille Braun Janzen, Daniela Sammler, and João R. Sato. 2019. fNIRS Responses in Professional Violinists While Playing Duets: Evidence for Distinct Leader and Follower Roles at the Brain Level. *Frontiers in Psychology* 10 (2019), 164. https://doi.org/10.3389/fpsyg.2019.00164

[31] Kevin J. Verdière, Raphaëlle N. Roy, and Frédéric Dehais. 2018. Detecting Pilot's Engagement Using fNIRS Connectivity Features in an Automated vs. Manual Landing Scenario. *Frontiers in Human Neuroscience* 12 (2018), 6. https://doi.org/10.3389/fnhum.2018.00006

[32] Glenn F. Wilson. 2002. An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *The International Journal of Aviation Psychology* 12, 1 (2002), 3–18. https://doi.org/10.1207/S15327108IJAP1201_2 arXiv:https://doi.org/10.1207/S15327108IJAP1201$_2$

[33] Glenn F Wilson and Thomas Hankins. 1994. EEG and subjective measures of private pilot workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 38. SAGE Publications Sage CA: Los Angeles, CA, 1322–1325.

[34] Fu Yang, Xingcong Zhao, Wenge Jiang, Pengfei Gao, and Guangyuan Liu. 2019. Multi-method Fusion of Cross-Subject Emotion Recognition Based on High-Dimensional EEG Features. *Frontiers in Computational Neuroscience* 13 (2019), 53. https://doi.org/10.3389/fncom.2019.00053