USAARL-TECH-TR--2023-11

UNITED STATES ARMY AEROMEDICAL RESEARCH LABORATORY

A Systematic Literature Review of Operator State Detection using Physiological Measures

Michelle J. Duffy & Kathryn A. Feltman

Notice

Qualified Requesters

Qualified requesters may obtain copies from the Defense Technical Information Center (DTIC), Fort Belvoir, Virginia 22060. Orders will be expedited if placed through the librarian or other person designated to request documents from DTIC.

Change of Address

Organizations receiving reports from the U.S. Army Aeromedical Research Laboratory on automatic mailing lists should confirm correct address when corresponding about laboratory reports.

Disposition

Destroy this document when it is no longer needed. Do not return it to the originator.

Disclaimer

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation. Citation of trade names in this report does not constitute an official Department of the Army endorsement or approval of the use of such commercial items.

IRB Determination and Number

N/A

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
08-12-2022	Technical Report	

4. TITLE AND SUBTITLE
A Systematic Literature Review of Operator State Detection using Physiological Measures

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Duffy, M. J.[1,2] & Feltman, K. A.[1]

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Aeromedical Research Laboratory
P.O. Box 620577
Fort Rucker, AL 36362

8. PERFORMING ORGANIZATION REPORT NUMBER
USAARL-TECH-TR--2023-11

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Development Command
Military Operational Medicine Research Program
504 Scott Street
Fort Detrick, MD 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)
USAMRDC MOMRP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
[1]U.S. Army Aeromedical Research Laboratory; [2]Oak Ridge Institute for Science and Education

14. ABSTRACT
A systematic literature review was conducted to examine the literature (published between 2010 and 2021) surrounding the use of physiological measures to identify operator's cognitive state. operator states of interest were those that pose significant risks to Army aviators. Specifically, workload, fatigue, inattention, stress, and hypoxia. Additionally, studies that took place in applied and/or mobile contexts were sought in order to ensure the greatest likelihood of operationally relevant work. From this review, thirty-two eligible studies were identified. From these studies, it was determined that workload, fatigue, and inattention show the greatest promise for detection through physiological metrics. However, this is in part due to number of available studies. For instance, only one study was eligible where hypoxia was the cognitive state of interest. Additionally, based on the papers reviewed, EEG and eye metrics appear the most promising for identifying these various operator cognitive states. Further work is needed to validate some of these measures within true operational contexts. Specifically, based on the literature to-date, it is unknown how well some of these measures would hold up in a rotary-wing environment, where the sensors would be exposed to vibration and extreme temperatures. Further work is needed to determine whether the findings from these papers generalize to a larger population.

15. SUBJECT TERMS
Operator state monitoring, aviation, rotary-wing

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	19b. TELEPHONE NUMBER *(Include area code)*
UNCLAS	UNCLAS	UNCLAS	SAR	33	Loraine St. Onge, PhD	334-255-6906

This page is intentionally blank.

Executive Summary

A systematic literature review was conducted to examine the current literature (published between 2010 and 2021) surrounding the use of physiological measures to identify operator's cognitive state. The operator states of interest were those that pose significant risks to Army aviators, specifically, workload, fatigue, inattention, stress, and hypoxia. Additionally, studies that took place in applied and / or mobile contexts were sought in order to ensure the greatest likelihood of operationally relevant work. From this review, thirty-two eligible studies were identified. From these studies, it was determined that workload, fatigue, and inattention show the greatest promise for detection through physiological metrics. However, this is in part due to the number of studies available for review. For instance, only one study was eligible where hypoxia was the cognitive state of interest. Additionally, based on the papers reviewed, electroencephalogram (EEG) and eye metrics appear the most promising metrics for identifying these various operator cognitive states. Further work, however, is needed to validate some of these measures within true operational contexts. Specifically, based on the literature to-date, it is unknown how well some of these measures would hold up in a rotary-wing environment, where the sensors would be exposed to vibrations and extreme temperatures. Also, many of the sample sizes were small. Further work is also needed to determine whether the findings from these papers generalize to a larger population.

This page is intentionally blank.

Acknowledgements

This research was supported in part by an appointment to the Research Participation Program at the U.S. Army Aeromedical Research Laboratory administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Army Medical Research and Development Command.

This page is intentionally blank.

Table of Contents

List of Tables

This page is intentionally blank.

Introduction

Operator state monitoring of Army aviators is an ongoing research area within the U.S. Army Aeromedical Research Laboratory (USAARL). The goal of this line of research is to identify physiological markers that can be used to identify an aviator's cognitive state in real-time. Much of the research to-date has focused on identifying operator workload, specifically when a state of overload is reached (e.g., Feltman et al., 2020; Aura et al., 2021). However, other states have also been explored, such as vigilance (e.g., Kelley et al., 2020) and hypoxia (e.g., Temme et al., 2016). Within these studies, the physiological measures used to correlate with cognitive state included electroencephalograph (EEG), electrocardiograph (ECG), electrooculography (EOG), eye tracking, pulse oximetry (SpO_2), and respiration. These measures have been chosen and used within USAARL studies due to their association with these cognitive states, as well as the ease of use within the flight simulator environment (e.g., wireless connectivity, integration with other equipment). However, recent years have seen an influx of research conducted on this topic, particularly research conducted within similar settings. These settings include a variety of flight and driving simulators, ambulatory studies, and even real driving and flight. Additionally, researchers are using a variety of physiological measures to capture varying operator states. Indeed, multiple recent reviews summarizing many of these topics are available, with the majority focused specifically on mental workload (e.g., Weelden et al., 2022; Charles & Nixon, 2019; Tao et al., 2019; Pagnotta et al., 2022).

Missing in the current literature, however, is a review of which physiological measures have demonstrated promise in identifying a range of operator cognitive states of interest to USAARL's research program. Such cognitive states that Army aviators are and will likely be prone to experiencing within current and future aircraft include but are not limited to mental workload (overload and underload), physiological and cognitive fatigue, distraction and / or inattention, and the need to maintain vigilance. Cognitive states included in this review such as hypoxia and stress are frequently considered physiological states, while fatigue can be interpreted from both a cognitive and physiological perspective. Due to the shared effects on operator performance, these types of physiological states and other traditional cognitive states were grouped together and referred to as operator states and/or cognitive states for the purposes of this review. In addition, which types of settings (e.g., driving simulator, real flight) these measures have been evaluated within is oftentimes overlooked in reviews. In order to identify which operator states have shown the most promise of identification through physiological measures, and which physiological measures hold promise for integration within the flight environment, a systematic literature review was conducted. The objective of this review was to identify the types of psychophysiological metrics that most reliably detect changes to operator state and use that information to determine sensor types to pursue for evaluation within USAARL's operator state monitoring research program.

Methods

Literature searches were conducted using Google Scholar, PubMed, and EBSCO. The use of Google Scholar aimed to capture "grey literature," such as proceedings papers, in order to have a more *expansive* search. Keywords used for literature searches included physiological terms (electrocardiogram, functional near infrared spectroscopy [fNIRS], etc.) and operator states such as workload, stress, and fatigue. Table 1 lists the search terms used. Due to inconsistencies of terminology across multiple scientific fields, additional terms were introduced to help limit the search results. The inclusion of grey literature from Google Scholar further contributed to inconsistencies with terminology.

Table 1. Keywords Used in Literature Search

Physiology Terms	Operator States	Additional Terms
Physiological indices	Workload	Operator state monitoring
Sensors	Underload	Cognitive state monitoring
Physiology	Overload	State monitoring
Psychophysiology	Fatigue	Real-time monitoring
Pulse measurement	Distraction	
Blood pressure	Inattention	
Respiration	Vigilance	
Eye tracking	Engagement	
Pupillometry	Stress	
Electrooculogram	Cognitive state	
Electrocardiogram	Mind-wandering	
Heart rate	Boredom	
Heart rate variability		
Electroencephalograph		
fNIRS		

Eligibility

To be included in the systematic review, studies needed to meet the following criteria: an experimental design with manipulation of a cognitive state, must have had a sample between the age range of 18-50, in a mobile context (i.e., aircraft, driving simulator), and of a healthy non-abnormal population. Exclusion criteria included studies using a non-mobile context (i.e., desktop simulator), non-English language, unhealthy or abnormal population, sample under the age of 18 or over the age of 50, or non-experimental design papers. Inclusion and exclusion criteria are summarized in Table 2.

Table 2. Inclusion and Exclusion Criteria

Criteria	Included	Excluded
Date Published	2010 - 2021	Any prior to 2010
Language	English Language	Non-English Language
Test Population	Age: 18 - 50 years	Age: under 18 years and above 50 years
	Race: Any	Race: None
	Gender: Males and Females	Gender: None
	Healthy	Unhealthy or abnormal
	Nationality: Any	Nationality: None
Study Design	Peer-reviewed articles, proceedings papers, conference papers, technical reports with experimental designs	Non-experimental designs, review papers (meta-analyses), descriptive or case studies; theoretical or position papers, editorials, book reviews
Study Context	Mobile context (flying, driving, simulators)	*Non-mobile context (desktop simulators)

Note. Approximately four non-mobile context articles were included due to similarity of study content.

Procedure

A research team located potentially relevant studies using search criteria above (see Table 1). The team then reviewed titles and abstracts for each search result to deem pre-eligibility, and then requested full-text versions of potentially relevant articles. Following the first screening, articles were then reviewed by multiple team members for eligibility. The initial search results included 8586 of items, followed by the removal of 2229 duplicate citations. Approximately 5230 of articles or papers were deemed ineligible by title and abstract. An additional 1037 results were judged to irrelevant or ineligible by abstract. A total of 90 full-text articles were retrieved for a full review of eligibility by multiple team members. The final number of included articles for this report was 32. Table 3 shows the literature search and review results.

This space is intentionally blank.

Table 3. Literature Search and Review Results

Search results (January 2022)	**8586**
Duplicated citations	2229
Judged irrelevant or ineligible by title	5230
Judged irrelevant or ineligible by abstract	1037
Full-text articles retrieved	90
Included studies	32

Data extracted from each of the 32 studies included target population, study location, cognitive state manipulation, mobile context details, study outcomes and tasks, use of physiological equipment, data analysis, and the overall study quality such as clarity of descriptions of methods and results sections, as well as transparency of analyses. Microsoft Excel spreadsheets were used to organize extracted study information.

Of special note, one study (Huang et al., 2016) was removed from the final report. This study evaluated a closed-loop EEG system where fatigue was monitored via EEG and an auditory alert was delivered when fatigue was detected. Although very similar to the goals of this report, Huang et al.'s article focused on evaluating the efficacy of the alerting system. Thus, it was excluded from the report, but deserves a special note.

This space is intentionally blank.

Results

Extracted data were summarized and evaluated for general themes. Of the 32 eligible pieces of literature, the majority examined workload (16) (three of which were within non-mobile contexts), followed by fatigue (8), inattention/distraction (4), and stress/overload (3) and one non-mobile context looking at hypoxia (Rice et al., 2019). Table 4 summarizes the overall characteristics of studies for each cognitive state, while the following sections describe the physiological measures used.

*Table 4.*Overview of Study Characteristics by Cognitive States

		Workload (n = 16)	Fatigue / Drowsiness (n = 8)	Inattention / Distraction (n = 4)	Stress / Overload (n = 3)	Hypoxia (n = 1)
Text Category	Peer Review	14	8	4	3	1
	Non-Peer Review	2	-	-	-	-
Subject Population	University Students	6	-	-	-	-
	Military Personnel	2	-	-	3	1
	Other/Unclear	9	8	4	-	-
Study Context	Simulator (i.e., vehicle, airframe)	11	8	3	-	-
	Real Vehicle / Airframe	1	-	-	-	-
	Non-Vehicle Mobile (i.e., walking)	1	-	1	3	-
	Non-Mobile Context	3	-	-	-	1

Note. Non-peer reviewed texts included proceedings/conference papers (4) and technical reports (1).

The majority of articles measured these cognitive states using EEG (17 total), followed by ECG/heart monitoring (14 total), and then EOG/eye tracking technology (9 total). Additional physiological metrics included across studies were fNIRS, Galvanic skin response (GSR)/electrodermal activity (EDA)/skin conductance, respiration rate, salivary amylase levels, blood pressure, blood glucose/blood lactate, and blood oxygen. Detailed results are reported by cognitive state within the following sections. The vast majority of studies used fixed-based simulators with only seven using motion-based or real vehicle monitoring.

Workload

A total of 16 included articles measured workload. Of the 16, three used non-mobile contexts but were included due to the similarities in study inclusion criteria (e.g., real-time workload monitoring). Seven of the 16 studies used EEG, nine used ECG/heart monitoring, four used eye tracking, two used skin conductance, two used respiration, and one used fNIRS. The *n*-back task was commonly used to manipulate workload conditions as a secondary task, and those studies used both visual and auditory versions (Unni et al., 2017; Tarabay & Abou-Zeid, 2018; Yang et al., 2021). The *n*-back task was typically used in motor vehicle simulator studies, which accounted for approximately half of the total number of studies assessing workload. Other commonly used manipulations for workload included varying amounts of traffic density, visual conditions such as fog, and external distractions such as music (Blonco et al., 2018; Heikoop et al., 2017; Stuiver et al., 2014).

The remaining studies consisted of flight simulation, and typically used varying flight maneuvers to manipulate workload conditions or used changes in flight conditions such as day vs. night and weather, which included limited visibility vs. clear skies (Feltman et al., 2020; Feng et al., 2018). Three studies used air traffic control simulators (Bernhardt et al., 2019; Arico et al., 2016; Raduntz et al., 2020). Two studies included real flight. One was in a fixed-wing aircraft (Wilson et al., 2021), and the second used real-life helicopter air rescue scenarios (Schoniger et al., 2020).

Only two (Blonco et al., 2018; Wilson et al., 2021) of the 16 studies developed and evaluated the performance of classifiers. Both used only EEG inputs for the classifiers and reported up to 100% accuracies in identifying workload conditions. The remaining studies evaluated whether the physiological measures significantly differed between workload conditions. While three studies (Feng et al., 2018; Unni et al., 2017; Feltman et al., 2021) also investigated the relationships between physiological measures and performance and / or subjective measures. Table 5 below summarizes all workload studies.

This space is intentionally blank.

Table 5. Summary of Workload Studies

Reference	Device Description	Task Used to Induce Workload	***n*** **(Gender), Age (*range* or *M[SD]*)**	Physiological Outcome Measures	Method of Analysis	Outcomes
Blonco et al. (2018)	Flight Simulator (Prepar 3D v1.4, Lockheed Martin Coporation)	Flight task with degraded weather	21 (21M), 19-23 yrs	EEG with 13 classifiers with features extracted from four sites: *Fz, FCz, Cz, Pz*	Evaluated performance of classifiers	Classification accuracies ranged from 45.91% to 100%
Tarabay & Abou-Zeid (2018)	DriveSafety DS-600c simulator (full Ford cab)	Urban road situations (pedestrians, sudden stops, traffic lights) with auditory *n*-back task	80 (53M), age not reported	Heart rate via ECG *(average, SD, min/max)* Skin Conductance *(average, SD, min/max)*	Wilcoxon Signed-Ranks test compaing physiological measures and task performance between conditions	Significant differences between treatment and condition phase for min/max heart rate and skin conductance levels indicating higher workload with *n*-back task
Liu et al. (2020)	Fixed-based flight simulator	Embedded secondary task on flight dispalys	21 (19M), 20.6 yrs	Eye tracking *(blink frequency, mean gaze time)*	RM ANOVAs comparing phases of flight	Blink frequency differed between take off and straight phase and between turns. Average gaze time had main effect of display and flight phase
Wilson et al. (2021)	Real aircraft (four-seat, single engine)	Directed flight maneuvers	10, gender and age not reported	EEG *(Delta [1-4 Hz], Theta [4-8 Hz], Alpha [8-13 Hz], Beta [13-30 Hz])*	Classified high and low workload based on SVM classifier with 5-fold cross-validation	SVM classifier had 100% precision
Heikoop et al. (2017)	Southampton University fixed-based driving simulator Jaguar XJ Saloon	Monotonous automated driving vs. voluntary reading, listening to music, etc., vs. detecting number of red cars during session	22 (14M), 19-45 yrs (M = 29.6, SD = 6.8)	ECG (*heart rate, heart rate variability, SDNN, LF/HF)* Eye tracking *(PERCLOS)*	RM ANOVAs compared physiological variables among three driving conditions	LF/HF ratios were lower in the voluntary task compared to the detection task
Feng et al. (2018)*	Flying rocker simulator with HUD	No manipulation, looked at naturally occurring workload	11 (11M), 20.6 yrs	Eye tracking (*fixation frequency and average time, mean saccade time, blink rate and mean, pupil*	RM ANOVA (3 flight phases); Pearson correlations of	Main effect of workload on physiological measures:

Reference	Device Description	Task Used to Induce Workload	*n* (Gender), Age (*range* or *M[SD]*)	Physiological Outcome Measures	Method of Analysis	Outcomes
		throughout phases of flight		*diameter)* ECG (*mean NN, LF element*) EDA (*mean phasic and tonic*)	physiological measures and flight performance	Eye Tracking: higher fixation time, lower saccade frequency & higher saccade time high workload; blink time decreased with increasing workload; smaller pupil diameter during low workload ECG variables: NN decreased with increasing workload; LF/LHF higher during low workload EDA phasic was lower during cruise, EDA tonic was higher during landing Positive correlations with pupil diameter, skin conductance, fixation time. Negative correlations with fixation frequency, saccade frequency and time, blink rate
Yang et al. (2021)	Monach University Accident Research Centre (MUARC) advanced fixed-based driving simulator	Secondary tasks while driving (*n*-bask task, texting, or both combined)	57 (38M), (*M* = 31, *SD* = 11.1)	ECG for heart rate and heart rate variability *(16 outcome variables)*	Linear mixed effect models of heart rate during four driving conditions of *n*-back, baseline, text-messaging, and *n*-back with text-messaging	10/16 HRV variables significantly increased with increased cognitive demand across conditions (SDNN, CVNN, triangular index, VLF power, LF power, HF power, total power, CSI, modified CSI, CVI)
Stuiver et al. (2014)	ST Software fixed-based driving simulator	Virtual driving in high and low traffic density with no fog or fog conditions	15 (7M), 20-25 yrs	ECG *(HR, HRV)* Systolic finger blood pressure	RM MANOVAs compared physiological variables among four different conditions	Blood pressure increased with fog in high density traffic condition and a decrease in HRV was seen with low traffic condition compared to high traffic with fog present
Unni et al. (2017)	German Aerospace Center virtual reality fixed-based driving simulator	Speed regulation *n*-back task (5 level) were completed during driving	19 (17M), 19-32 yrs (*M* = 25.2, *SD* = 3.7)	ECG (*heart rate, root mean square of successive differences [RMSSD])*	Linear mixed effects analysis evaluated *n*-back conditions with HR & HR. Multivariate lasso	Heart rate increased wile heart rate variability decreased with increasing *n*-back levels. Pearson's correlation = 0.61 for fNIRS (group level;

Reference	Device Description	Task Used to Induce Workload	***n* (Gender), Age (*range* or *M[SD]*)**	Physiological Outcome Measures	Method of Analysis	Outcomes
				fNIRS *(HbO, HbR)*	regressions using fNIRS HbR to predict *n*-level, Pearson correlations from regressions	individual level analyses completed but not reported here)
Feltman et al. (2021)	UH-60 Black Hawk full-motion simulator	Two sets of flight scenarios: 1) light vs. dark conditions and 2) high volume radio calls vs. low volume radio calls	23 (23M), age (M = 36, SD = 4.99)	EEG *(PSD values for theta, alpha, beta, engagement index ratio of beta to [alpha + theta])*	RM ANOVAs and paired samples *t*-tests compared physiological and performance variables between conditions. Correlational analyses evaluated relationships between physiological and performance variables.	Worse performance including airspeed and altitude deviations in high workload conditions with beta values being sensitive between workload conditions
Schöniger (2020)	No device, real-life rescue missions of a German helicopter emergency medical service	Real-life air rescue missions with 3 emergency operations per day	20 (17M), (M = 44.95, SD = 4.80)	ECG *(HR, HRV, SDNN, RMSSD LF/HF ratio)*	RM ANOVAs comparing ECG values across phases of emergency operations	No significant differences in HR indices between three operations, while HRV rises significantly at beginning stages of an emergency up until the landing at an operation site
Feltman et al. (2019)	UH-60 Black Hawk full-motion simulator	Manipulated visibility to create high and low workload conditions during flight scenarios	32 (29M), 22-47 yrs (M = 31.31, SD = 6.82)	EEG *(PSD values for theta, alpha, beta, workload metric, engagement index ratio of beta to [alpha + theta]),* ECG (*HR, HRV, LF/HF ratio)* Respiration monitor (*RR)*	Series of hierarchical multiple linear regression models using workload manipulation and individual difference variables to predict flight performance and psychophysiological measures	Individual difference measures to include daytime sleepiness, sleep quality, and chronotype predicted multiple performance and psychophysiological measures
Hidalgo-Munoz et al. (2019)	IFSTTAR-LEPSIS driving simulator (Peugeot 308 cabin surrounded by video projection screens)	Performed two workload conditions (high vs. low) in either a sitting condition, or simulated driving condition	18 (10M), 22.7 yrs (SD = 1.4)	ECG (*HR, SDNN, RMSSD, pNN20, pNN50, LF, HF, LF/HF*) Respiration (*BR [inspirations/minute], mid-*	Two-way ANOVA (sitting vs. driving; high vs. low workload) comparing physiological measures	Increase in HR while driving compared to sitting for high-workload condition. RMSSD, pNN20, and pNN50 all decreased in high workload condition for driving compared to sitting. LF and HF

Reference	Device Description	Task Used to Induce Workload	***n*** **(Gender), Age (*range* or *M[SD]*)**	Physiological Outcome Measures	Method of Analysis	Outcomes
				band spectral power [0.07-0.14 Hz], high band spectral power [0.15-0.50 Hz])		power decreased under high-workload condition while driving, while LF/HF ratio increased with driving. No difference between workload conditions was seen for BR and there was a significant main effect for mid-band between workload conditions
Aricó et al. (2016)*	Air traffic controller simulator	Number of aircraft and type of clearances, and number/trajectory of interfering flights created easy, medium, and hard conditions	12 (gender not reported), age (M = 40.41, SD = 5.54)	EEG (*PSD values – theta and alpha*) were used to develop an *EEG-based workload index*	ANOVA comparing the *workload index* across workload conditions	Significant effect between the three levels where score related to difficulty conditions (i.e., lower score for easy compared to medium, medium lower compared to hard, and easy lower compared to hard)
Bernhardt et al. (2019)	High fidelity air traffic controller simulator	Number of aircraft arriving/departing, and presence/absence of issuing uncontrolled aircraft IFR clearances	47 (45M), 20.97 yrs	EEG (*commerical engagement index, mean workload index*) Eye tracking (*pupil diameter*)	Linear mixed effects models incorporating individual difference measures to determine differences in workload phases	The commercial workload index varied with the workload phases; the engagement index differentiated participant experience levels; pupil diameter varied with workload phases
Radüntz et al. (2020)	Air Traffic Controller Simulator	Eight scenarios with varying workload; number of aircraft and an exceptional event (pilot request for prioritization)	21(19M), 38 yrs (SD = 11)	EEG (*developed Dual Frequency Head Maps-workload index [DFHM-index]*)	Correlations of DFHM-index across scenarios with same traffic loads; RM ANOVA comparing scenarios with and without exceptional event	Correlations demonstrated stability of DFHM-index within scenarios of same traffic loads; DFHM-index increased with increased workload

*Note. We highlighted only the outcomes of interest for this review. SVM = support vector machine, CVNN = coefficient of variation of successive NN intervals, SDNN = standard deviation of NN intervals, NN = normal-to-normal, RMSSD = root mean square of successive NN intervals, LF/HF = low frequency/high frequency, VLF = absolute power of the very-low-frequency band, CSI = cardiac sympathetic index, CVI = cardiac vagal index, PERCLOS = percentage eyes closed, HUD = heads-up display, CVI = , CSI = , RR = respiration rate, HR = heart rate, HRV = heart rate variability, Hz = hertz, BR = breathing rate, , RM MANOVA = repeated measures multvariate analysis of variance, RM ANOVA = repeated measures analysis of varaince, HbO = Oxyhemoglobin , HbR = Deoxyhemoglobin, IFR = instrument flight rules.

Fatigue

A total of eight studies assessing fatigue were included in the review. A large portion of articles synonymously used other terms such as drowsiness and alertness. Most studies looking at fatigue used real-vehicle simulators (e.g., actual vehicles made into simulators) and manipulated fatigue by using various types of sleep deprived conditions (Jackson et al., 2016; Zhang et al., 2017; Hu et al., 2012; Sommer & Golz, 2010; Ahn et al., 2016). The remaining studies used monotonous driving conditions or controlled driving scenarios such as steady traffic (Chaung et al., 2010; Xu et al., 2018; Wang, Guragain, et al., 2019). One study utilized a train simulator (Zhang et al., 2017), while none of the included studies examined flight.

Six of the included studies used EEG, one used ECG, four used ocular metrics (e.g., eye tracking and/or EOG data), and one used fNIRS. Based on the overall findings of the review, eye tracking data as an operator monitoring technique was used most frequently to assess fatigue compared to all other cognitive states.

Five of the eight studies developed classifiers or algorithms from the collected physiological data. The goal of all classifiers/algorithms was to determine the likelihood of accurately classifying the participants' fatigued state. Classification accuracies were generally good, ranging from 62 % (Hu et al., 2012) to 90.70% (Zhang et al., 2017). The remaining studies examined the relationships between the physiological measures and subjective and performance measures (Jackson et al., 2016; Sommer et al., 2010; Wang et al., 2019a). Notably, Zhang et al. (2017) included the assessment of a wearable EEG device that could potentially be integrated within a working environment.

This space is intentionally blank.

This page is intentionally blank.

Table 6. Fatigue (Includes Drowsiness/Alertness)

Reference	**Task Used to Induce Fatigue**	**Device Description**	***n* (Gender), Age [*range* or *M(SD)*]**	**Physiological Outcome Measures**	**Method of Analysis**	**Outcomes**
Chuang et al. (2010)	Simulated monotonous freeway driving	Motion driving simulator	6 (gender not reported), 19-23 yrs	EEG (*30 independent components consisting of power spectra from occipital area*)	Assessed performance of 5 classifiers in relation to driver reaction time: ML = *Gaussian Maximum Likelihood;* kNN = *k-nearest-neighbor classifier;* SVM = *support vector machine;* BPN = *back propagation neural network;* RBFNN = *Radial basis function neural network*	kNN = highest accuracy with 89.4% +/- 2.7
Jackson et al. (2016)	24-hour sleep deprivation	AusEd fixed-based driving simulator	22 (3M), 18-26 yrs, (*M* = 20.8, *SD* = 1.9)	Ocular metrics *(PERCLOS, % time with eyes closed [%TEC], John's drowsiness scale [JDS])*	Friedman's test (Baseline vs. sleep deprivation) of ocular metrics; Spearman's correlations between ocular, performance and subjective measures	Significant increase in time period of long eyelid closure in the sleep deprivation condition, no difference seen for PERCLOS; % TEC correlated with PVT and crashes in simulator; JDS correlated with subjective fatigue ratings
Zhang et al. (2017)	Unspecified sleep deprivation with testing between 0400-0600 hrs	High speed fixed-based train simulator	10 (7M), 19-42 yrs	EEG (*Developed a drowsiness index using theta, alpha and beta PSD values from sites O1 and O2*)	Evaluated classification accuracy, sensitivity and false positive rate of drowsiness index across two study conditions (drowsiness vs. alertness)	Accuracy: up to 90.70%, Sensitivity: up to 86.80%, False positives: down to 5.40%
Xu et al. (2018)	Highway driving with medium traffic starting at 1400 hrs	reZY-31D fixed-based car simulator	10 (10M), 22.4 yrs	Ocular Metrics *(duration of fixation in AOI, pupil diameter)*	Evaluated classifier using fuzzy *k*-nearest neighbor with average fixation time and pupil area as single features, then combined. Compared performance in a normal and fatigued-states	Mean accuracy of classifier with the combined features (AOI and pupil diameter) was 88.75%

Reference	Task Used to Induce Fatigue	Device Description	*n* (Gender), Age [*range* or *M(SD)*]	Physiological Outcome Measures	Method of Analysis	Outcomes
Hu et al. (2012)	Sleep deprivation for 1 night; completed 45-90 mins of driving	Swedish National Road and Transport Research Institute (VTI) moving based driving simulator (3rd generation)	40, (gender and age not reported)	EEG *(delta, theta, beta, alpha, dominant frequency, average power of dom. Peak, central of gravity frequency, frequency variability, mean power frequency)*	Evaluated classifiers that included 75 EEG features to predict fatigued state	The 75-features model had sensitivity of 62%, and specificity of 74% in detecting fatigue
Sommer & Golz (2010)	Drove overnight (2330-0830 hrs) for 40 min segments	Real car simulated driving	16, (gender and age not reported)	EEG *(PSD values ranging from 1-23 Hz from sites FP1, FP2, C3, Cz, C4, O1, O2, A1, A2)* EOG *(PERCLOS)*	Correlations between KSS scores with PERCLOS, lane deviations and PERCLOS, and PERCLOS vs. EEG/EOG for KSS via discriminate analysis	Significant correlations and EEG/EOG better predicted mild or strong fatigue than PERCLOS
Wang, Guragain, et al. (2019)	Man-machine intervention while auto-driving vs. no intervention; auto-driving manipulation was to induce fatigue	JT/T378 fixed-based vehicle simulator	12 (10M), 31 yrs (*SD* = 1.6)	EEG *(beta, alpha, theta from 14 total sites within frontal, central, and posterior regions)*	*t*-test between ratio of beta/(theta + alpha) during normal driving and man-machine response mode (MRM)	There was a smaller downward trend of the beta ratio in the MRM condition compared to non-MRM, suggesting the MRM intervention reduced onset of fatigue
Ahn et al. (2016)	Well-rested condition with 7+ hours of sleep vs. sleep deprived condition where participants stay up all night	Fixed-based driving simulator	11 (10M), age (*M* = 26.6, SD = *1.4*)	EEG (*PSD values for delta, theta, alpha, beta, gamma; ratio of beta to alpha)* EOG *(rate of eye blinks)* fNIRS *(amplitudes of HbO, HbR)* ECG *(QRS-complex, RR-peal intervals)*	Developed algorithm from EEG, ECG and fNIRS features to classify well-rested vs. sleep-deprived conditions *t*-test between rested and fatigued condition had significant differences for alpha and beta	Found highest classification accuracies for combination of EEG, ECG, and fNIRS. No difference seen for EOG.

Note. ECG = electrocardiography, EEG = electroencephalography, GSR = Galvanic skin response, fNIRS/NIRS = functional , PSD = power spectral density, PVT = psychomotor vigilance task, HR = heart rate, HRV = heart rate variability, LF = Low frequency, HF = High Frequency, PERCLOS = percentage eyes closed, RMSSD = root mean square of successive differences, SDNN = standard deviation of normal-to-normal intervals, HbO = oxy-hemoglobin concentration, HbR = deoxy-hemoglobin concentration, AOI = area of interest

Inattention/distraction/vigilance

Four studies assessing inattention or synonymously distraction and/or vigilance were included based on eligibility. Attention was typically manipulated by the interjection of intermittent tasks such as texting, or flight maneuver instructions, or with changes in typical traffic patterns (Dehzangi & Taherisadr, 2018; Harrivel et al., 2016). The most common physiological metric/device used within the inattention literature was GSR (two studies) and EEG (two studies). Respiration was also used by Harrivel et al. (2016), whereas Rahman et al. (2021) only used fNIRS.

Three of the studies developed and evaluated the performance of classifiers in predicting inattention. Performance ranged from 43% to 95% (Wang, Li, et al., 2019) for an EEG-only classifier. Other classifiers included one using GSR (maximum accuracy of 93%; Dehzangi et al., 2018), and a combined EEG and GSR model (89% accuracy; Harrivel et al., 2016). Rahman et al. (2021) found differences in oxy-hemoglobin concentration (HbO_2) based on task conditions, suggesting a relationship between HbO_2 and inattention/vigilance performance. Table 7 below summarizes inattention studies.

This space is intentionally blank.

This page is intentionally blank.

Table 7. Inattention/Distraction/Vigilance

Reference	Task Used to Induce Inattention	Device Description	***n*** **(Gender), Age [*range* or *M(SD)*]**	Physiological Outcome Measures	Method of Analysis	Outcomes
Dehzangi & Taherisadr (2018)	Participants texted while driving during scenarios with 2 minute durations	Real vehicle	7 (7M), 20-40 yrs	Galvanic Skin Responses (GSR) (*used convolutional neural networks [CNN] to identify features for inclusion in classifier)*	Evaluated performance of developed classifier (*note*, it is unclear what features were extracted and used in the classifier)	Achieved maximum prediction accuracies of 93.28% in detecting inattention
Harrivel et al. (2016)	Flight scenarios in fixed-based simulator with benchmark tasks	Fixed-based flight simulator	12 (11M), age not reported	EEG *(8 mono-polar EEG signals)* ECG *(HRV)* GSR *(skin conductance)* Respiration	EEG, ECG, GSR, and resp. were used to generate normalized classifier input features	Model with EEG + GSR = 89% accuracy, Model with EEG only = 82%. Both models were reliable for distinguishing lower workload and channelized attention benchmark states
Rahman et al. (2021)	Completed cognitive tasks with or without self-paced walking	No device used	19 (6M), 18-35 yrs (*M* = 21.5, *SD = 3.6*)	fNIRS *(ΔHbO_2, ΔHbR)*	RM ANOVA for ΔHbO_2 and ΔHbR across single motor condition, single cognitive condition, and dual condition	Significant differences in ΔHbO_2 where dual task was greater than single motor in right hemisphere.
Wang , Li, et al. (2019)	Construction-based vigilance task of moving two metal tubes from one location to another with obstacles presented	No device used	10 (10M), age not reported	EEG (*PSD values from 14 frontal, temporal, parietal, occipital sites*)	Evaluated 30 EEG vigilance indicators (e.g., various ratios of PSD values) by using NASA-TLX as ground truth measure and EEG-vigilance stage model as a benchmark; correlation analyses were used to evaluate performance of vigilance indicators	Three indices resulted in the highest correlation coefficients for Index 6 (θ/β), Index 19 ($(\theta+\beta)/(\alpha+y)$), and Index 26 ($\alpha/(\beta+y)$); coefficients ranged from 0.43 to 0.95

Note. ECG = electrocardiography, EEG = electroencephalography, GSR = Galvanic skin response, fNIRS/NIRS = functional , PSD = power spectral density, HR = heart rate, HRV = heart rate variability, LF = Low frequency, HF = High Frequency, PERCLOS = percentage eyes closed, RMSSD = root mean square of successive differences, SDNN = standard deviation of NN intervals, HbO = Oxyhemoglobin , HbR = Deoxyhemoglobin

This page is intentionally blank.

Stress/Overload

The three included articles for stress/overload assessed conditions through high-stress scenarios in order to manipulate the cognitive state. For example, in both of their articles, Tornero-Aguilera et al. (2017; 2018) use pre- and post-measures of physiological metrics for combat Soldiers during a mock rescue mission. These physiological metrics included ECG/heart rate measures, blood oxygenation saturation, blood glucose levels, blood lactate, and assessed critical flicker fusion threshold (CFFT) (Tornero-Aguilera et al., 2017; Tornero-Aguilera et al., 2018; Sanchez-Molina et al., 2018). Locating literature proved somewhat difficult compared to the other included cognitive states possibly due to stark differences in terminology used for the construct of stress. For example, the terms *overload*, *arousal*, and *high workload stress* were in studies that did not accurately measure stress. The three included studies were the only ones located that specifically used stress as a state of interest.

Of the literature assessing stress/overload cognitive states, no known study within the timeframe of this review has used vehicle operator state monitoring, either real-life or via simulation. Nor did any of the included studies used other common psychophysiological equipment such as EEG or eye tracking. All three studies compared physiological measures pre- and post-missions. An increase in heart rate variables, as well as blood lactate levels, were consistently demonstrated for post-physiological measures (Tornero-Aguilera et al., 2017; Tornero-Aguilera et al., 2018; Sanchez-Molina et al., 2018).

This space is intentionally blank.

This page is intentionally blank.

Table 8. Stress/Overload

Reference	Task Used to Induce Stress	Device Description	***n*** **(Gender), Age [*range* or *M(SD)*]**	Physiological outcome Measures	Method of Analysis	Outcomes
Tornero-Aguilera et al. (2017)	Pre- and Post-combat simulation rescue mission	No device used; combat simulation performed in an urban area with one-floor buildings	40 (gender not reported) Elite = 28.5 yrs (*SD* = 6.38), non-Elite = 31.94 yrs (*SD* = 6.24)	ECG *(HR, HRV, RMSSD, LF, HF)* *Blood oxygen saturation, blood glucose, blood lactate, lower body muscular strength*	Compared pre- and post-measures on Soldiers during combat mission with respective bivariate correlations	HR variables, lower body muscular strength, and blood lactate increased after simulation in both elite and non-elite groups. Pre skin temperature was positively correlated with lower muscular strength and negatively correlated with pre-HR
Tornero-Aguilera et al. (2018)	Pre- and Post-combat simulation rescue mission for highly trained vs. lower trained groups.	No device used; simulated combat mission	49 (gender not reported), 30.6 yrs (*SD* = 5.7)	ECG (*HR, HRV, RMSSD, LF, HF*), blood oxygen saturation, blood glucose, blood lactate, lower body muscular strength, body temperature, cortical arousal (CFFT)	Compared pre- and post-measures on Soldiers during combat mission with respective bivariate correlations	Variables of HR, lower body muscular strength, blood lactate and glucose increased after combat simulation. Highly trained group only showed significant increase in RMSSD, LF after simulation and a decrease in HF compared to the lower trained group.
Sanchez-Molina et al. (2018)	Four-person simulated combat training rescue mission	No device, simulated combat situation	19 (19M), 30.1 yrs *(SD* = 5.25*)*	ECG (*HR, HRV, RMSSD, LF, HF*), blood oxygen saturation, blood glucose, blood lactate, critical flicker frequency	Pre- and post- analysis before and after simulated combat situation	HR and blood lactate significantly increased after combat simulation, LF increased while HD and RMSSD decreased.

Note. ECG = electrocardiography, EEG = electroencephalography, GSR = Galvanic skin response, fNIRS/NIRS = functional , PSD = power spectral density, HR = heart rate, HRV = heart rate variability, LF = Low frequency, HF = High Frequency, PERCLOS = percentage eyes closed, RMSSD = root mean square of successive differences, SDNN = standard deviation of NN intervals, HbO = Oxyhemoglobin , HbR = Deoxyhemoglobin

This page is intentionally blank.

Hypoxia

Only one study was identified where hypoxia was the cognitive state of interest. Rice et al. (2019) used EEG, HR and SpO_2 to evaluate differences between altitude conditions. They found that all three measures significantly differed between altitude conditions under assessment.

This space is intentionally blank.

This page is intentionally blank.

Table 9. Hypoxia

Reference	**Task Used to Induce Hypoxia**	**Device Description**	***n* (Gender), Age [*range* or *M(SD)*]**	**Physiological outcome Measures**	**Method of Analysis**	**Outcomes**
Rice et al. (2019)	Oxygen manipulations through ROBD	Desktop X-Plane v 10.5 simulator	60(30M), Female age (*M* = 24, *SD* = 2.5), Male age (*M* = 23.8, *SD* = 1.7)	EEG (*alpha, beta, gamma theta from six frontal, central, and parietal sites*) Heart Rate SPO_2	EEG - RM ANOVAs between electrode channels and altitude conditions (sea level, 25,000 feet, 20,000 feet); HR and SpO_2 – RM ANOVAs between altitude conditions	Alpha, beta, gamma, and theta were all significant across all channels between acute and insidious altitude conditions; similar patterns for HR and SpO_2

Note. ROBD = reduced oxygen breathing device

This space is intentionally blank.

This page is intentionally blank.

Discussion

The objective of this systematic review was to identify types of psychophysiological metrics that most reliably detect changes to operator state and use that information to determine sensor types to pursue for evaluation within USAARL's operator state monitoring research program. A variety of different literature was included such as proceedings papers, technical reports, conference papers, and peer-reviewed articles. Of the literature included, the vast majority of studies used aircraft simulators and real-vehicle car simulators. To the best of our knowledge, only one known article used real in-flight data collection and no known studies used operator state monitoring with psychophysiological metrics during a real-life driving scenario. The number of articles using real vehicles, such as aircraft, is most likely limited potentially due to costs and logistics of conducting such experiments, as well as potential safety concerns. The current review found only two articles using a real-life vehicle based on exclusion criteria. Wilson et al. (2021) used in-flight state monitoring via EEG during flight maneuvers requiring varying workload demands. In-flight state was classified into high-workload and low-workload based on the EEG indices and data was collect for a total of 10 aviation students (Wilson et al., 2021). Schöniger et al. (2020) conducted a series of real-life air rescue missions via a German emergency medical service helicopter, though workload was not investigated via pilot performance, but for the emergency physicians conducting operations.

In terms of measured cognitive states, most studies looked at workload followed by aspects of fatigue/drowsiness. The most commonly reported physiological metrics across all cognitive states were ECG and EEG, followed by eye tracking data. Studies looking at workload used a mixture of within- and between-group comparisons on flight maneuvers and tasks such as the *n*-back task, and a total of 2 studies used algorithm development for classifier performances during flight maneuvers that ranged from 45.91-100% (Blonco et al., 2018; Wilson et al., 2021). The general pattern of changes in physiological measures within the workload studies included increases in physiological metrics like heart rate, blink rate, EDA, during tasks/scenarios that required high workload conditions (Feng et al., 2018; Unni et al., 2017; Yang et al., 2021, etc.).

Studies looking at fatigue and/or drowsiness typically used EEG and eye tracking data. As a result, a large portion of the studies looked at algorithm development to predict a drowsiness state with classification accuracies ranging from 89.4% to 95% (Chaung et al., 2010; Yin et al., 2016; Hu et al., 2013; Zhang et al., 2017). In order to manipulate fatigue conditions, some studies included overnight driving, and some had prolonged periods of sleep deprivation of up to 24 hours (Sommer et al., 2014; Jackson et al., 2016; Ahn et al., 2016). Jackson et al. (2016) did not find any differences in PERCLOS between sleep deprivation conditions but did see a significant increase in the time of long eyelid closure in their sleep deprivation condition (24 hours). Sommer et al. (2014) found that their EEG and EOG results were better predictors of mild or strong fatigue than PERCLOS. Ahn et al.'s (2016) sleep deprived condition of participants staying up throughout the night did not show any differences compared to the well-rest condition for rates of blinking, but did find the combination of EEG, ECG, and fNIRS to provide the highest classification accuracies. They also found significant differences in alpha and beta between rested and fatigue conditions (Ahn et al., 2016).

Conclusion

Based on the findings from the literature review, workload, fatigue, and inattention all show promise for detection through physiological metrics. Due to the few number of studies looking at stress and hypoxia, further research is needed to conclude the usefulness of physiological measures for detecting these states on operator cognition and performance. This lack of research may in part be due to differences in how researchers interpret physiological states vs cognitive states. EEG was the most commonly used physiological measure and was especially successful for detecting workload and fatigue. In addition, multiples aspects of eye-tracking data (i.e., PERCLOS, blink rate) proved useful for fatigue detection. Though literature on stress was scarce, heart rate metrics showed promise. To the best of our knowledge, only one study used fNIR data, demonstrating a large gap in the literature that should be pursued for future investigations in the detection of cognitive states.

Additionally, this review highlighted some current limitations in the field. For example, to the best of our knowledge, only one study used a real-time monitoring while operating an actual vehicle (e.g., Wilson et al., 2021). The cost and logistical concerns of using in-flight monitoring or driving most likely contribute to the lack of studies available. It is our recommendation that studies shift to using real-vehicle studies. The current literature has a large number of studies that use fixed-based simulators with very few that use motion-based simulators. Keeping cost and other logistical concerns in mind regarding the use of real-vehicle operator monitoring, a next logical step for the field would be to increase the number of studies conducted with motion-based simulators to increase the generalizability of findings.

Lastly, the lack of consistency in cognitive state terminology may be problematic for the field of operator state monitoring moving forward. For example, some studies use the term *fatigue* while others use *drowsiness*. However, possibly the least concerning recommendation from this review, consistent use of terminology may contribute to more uniformity in testing of these cognitive states and interpretation of results. Explanations of differences between constructs can contribute to differences in theoretical understandings as well as execution of testing these cognitive constructs. In their review, Weelden et al. (2022) expressed similar concerns regarding heterogeneity of terminology in the aviation and neurophysiology field. A highlighted example of inconsistent terminology was workload with phrases such as engagement or arousal (Weelden et al., 2022). The operator state monitoring literature encompasses many disciplines of study including individuals from psychology, human factors, and neuroscience. More uniformity in terminology may be helpful for an area of research conducted by such a diverse population of investigators.

References

Ahn, S., Nguyen, T., Jang, H., Kim, J. G., & Jun, S. C. (2016). Exploring neuro-physiological correlates of driver's mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and fNIRS data. *Frontiers in Human Neuroscience, 10*, 219. https://doi.org/10.3389/fnhum.2016.00219

Arico, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., & Babiloni., F. (2016). A passive brain-computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. *Progress in Brain Research, 228*. http://dx.doi.org/10.1016/bs.pbr.2016.04.021

Aura, C.J., Temme, L.A., St. Onge, P., McAtee, A., Wilson, M., Bowers, B., Baugher, K., Hayes, A. (2021) Operator state monitoring via pupillometry: measuring mental workload under varying luminance conditions. *Proc. SPIE 11757, Smart Biomedical and Physiological Sensor Technology XVIII, 117570M (12 April 2021).* https://doi.dx.org10.1117/12.2585989

Bernhardt, K. A., Poltavski, D., Petros, T., Ferraro, F. R., Jorgenson, T., Carlson, C., Drechsel, P., & Iseminger, C. (2019). The effects of dynamic workload and experience on commercially available EEG cognitive state metrics in a high-fidelity air traffic control environment. *Applied Ergonomics, 77,* 83–91. https://doi.org/10.1016/j.apergo.2019.01.008

Blonco, J. A., Johnson, M. K., Jaquess, K. J., Oh, H., Lo, L., Gentili, R. J., & Hatfield, B. D. (2018). Quantifying cognitive workload in simulated flight using passive, dry EEG measurements. *IEEE Transactions on Cognitive and Developmental Systems, 10*(2). https://doi.org/10.1109/TCDS.2016.2628702

Charles, R. L., & Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics, 74*, 221–232.

Chaung, C., Lai, P., Ko, L., Kuo, B., & Lin, C. (2010). Driver's cognitive state classification toward brain computer interface via using a generalized and supervised technology. (2010). *The 2010 International Joint Conference on Neural Networks (IJCNN)*, 1–7, https://doi.org/10.1109/IJCNN.2010.5596835.

Dehzangi, O., & Taherisadr, M. (2018). Driver distraction detection using MEL cepstrum representation of galvanic skin responses convolutional neural networks. *The 24th International Conference on Patter Recognition (ICPR*).

Feltman, K. A., Bernhardt, K. A., & Kelley, A. M. (2021). Measuring the domain specificity of workload using EEG: Auditory and visual domains in rotary-wing simulated flight. *Human Factors, 63*(7), 1271–1283. DOI: 10. 1177/ 0018 7208 20928626

Feltman, K. A., Basso, J., Bernhardt, K. A., Hayes, A., Colby, M., Chiaramonte, J., King, M., Kelley, A. M., Erickson, B., & Curry, I. (2019). *Individual differences in aviation performance and psychophysiological indices during early morning simulate flight* (Report no. 2019-06). United States Army Aeromedical Research Laboratory.

Feng, C., Wanya, X., Yang, K., Zhuang, D., & Wu, X. (2018). A comprehensive prediction and evaluation method of pilot workload. *Technology and Health Care, 26,* S65–S78. DOI 10.3233/THC-174201

Harrivel, A. R. Harrivel, C. L., Stephens, C. L., Ellis, K. K., Prinzel, L. J. &. Pope, A. T. (2016). Psychophysiological sensing and state classification for attention management in commercial aviation. AIAA 2016-1490. *AIAA Infotech @ Aerospace*.

Heikoop, D. D., de Winter, J. C. F., van Arem, B., & Stanton, N. A. (2017). Effects of platooning on signal-detection performance, workload, and stress: A driving simulator study. *Applied Ergonomics, 60,* 116–127. http://dx.doi.org/10.1016/j.apergo.2016.10.016

Hidalgo-Muñoz, A. R., Béquet, A. J., Astier-Juvenon, M., Pépin, G., Fort, A., Jallais, C., Tattegrain, H., & Gabaude C. (2019). Respiration and heart rate modulation due to competing cognitive tasks while driving. *Frontiers in Human Neuroscience, 12*(525). https://doi.org/10.3389/fnhum.2018.00525

Hu, S., Zheng, G., & Peters, B. (2012). Driver fatigue detection from electroencephalogram spectrum after electrooculography artefact removal. *IET Intelligent Transportation Systems, 7*(1), 105–113. https://doi.org/10.1049/iet-its.2012.0045

Huang, K. C., Huang, T. Y., Chuang, C. H., King, J. T., Wang, Y. K., Lin, C. T., & Jung, T. P. (2016). An EEG-based fatigue detection and mitigation system. *International Journal of Neural Systems, 26*(4). https//doi.org/10.1142/S0129065716500180

Jackson, M. L., Kennedy, G. A., Clarke, C., Gullo, M., Swann, P., Downey, L. A., Hayley, A. C., Pierce, R. J., & Howard, M. E. (2016). The utility of automated measures of ocular metrics for detecting driver drowsiness during extended wakefulness. *Accident Analysis and Prevention, 87,* 127–133. http://dx.doi.org/10.1016/j.aap.2015.11.033

Kelley, A., Hayes, A., Mathews, C., & Bernhardt, K. (2020). Sustaining performance and vigilance during extended UAS operations. (USAARL Report No. 2020-43).

Liu, C., Wanyan, X., Xiao, X., Zhao, J., & Duan. Y. (2020). Pilots' mental workload prediction based on timeline analysis. *Technology and Health Care, 28,* S207–S216. https://doi.dx.org/10.3233/THC-209021

Pagnotta, M., Jacobs, D. M., de Frutos, P. L., Rodríguez, R., Ibáñez-Gijón, J., & Travieso, D. (2022). Task difficulty and physiological measures of mental workload in air traffic control: A scoping review. *Ergonomics, 65*(8). https://doi.org/10.1080/00140139.2021.2016998

Raduntz, T., Furstenau, N., Muhlhausen, T., & Meffert, B. (2020). Indexing mental workload during simulated air traffic control tasks by means of dual frequency head maps. *Frontiers in Physiology, 11,* 1 – 13. https://doi.org/10.3389/fphys.2020.00300

Rahman, T. T., Polskaia, N., St-Amant, G., Salzman, T., Vallejo, D. T., Lajoie, Y., & Fraser, S. A. (2021). An fNIRS investigation of discrete and continuous cognitive demands during dual-task walking in young adults. *Frontiers in Human Neuroscience, 15.* https://doi.org/10.3389/fnhum.2021.711054

Rice, G. M., Snider, D., Drollinger, S., Greil, C., Bogni, F., Phillips, J., Raj, A., Marco, K., & Linnville, S. (2019). Dry-EEG manifestations of acute and insidious hypoxia during simulated flight. *Aerospace Medicine and Human Performance, 90*(2), 92–100. https://doi.org/10.3357/AMHP.5228.2019

Sanchez-Molina, J., Robles-Perez, J. J., & Clemente-Suarez, V. J. (2018). Assessment of psychophysiological response and specific fine motor skills in combat units. *Journal of Medical Systems, 42(67),* 1–8. https://doi.org/10.1007/s10916-018-0922-9

Schöngler, C., Braun, D., Sleepman, M., Petrowski, K. (2020). Comparison of the HRV of emergency physicians in the HEMS during helicopter operations: Analysis of differences as a function of number of operations and workload. *Applied Psychophysiology and Biofeedback, 45,* 249-257. https://doi.org/10.1007/s10484-020-09480-1

Sommer, D., & Golz, M. (2014). Evaluation of PERCLOS based current fatigue monitoring technologies. *Annual International Conference of the IEEE Engineering in Medicine & Biology Society,* 4456-4459. https://doi.dx.org/10.1109/IEMBS.2010.5625960

Stuiver, A., Brookhuis, K. A., de Waard, D., & Mulder, B. (2014). Short-term cardiovascular measures for driver support: Increasing sensitivity for detecting changes in mental workload. *International Journal of Psychophysiology, 92,* 35–41. http://dx.doi.org/10.1016/j.ijpsycho.2014.01.010

Tarabay, R., & Abou-Zeid, M. (2018). Assessing the effects of auditory-vocal distraction on driving performance and physiological measures using a driving simulator. *Transportation Research Part F, 58,* 351–364. https://doi.org/10.1016/j.trf.2018.06.026

Tao, D., Tan, H., Wang, H., Zhang, X., Qu, X., & Zhang, T. (2019). A systematic review of physiological measures of workload. *International Journal of Environmental Research and Public Health, 16*(15), 2716. https://doi.org/10.3390/ijerph16152716

Temme, L., St. Onge, P., & Bleiberg, J. (2016). A history of mild traumatic brain injury affects peripheral pulse oximetry during normobaric hypoxia. *Frontiers in Neurology.* https://doi.org/10.3389/fneur.2016.00149

Torneo-Aguilera, J. K., Robles-Perez, J. J., & Clemente-Suarez, V. J. (2017). Effect of combat stress in the psychophysiological response of elite and non-elite soldiers. *Journal of Medical System, 41. 2–6.* https://doi.org/10.1007/s10916-017-0748-x

Torneo-Aguilera, J. K., Robles-Perez, J. J., & Clemente-Suarez, V. J. (2018). Use of psychophysical portable devices to analyze stress response in different experienced Soldiers. *Journal of Medical System, 42(75), 2–9.* https://doi.org/10.1007/s10916-018-0929-2

Unni, A., Ihme, K., Jipp, M., & Rieger, J. W. (2017). Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study. *Frontiers in Human Neuroscience, 11*(167). https://doi.org/10.3389/fnhum.2017.00167

Wang, C., Guragain, B., Verma, A. K., Archer, L., Shubha, M., Mohamud, A., Flaherty-Woods, E., Shapiro, G., Almashor, M., Lenn(e), M., Myers, R., Kuo, J., Yang, S., Wilson, N., & Tavakolian, K. (2019). Spectral analysis of EEG during microsleep events annotated via driver monitoring system to characterize drowsiness. *IEEE Transactions on Aerospace and Electronic Systems, 56(2). 1346–1356.* https://doi.org/10.1109/TAES.2019.2933960

Wang, D., Li, H., & Chen, J. (2019). Detecting and measuring construction workers' vigilance through hybrid kinematic-EEG signals. *Automation in Construction, 100,* 11–23. https://doi.org/10.1016/j.autcon.2018.12.018

Weelden, E., Alimardani, M., Wiltshire, T. J., & Louwerse, M. M. (2022). Aviation and neurophysiology: A systematic review. *Applied Ergonomics, 105*, 103838.

Wilson, N., Gorji, H. T., Van Bree, J., Hoffman, B., Tavakolian, K. & Petros, T. (2021). Identifying opportunities for augmented cognition during live flight scenario: An analysis of pilot mental workload using EEG. *94th International Symposium on Aviation Psychology,* 444–449.

Xu, J., Min, J., & Hu, J. (2018). Real-time eye tracking for the assessment of driver fatigue. *Healthcare Technology Letters, 5*(2), 54–58. https://doi.org/10.1049/htl.2017.0020

Yang, S., Kuo, J., Lenné, M., Fitzharris, M., Horbery, T., Blay, K., Wood, D., Mulvihill, C. & Truche, C. (2021). The impacts of temporal variation and individual differences in driver cognitive workload on ECG-based detection. *Human factors, 63.* https://doi.org/10.1177/0018720821990484.

Yin, J., Hu, J., & Mu, Z. (2016). Developing and evaluating a mobile drive fatigue detection network based on electroencephalography signals. *Healthcare Technology Letters, 4*(1), 34–38. https://doi.org/10.1049/htl.2016.0053

Zhang, X., Li, J., Liu, Y., Zhang, Z., Luo, D., Zhou, X., Zhu, M., Salman, W., Hu, G., & Wang, C. (2017). Design of a fatigue detection system for high-speed trains based on driver vigilance using a wireless wearable EEG. *Sensors, 17,* 2 – 21. https://doi.org/10.3390/s17030486

U.S. Army Aeromedical Research Laboratory
Fort Rucker, Alabama

All of USAARL's science and technical
information documents are available for
download from the
Defense Technical Information Center.

https://discover.dtic.mil/results/?q=USAARL

Army Futures Command
U.S. Army Medical Research and Development Command