

UNITED STATES ARMY AEROMEDICAL RESEARCH LABORATORY



The Feasibility of Operator State Monitoring in Army Aviation: A Survey of Subject Matter Experts

Jordayne C. Wilkins, Katie Feltman, & Christopher Aura

DISTRIBUTION STATEMENT A. Approved for public release; distribution unlimited.

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.

Human Subjects Use Statement

In the conduct of research involving human subjects, the investigator(s) adhered to the policies regarding the protection of human subjects as prescribed by Department of Defense Instruction 3216.02 (Protection of Human Subjects and Adherence to Ethical Standards in DoD-Supported Research) dated 8 November 2011.

IRB Determination and Number

This study, USAARL 2022-009, was approved as a non-research activity by the U.S. Army Aeromedical Research Laboratory's Research Compliance Office on 21 March 2022.

REPORT DOCUMENTATION PAGE					<i>Form Approved</i> <i>OMB No. 0704-0188</i>	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.						
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.						
1. REPORT DATE (DD-MM-YYYY) 13-10-2022		2. REPORT TYPE Technical Report			3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE The Feasibility of Operator State Monitoring in Army Aviation: A Survey of Subject Matter Experts					5a. CONTRACT NUMBER	
					5b. GRANT NUMBER	
					5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Wilkins, J. C. ^{1,2} , Feltman, K. A. ¹ , & Aura, C. 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--2023-01	
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 The U.S. Army's Future Vertical Lift (FVL) program is developing a new fleet of aircraft that is set to replace the UH-60 Black Hawk and the AH-64 Apache. These new aircraft will include autonomous systems and have increased capabilities of speed, maneuverability, and operational range. With this increase in capability, there will be an increased overload in Aviators' cognitive state leading to mishaps and accident fatalities. Operator state monitoring through psychophysiological indices and biomarkers are used to identify cognitive states. Subject matter experts completed two rounds of a Delphi-style survey to determine the reliability of selected sensors and biomarkers when incorporated in aviator operational environments.						
15. SUBJECT TERMS Delphi, survey, future vertical lift, operator state monitoring						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 85	19a. NAME OF RESPONSIBLE PERSON Loraine St. Onge, PhD	
a. REPORT UNCLAS	b. ABSTRACT UNCLAS	c. THIS PAGE UNCLAS			19b. TELEPHONE NUMBER (Include area code) 334-255-6906	

This page is intentionally blank.

Executive Summary

A Delphi-style survey was completed to gain the perspective of subject matter experts regarding the current state of operator state monitoring and likelihood of integration within an operational, aviation environment. Two rounds of surveys were completed. The findings suggest that operator state detection from physiological measures is believed to be feasible by respondents. Specifically, the following three operator states were identified as most likely to be detectable in real-time using physiological measures: fatigue / sleepiness, mental workload, and hypoxia. Additionally, physiological measures perceived as most likely capable of detecting these operator states, and others, include electroencephalogram, eye tracking, and pupillometry. However, presently, none of these measures are ready for integration within the operational aviation environment. Further research is needed to determine whether sensors can be developed that are able to withstand the unique environmental effects within such an environment (e.g., vibrations, bright light exposure). Ongoing research at the U.S. Army Aeromedical Research Laboratory aims to further evaluate some of these factors.

Acknowledgements

The authors would like to thank Mr. Andrew Alvarado and Mr. Travis Zlateff, Information Technology Systems and Management at the U.S. Army Aeromedical Research Laboratory, for their assistance with maintaining the online survey.

This research was supported in part by an appointment to the Biomedical Sciences and Engineering Research Intern 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

	Page
Executive Summary	iii
Acknowledgements.....	iv
Introduction.....	1
Methods.....	2
Survey Participants	2
Procedure	4
Results.....	5
Round One Results	5
Operator states.	6
Device instrumentation.	14
Compatibility with environment.	19
Sensitivity	23
Specificity.	25
Physical characteristics.	26
Round Two Results.....	27
Sensitivity.	27
Operator states.	27
Physiological measures.	28
Discussion	29
Operator States and State Detection	30
Instrumentation.	30
Environment.....	30
Sensitivity and specificity.	31
Limitations	31
Conclusions.....	32
References.....	33
Appendix A. Acronyms and Abbreviations.....	34
Appendix B. Initial Survey	35
Demographics Items	35
Operator States Items.....	39
Device Set Up Items	48
Compatibility with the Environment Items.....	52
Signal Quality and Reliability Items.....	55
Device Measurements Item.....	57
Appendix C. Second Survey	58
Appendix D. Additional Comments	68
Signal Sensitivity – Additional Comments.....	68
Signal Specificity – Additional Comments.....	68
Physical Characteristics - Participant Comments	70
Appendix E. Round 2 Survey	73

Table of Contents (continued)

List of Figures

	Page
1. Current state of the technology – ratings for ease of application.	14
2. Anticipated 5-year state of the technology – ratings for ease of application.	15
3. Ratings for ease of maintenance.	16

List of Tables

1. Survey Participant Degrees	3
2. Experience with Psychophysiological Measures	3
3. Experience with Operator States	4
4. Likelihood of Each Operator State as a Candidate for Detection in Real-Time	6
5. Mental Workload	7
6. Fatigue / Sleepiness	8
7. Distraction	9
8. Mental Fatigue	10
9. Hypoxia	11
10. Spatial Disorientation	12
11. Autonomic Tone	13
12. Calibration Estimates for Physiological Measures	17
13. Likelihood of Withstanding Vibrations	19
14. Likelihood of Withstanding Extreme Temperatures	20
15. Likelihood of Withstanding Bright Environments and Direct Sunlight	21
16. Likelihood of Maintaining Adhesion	22
17. Signal Sensitivity	24
18. Signal Specificity	25
19. Themes of Physical Characteristics	26
D1. Physical Characteristics Respondent Comments	70
E1. Summary of Operator State Responses	75
E2. Summary of Physiological Measures Responses	77
E4. Summary of Physiological Measures for Fatigue / Sleepiness Responses	79
E5. Summary of Physiological Measures for Distraction Responses	80
E6. Summary of Physiological Measures for Mental Fatigue Responses	81
E7. Summary of Physiological Measures for Hypoxia Responses	82
E8. Summary of Physiological Measures for Spatial Disorientation Responses	83
E9. Summary of Physiological Measures for Autonomic Tone Responses	84

Introduction

The U.S. Army's new fleet of aircraft are in development under the Future Vertical Lift (FVL) program. The FVL is set to replace the UH-60 Black Hawk and the AH-64 Apache with the Future Long Range Assault Aircraft (FLRAA) and the Future Attack Reconnaissance Aircraft (FARA). These new aircraft will have increased speed, maneuverability, and operational range, while supporting autonomous systems and advanced sensors. With this increase in the aircraft's capabilities, there is potential for a significant increase in complexity and an increased risk of overloading aviators operating the aircraft. In an overloaded cognitive state, mission critical or flight critical tasks may be inadvertently shed, leading to serious mistakes and deadly accidents (e.g., Caldwell, 2005; Matthews & Desmond, 2002; Saxby et al., 2013). A degraded visual environment (DVE) is known to increase cognitive workload levels subsequently increasing the likelihood of accidents (Stanton et al., 2019). Pilots flying in environmental conditions such as brown-outs, dust, wind, and meteorological conditions were reported to have an increase in accidents due to spatial disorientation (SD). For the fiscal year (FY) 2022 to FY 2011 period, of 100 rotary-wing flight mishaps caused by SD, 22% involved fatalities and 39% involved fatalities and/or injuries (Gaydos et al., 2013). Soldiers also demonstrate significantly diminished cognitive abilities while maintaining vigilance during the monitoring of unmanned aerial systems (UAS) for extended periods of time (Gunn et al., 2005; Kelley et al., 2020). An intervention method that can monitor aviator's cognitive state in real-time during flight could assist in mitigating these outcomes. Operator state monitoring (OSM) through psychophysiological indices has been introduced as a solution.

OSM, using psychophysiological indices and biomarkers to identify cognitive states, will need to be compatible with the ever-evolving demands of operational environments. Incorporating wearable sensors into a wide range of military and civilian operational settings and thus, applying these technologies to the development of OSM sensor suites compatible with military aviation operational settings and scenarios is propitious. As such, there are a few key steps that must be accomplished to meet this goal. These include guaranteeing the reliability and validity of selected sensors, combating issues in relation to end-user compliance, and assessing the ability of sensors to withstand military operational demands. The U.S. Army Aeromedical Research Laboratory (USAARL) has ongoing studies aimed at evaluating the suitability of selected sensors and biomarkers. USAARL researchers endeavor to identify the requirements that OSM systems will need to meet, by surveying a diverse cohort of experts drawn from inside the Department of Defense, and from academic researchers, as well as commercial vendors.

The purpose of the present study was to survey experts within the fields of operator state detection and wearable sensor technologies in order to:

- Identify psychophysiological metrics likely to reliably detect operator state changes;
- Characterize the current state of physiological sensors regarding time required for application and compatibility with the aviation environment;
- Determine the criteria sensors need to meet for sensitivity and specificity; and
- Define any physical characteristics that may impact utility of sensors.

To accomplish these goals, the Delphi method for surveying experts was used. The Delphi method is a process used for anonymously querying a group. The overall purpose of using the Delphi method is generally to facilitate the ability to reach a group consensus on a particular complex topic (Belton et al., 2019). This method uses an iterative survey approach, whereby participants complete an initial survey, are then given feedback on the group's responses, and additional questions are posed in an attempt to reach group consensus. In this expert survey, two iterations were completed.

Methods

The survey was determined to be a non-research activity by the U.S. Army Aeromedical Research Laboratory's determination official.

Survey Participants

The aim of the survey was to query experts across a broad range of relevant research fields. However, there is no clear definition of what constitutes an expert. Hopkins and Unger (2017) defined subject matter expert (SME) as someone who "has specialist skills, knowledge, or experience in a field." Thus, to quantify expertise in our study, we set criteria that the identified experts must meet. An individual was considered an SME for the purpose of this survey if he/she met one of the following criteria:

- At least 10 years of experience working within the field;
- At least five peer-reviewed publications on a relevant topic;
- A patent for relevant technology; and/or
- Recommended for other reasons not listed.

The fourth criteria was included to extend the survey invitation to individuals who were recommended by the identified SMEs, such as a post-doctoral fellow who may not have met the established criteria, but who is extensively working within the field.

An initial invitation to participate in the survey process was emailed to each individual SME (41 total). Following these invitations, additional participants were suggested by the invitees. In order to increase the sample size, these SME recommended individuals who were added to the list of invitees, and subsequently sent invitations. The final count for invited participants was sixty. Of the 60 invited, 44 agreed to participate. Twenty-four individuals completed the initial survey. Of these 24 individuals, 17 held a Doctor of Philosophy, one a Medical Doctor degree, two Bachelor's degrees, and two Master's degrees. Two of the participants did not indicate which degrees they hold. Table 1 below summarizes the degree fields. Degree dates ranged from 1981 to 2021. Participants consisted of individuals from academia ($n = 4$), industry ($n = 2$), and Department of Defense research laboratories ($n = 18$).

This space is intentionally blank.

Table 1. Survey Participant Degrees

Degree Fields	<i>n</i>
Human Factors	3
Cognitive Psychology or Experimental Psychology	6
Engineering (Electrical, industrial, biomedical)	6
Medicine	1
Vision Science / Neuroscience	1
Cognitive Neuroscience	2
Mathematics	1
Physiology	3
Communicative Disorders	1
Physical Chemistry	1

All participants had experience in using various psychophysiological measures for data collection. Experience with the types of measures are summarized in Table 2 below.

Table 2. Experience with Psychophysiological Measures

Measure	<i>n</i>	Years Experience (Range)	Years Experience (<i>M</i> (<i>SD</i>))
Actigraphy	13	1 – 20	7.85 (7.52)
Body/face Temperature	10	2 – 20	8.1 (7.34)
ECG Heart Rate (HR)	19	0.5 – 35	9.97 (9.11)
ECG Heart Rate Variability (HRV)	18	0.5 – 20	7.47 (5.64)
Electroencephalography	14	0.5 – 24	8.68 (6.63)
Electromyography	6	2 – 16	7 (6.42)
Electrooculogram	7	5 – 24	9.57 (6.80)
Excreted Metabolites	9	1 – 20	6.44 (5.98)
Eye Tracking	14	2 – 20	9.36 (5.76)
fNIRS	8	3 – 10	4.88 (2.36)
Photoplethysmography	9	3 – 15	6.78 (3.87)
Pupillometry	11	3 – 15	8.36 (4.11)
Respiration Dynamics	11	1 – 30	8.91 (10.10)

Note. ECG = electrocardiogram; fNIRS = functional near-infrared spectroscopy

In addition to the measures that were listed within the survey, participants reported on other measures they held experience with. These included:

- Breath dynamics / respiratory output (*n* = 1) (8 years)
- Continuous blood pressure / cardiac output (*n* = 1) (approx. 10 years)
- Electrodermal activity (*n* = 3) (1 participant reported 6 years; remaining did not report)
- Functional magnetic resonance imaging (fMRI) (*n* = 1) (experience not reported)
- Head tracking / hand motions (*n* = 1) (experience not reported)
- Movement (i.e., via worn accelerometer), speech (*n* = 1) (experience not reported)
- Power meters and ergonomics in cyclists (*n* = 1) (22 years)
- End-tidal gas measurements (*n* = 1) (over 30 years)

Participants also reported on the operator states they have experience working with. These are summarized in Table 3 below.

Table 3. Experience with Operator States

Operator State	N
Autonomic Tone	6
Mental Workload	18
Fatigue / Sleepiness	16
Distraction	8
Mental Fatigue	13
Hypoxia	8
Spatial Disorientation	2

The states listed within the table were prechosen by the research team, as these states are of the greatest interest for detection within Army aviation. Participants were also requested to indicate other states they have experience with. These states included:

- Attention / mental engagement / impulse response ($n = 4$)
- Arousal / stress / emotion ($n = 5$)
- Vigilance / target detection ($n = 3$)
- Decision making / uncertainty ($n = 2$)
- Physiological load / physical fatigue ($n = 4$)
- Hypercapnia ($n = 1$)
- Out of the loop ($n = 1$)
- Readiness ($n = 1$)
- Exploration / exploitation ($n = 1$)
- Communication / perceptual tasks ($n = 1$)

Procedure

An initial survey was developed consisting of 52 items. The items addressed six different topic categories: demographics, operator states, device set up, compatibility with environment, signal quality and reliability, and device measurements. The questions included within the survey are available in the Appendix B. The survey questions were developed by members of the research team and sent to SMEs within the USAARL for feedback and refinement prior to sending out to the participants.

Following responses to the initial survey, a second survey was developed. The second survey included clarification questions following the initial survey. More specifically, the responses to the initial survey included a wide range of responses (ranging from “very unlikely” to “very likely”) on some items. The purpose of the second survey was to request participants to provide justification regarding why they may have responded at one extreme, or the other, to those items. The questions included are available in the Appendix C.

Results

Survey responses were collated for each item. Below, we report the frequencies of responses to each item, within each survey round. Participants were encouraged to only respond to items that they are familiar with. Therefore, the total respondents (n) for each item are reported as well.

Round One Results

Results from the first round of the survey are summarized in the following sections based on category.

This space is intentionally blank.

Operator states.

Amongst the seven operator states presented to participants, three states received high frequencies of “likely” and “very likely” ratings, amongst those who rated each, as being capable for detection in real-time (see Table 4). These included fatigue / sleepiness (95.5%), mental workload (59.1%), and hypoxia (54.6%). Moreover, only four states received “unlikely ratings” (autonomic tone, distraction, mental fatigue, and spatial disorientation), while several received “unsure” ratings (autonomic tone, mental workload, distraction, hypoxia, and spatial disorientation). This suggests further research may be needed to determine the feasibility of some of these additional states for real-time detection.

Table 4. Likelihood of Each Operator State as a Candidate for Detection in Real-Time

Operator State	Very Unlikely n (%)	Unlikely n (%)	Neutral n (%)	Likely n (%)	Very Likely n (%)	Unsure n (%)
Autonomic Tone (n = 21)	-	3 (14.3%)	1 (4.8%)	5 (23.8%)	4 (19.0%)	8 (38.1%)
Mental Workload (n = 22)	-	-	7 (31.8%)	6 (27.3%)	7 (31.8%)	2 (9.1%)
Fatigue / Sleepiness (n = 22)	-	-	1 (4.5%)	8 (36.4%)	13 (59.1%)	-
Distraction (n = 21)	-	1 (4.8%)	3 (14.3%)	10 (47.6%)	4 (19.0%)	3 (14.3%)
Mental Fatigue (n = 20)	-	2 (10.0%)	5 (25.0%)	10 (50.0%)	3 (12.5%)	-
Hypoxia (n = 22)	-	-	2 (9.1%)	4 (18.2%)	8 (36.4%)	8 (36.4%)
Spatial Disorientation (n = 21)	-	1 (4.8%)	6 (28.6%)	5 (23.8%)	2 (9.5%)	7 (33.3%)

Below are the responses for likelihood of each physiological measures’ ability to detect each operator state. Across the seven operator states, electroencephalogram, eye tracking, and pupillometry were most frequently rated as being likely of capable in detecting operator state.

This space is intentionally blank.

Mental workload.

Within the mental workload ratings, four physiological measurements received high frequencies of “likely” or “very likely” by those familiar with these measurements for mental workload. These included ECG HRV (54.5% of ratings), electroencephalograph (EEG) (80.9% of ratings), eye tracking (76.1% of ratings), and pupillometry (72.7% of ratings). However, three of these also received ratings of “unlikely” (ECG – HRV [13.6%], eye tracking [4.8%], and pupillometry [9.1%]), indicating a lack of agreement on this topic. Table 5 below provides the frequencies and total responses for each measure.

Table 5. Mental Workload

Measure	Unsure <i>n</i> (%)	Very Unlikely <i>n</i> (%)	Unlikely <i>n</i> (%)	Neutral <i>n</i> (%)	Likely <i>n</i> (%)	Very Likely <i>n</i> (%)
Actigraphy (<i>n</i> = 20)	8 (40.0 %)	3 (15.0 %)	2 (10.0 %)	3 (15.0%)	3 (15.0 %)	1 (5.0%)
Temperature (<i>n</i> = 21)	6 (28.6 %)	1 (4.8 %)	5 (23.8 %)	4 (19.0 %)	5 (23.8 %)	-
ECG HR (<i>n</i> = 22)	4 (18.2%)	1 (4.5 %)	4 (18.2 %)	7 (31.8 %)	4 (18.2 %)	2 (9.1 %)
ECG HRV (<i>n</i> = 22)	3 (13.6%)	-	3 (13.6 %)	4 (18.2 %)	9 (40.9 %)	3 (13.6 %)
EEG (<i>n</i> = 21)	2 (9.5 %)	-	-	2 (9.5 %)	7 (33.3 %)	10 (47.6 %)
EMG (<i>n</i> = 21)	6 (28.6 %)	-	6 (28.6%)	4 (19.3 %)	5 (23.8 %)	-
EOG (<i>n</i> = 21)	3 (14.3 %)	-	2 (9.5 %)	6 (28.6 %)	7 (33.3 %)	3 (14.3 %)
Metabolites (<i>n</i> = 21)	4 (19.0 %)	3 (14.3 %)	6 (28.6 %)	3 (14.3 %)	4 (19.0 %)	1 (4.8 %)
Eye Tracking (<i>n</i> = 21)	-	-	1 (4.8 %)	4 (19.0 %)	12 (57.1 %)	4 (19.0 %)
fNIRS (<i>n</i> = 21)	7 (33.3 %)	-	1 (4.8 %)	3 (14.3 %)	5 (23.8 %)	5 (23.8 %)
PPG (<i>n</i> = 21)	7 (33.3 %)	1 (4.8 %)	3 (14.3 %)	6 (28.6 %)	4 (19.0 %)	-
Pupillometry (<i>n</i> = 22)	1 (4.5 %)	-	2 (9.1 %)	3 (13.6 %)	11 (50.0 %)	5 (22.7 %)
Respiration (<i>n</i> = 21)	6 (28.6 %)	-	7 (33.3%)	6 (28.6 %)	1 (4.8 %)	1 (4.8 %)

Note. ECG HR = electrocardiogram – heart rate; ECG HRV = electrocardiogram – heart rate variability; EEG = electroencephalograph; EMG = electromyography; EOG = electrooculogram; PPG = photoplethysmography

Participants provided two additional responses for detecting mental workload, head movement, rated as very likely (*n* = 1), and electrodermal activity (EDA), rated as neutral (*n* = 1).

Fatigue / sleepiness.

Within the fatigue / sleepiness ratings, six physiological measurements received high frequencies of “likely” or “very likely” by those familiar with these measurements for fatigue / sleepiness. These included actigraphy (60% of ratings), ECG HRV (54.5% of ratings), EEG (75% of ratings), EOG (61.9% of ratings), eye tracking (73.7% of ratings), and pupillometry (66.7% of ratings). However, three of these also received ratings of “unlikely” (ECG HRV [18.2%], EOG [4.8%], and pupillometry [9.5%]), indicating a lack of agreement on this topic. Table 6 below provides the frequencies and total responses for each measure.

Table 6. Fatigue / Sleepiness

Measure	Unsure <i>n</i> (%)	Very Unlikely <i>n</i> (%)	Unlikely <i>n</i> (%)	Neutral <i>n</i> (%)	Likely <i>n</i> (%)	Very Likely <i>n</i> (%)
Actigraphy (<i>n</i> = 20)	4 (20.0%)	-	-	4 (20.0%)	8 (40.0%)	4 (20.0%)
Temperature (<i>n</i> = 21)	7 (33.3%)	-	1 (4.8%)	5 (23.8%)	6 (28.6%)	2 (9.5%)
ECG HR (<i>n</i> = 22)	2 (9.1%)	-	5 (22.7%)	5 (22.7%)	10 (45.5%)	-
ECG HRV (<i>n</i> = 22)	3 (13.6%)	-	4 (18.2%)	3 (13.6%)	9 (40.9%)	3 (13.6%)
EEG (<i>n</i> = 20)	3 (15.0%)	-	-	2 (10.0%)	6 (30.0%)	9 (45.0%)
EMG (<i>n</i> = 21)	4 (19.0%)	-	6 (28.6%)	5 (23.8%)	5 (23.8%)	1 (4.8%)
EOG (<i>n</i> = 21)	5 (23.8%)	-	1 (4.8%)	2 (9.5%)	7 (33.3%)	6 (28.6%)
Metabolites (<i>n</i> = 22)	6 (27.3%)	2 (9.1%)	6 (27.3%)	5 (22.7%)	3 (13.6%)	-
Eye Tracking (<i>n</i> = 19)	1 (5.3%)	-	-	4 (21.1%)	5 (26.3%)	9 (47.4%)
fNIRS (<i>n</i> = 21)	6 (28.6%)	-	1 (4.8%)	4 (19.0%)	7 (33.3%)	3 (14.3%)
PPG (<i>n</i> = 20)	7 (35.0%)	1 (5.0%)	2 (10.0%)	7 (35.0%)	3 (15.0%)	-
Pupillometry (<i>n</i> = 21)	2 (9.5%)	-	2 (9.5%)	3 (14.3%)	8 (38.1%)	6 (28.6%)
Respiration (<i>n</i> = 21)	3 (14.3%)	-	6 (28.6%)	8 (38.1%)	4 (19.0%)	-

Participants provided two additional responses for detecting fatigue / sleepiness, head-worn infrared cameras for eye area-based computer vision, rated as very likely (*n* = 1); and EDA, rated as neutral (*n* = 1).

Distraction.

Within the distraction ratings, four physiological measurements received high frequencies of “likely” or “very likely” by those familiar with these measurements for distraction. These included EEG (47.6% of ratings), EOG (57.1% of ratings), eye tracking (85.7% of ratings), and pupillometry (47.6% of ratings). However, three of these also received ratings of “very unlikely” (EEG [9.5%], EOG [4.8%], and pupillometry [4.8%]), indicating a lack of agreement on this topic. Interestingly, all but one device (eye tracking) also received at least one rating of “Very Unlikely.” Table 7 below provides the frequencies and total responses for each measure.

Table 7. Distraction

Measure	Unsure <i>n</i> (%)	Very Unlikely <i>n</i> (%)	Unlikely <i>n</i> (%)	Neutral <i>n</i> (%)	Likely <i>n</i> (%)	Very Likely <i>n</i> (%)
Actigraphy (<i>n</i> = 20)	7 (35.0%)	3 (15.0 %)	2 (10.0%)	5 (25.0 %)	3 (15.0 %)	-
Temperature (<i>n</i> = 20)	9 (45.0%)	3 (15.0%)	3 (15.0%)	5 (25.0%)	-	-
ECG HR (<i>n</i> = 22)	4 (18.2%)	3 (13.6%)	7 (31.8%)	7 (31.8%)	1 (4.5 %)	-
ECG HRV (<i>n</i> = 22)	5 (22.7%)	3 (13.6%)	5 (22.7%)	6 (27.3%)	3 (13.6%)	-
EEG (<i>n</i> = 21)	5 (23.8%)	2 (9.5%)	-	4 (19.0%)	5 (23.8%)	5 (23.8%)
EMG (<i>n</i> = 21)	6 (28.6%)	2 (9.5%)	7 (33.3%)	5 (23.8%)	1 (4.8%)	-
EOG (<i>n</i> = 21)	5 (23.8%)	1 (4.8%)	-	3 (14.3%)	10 (47.6%)	2 (9.5%)
Metabolites (<i>n</i> = 22)	9 (40.9%)	6 (27.3%)	4 (18.2%)	2 (9.1%)	1 (4.5%)	-
Eye Tracking (<i>n</i> = 21)	1 (4.8%)	-	-	2 (9.5%)	8 (38.1%)	10 (47.6%)
fNIRS (<i>n</i> = 21)	8 (38.1%)	4 (9.5%)	4 (9.5%)	5 (23.8%)	1 (4.8%)	3 (14.3%)
PPG (<i>n</i> = 21)	11 (52.4%)	3 (14.3%)	4 (19.0%)	3 (14.3%)	-	-
Pupillometry (<i>n</i> = 21)	3 (14.3%)	1 (4.8%)	-	7 (33.3%)	10 (47.6%)	-
Respiration (<i>n</i> = 21)	8 (38.1%)	2 (9.5%)	6 (28.6%)	4 (19.0%)	1 (4.8 %)	-

Participants provided one additional response for detecting distraction, EDA, rated as very unlikely (*n* = 1).

Mental fatigue.

Within the mental fatigue ratings, three physiological measurements received high frequencies of “likely” or “very likely” by those familiar with these measurements for mental fatigue. These included EEG (71.5% of ratings), eye tracking (57.1% of ratings), and pupillometry (57.2% of ratings). None of these three devices received any ratings of “unlikely” or “very unlikely.” Table 8 below provides the frequencies and total responses for each measure.

Table 8. Mental Fatigue

Measure	Unsure <i>n</i> (%)	Very Unlikely <i>n</i> (%)	Unlikely <i>n</i> (%)	Neutral <i>n</i> (%)	Likely <i>n</i> (%)	Very Likely <i>n</i> (%)
Actigraphy (<i>n</i> = 20)	8 (40.0 %)	1 (5.0 %)	1 (5.0%)	5 (25.0 %)	4 (20.0%)	1 (5.0 %)
Temperature (<i>n</i> = 20)	8 (40.0%)	-	7 (35.0%)	2 (10.0%)	3 (15.0%)	-
ECG HR (<i>n</i> = 22)	3 (13.6%)	-	9 (40.9%)	6 (27.3%)	4 (18.2%)	-
ECG HRV (<i>n</i> = 22)	4 (18.2%)	-	5 (22.7%)	6 (27.3%)	6 (27.3%)	1 (4.5%)
EEG (<i>n</i> = 21)	2 (9.5%)	-	-	4 (19.0%)	6 (28.6%)	9 (42.9%)
EMG (<i>n</i> = 21)	4 (19.0%)	1 (4.8%)	5 (23.8%)	10 (47.6%)	1 (4.8%)	-
EOG (<i>n</i> = 21)	4 (19.0%)	-	2 (9.5%)	8 (38.1%)	4 (19.0%)	3 (14.3%)
Metabolites (<i>n</i> = 22)	6 (27.3%)	3 (13.6%)	5 (22.7%)	5 (22.7%)	3 (13.6%)	-
Eye Tracking (<i>n</i> = 21)	2 (9.5%)	-	-	7 (33.3%)	7 (33.3%)	5 (23.8%)
fNIRS (<i>n</i> = 21)	6 (28.6%)	-	2 (9.5%)	3 (14.3%)	7 (33.3%)	3 (14.3%)
PPG (<i>n</i> = 21)	9 (42.9%)	1 (4.8%)	4 (19.0%)	5 (23.8%)	2 (9.5%)	-
Pupillometry (<i>n</i> = 21)	2 (9.5%)	-	-	7 (33.3%)	6 (28.6%)	6 (28.6%)
Respiration (<i>n</i> = 22)	6 (27.3%)	1 (4.5%)	5 (22.7%)	8 (36.4%)	2 (9.1 %)	-

Participants provided one additional response for detecting mental fatigue, EDA, rated as unlikely (*n* = 1).

Hypoxia.

Overall, the ratings for hypoxia favored “unsure” for nearly all measures. These ranged from 40.0% of respondents who rated respiration dynamics to 55.0% of respondents who rated body / face temperature and pupillometry. Three physiological measurements received high frequencies of “likely” or “very likely” by those familiar with these measurements for hypoxia. These included respiration dynamics (55.0% of ratings), fNIRS (42.9% of ratings), and photoplethysmography (42.8% of ratings). However, two of these also received ratings of “unlikely” (fNIRS [4.8%], and photoplethysmography [9.5%]), indicating a lack of agreement on this topic. Table 9 below provides the frequencies and total responses for each measure.

Table 9. Hypoxia

Measure	Unsure <i>n</i> (%)	Very Unlikely <i>n</i> (%)	Unlikely <i>n</i> (%)	Neutral <i>n</i> (%)	Likely <i>n</i> (%)	Very Likely <i>n</i> (%)
Actigraphy (<i>n</i> = 21)	11 (52.4%)	3 (14.3%)	2 (9.5 %)	4 (19.0 %)	-	1 (4.8 %)
Temperature (<i>n</i> = 20)	11 (55.0%)	2 (10.0%)	1 (5.0%)	3 (15.0%)	3 (15.0%)	-
ECG HR (<i>n</i> = 21)	9 (42.9%)	1 (4.8%)	2 (9.5%)	3 (14.3%)	3 (14.3%)	3 (14.3%)
ECG HRV (<i>n</i> = 22)	11 (50.0%)	2 (9.1%)	1 (4.5%)	2 (9.1%)	4 (18.2%)	2 (9.1%)
EEG (<i>n</i> = 22)	10 (45.5%)	2 (9.1%)	2 (9.1%)	4 (18.2%)	3 (13.6%)	1 (4.5%)
EMG (<i>n</i> = 21)	10 (47.6%)	2 (9.5%)	3 (14.3%)	4 (19.0%)	2 (9.5%)	-
EOG (<i>n</i> = 21)	11 (52.4%)	1 (4.8%)	1 (4.8%)	2 (9.5%)	6 (28.6%)	-
Metabolites (<i>n</i> = 22)	11 (50.0%)	3 (13.6%)	5 (22.7%)	-	3 (13.6%)	-
Eye Tracking (<i>n</i> = 21)	11 (52.4%)	1 (4.8%)	-	4 (19.0%)	5 (23.8%)	-
fNIRS (<i>n</i> = 21)	10 (47.6%)	-	1 (4.8%)	1 (4.8%)	3 (14.3%)	6 (28.6%)
PPG (<i>n</i> = 21)	10 (47.6%)	-	2 (9.5%)	-	2 (9.5%)	7 (33.3%)
Pupillometry (<i>n</i> = 20)	11 (55.0%)	1 (5.0%)	1 (5.0%)	3 (15.0%)	4 (20.0%)	-
Respiration (<i>n</i> = 20)	8 (40.0%)	-	-	1 (5.0 %)	6 (30.0%)	5 (25.0%)

Participants provided three additional responses for detecting hypoxia, EDA, rated as unsure (*n* = 1); inhaled, exhaled gas analysis, rated as likely (*n* = 1); and oximetry (SpO₂ %, dual wavelength variant of PPG), rated as very likely (*n* = 1).

Spatial disorientation.

Overall, the ratings for spatial disorientation favored “unsure” for nearly all measures. These ranged from 30.0% of respondents who rated eye tracking to 55.0% of respondents who rated actigraphy. Three physiological measurements received high frequencies of “likely” or “very likely” by those familiar with these measurements for spatial disorientation. These included EEG (40.0% of ratings), EOG (40.0% of ratings), and eye tracking (65.0% of ratings). However, two of these also received ratings of “unlikely” (EEG [5.0%], and EOG [5.0%]), indicating a lack of agreement on this topic. Table 10 below provides the frequencies and total responses for each measure.

Table 10. Spatial Disorientation

Measure	Unsure <i>n</i> (%)	Very Unlikely <i>n</i> (%)	Unlikely <i>n</i> (%)	Neutral <i>n</i> (%)	Likely <i>n</i> (%)	Very Likely <i>n</i> (%)
Actigraphy (<i>n</i> = 20)	11 (55.0%)	2 (10.0%)	2 (10.0%)	1 (5.0%)	4 (20.0%)	-
Temperature (<i>n</i> = 20)	11 (50.0%)	1 (5.0%)	4 (20.0%)	3 (15.0%)	2 (10.0%)	-
ECG HR (<i>n</i> = 21)	10 (47.6%)	-	3 (14.3%)	6 (28.6%)	2 (9.5%)	-
ECG HRV (<i>n</i> = 21)	10 (47.6%)	-	4 (19.0%)	4 (19.0%)	3 (14.3%)	-
EEG (<i>n</i> = 20)	8 (40.0%)	-	1 (5.0%)	3 (15.0%)	6 (30.0%)	2 (10.0%)
EMG (<i>n</i> = 20)	9 (45.0%)	-	5 (25.0%)	4 (20.0%)	2 (10.0%)	-
EOG (<i>n</i> = 20)	10 (50.0%)	-	1 (5.0%)	1 (5.0%)	8 (40.0%)	-
Metabolites (<i>n</i> = 21)	9 (42.9%)	4 (19.0%)	2 (9.5%)	5 (23.8%)	1 (4.8%)	-
Eye Tracking (<i>n</i> = 20)	6 (30.0%)	-	-	1 (5.0%)	9 (45.0%)	4 (20.0%)
fNIRS (<i>n</i> = 20)	9 (45.0%)	-	1 (5.0%)	4 (20.0%)	5 (25.0%)	1 (5.0%)
PPG (<i>n</i> = 20)	10 (50.0%)	1 (5.0%)	4 (20.0%)	5 (25.0%)	-	-
Pupillometry (<i>n</i> = 20)	9 (45.0%)	-	2 (10.0%)	3 (15.0%)	5 (25.0%)	1 (5.0%)
Respiration (<i>n</i> = 21)	9 (42.9%)	1 (4.8%)	3 (14.3%)	7 (33.3%)	1 (4.8%)	-

Respondents provided one additional response for detecting spatial disorientation, EDA, rated as unsure (*n* = 1).

Autonomic tone.

Within the autonomic tone ratings, three physiological measurements received high frequencies of “likely” or “very likely” by those familiar with these measurements for autonomic tone. These included ECG HR (54.5% of ratings), ECG HRV (68.2% of ratings), and excreted metabolites (45.4% of ratings). However, two of these also received ratings of “unlikely” (ECG HR [4.5%], and excreted metabolites [9.1%]), and one received a rating of “very unlikely” (excreted metabolites [4.5%]) indicating a lack of agreement on this topic. Table 11 below provides the frequencies and total responses for each measure.

Table 11. Autonomic Tone

Measure	Unsure <i>n</i> (%)	Very Unlikely <i>n</i> (%)	Unlikely <i>n</i> (%)	Neutral <i>n</i> (%)	Likely <i>n</i> (%)	Very Likely <i>n</i> (%)
Actigraphy (<i>n</i> = 21)	12 (57.1%)	1 (4.8 %)	3 (14.3 %)	3 (14.3 %)	1 (4.8%)	1 (4.8%)
Temperature (<i>n</i> = 21)	9 (42.9%)	-	2 (9.5%)	3 (14.3%)	7 (33.3%)	-
ECG HR (<i>n</i> = 22)	6 (27.3%)	-	1 (4.5%)	3 (13.6%)	9 (40.9%)	3 (13.6%)
ECG HRV (<i>n</i> = 22)	6 (27.3%)	-	-	1 (4.5%)	7 (31.8%)	8 (36.4%)
EEG (<i>n</i> = 21)	11 (52.4%)	-	3 (14.3%)	3 (14.3%)	3 (14.3%)	1 (4.8%)
EMG (<i>n</i> = 21)	10 (47.6%)	-	5 (23.8%)	1 (4.8%)	4 (10.0%)	1 (4.8%)
EOG (<i>n</i> = 21)	11 (52.4%)	-	3 (14.3%)	3 (14.3%)	3 (19.0%)	-
Metabolites (<i>n</i> = 22)	8 (36.4%)	1 (4.5%)	2 (9.1%)	1 (4.5%)	5 (22.7%)	5 (22.7%)
Eye Tracking (<i>n</i> = 21)	9 (42.9%)		3 (14.3%)	3 (14.3%)	6 (28.6%)	-
fNIRS (<i>n</i> = 21)	10 (47.6%)	1 (4.8%)	3 (14.3%)	4 (19.0%)	3 (14.3%)	-
PPG (<i>n</i> = 21)	10 (47.6%)	-	2 (9.5%)	3 (14.3%)	3 (14.3%)	3 (14.3%)
Pupillometry (<i>n</i> = 21)	10 (47.6%)	-	3 (14.3%)	3 (14.3%)	3 (14.3%)	2 (9.5%)
Respiration (<i>n</i> = 22)	9 (40.9%)	-	3 (13.6%)	3 (13.6%)	5 (22.7%)	2 (9.1%)

Respondents provided one additional response for detecting spatial disorientation, EDA, rated as very likely (*n* = 1).

Other states.

Respondents provided ratings for additional states. These included engagement / arousal, stress, individual training load, and hypercapnia / hypocapnia. For engagement / arousal, ECG HR, ECG HRV, EEG, eye tracking, and pupillometry were rated as “likely” ($n = 1$). For stress, ECG HR, ECG HRV, metabolites, and EDA were rated as “very likely” ($n = 1$). For individual training load, ECG HR and ECG HRV were rated as “very likely” ($n = 1$). Finally, for hypercapnia / hypocapnia, exhaled gas analysis was provided as a method for detection ($n = 1$).

Device instrumentation.

Respondents ranked the thirteen physiological measures in the order of quickest to slowest application time (i.e., time to place sensor), based on current technology. Respondents were asked to only rank those familiar with and skip those unfamiliar with. Total number of respondents are included with rankings in Figure 1 below.

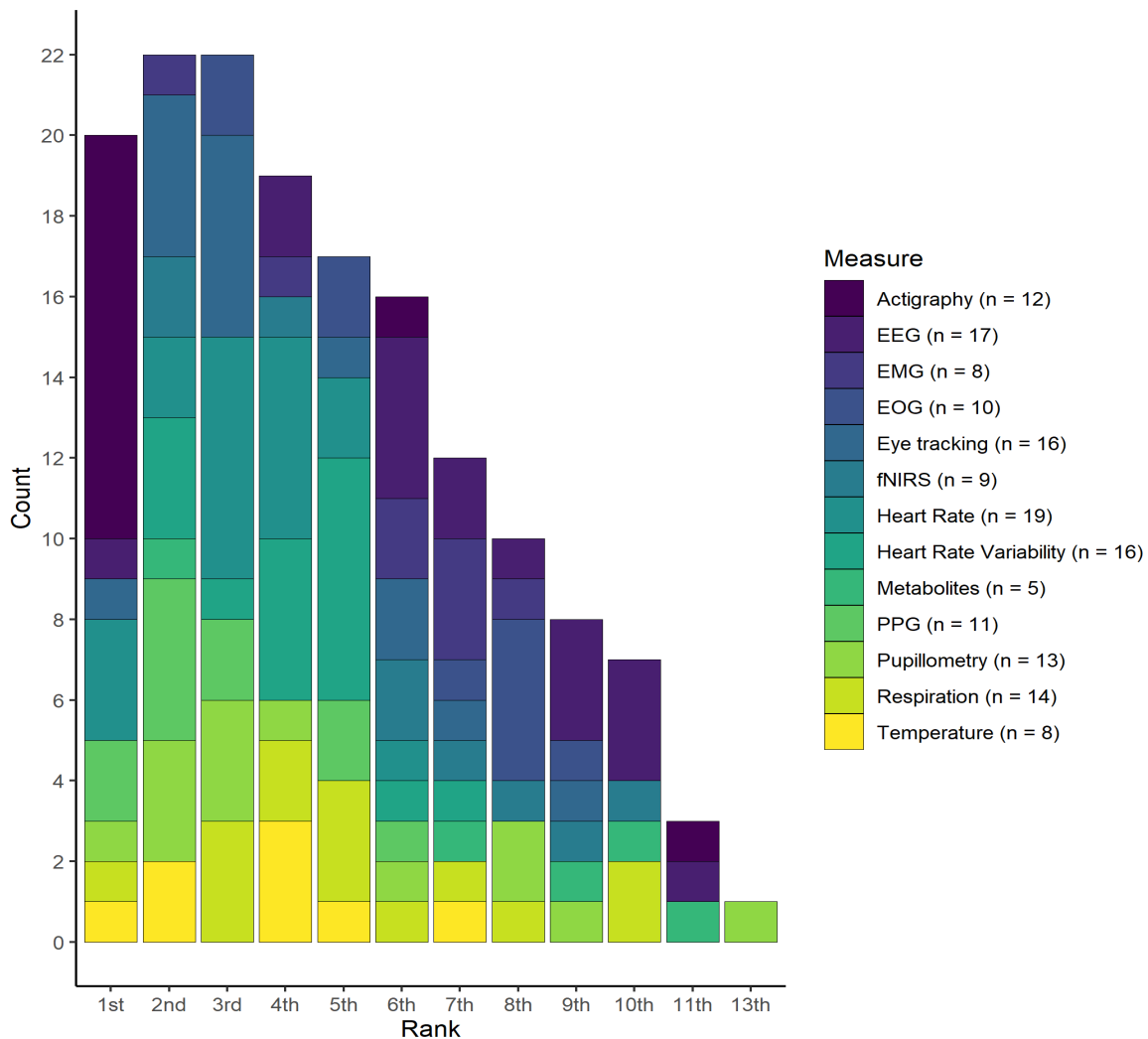


Figure 1. Current state of the technology – ratings for ease of application.

Respondents also ranked the thirteen physiological measures in the order of quickest to slowest application time (i.e., time to place sensor), based on the anticipated state of the technology within the next 5 years. Respondents were asked to only rank those familiar with and skip those unfamiliar with. Total number of respondents are included with rankings in Figure 2 below.

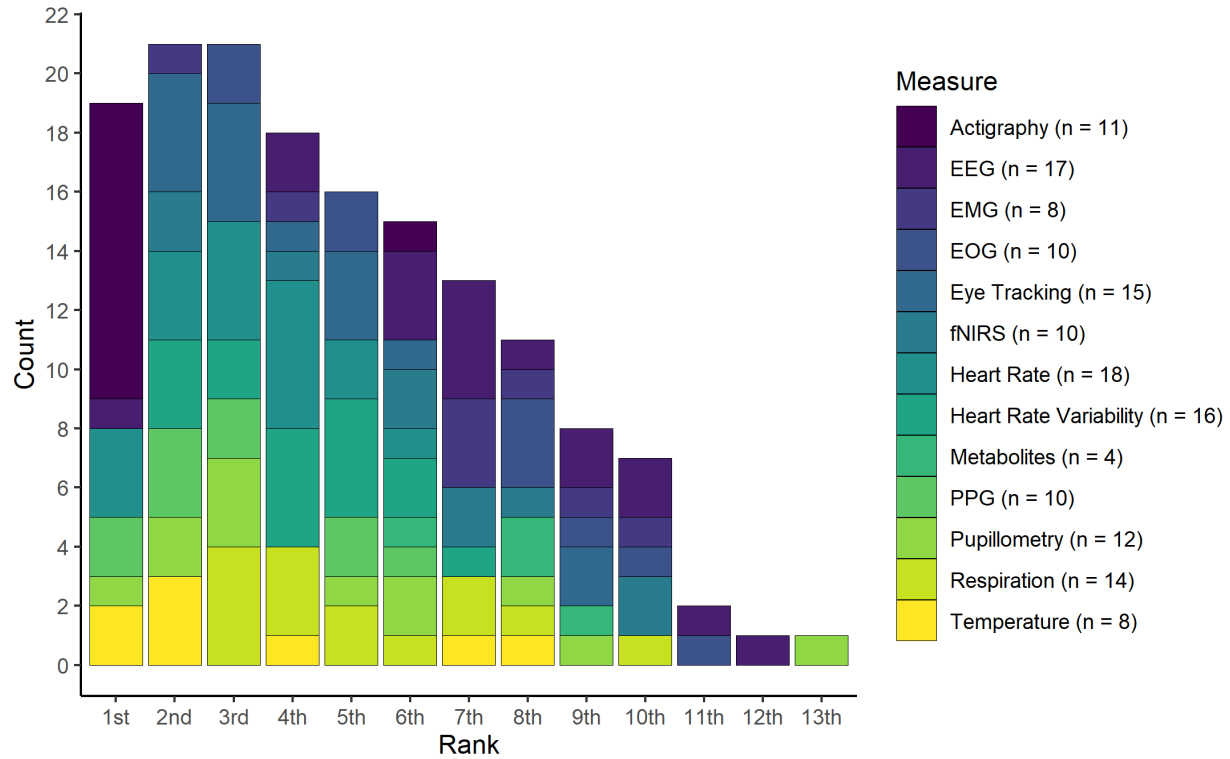


Figure 2. Anticipated 5-year state of the technology – ratings for ease of application.

Finally, respondents ranked the thirteen physiological measures in the order of ease of replacing, cleaning, or otherwise maintaining sensors to ensure an effective signal quality, based on the current state of the technology. Respondents were asked to only rank those familiar with and skip those unfamiliar with. Total number of respondents are included with rankings in Figure 3 below.

This space is intentionally blank.

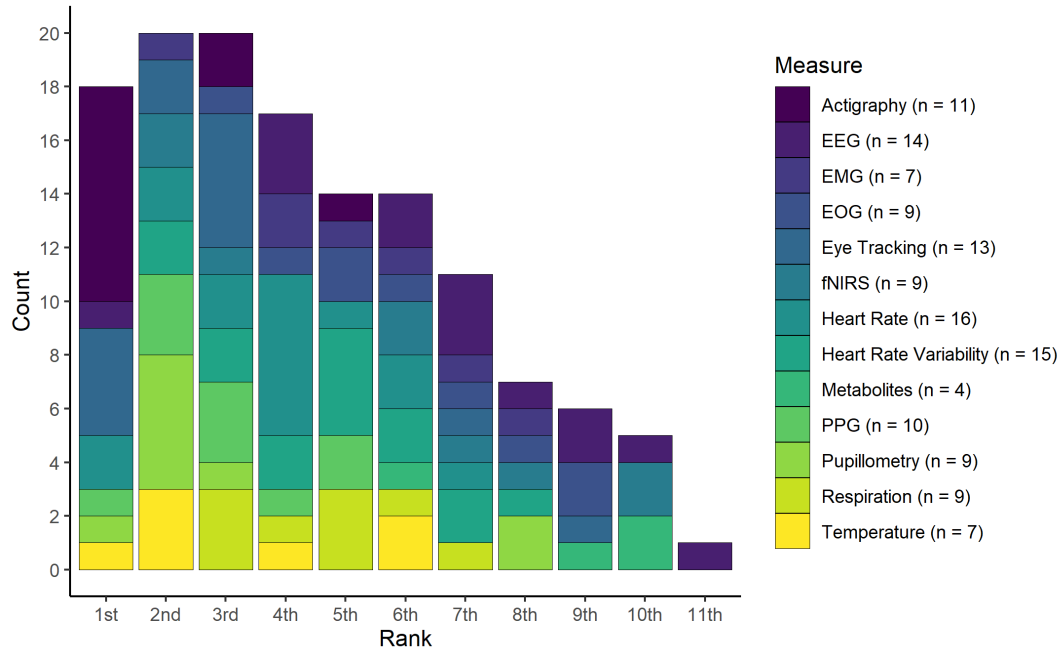


Figure 3. Ratings for ease of maintenance.

Calibration estimates.

Respondents provided calibration estimates for each of the physiological measures. Specifically, respondents were requested to: “For each physiological measure that you have experience using, please indicate the minimal acceptable criteria for calibration (e.g., at least 1 minute of baseline data). If the measure you have experience with requires no calibration, please indicate so.”

A total of 71 responses were given across measures for calibration estimates. Responses for calibration estimates across measure types generally ranged between “none” and less than 1 minute ($n = 18$) and between 1 and 5 minutes ($n = 32$). Table 12 below summarizes responses received for this item.

This space is intentionally blank.

Table 12. Calibration Estimates for Physiological Measures

Measure	Calibration Estimates		
Actigraphy (<i>n</i> = 6)	▪ None (<i>n</i> = 2) ▪ 5 seconds (s) (<i>n</i> = 1)	▪ < 1 minute (min) (<i>n</i> = 1) ▪ 1 min (<i>n</i> = 1)	▪ 1 day (<i>n</i> = 1)
Temperature (<i>n</i> = 2)	▪ 1 min (<i>n</i> = 1) ▪ 5 min (<i>n</i> = 1)		
ECG HR (<i>n</i> = 7)	▪ None (<i>n</i> = 2) ▪ 10 s (<i>n</i> = 1)	▪ 30 s (<i>n</i> = 1) ▪ <1 min (<i>n</i> = 1)	▪ 1 min (<i>n</i> = 1) ▪ 5 min (<i>n</i> = 1)
ECG HRV (<i>n</i> = 7)	▪ None (<i>n</i> = 1) ▪ 2 min (<i>n</i> = 1)	▪ 5 min (<i>n</i> = 4) ▪ 10 min (<i>n</i> = 1)	
EEG (<i>n</i> = 8)	▪ None (<i>n</i> = 1) ▪ 30 s (<i>n</i> = 1)	▪ 1 min (<i>n</i> = 1) ▪ 5 min (<i>n</i> = 4)	▪ Low impedances (<10 kiloOhms) (<i>n</i> = 1)
EMG (<i>n</i> = 4)	▪ None (<i>n</i> = 1) ▪ 30 s (<i>n</i> = 1)	▪ 5 min (<i>n</i> = 2)	
EOG (<i>n</i> = 4)	▪ None (<i>n</i> = 1) ▪ 30 s (<i>n</i> = 1)	▪ 1 min (<i>n</i> = 1) ▪ 5 min (<i>n</i> = 1)	
Metabolites (<i>n</i> = 1)	▪ Non-continuous, discrete baseline (<i>n</i> = 1)		
Eye Tracking (<i>n</i> = 12)	▪ 10 s (<i>n</i> = 2) ▪ 30 s (<i>n</i> = 3)	▪ 1 min (<i>n</i> = 3) ▪ Approx. 20 min (<i>n</i> = 1)	▪ Other (<i>n</i> = 3)*
fNIRS (<i>n</i> = 4)	▪ 10 s (<i>n</i> = 1) ▪ 5 min (<i>n</i> = 2)	▪ “Long calibration” (<i>n</i> = 1)	
PPG (<i>n</i> = 4)	▪ 10 s (<i>n</i> = 1) ▪ 30 s (<i>n</i> = 1)	▪ 5 min (<i>n</i> = 2)	
Pupillometry (<i>n</i> = 8)	▪ 1 min (<i>n</i> = 2) ▪ 2 min (<i>n</i> = 1)	▪ Other (<i>n</i> = 5)**	
Respiration (<i>n</i> = 3)	▪ 1 min (<i>n</i> = 2) ▪ 5 min (<i>n</i> = 1)		

*Note. Other responses included: 1) Calibration standard deviation < 2 degrees of visual angle; 2) Need world calibration to get coordinate system of the world or screen; 3) 5 – 9 point calibration for gaze estimation and should also be calibrated occasionally throughout if it is a long study. It also depends on the device how frequently recalibration is needed.

**Note. Other responses included: 1) Long calibration with standard deviation < 2 degrees of visual angle; 2) Baseline not necessarily needed, but quiet [sic] rest of 3 minutes is sufficient; 3) Use either a fiducial marker affixed on glasses or measure a reference point on the face, i.e., interpupillary distance for approximately 30 seconds; 4) Minimal (a couple of typical short operator tasks); 5) Could benefit from a baseline (at least 10 seconds) in some instances, but in others we are more interested in short term relative changes (i.e., phasic responses).

This space is intentionally blank.

Other measures reported:

- Speckleplethysmography (SPG): 10 s ($n = 1$)
- Saccadic velocity: At least 5 min baseline data ($n = 1$)
- Thermography: At least 1 min of recording ($n = 1$)
- Inertial measurement unit (IMU): Minimal (a couple of typical short operator tasks) ($n = 1$)
- Carbon dioxide (CO₂)/Respiratory: This sensor had to be calibrated by the vendor; it lasted for a period of several weeks ($n = 1$)

Miscellaneous responses:

- “It's hard at our stage of research to make a call as to what minimal acceptable criteria, as yet; obviously, less is more; on the other hand, a multi-modal / multi-sensor holistic suite that could provide state index and more nuanced info may be worth taking much longer to calibrate.” ($n = 1$)
- “We prefer to get 5-10 minutes of baseline (depending on protocol) before starting an experiment across all devices.” ($n = 1$)

This space is intentionally blank.

Compatibility with environment.

Vibrations.

Based on the current state of technology, the devices that most frequently received ratings of “likely” or “very likely” for ability to withstand vibrations were ECG HR (66.7 % of ratings) and respiration dynamics (73.7 % of ratings). However, body/face temperature, ECG HRV and excreted metabolites also received similarly high ratings (55.6 %, 55.5 %, 58.8 %, respectively). Disagreement in ratings was noted, with ECG HR receiving 5.6 % of ratings as “very unlikely.” Table 13 below provides the frequencies and total responses for each measure.

Table 13. Likelihood of Withstanding Vibrations

Measure	Unsure n (%)	Very Unlikely n (%)	Unlikely n (%)	Neutral n (%)	Likely n (%)	Very Likely n (%)
Actigraphy (n = 17)	9 (52.9%)	-	3 (17.6 %)	2 (11.8 %)	3 (17.6 %)	-
Temperature (n = 18)	5 (27.8 %)	-	2 (11.1 %)	1 (5.6 %)	7 (38.9 %)	3 (16.7 %)
ECG HR (n = 18)	2 (11.1 %)	1 (5.6 %)	-	3 (16.7 %)	10 (55.6 %)	2 (11.1 %)
ECG HRV (n = 18)	2 (11.1 %)	-	2 (11.1 %)	4 (22.2 %)	8 (44.4 %)	2 (11.1 %)
EEG (n = 17)	5 (29.4 %)	1 (5.9 %)	5 (29.4 %)	3 (17.6 %)	3 (17.6 %)	-
EMG (n = 17)	4 (23.5 %)	1 (5.9 %)	2 (11.8 %)	4 (23.5 %)	6 (35.3 %)	-
EOG (n = 17)	5 (29.4 %)	1 (5.9 %)	1 (5.9 %)	2 (11.8 %)	8 (47.1 %)	-
Metabolites (n = 17)	6 (35.3 %)	-	-	1 (5.9 %)	6 (35.3 %)	4 (23.5 %)
Eye Tracking (n = 20)	3 (15.0 %)	1 (5.0 %)	4 (20.0 %)	4 (20.0 %)	7 (35.0 %)	1 (5.0 %)
fNIRS (n = 18)	6 (33.3 %)	-	5 (27.8 %)	3 (16.7 %)	4 (22.2 %)	-
PPG (n = 18)	8 (44.4 %)	-	3 (16.7 %)	2 (11.1 %)	5 (27.8 %)	-
Pupillometry (n = 20)	7 (35.0 %)	1 (5.0 %)	4 (20.0 %)	4 (20.0 %)	4 (20.0 %)	-
Respiration (n = 19)	3 (15.8 %)	-	-	2 (10.5 %)	11 (57.9 %)	3 (15.8 %)

Extreme temperatures.

Based on the current state of technology, the devices that most frequently received ratings of “likely” or “very likely” for ability to withstand extreme temperatures were ECG HR (68.4 % of ratings), ECG HRV (66.7 % of ratings), and eye tracking (60.0 % of ratings). No disagreement in these ratings were noted. Table 14 below provides the frequencies and total responses for each measure.

Table 14. Likelihood of Withstanding Extreme Temperatures

Measure	Unsure <i>n</i> (%)	Very Unlikely <i>n</i> (%)	Unlikely <i>n</i> (%)	Neutral <i>n</i> (%)	Likely <i>n</i> (%)	Very Likely <i>n</i> (%)
Actigraphy (<i>n</i> = 17)	8 (47.1 %)	-	-	1 (5.9 %)	3 (17.6 %)	5 (29.4 %)
Temperature (<i>n</i> = 17)	7 (41.2 %)	2 (11.8 %)	3 (17.6 %)	-	2 (11.8 %)	3 (17.6 %)
ECG HR (<i>n</i> = 19)	6 (31.6 %)	-	-	-	10 (52.6 %)	3 (15.8 %)
ECG HRV (<i>n</i> = 18)	6 (33.3 %)	-	-	-	10 (55.6 %)	2 (11.1 %)
EEG (<i>n</i> = 17)	8 (47.1 %)	2 (11.8 %)	3 (17.6 %)	1 (5.9 %)	3 (17.6 %)	-
EMG (<i>n</i> = 17)	9 (52.9 %)	1 (5.9 %)	-	2 (11.8 %)	4 (23.5 %)	1 (5.9 %)
EOG (<i>n</i> = 17)	7 (41.2 %)	-	2 (11.8 %)	3 (17.6 %)	3 (17.6 %)	2 (11.8 %)
Metabolites (<i>n</i> = 17)	8 (47.1 %)	-	-	1 (5.9 %)	6 (35.3 %)	2 (11.8 %)
Eye Tracking (<i>n</i> = 20)	7 (35.0 %)	-	-	1 (5.0 %)	5 (25.0 %)	7 (35.0 %)
fNIRS (<i>n</i> = 17)	6 (35.3 %)	2 (11.8 %)	-	3 (17.6 %)	5 (29.4 %)	1 (5.9 %)
PPG (<i>n</i> = 17)	10 (58.8 %)	-	-	2 (11.8 %)	3 (17.6 %)	2 (11.8 %)
Pupillometry (<i>n</i> = 19)	8 (42.1 %)	-	-	1 (5.3 %)	3 (15.8 %)	7 (36.8 %)
Respiration (<i>n</i> = 18)	7 (38.9 %)	-	-	1 (5.3 %)	5 (27.8 %)	5 (27.8 %)

Bright environments and direct sunlight.

Based on the current state of technology, the devices frequently received ratings of “likely” or “very likely” for ability to withstand bright environments and direct sunlight were ECG HR (94.1 % of ratings), ECG HRV (94.5 % of ratings), and respiration dynamics (82.4 % of ratings). No disagreement in these ratings were noted. Table 15 below provides the frequencies and total responses for each measure.

Table 15. Likelihood of Withstanding Bright Environments and Direct Sunlight

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy (n = 16)	4 (25.0 %)	-	-	1 (6.3 %)	1 (6.3 %)	10 (62.5 %)
Temperature (n = 16)	5 (31.3 %)	-	4 (25.0 %)	-	3 (18.8 %)	4 (25.0 %)
ECG HR (n = 17)	1 (5.9 %)	-	-	-	3 (17.6 %)	13 (76.5 %)
ECG HRV (n = 18)	1 (5.6 %)	-	-	-	3 (16.7 %)	14 (77.8 %)
EEG (n = 16)	4 (25.0 %)	-	-	-	3 (18.8 %)	5 (56.3 %)
EMG (n = 16)	4 (25.0 %)	-	-	-	3 (18.8 %)	5 (56.3 %)
EOG (n = 16)	5 (31.3 %)	-	-	-	3 (18.8 %)	8 (50.0 %)
Metabolites (n = 16)	6 (37.5 %)	-	-	-	3 (18.8 %)	7 (43.8 %)
Eye Tracking (n = 20)	3 (15.0 %)	4 (20.0 %)	7 (35.0 %)	3 (15.0 %)	2 (10.0 %)	1 (5.0 %)
fNIRS (n = 17)	5 (29.4 %)	3 (17.6 %)	3 (17.6 %)	2 (11.8 %)	3 (17.6 %)	1 (5.9 %)
PPG (n = 17)	5 (29.4 %)	-	2 (11.8 %)	2 (11.8 %)	5 (29.4 %)	3 (17.6 %)
Pupillometry (n = 18)	2 (11.1 %)	5 (27.8 %)	7 (38.9 %)	2 (11.1 %)	-	2 (11.1 %)
Respiration (n = 17)	3 (17.6 %)	-	-	-	6 (35.3 %)	8 (47.1 %)

Likelihood of maintaining adhesion.

Based on the current state of technology, the devices that received the highest frequency of “likely” or “very likely” ratings for ability to maintain adhesion were ECG HR (54.2% of ratings), ECG HRV (45.8% of ratings), and respiration dynamics (54.2% of ratings). Disagreement in these ratings were noted. Each of these measures received ratings of “unlikely” (ECG HR [8.3%], ECG HRV [8.3%], and respiration dynamics [4.2%]) and “very unlikely” (ECG HR [4.2%], ECG HRV [4.2%], and respiration dynamics [4.2%]). Table 16 below provides the frequencies and total responses for each measure.

Table 16. Likelihood of Maintaining Adhesion

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy (n = 16)	5 (31.3 %)	-	-	2 (12.5 %)	5 (31.3 %)	4 (25.0 %)
Temperature (n = 17)	6 (35.3 %)	2 (11.8 %)	-	2 (11.8 %)	6 (35.3 %)	1 (5.9 %)
ECG HR (n = 19)	2 (10.5 %)	1 (5.3 %)	2 (10.5 %)	1 (5.3 %)	10 (52.6 %)	3 (15.8 %)
ECG HRV (n = 17)	2 (11.8 %)	1 (5.9 %)	2 (11.8 %)	1 (5.9 %)	9 (52.9 %)	2 (11.8 %)
EEG (n = 16)	5 (31.3 %)	-	6 (37.5 %)	4 (25.0 %)	1 (6.3 %)	-
EMG (n = 16)	3 (18.8 %)	-	3 (18.8 %)	2 (12.5 %)	4 (43.8 %)	1 (6.3 %)
EOG (n = 15)	4 (26.7 %)	-	4 (26.7 %)	3 (20.0 %)	3 (20.0 %)	1 (6.7 %)
Metabolites (n = 15)	6 (40.0 %)	-	1 (6.7 %)	4 (26.7 %)	3 (20.0 %)	1 (6.7 %)
Eye Tracking (n = 19)	3 (15.8 %)	2 (10.5 %)	4 (21.1 %)	2 (10.5 %)	5 (26.3 %)	3 (15.8 %)
fNIRS (n = 17)	6 (35.3 %)	1 (5.9 %)	2 (11.8 %)	3 (17.6 %)	5 (29.4 %)	-
PPG (n = 17)	6 (35.3 %)	1 (5.9 %)	1 (5.9 %)	1 (5.9 %)	8 (47.1 %)	-
Pupillometry (n = 18)	6 (33.3 %)	1 (5.6 %)	2 (11.1 %)	1 (5.6 %)	5 (27.8 %)	3 (16.7 %)
Respiration (n = 19)	3 (15.8 %)	1 (5.3 %)	1 (5.3 %)	1 (5.3 %)	10 (52.6 %)	3 (15.8 %)

Sensitivity

Respondents were given the following definition of sensitivity: "proportion of truly positive cases that correspond to positive results." Based on this definition, respondents were asked: *Regarding the ability of a system or device to collect physiological signals, how sensitive does the system need to be in rejecting unreliable and retaining reliable signals?*

Responses are summarized in Table 17 below. Only those physiological measures that received responses are reported, along with the number of respondents.

Responses for this item varied based on the physiological measure type, as well as within the measure. For example, two respondents stated ECG HRV "requires minimal noise and at least 2 min of data", quantified as 90 %. Alternatively, one respondent stated EEG requires "low noise and 5 min of data", quantified as 80 %. Within EEG responses, another respondent stated EEG requires "minimal noise, at least 1 min of data", quantified as 90 %. The measure that was given the lowest sensitivity value was pupillometry, where one participant stated a minimum of 50 % quality signal from a single eye.

Besides those summarized in the table below, additional comments (Appendix D) by two participants included discussion that sensitivity needs would be dependent on what the signals were used for. For example, "requirements would differ if the output were interpreted by another individual versus used by the system to make decisions such as firing a weapon." Overall, there was a range of opinions provided for signal sensitivity requirements.

This space is intentionally blank.

Table 17. Signal Sensitivity

Physiological Measure	Signal Quality Requirements (Specific measurement if given)	Quantity
Actigraphy (<i>n</i> = 1)	▪ Clean signal, at least 2 min of data (motion)	▪ 70%
ECG HR (<i>n</i> = 3)	▪ Minimal noise, at least 1 min of data (<i>n</i> = 1) ▪ Minimal noise, at least 5 min of data (<i>n</i> = 1) ▪ HR is robust (<i>n</i> = 1)	▪ 80% (<i>n</i> = 1) ▪ 90% (<i>n</i> = 1) ▪ 99% (<i>n</i> = 1)
ECG HRV (<i>n</i> = 4)	▪ 1 min of data, but no missing beat allowable (<i>n</i> = 1) ▪ Minimal noise, at least 2 min of data (<i>n</i> = 2) ▪ Minimal noise, at least 3 min of data (<i>n</i> = 1)	▪ 80% (<i>n</i> = 1) ▪ 90% (<i>n</i> = 2) ▪ 95% (<i>n</i> = 1)
EEG (<i>n</i> = 2)	▪ Low noise, 5 min of data (<i>n</i> = 1) ▪ Minimal noise, at least 1 min of data (<i>n</i> = 1)	▪ 80% (<i>n</i> = 1) ▪ 90% (<i>n</i> = 1)
EOG (<i>n</i> = 1)	▪ Minimal noise and few artefacts, at least 1 minute of data	▪ 70%
Eye Tracking (<i>n</i> = 3)	▪ Not specified: some noise, 30 s of data (<i>n</i> = 1) ▪ Gaze-based measures: minimal noise, at least 5-point calibration (<i>n</i> = 1) ▪ Other (<i>n</i> = 1)*	▪ NR (<i>n</i> = 2) ▪ 80% of data pts present; <1 degree of visual angle error (<i>n</i> = 1)
fNIRS (<i>n</i> = 3)	▪ Minimal noise, at least 5 min of data (deoxygenated hemoglobin (Hhb), hemoglobin (Hb), tissue oxygenation index (TOI)) (<i>n</i> = 1) ▪ Minimal noise, at least 2 min of data (regional oxygen saturation (rSO ₂)) (<i>n</i> = 1) ▪ Clean signal, at least 2 min of data (cerebral hemodynamics (<i>n</i> = 1)	▪ 90% (<i>n</i> = 3)
PPG (<i>n</i> = 4)	▪ Minimal noise, at least 1 min of data for HR and at least 2 min of data for HRV (<i>n</i> = 1) ▪ Minimal noise, 5 s (HRV) (<i>n</i> = 1) ▪ Minimal noise/clean signal, at least 2 min of data (SpO ₂) (<i>n</i> = 2)	▪ 60% (<i>n</i> = 1) ▪ 95% (<i>n</i> = 1) ▪ 90% (<i>n</i> = 2)
Pupillometry (<i>n</i> = 3)	▪ 50% quality data from one eye is the minimum (<i>n</i> = 1) ▪ Baseline: pre-stimulus period of at least 3 s since previous stimulus, or at least 3 min of data (<i>n</i> = 1) ▪ Phasic response: no dropout during phasic response (<i>n</i> = 1)	▪ 50% one eye (<i>n</i> = 1) ▪ 70% (<i>n</i> = 1) ▪ 90% (<i>n</i> = 1)
Respiration (<i>n</i> = 2)	▪ Respiration rate: minimal noise, at least 5 min of data (<i>n</i> = 1) ▪ Breathing depth: minimal noise, at least 5 min of data (<i>n</i> = 1)	▪ 85% (<i>n</i> = 2)

*Note. NR = no response. * Other response: For eye tracking data, it depends a lot on what you want to do with it. If you're mainly interested in some area-of-interest analysis (different parts of a control panel, etc.) then this could be quite robustly. The problem is a bit harder if a mobile eye tracker is used (e.g., to allow for head movements), but I think this is solvable using Computer Vision to help assign fixation information to control area of interest (AOIs).

Specificity.

Respondents were given the following definition of specificity: "proportion of truly negative cases that correspond to negative test results." Based on this definition, respondents were asked: *Regarding the ability of a system or device to collect physiological signals, how specific does the system need to be in rejecting unreliable and retaining reliable signals?*

Responses are summarized in Table 18 below. Only those physiological measures that received responses are reported, along with the number of respondents. Overall, responses to this item followed a similar pattern as that for sensitivity. However, within the additional comments provided (Appendix D), several participants indicated the need for the system to be robust in its ability to reject unreliable signals.

Table 18. Signal Specificity

Physiological Measure	Signal Quality Requirements (specific measurement if given)	Quantify
ECG HR (<i>n</i> = 2)	- High specificity (<i>n</i> = 1) - Minimal noise, at least 5 min of data (<i>n</i> = 1)	90% (<i>n</i> = 1)
ECG HRV (<i>n</i> = 1)	High specificity (<i>n</i> = 1)	NR
EEG (<i>n</i> = 2)	- Low noise, 30 min of data (<i>n</i> = 1) - Medium specificity (<i>n</i> = 1)	70% (<i>n</i> = 1)
EOG (<i>n</i> = 1)	High specificity (<i>n</i> = 1)	NR
Eye Tracking (<i>n</i> = 2)	- Some noise, 5 min of data (<i>n</i> = 1) - Minimal noise, at least 5-point calibration (<i>n</i> = 1)	80% (<i>n</i> = 2*)
fNIRS (<i>n</i> = 1)	Minimal noise, at least 2 min of data (rSO₂) (<i>n</i> = 1)	90% (<i>n</i> = 1)
PPG (<i>n</i> = 2)	- High specificity (HR/HRV) (<i>n</i> = 1) - Minimal noise, at least 2 min of data (SpO₂) (<i>n</i> = 1)	90% (<i>n</i> = 1)
Pupillometry (<i>n</i> = 1)	50% good quality data from one eye is the minimum (<i>n</i> = 1)**	50% one eye (<i>n</i> = 1)
Respiration Dynamics	- Minimal noise, at least 5 min of data (respiration rate from ECG) (<i>n</i> = 1) - Minimal noise, at least 5 min of data (breathing depth from ECG) (<i>n</i> = 1)	85% (<i>n</i> = 1)

*Note. NR = no response; *One respondent also stated: 80% of data points present. <1 degree of visual angle error; Drift is an issue with gaze-based studies over a long period of time, but you can insert additional calibrations to correct drift. **Respondent added, very context dependent.

Physical characteristics.

Respondents were asked: *Based on your personal experience in working with physiological data, please indicate the physical characteristics of pilots that may impact the signal quality and/or ability to secure the sensor to the body. Please include only those metrics you have direct experience in using.*

Several consistent themes emerged within the physiological measures evaluated. Body hair was considered a concern for temperature, ECG HR, ECG HRV, and PPG. Body fat was considered a concern for both ECG HR and HRV. Scalp, hair thickness, and sweat were identified as concerns for both EEG and fNIRS. Eye tracking and pupillometry share concerns of eye shape and the presence of glasses or contacts.

Table 19. Themes of Physical Characteristics

Physiological Measure	Respondent Themes on Physical Characteristic(s)
Actigraphy ($n = 2$)	▪ Movement disorders ▪ None
Temperature ($n = 1$)	▪ Body hair
ECG HR ($n = 6$)	▪ Body hair ▪ Body fat
ECG HRV ($n = 6$)	▪ Body hair ▪ Body fat
EEG ($n = 5$)	▪ Scalp condition (e.g., oily, dry) ▪ Hair thickness ▪ Sweat ▪ Scalp thickness ▪ Aviation equipment (e.g., headphones)
EMG ($n = 1$)	▪ Botox / other dermal injections
EOG ($n = 2$)	▪ Botox / other dermal injections
Eye Tracking ($n = 7$)	▪ Eye shape ▪ Eye color ▪ Presence of glasses or contact lenses ▪ Individual differences in ability to track ▪ Eye makeup (e.g., eyeliner)
Metabolites ($n = 1$)	▪ Saliva volume
fNIRS ($n = 3$)	▪ Hair thickness ▪ Skin tone ▪ Sweating ▪ Scalp thickness
PPG ($n = 2$)	▪ Placement location ▪ Hair
Pupillometry ($n = 3$)	▪ Older age, high expertise ▪ Eyelashes ▪ Presence of glasses or contact lenses ▪ Eye shape
Respiration Dynamics ($n = 1$)	▪ None

Additional comments from respondents included the following themes:

- Aircrew equipment (e.g., helmets, gloves) interfering with placement of sensors and introduction of artefacts.
- Sweat being a concern for attaching and maintaining adhesion of a variety of sensor types.

Round Two Results

The second round of the survey aimed to clarify some of the diverging responses from the first round.

Sensitivity.

The first set of diverging results that were explored were regarding sensitivity. Respondents were provided with two use case scenarios and asked to determine sensitivity needed for each. The full use case descriptions and participant responses are in Appendix E.

Use Case One described a scenario where the physiological information collected from the aviator was being relayed to a mission commander. Regarding quantifiable responses, respondents ranged from 75 % to 95 %. However, response themes also included the ambiguity of the term “operator state,” variability within human physiology, and the importance of what is being done with the information at the commander level.

Use Case Two described a scenario where the physiological information collected from the aviator was used to trigger an adaptive automation system. Regarding quantifiable responses, respondents ranged from 80 % to 95 %, with the majority reporting 90 % ($n = 4$). Response themes also included considerations for whether the human would be able to override the autonomy, and what tasks would be automated (e.g., straight and level flight vs. something more involved).

Operator states.

Responses regarding which operator states are candidates for detection in real-time also varied. In this round, respondents were presented the results from the previous survey iteration and asked for justification regarding why or why not a specific state may be a candidate for detection.

Responses included the following themes:

- Recommendation for a current, university-developed, tool for assessing cognitive workload in real-time.
- Lack of definitive criteria for identifying various operator states (e.g., multiple “states” may contribute to spatial disorientation, such as fatigue and inattention).
- Insufficient evidence for environmental causes associated with operator state, thus leading to inability to determine subsequent actions for an autonomous system to take

- Surprise regarding autonomic tone ratings. Specifically, respondents were surprised by the number of “unsure” ratings for autonomic tone.

Physiological measures.

Similar to which operator states may be candidates for detection in real-time, varying responses were received regarding *which physiological measures* are capable of detecting various the operator states. Respondents were again invited to provide justification for why a physiological measure may or may not be capable of detecting each state.

Mental workload. Respondent themes included:

- ECG HR and HRV were both indicated as good measures to include in an algorithm for detecting mental workload.
- Pupillometry was indicated as promising but is in need of further study before moving into operational contexts.
- Actigraphy, temperature, ECG, EMG, EOG, PPG, pupillometry, and respiration were all indicated as too insensitive / nonspecific.
- EEG, fNIRS, and eye tracking were listed as being too prone to artifacts for use in operational settings.
- Combining physiological measures (e.g., actigraphy, ECG HR, EMG, EOG, eye tracking, PPG, pupillometry, respiration) with other information may be beneficial in improving accuracy.

Fatigue / sleepiness. Respondent themes included:

- Cost-benefit of including physiological sensors when other well-established indicators are available (e.g., time on task).
- Pupillometry was indicated as promising, but in need of further study before moving into operational contexts.
- Surprise with EOG ratings, as blink rate is well-known for its sensitivity to sleepiness.
- Alpha power and sleep spindles in EEG are also sensitive to sleepiness.
- EMG is useful for muscle fatigue but may be less specific for aviation settings.

Distraction. Respondent themes included:

- Actigraphy, temperature, ECG, EMG, EOG, PPG, pupillometry, and respiration were all indicated as too insensitive / nonspecific.
- EEG, fNIRS, and eye tracking as being too prone to artefacts for use in operational settings.
- Eye tracking may be considered the “best” indicator, as you can assess where they are looking.
- Measures of brain activity (EEG, fNIRS) and measures tracking motor-behavioral characteristics of distraction (actigraphy, EMG, EOG) may be beneficial.
- Measures used to characterize mental workload are also promising.

Mental fatigue. Respondent themes included:

- Same comment themes as workload.
- Mental fatigue is less likely to be accompanied by specific physiological reactions; measures listed are too non-specific.
- Surprise that EDA was not viewed as capable of detecting mental fatigue and a reference provided.

Hypoxia. Respondent themes included:

- Inhaled/exhaled gas analysis and oximetry are capable of assessing oxygen availability, and thus well-suited to detect hypoxia.
- Other measures, such as actigraphy, could help detect onset / loss of consciousness, but would likely not be helpful in detecting to prevent occurrence of this state.
- HR may be able to detect *possible* onset of hypoxia.

Spatial disorientation. Respondent themes included:

- Spatial disorientation is less likely to be accompanied by specific physiological reactions; measures listed are too non-specific.
- Agreement that eye behavioral measures (eye tracking, EOG, pupil measures) are viewed as likely feasible.
- Head movement-related measures, such as neck EMG, may also be useful.

Discussion

The purpose of the present study was to survey experts within the fields of operator state detection and wearable sensor technologies in order to identify operator states likely to be detectable by physiological measures, and which physiological measures are likely able to detect each operator state. In addition, we sought to characterize the current state of physiological sensors in terms of time required for set up, and compatibility with the aviation environment. Finally, we wanted to determine the criteria in terms of sensitivity and specificity the sensors need to meet and any additional measures in terms of physical characteristics that need to be considered.

To fulfill the goals of this study, the Delphi method of surveying SMEs was used. Two rounds of the survey were completed. Within the first round, we had a response rate of 54.54 %, while in the second round the response rate dropped to 42.86 %. Both response rates are considered to be in-line with response rates of survey efforts, which fall around 53% (Baruch & Holtom, 2008).

Operator States and State Detection

Despite some differing responses that required clarity with a second-round survey, several themes emerged within the categories that were surveyed. Overall, the operator states that received the highest number of positive responses for likelihood of detection in real-time included: fatigue / sleepiness, mental workload, and hypoxia. Within these three operator states, fatigue / sleepiness and mental workload shared some commonalities with physiological measures that may be capable of detecting these states in real-time. These included ECG HRV, EEG, eye tracking, and pupillometry. Measures capable of detecting fatigue / sleepiness also included actigraphy and EOG. Regarding hypoxia, respiration dynamics, fNIRS, and PPG were frequently indicated as likely to detect this state.

Across the seven states that were presented to respondents (autonomic tone, mental workload, fatigue / sleepiness, distraction, mental fatigue, hypoxia, and spatial disorientation), three physiological measures were repeatedly rated as likely to detect the various states. These measures included EEG, eye tracking, and pupillometry. Both EEG and eye tracking received high frequencies of ratings as likely to detect physiological states ratings across five states, while pupillometry did so across four states. Next, ECG HRV and EOG received high frequencies of likely ratings across three states. The remaining physiological measures only received high frequencies of ratings for singular states, suggesting they may be state-specific (e.g., respiration measures for detecting hypoxia). The remainder of the discussion will focus on these three measures (EEG, eye tracking, and pupillometry). Of note, themes that emerged in the second round of the survey included concerns for how well these measures would work in an operational setting (e.g., artifacts, lack of demonstration of use in similar settings) and that a combination of measures would likely yield the most useful results.

Instrumentation.

Devices that will be considered for use in an operational environment will need to feature minimal time commitments for set up. Based on the current and anticipated future state of the technology, EEG did not rank well for ease of setup (e.g., most frequently ranked as 6th). Both eye tracking and pupillometry were ranked as easier to setup compared to EEG, but actigraphy was most frequently ranked as easiest. In addition to ease of use, it will be important for devices to be easy to maintain. Regarding this, similar patterns were identified, where other measures may be easier to maintain (e.g., actigraphy), and eye tracking and pupillometry ranked as easier than EEG.

Environment.

One limiting factor of identifying physiological measures to use for monitoring operator state is whether they are able to withstand the demands of the aviation environment. This environment includes vibrations, extreme temperatures, and exposure to bright environments and / or direct sunlight. Additionally, the likelihood of whether sensor(s) can maintain adhesion under such conditions is also of interest.

Measures that were viewed as likely to be able to withstand the vibrations of the aircraft included ECG HR, respiration dynamics, temperature, ECG HRV, and metabolites. Regarding the three measures of interest, most rated EEG as unlikely to withstand the vibrations, while one-third were unsure. Eye tracking and pupillometry received more mixed responses between likely and unlikely, with approximately one-third of respondents unsure of pupillometry in such conditions. Regarding extreme temperatures, ECG HR and HRV and eye tracking were all rated as likely to withstand this condition. The majority of respondents were unsure of EEG's ability. Pupillometry followed a similar pattern as eye tracking, where most respondents rated it as likely to withstand extreme temperatures, and a few unsure. For bright environments and / or direct sunlight, ECG HR and HRV, and respiration all received high frequencies of likely ratings. EEG also received a high number of likely ratings, with one-fourth unsure. Eye tracking and pupillometry mostly received unlikely ratings. The likelihood of sensors maintaining adhesion in these environments was evaluated. ECG HR and HRV, and respiration were rated most frequently as likely to maintain adhesion. EEG received mostly unlikely ratings, with one-third of respondents unsure. Eye tracking mostly received likely ratings, with one-third of respondents rating very unlikely. Finally, pupillometry mostly received likely ratings, with one-third of respondents unsure.

Sensitivity and specificity.

Overall, the ranges given for sensitivity and specificity varied quite a bit. For signal sensitivity the lowest value given was 50% quality data from one eye as the minimum for pupillometry, whereas the highest value given was 99% for ECG HR. Similar patterns were seen with regard to specificity, with the same value given for pupillometry and the highest value given was 90% for a number of measures (ECG HR, fNIRS, PPG). Given the variability in sensitivity responses, we included questions related to this in the second round of the survey. Within the second round of the survey, we provided two use-case scenarios from which respondents were requested to consider their responses for. With these two case scenarios, respondents indicated that sensitivity of a system should be around 90% if the information is to be used for adaptive automation. However, several items for concern were also pointed out, such as whether the aviator would have the ability to override the system.

Limitations

The study had several limiting factors. One limitation was that we did not provide the SMEs with operational definitions of the operator states we used in the survey. This may have resulted in the SMEs interpreting the states differently. However, we did not wish to attribute a specific definition to each state, and rather wished to allow the SMEs to interpret based on their own experiences. Additionally, responses within the survey may have been biased based on the operator states and / or devices the SMEs are familiar with. For example, only two respondents reported having experience with spatial disorientation. Thus, findings may be biased based on what the respondents are most familiar with.

Conclusions

Operator state detection through physiological measures has promise as a means to monitor an aviator's cognitive state in real-time. From this survey effort, fatigue / sleepiness, mental workload, and hypoxia emerged as the states most likely to be detectable in real-time from physiological measures. The physiological measures that emerged as most likely to be able to detect these, and a variety of other states, in real-time included EEG, eye tracking and pupillometry. However, based on the current state of technology, none of these measures can be considered ready to deploy in an operational setting. This conclusion is based on the amount of time required for setting up / calibrating, and the likelihood of withstanding various environmental considerations such as vibrations, extreme temperature, maintaining adhesion, and bright light / direct sunlight. Additionally, the required sensitivity and specificity needs further exploration, and will depend mainly on how the physiological signals are used (e.g., informing mission commander). In conclusion, while operator state detection appears to be feasible through the use of physiological measures, significant work remains in determining the combination of measures to use, how those measures will be used (in order to inform sensitivity and specificity requirements), and development and / or testing of sensors that can be deployed to an operational environment.

References

- Baruch, Y., & Holtom, B. C. (2008). Survey response rate levels and trends in organizational research. *Human Relations*, 61(8). <https://doi.org/10.1177/0018726708094863>
- Belton, I., MacDonald, A., Wright, G., & Hamlin, I. (2019). Improving the practical application of the Delphi method in group-based judgment: A six-step prescription for a well-founded and defensible process. *Technological Forecasting and Social Change*, 147, 72–82.
- Caldwell, J. A. (2005). Fatigue in aviation. *Travel Medicine and Infectious Disease*, 3(2), 85–96.
- Gunn, D. V., Warm, J. S., Nelson, W. T., Bolia, R. S., Schumsky, D. A., & Corcoran, K. J. (2005). Target acquisition with UAVs: Vigilance displays and advanced cuing interfaces. *Human Factors*, 47(3), 488–497.
- Hopkins, P. & Unger, M. (2017). What is a ‘subject-matter expert’? *Journal of Pipeline Engineering*, 16, 227–230.
- Kelley, A. M., Hayes, A. M., Bernhardt, K., & Mathews, C. (2020). *Sustaining performance and vigilance in extended UAS operations* (Report No. 2020-043). U.S. Army Aeromedical Research Laboratory.
- Matthews, G. & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology*, 55A(2), 659–686.
- Saxby, D. J., Matthews, G., Warm, J. S., Hitchcock, E. M., & Neubauer, C. (2013). Active and passive fatigue in simulated driving: Discriminating styles of workload regulation and their safety impacts, *Journal of Experimental Psychology: Applied*, 19(4), 287–300. doi: 10.1037/a0034386
- Stanton, N. A., Plant, K. L., Roberts, A. P., Allison, C. K., & Howell, M. (2019). Seeing through the mist: An evaluation of an iteratively designed heads-up display, using a simulated degraded visual environment, to facilitate rotary-wing pilot situation awareness and workload. *Cognition, Technology, and Work*.

Appendix A. Acronyms and Abbreviations

ECG HR	Electrocardiogram Heart Rate
ECG HRV	Electrocardiogram Heart Rate Variability
EDA	Electro-Dermal Activity
EEG	Electroencephalogram
EMG	Electromyography
EOG	Electrooculogram
FLRAA	Future Long Range Assault Aircraft
fMRI	Functional Magnetic Resonance Imaging
FNIRS	Functional Near-Infrared Spectroscopy
FVL	Future Vertical Lift
IMU	Internal Measurement Unit
OSM	Operator State Monitoring
PPG	Photoplethysmography
SME	Subject Matter Expert
SPG	Speckleplethysmography
USAARL	U.S. Army Aeromedical Research Laboratory

Appendix B. Initial Survey

Demographics Items

1. **Please provide your name (this is to ensure we send follow-on surveys only to those who complete the initial survey).**

2. **Please provide your email.**

3. **Please indicate what field your degree is in.**

4. **What degree do you hold?**

5. **When did you complete your degree?**

6. **Please indicate which physiological measures you have experience using within a research / professional environment.**

Please select all that apply.

- ☐ Actigraphy
- ☐ Body and/or face temperature
- ☐ Electrocardiography (ECG) heart rate
- ☐ Electrocardiography (ECG) heart rate variability
- ☐ Electroencephalography (EEG)
- ☐ Electromyography (EMG)
- ☐ Electrooculography (EOG)
- ☐ Excreted metabolites (e.g., cortisol)
- ☐ Eye tracking
- ☐ Functional near-infrared spectroscopy (fNIRS)
- ☐ Photoplethysmography (PPG)
- ☐ Pupillometry
- ☐ Respiration Dynamics
- ☐ Other, please specify

7. Please indicate which operator states you have research experience with.

- ☐ Autonomic tone
- ☐ Distraction
- ☐ Fatigue / sleepiness
- ☐ Hyperoxia
- ☐ Hypoxia
- ☐ Mental fatigue
- ☐ Mental workload
- ☐ Spatial disorientation
- ☐ Other, please specify

8. Please indicate which operator states you are familiar with through knowledge of the literature.

- ☐ Autonomic tone
- ☐ Distraction
- ☐ Fatigue / sleepiness
- ☐ Hyperoxia
- ☐ Hypoxia
- ☐ Mental fatigue
- ☐ Mental workload
- ☐ Spatial disorientation
- ☐ Other, please specify

9. Please indicate the number of years you have direct experience working with each sensor type.

Actigraphy	
Body and/or face temperature	
Electrocardiography (ECG) heart rate	
Electrocardiography (ECG) heart rate variability	
Electroencephalography (EEG)	
Electromyography (EMG)	
Electrooculography (EOG)	
Excreted metabolites (e.g., cortisol)	
Eye tracking	
Functional near-infrared spectroscopy (fNIRS)	
Photoplethysmography (PPG)	
Pupillometry	
Respiration dynamics	
Other, please state	

10. What devices do you use / what do you plan to use ?

Please indicate device and physiological measurement.

11. Which devices have you had the best experiences with, and why?

12. Which devices have you had the worst experiences with, and why?

Operator States Items

- 13. Please indicate the likelihood of each operator state being a candidate for detection in real-time through physiological measures. Please select one option per row.**

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Autonomic tone	☐	☐	☐	☐	☐	☐
Distraction	☐	☐	☐	☐	☐	☐
Fatigue / sleepiness	☐	☐	☐	☐	☐	☐
Hypoxia	☐	☐	☐	☐	☐	☐
Mental fatigue	☐	☐	☐	☐	☐	☐
Mental workload	☐	☐	☐	☐	☐	☐
Spatial disorientation	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

- 14. If you selected Other in previous question, please indicate.**

- 15. Please indicate the likelihood of each physiological measure for being capable of detecting MENTAL WORKLOAD.**
Please select one option per row.

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Actigraphy	☐	☐	☐	☐	☐	☐
Body and/or face temperature	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate variability	☐	☐	☐	☐	☐	☐
Electroencephalography (EEG)	☐	☐	☐	☐	☐	☐
Electromyography (EMG)	☐	☐	☐	☐	☐	☐
Electrooculography (EOG)	☐	☐	☐	☐	☐	☐
Excreted metabolites (e.g., cortisol)	☐	☐	☐	☐	☐	☐
Eye tracking	☐	☐	☐	☐	☐	☐
Functional near-infrared spectroscopy (fNIRS)	☐	☐	☐	☐	☐	☐
Photoplethysmography (PPG)	☐	☐	☐	☐	☐	☐
Pupillometry	☐	☐	☐	☐	☐	☐
Respiration dynamics	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

- 16. If you selected Other in previous question, please indicate.**

- 17. Please indicate the likelihood of each physiological measure for being capable of detecting FATIGUE/ SLEEPINESS. Please select one option per row.**

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Actigraphy	☐	☐	☐	☐	☐	☐
Body and/or face temperature	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate variability	☐	☐	☐	☐	☐	☐
Electroencephalography (EEG)	☐	☐	☐	☐	☐	☐
Electromyography (EMG)	☐	☐	☐	☐	☐	☐
Electrooculography (EOG)	☐	☐	☐	☐	☐	☐
Excreted metabolites (e.g., cortisol)	☐	☐	☐	☐	☐	☐
Eye tracking	☐	☐	☐	☐	☐	☐
Functional near-infrared spectroscopy (fNIRS)	☐	☐	☐	☐	☐	☐
Photoplethysmography (PPG)	☐	☐	☐	☐	☐	☐
Pupillometry	☐	☐	☐	☐	☐	☐
Respiration dynamics	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

- 18. If you selected Other in previous question, please indicate.**

**19. Please indicate the likelihood of each physiological measure for being capable of detecting DISTRACTION.
Please select one option per row.**

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Actigraphy	☐	☐	☐	☐	☐	☐
Body and/or face temperature	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate variability	☐	☐	☐	☐	☐	☐
Electroencephalography (EEG)	☐	☐	☐	☐	☐	☐
Electromyography (EMG)	☐	☐	☐	☐	☐	☐
Electrooculography (EOG)	☐	☐	☐	☐	☐	☐
Excreted metabolites (e.g., cortisol)	☐	☐	☐	☐	☐	☐
Eye tracking	☐	☐	☐	☐	☐	☐
Functional near-infrared spectroscopy (fNIRS)	☐	☐	☐	☐	☐	☐
Photoplethysmography (PPG)	☐	☐	☐	☐	☐	☐
Pupillometry	☐	☐	☐	☐	☐	☐
Respiration dynamics	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

20. If you selected Other in previous question, please indicate.

21. Please indicate the likelihood of each physiological measure for being capable of detecting MENTAL FATIGUE.
Please select one option per row.

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Actigraphy	☐	☐	☐	☐	☐	☐
Body and/or face temperature	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate variability	☐	☐	☐	☐	☐	☐
Electroencephalography (EEG)	☐	☐	☐	☐	☐	☐
Electromyography (EMG)	☐	☐	☐	☐	☐	☐
Electrooculography (EOG)	☐	☐	☐	☐	☐	☐
Excreted metabolites (e.g., cortisol)	☐	☐	☐	☐	☐	☐
Eye tracking	☐	☐	☐	☐	☐	☐
Functional near-infrared spectroscopy (fNIRS)	☐	☐	☐	☐	☐	☐
Photoplethysmography (PPG)	☐	☐	☐	☐	☐	☐
Pupillometry	☐	☐	☐	☐	☐	☐
Respiration dynamics	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

22. If you selected Other in previous question, please indicate.

**23. Please indicate the likelihood of each physiological measure for being capable of detecting HYPOXIA.
Please select one option per row.**

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Actigraphy	☐	☐	☐	☐	☐	☐
Body and/or face temperature	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate variability	☐	☐	☐	☐	☐	☐
Electroencephalography (EEG)	☐	☐	☐	☐	☐	☐
Electromyography (EMG)	☐	☐	☐	☐	☐	☐
Electrooculography (EOG)	☐	☐	☐	☐	☐	☐
Excreted metabolites (e.g., cortisol)	☐	☐	☐	☐	☐	☐
Eye tracking	☐	☐	☐	☐	☐	☐
Functional near-infrared spectroscopy (fNIRS)	☐	☐	☐	☐	☐	☐
Photoplethysmography (PPG)	☐	☐	☐	☐	☐	☐
Pupillometry	☐	☐	☐	☐	☐	☐
Respiration dynamics	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

24. If you selected Other in previous question, please indicate.

25. Please indicate the likelihood of each physiological measure for being capable of detecting SPATIAL DISORIENTATION. Please select one option per row.

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Actigraphy	☐	☐	☐	☐	☐	☐
Body and/or face temperature	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate variability	☐	☐	☐	☐	☐	☐
Electroencephalography (EEG)	☐	☐	☐	☐	☐	☐
Electromyography (EMG)	☐	☐	☐	☐	☐	☐
Electrooculography (EOG)	☐	☐	☐	☐	☐	☐
Excreted metabolites (e.g., cortisol)	☐	☐	☐	☐	☐	☐
Eye tracking	☐	☐	☐	☐	☐	☐
Functional near-infrared spectroscopy (fNIRS)	☐	☐	☐	☐	☐	☐
Photoplethysmography (PPG)	☐	☐	☐	☐	☐	☐
Pupillometry	☐	☐	☐	☐	☐	☐
Respiration dynamics	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

26. If you selected Other in previous question, please indicate.

**27. Please indicate the likelihood of each physiological measure for being capable of detecting AUTONOMIC TONE.
Please select one option per row.**

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Actigraphy	☐	☐	☐	☐	☐	☐
Body and/or face temperature	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate variability	☐	☐	☐	☐	☐	☐
Electroencephalography (EEG)	☐	☐	☐	☐	☐	☐
Electromyography (EMG)	☐	☐	☐	☐	☐	☐
Electrooculography (EOG)	☐	☐	☐	☐	☐	☐
Excreted metabolites (e.g., cortisol)	☐	☐	☐	☐	☐	☐
Eye tracking	☐	☐	☐	☐	☐	☐
Functional near-infrared spectroscopy (fNIRS)	☐	☐	☐	☐	☐	☐
Photoplethysmography (PPG)	☐	☐	☐	☐	☐	☐
Pupillometry	☐	☐	☐	☐	☐	☐
Respiration dynamics	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

28. If you selected Other in previous question, please indicate.

29. **Please indicate the likelihood of each physiological measure for being capable of detecting OTHER (please state below). Please select one option per row.**

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Actigraphy	☐	☐	☐	☐	☐	☐
Body and/or face temperature	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate variability	☐	☐	☐	☐	☐	☐
Electroencephalography (EEG)	☐	☐	☐	☐	☐	☐
Electromyography (EMG)	☐	☐	☐	☐	☐	☐
Electrooculography (EOG)	☐	☐	☐	☐	☐	☐
Excreted metabolites (e.g., cortisol)	☐	☐	☐	☐	☐	☐
Eye tracking	☐	☐	☐	☐	☐	☐
Functional near-infrared spectroscopy (fNIRS)	☐	☐	☐	☐	☐	☐
Photoplethysmography (PPG)	☐	☐	☐	☐	☐	☐
Pupillometry	☐	☐	☐	☐	☐	☐
Respiration dynamics	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

30. If applicable, please indicate OTHER is previous question.

31. **Based on your own research, please indicate which physiological measure(s) you have found to be the most effective in discriminating operator state.**

Please provide a citation (if available).

Physiological Measure	
Operator State	
Citation	

Device Set Up Items

- 32. Please rank the physiological measures in order of quickest to slowest current application time (i.e., time to place sensor). Please rank only the ones you have experience with. Only use each number once and skip the items you are not familiar with.**

Rank the items below, using numeric values starting with 1.

Actigraphy	
Body and/or face temperature	
Electrocardiography (ECG) heart rate	
Electrocardiography (ECG) heart rate variability	
Electroencephalography (EEG)	
Electromyography (EMG)	
Electrooculography (EOG)	
Excreted metabolites (e.g., cortisol)	
Eye tracking	
Functional near-infrared spectroscopy (fNIRS)	
Photoplethysmography (PPG)	
Pupillometry	
Respiration dynamics	
Other, Please state	

- 33. Based on the anticipated state of the technology in 5 years, please rank the physiological measures in order of quickest to slowest application time (i.e., time to place sensor). Please rank only the ones you have experience with. Only use each number once and skip the items you are not familiar with.**

Rank the items below, using numeric values starting with 1.

Actigraphy	
Body and/or face temperature	
Electrocardiography (ECG) heart rate	
Electrocardiography (ECG) heart rate variability	
Electroencephalography (EEG)	
Electromyography (EMG)	
Electrooculography (EOG)	
Excreted metabolites (e.g., cortisol)	
Eye tracking	
Functional near-infrared spectroscopy (fNIRS)	
Photoplethysmography (PPG)	
Pupillometry	
Respiration dynamics	
Other, Please state	

34. **Based on your own experiences, please rank the physiological measures in order of ease of replacing, cleaning, or otherwise maintaining sensors to ensure an effective signal quality. Please refer to a standard application. Please rank only the ones you have experience with. Only use each number once and skip the items you are not familiar with.**

Rank the items below, using numeric values starting with 1.

Actigraphy	
Body and/or face temperature	
Electrocardiography (ECG) heart rate	
Electrocardiography (ECG) heart rate variability	
Electroencephalography (EEG)	
Electromyography (EMG)	
Electrooculography (EOG)	
Excreted metabolites (e.g., cortisol)	
Eye tracking	
Functional near-infrared spectroscopy (fNIRS)	
Photoplethysmography (PPG)	
Pupillometry	
Respiration dynamics	
Other, Please state	

35. For each physiological measure that you have experience using, please indicate the minimal acceptable criteria for calibration (e.g., at least 1 minute of baseline data). If the measure you have experience with requires no calibration, please indicate so.

Please state which measure your answer is referring to.

36. For each physiological measure that you have experience using, please indicate the level of expertise required to apply the sensor.

37. For each physiological measure that you have experience using, please indicate whether it is possible for a person to apply the sensor without assistance.

38. For each physiological measure that you have experience using, please indicate the level of expertise required to evaluate and/or trouble shoot common issues (e.g., impedance issues, calibration issues). Please indicate which common issue you are referencing.

39. The rotary-wing flight environment routinely encounters significant vibration. Based on the current state of the technology, how likely do you see each of the sensors as being resilient to the vibrations of the flight environment?

Please respond to only those you have experience with.

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Actigraphy	☐	☐	☐	☐	☐	☐
Body and/or face temperature	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate variability	☐	☐	☐	☐	☐	☐
Electroencephalography (EEG)	☐	☐	☐	☐	☐	☐
Electromyography (EMG)	☐	☐	☐	☐	☐	☐
Electrooculography (EOG)	☐	☐	☐	☐	☐	☐
Excreted metabolites (e.g., cortisol)	☐	☐	☐	☐	☐	☐
Eye tracking	☐	☐	☐	☐	☐	☐
Functional near-infrared spectroscopy (fNIRS)	☐	☐	☐	☐	☐	☐
Photoplethysmography (PPG)	☐	☐	☐	☐	☐	☐
Pupillometry	☐	☐	☐	☐	☐	☐
Respiration dynamics	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

40. If you selected Other in previous question, please indicate.

Compatibility with the Environment Items

41. Based on the current state of the technology, how likely do you see each of the sensors being able to withstand extreme temperatures (ranging -13 to 49° C) of the flight environment?

Please respond to only those you have experience with.

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Actigraphy	☐	☐	☐	☐	☐	☐
Body and/or face temperature	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate variability	☐	☐	☐	☐	☐	☐
Electroencephalography (EEG)	☐	☐	☐	☐	☐	☐
Electromyography (EMG)	☐	☐	☐	☐	☐	☐
Electrooculography (EOG)	☐	☐	☐	☐	☐	☐
Excreted metabolites (e.g., cortisol)	☐	☐	☐	☐	☐	☐
Eye tracking	☐	☐	☐	☐	☐	☐
Functional near-infrared spectroscopy (fNIRS)	☐	☐	☐	☐	☐	☐
Photoplethysmography (PPG)	☐	☐	☐	☐	☐	☐
Pupillometry	☐	☐	☐	☐	☐	☐
Respiration dynamics	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

42. If you selected Other in previous question, please indicate.

43. Based on the current state of the technology, how likely do you see each of the sensors being able to withstand bright environments / direct sunlight?

Please respond to only those you have experience with.

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Actigraphy	☐	☐	☐	☐	☐	☐
Body and/or face temperature	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate variability	☐	☐	☐	☐	☐	☐
Electroencephalography (EEG)	☐	☐	☐	☐	☐	☐
Electromyography (EMG)	☐	☐	☐	☐	☐	☐
Electrooculography (EOG)	☐	☐	☐	☐	☐	☐
Excreted metabolites (e.g., cortisol)	☐	☐	☐	☐	☐	☐
Eye tracking	☐	☐	☐	☐	☐	☐
Functional near-infrared spectroscopy (fNIRS)	☐	☐	☐	☐	☐	☐
Photoplethysmography (PPG)	☐	☐	☐	☐	☐	☐
Pupillometry	☐	☐	☐	☐	☐	☐
Respiration dynamics	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

44. If you selected Other in previous question, please indicate.

45. Based on the current state of the technology, how likely do you see each of the sensors being able to maintain adhesion / sensor sensitivity after vigorous exercise?

Please respond to only those you have experience with.

	Very Unlikely	Unlikely	Neutral	Likely	Very likely	Unsure
Actigraphy	☐	☐	☐	☐	☐	☐
Body and/or face temperature	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate	☐	☐	☐	☐	☐	☐
Electrocardiography (ECG) heart rate variability	☐	☐	☐	☐	☐	☐
Electroencephalography (EEG)	☐	☐	☐	☐	☐	☐
Electromyography (EMG)	☐	☐	☐	☐	☐	☐
Electrooculography (EOG)	☐	☐	☐	☐	☐	☐
Excreted metabolites (e.g., cortisol)	☐	☐	☐	☐	☐	☐
Eye tracking	☐	☐	☐	☐	☐	☐
Functional near-infrared spectroscopy (fNIRS)	☐	☐	☐	☐	☐	☐
Photoplethysmography (PPG)	☐	☐	☐	☐	☐	☐
Pupillometry	☐	☐	☐	☐	☐	☐
Respiration dynamics	☐	☐	☐	☐	☐	☐
Other (please list below)	☐	☐	☐	☐	☐	☐

46. If you selected Other in previous question, please indicate.

Signal Quality and Reliability Items

- 47. Using the following definition of sensitivity: "proportion of truly positive cases that correspond to positive results"**

Regarding the ability of a system or device to collect physiological signals, how sensitive does the system need to be in rejecting unreliable and retaining reliable signals?

- 48. Using the following definition of specificity: "proportion of truly negative cases that correspond to negative test results"**

Regarding the ability of a system or device to collect physiological signals, how specific does the system need to be in rejecting unreliable and retaining reliable signals?

- 49. Assuming some of the candidate sensor systems include pre-developed algorithms for identifying operator state, what is an acceptable error rate for correctly identifying operator state?**

50. **Please list which systems, that you commonly use, rely on patented/protected algorithms for analysis/interpretation/signal pre-processing.**

51. **What is the level of expertise required to process and evaluate the system?**

Please specify a level of training/education, from No Training to Post-Doctoral expertise.

Device Measurements Item

52. **Based on your personal experience in working with physiological data, please indicate the physical characteristics of pilots that may impact the signal quality and/or ability to secure the sensor to the body. Please include only those metrics you have direct experience in using.**



Appendix C. Second Survey

1. **Assuming a hypothetical physiological sensor system has been developed and integrated into the aircraft, within this use-case the information regarding operator state is being relayed to a mission commander. The mission commander is using this information to make determinations regarding the aviator's readiness level to continue mission. How sensitive would such a system need to be in determining operator state?**

2. **Assuming a hypothetical physiological sensor system has been developed and integrated into the aircraft, within this use-case the information regarding operator state is being used to trigger an adaptive automation system. The adaptive automation system is able to take over tasks (such as maintaining a specified flight route) for the aviator. How sensitive would such a system need to be in determining operator state?**

3. We received varying responses regarding whether some of the operator states are candidates for detection in real-time. These are indicated below. We invite you to provide any justification for why or why not a specific state (or states) may **not** be feasible for detection in real-time. Alternatively, provide justification for why you may be **unsure** of whether or not a state (or states) are feasible for detection in real-time.

Operator State	Very Unlikely	Unlikely	Neutral	Likely	Very Likely	Unsure
Autonomic Tone		12.50 %	4.20 %	20.80 %	16.70 %	33.33 %
Mental Workload			29.20 %	25.00 %	29.20 %	8.30 %
Fatigue / Sleepiness			4.20 %	33.33 %	54.20 %	
Distraction		4.20 %	12.50 %	41.70 %	16.70 %	12.50 %
Mental Fatigue		8.30 %	20.80 %	41.70 %	12.50 %	
Hypoxia			8.30 %	16.70 %	33.33 %	33.33 %
Spatial Disorientation		4.20 %	25.00 %	20.80 %	8.30 %	29.20 %

Note, you may choose to skip this item.

4. Likelihood of Each Physiological Measure Capable of Detecting Mental Workload.

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	33.3 %	12.5 %	8.3 %	12.5 %	12.5 %	4.2 %
Body/face Temperature	25.0 %	4.2 %	20.8 %	16.7 %	20.8 %	
ECG – Heart Rate	16.7 %	4.2 %	16.7 %	29.2 %	16.7 %	8.3 %
ECG – Heart Rate Variability	12.5 %		12.5 %	16.7 %	37.5 %	12.5 %
Electroencephalography	8.3 %			8.3 %	29.2 %	41.7 %
Electromyography	25.0 %		25.0 %	16.7 %	20.8 %	
Electrooculogram	12.5 %		8.3 %	25.0 %	29.2 %	12.5 %
Excreted Metabolites	16.7 %	12.5 %	25.0 %	12.5 %	16.7 %	4.2 %
Eye Tracking			4.2 %	16.7 %	50.0 %	16.7 %
fNIRS	29.2 %		4.2 %	12.5 %	20.8 %	20.8 %
Photoplethysmography	29.2 %	4.2 %	12.5 %	25.0 %	16.7 %	
Pupillometry	4.2 %		8.3 %	12.5 %	45.8 %	20.8 %
Respiration Dynamics	25.0 %		29.2 %	25.0 %	4.2 %	4.2 %

Other responses received include: head movement (very likely); electrodermal activity (neutral)

Please provide justification for why or why not a specific measure (or measures) may not be feasible for detecting Mental Workload. Or, explain why you are unsure.

Note, you may choose to skip this item.

5. **Likelihood of Each Physiological Measure Capable of Detecting Fatigue / Sleepiness.**

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	16.7 %			16.7 %	33.3 %	16.7 %
Body/face Temperature	29.2%		4.2%	20.8%	25.0%	8.3%
ECG – Heart Rate	8.3%		20.8%	20.8%	41.7%	
ECG – Heart Rate Variability	12.5%		16.7%	12.5%	37.5%	12.5%
Electroencephalography	12.5%			8.3%	25.0%	37.5%
Electromyography	16.7%		25.0%	20.8%	20.8%	4.2%
Electrooculogram	20.8%		4.2%	8.3%	29.2%	25.0%
Excreted Metabolites	25.0%	8.3%	25.0%	20.8%	12.5%	
Eye Tracking	2.4%			16.7%	20.8%	37.5%
fNIRS	25.0%		4.2%	16.7%	29.2%	12.5%
Photoplethysmography	29.2%	4.2%	8.3%	29.2%	12.5%	
Pupillometry	8.3%		8.3%	12.5%	33.3%	25.0%
Respiration Dynamics	12.5%		25.0%	33.3%	16.7%	

Other responses included: head-worn infrared cameras for eye area based computer vision (very likely); electrodermal activity (neutral)

Please provide justification for why or why not a specific measure (or measures) may not be feasible for detecting Fatigue/ Sleepiness. Or, explain why you are unsure.

Note, you may choose to skip this item.

6. Likelihood of Each Physiological Measure Capable of Detecting Distraction.

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	29.2 %	12.5 %	8.3 %	20.8 %	12.5 %	
Body/face Temperature	37.5%	12.5%	12.5%	20.8%		
ECG – Heart Rate	16.7%	12.5%	29.2%	29.2%	4.2%	
ECG – Heart Rate Variability	20.8%	12.5%	20.8%	25.0%	12.5%	
Electroencephalography	20.8%	8.3%		16.7%	20.8%	20.8%
Electromyography	25.0%	8.3%	29.2%	20.8%	4.2%	
Electrooculogram	20.8%	4.2%		12.5%	41.7%	8.3%
Excreted Metabolites	37.5%	25.0%	16.7%	8.3%	4.2%	
Eye Tracking	4.2%			8.3%	33.3%	41.7%
fNIRS	33.3%	8.3%	8.3%	20.8%	4.2%	12.5%
Photoplethysmography	45.8%	12.5%	16.7%	12.5%		
Pupillometry	12.5%	4.2%		29.2%	41.7%	
Respiration Dynamics	33.3%	8.3%	25.0%	16.7%	4.2%	

Other responses included: electrodermal activity (very unlikely)

Please provide justification for why or why not a specific measure (or measures) may not be feasible for detecting Distraction. Or, explain why you are unsure.

Note, you may choose to skip this item.

7. Likelihood of Each Physiological Measure Capable of Detecting Mental Fatigue.

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	33.3 %	4.2 %	4.2 %	20.8 %	16.7 %	4.2 %
Body/face Temperature	33.3%		29.2%	8.3%	12.5%	
ECG – Heart Rate	12.5%		37.5%	25.0%	16.7%	
ECG – Heart Rate Variability	16.7%		20.8%	25.0%	25.0%	4.2%
Electroencephalography	8.3%			16.7%	25.0%	37.5%
Electromyography	16.7%	4.2%	20.8%	41.7%	4.2%	
Electrooculogram	16.7%		8.3%	33.3%	16.7%	12.5%
Excreted Metabolites	25.0%	12.5%	20.8%	20.8%	12.5%	
Eye Tracking	8.3%			29.2%	29.2%	20.8%
fNIRS	25.0%		8.3%	12.5%	29.2%	12.5%
Photoplethysmography	37.5%	4.2%	16.7%	20.8%	8.3%	
Pupillometry	8.3%			29.2%	25.0%	25.0%
Respiration Dynamics	25.0%	4.2%	20.8%	33.3%	8.3%	

Other responses included: electrodermal activity (unlikely)

Please provide justification for why or why not a specific measure (or measures) may not be feasible for detecting Mental Fatigue. Or, explain why you are unsure.

Note, you may choose to skip this item.

8. Likelihood of Each Physiological Measure Capable of Detecting Hypoxia.

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	45.8 %	12.5 %	8.3 %	16.7 %		4.2 %
Body/face Temperature	45.8%	8.3%	4.2%	12.5%	12.5%	
ECG – Heart Rate	37.5%	4.2%	8.3%	12.5%	12.5%	12.5%
ECG – Heart Rate Variability	45.8%	8.3%	4.2%	8.3%	16.7%	8.3%
Electroencephalography	41.7%	8.3%	8.3%	16.7%	12.5%	4.2%
Electromyography	41.7%	8.3%	12.5%	16.7%	8.3%	
Electrooculogram	45.8%	4.2%	4.2%	8.3%	25.0%	
Excreted Metabolites	45.8%	12.5%	20.8%		12.5%	
Eye Tracking	45.8%	4.2%		16.7%	20.8%	
fNIRS	41.7%		4.2%	4.2%	12.5%	25.0%
Photoplethysmography	41.7%		8.3%		8.3%	29.2%
Pupillometry	45.8%	4.2%	4.2%	12.5%	16.7%	
Respiration Dynamics	33.3%			4.2%	25.0%	20.8%

Other responses included: electrodermal activity (unlikely); inhaled/exhaled gas analysis (likely); Oximetry – SpO2% (dual wavelength variant of PPG – note, not all oximetry systems are equal) (very likely)

Please provide justification for why or why not a specific measure (or measures) may not be feasible for detecting Hypoxia. Or, explain why you are unsure.

Note, you may choose to skip this item.

9. Likelihood of Each Physiological Measure Capable of Detecting Spatial Disorientation.

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	45.8 %	8.3 %	8.3 %	4.2 %	16.7 %	
Body/face Temperature	41.7%	4.2%	16.7%	12.5%	8.3%	
ECG – Heart Rate	41.7%		12.5%	25.0%	8.3%	
ECG – Heart Rate Variability	41.7%		16.7%	16.7%	12.5%	
Electroencephalography	33.3%		4.2%	12.5%	25.0%	8.3%
Electromyography	37.5%		20.8%	16.7%	8.3%	
Electrooculogram	41.7%		4.2%	4.2%	33.3%	
Excreted Metabolites	37.5%	16.7%	8.3%	20.8%	4.2%	
Eye Tracking	25.0%			4.2%	37.5%	16.7%
fNIRS	37.5%		4.2%	16.7%	20.8%	4.2%
Photoplethysmography	41.7%	4.2%	16.7%	20.8%		
Pupillometry	37.5%		8.3%	12.5%	20.8%	4.2%
Respiration Dynamics	37.5%	4.2%	12.5%	29.2%	4.2%	

Other responses included: pupillary light response (very likely)

Please provide justification for why or why not a specific measure (or measures) may not be feasible for detecting Spatial Disorientation. Or, explain why you are unsure

Note, you may choose to skip this item.

10. Likelihood of Each Physiological Measure Capable of Detecting Autonomic Tone.

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	50.0 %	4.2 %	12.5 %	12.5 %	4.2 %	4.2 %
Body/face Temperature	37.5%		8.3%	12.5%	29.2%	
ECG – Heart Rate	25.0%		4.2%	1.25%	37.5%	12.5%
ECG – Heart Rate Variability	25.0%			4.2%	29.2%	33.3%
Electroencephalography	45.8%		12.5%	12.5%	12.5%	4.2%
Electromyography	41.7%		20.8%	4.2%	16.7%	4.2%
Electrooculogram	45.8%		12.5%	12.5%	16.7%	
Excreted Metabolites	33.3%	4.2%	8.3%	4.2%	20.8%	20.8%
Eye Tracking	37.5%		12.5%	12.5%	25.0%	
fNIRS	41.7%	4.2%	12.5%	16.7%	12.5%	
Photoplethysmography	41.7%		8.3%	12.5%	12.5%	12.5%
Pupillometry	41.7%		12.5%	12.5%	12.5%	8.3%
Respiration Dynamics	37.5%		12.5%	12.5%	20.8%	8.3%

Other responses included: electrodermal activity (very likely)

Please provide justification for why or why not a specific measure (or measures) may not be feasible for detecting Autonomic Tone. Or, explain why you are unsure.

Note, you may choose to skip this item.

11. In the past year, what would you consider the most important finding or paper related to operator state monitoring? Please indicate why.



Appendix D. Additional Comments

Signal Sensitivity – Additional Comments

Other comments included:

- That really depends on what the signal will be used for. If the signal will be further interpreted by another human operator, such as a mission commander, who can use context and previous experience as a further filter, then it might be OK for the system to be fairly unreliable (maybe as high as 20-30% unreliable). On the other hand, if the signal will be used to automatically activate some kind of aid system, then it needs to be very, very good at rejecting unreliable signals (maybe 1-5% unreliable).
- This is very dependent on what application the signals are being used for, trade-off of type I vs II error rates, etc. If it's for detecting workload state it's relatively low sensitivity needed, but if it's detecting intent or commands (e.g., fire my weapon) then very high. So I guess I'll just give a middle-range of 0.8.
- Unable to comment based on experimental work to date.
- Significantly above chance.
- Depends on the application. In general, I'd aim to capture at least 70% of the positives.
- Regarding eye tracking, "Drift is an issue with gaze-based studies over a long period of time, but you can insert additional calibrations to correct drift."

Other Measurements

- EDA (skin conductance level (SCL)/skin conductance response (SCR)), minimal noise, at least 1 min of data. 70%, high sensitivity ($n = 1$)
- Respiratory flow for peak flow, tidal volume, minute ventilation: Minimal noise across at least 6 consecutive breaths; 100%; No drift in any zero offset, and preferably 0 offset. ($n = 1$)
- Mask pressure for peak pressures, pressures at peak flow: minimal noise across at least 6 consecutive breaths; 100%; No drift in any zero offset, and preferably 0 offset; time-synchronous with flow to within 0.005 s. ($n = 1$)
- Gas composition in the mask to extract end-tidal values: Minimal noise across at least 6 consecutive breaths; Inspired values must match the supplied gas composition for at least 5 of the 6 breaths; maxima clean on at least 5 of 6 breaths; breaths with unclear maxima must be omitted. ($n = 1$)
- Finger blood pressure: minimal noise at least 1 min of data; 90% ($n = 1$)

Signal Specificity – Additional Comments

Other responses:

- That really depends on what the signal will be used for. If the signal will be further interpreted by another human operator, such as a mission commander, who can use context and previous experience as a further filter, then it might be OK for the system to be fairly unreliable (maybe as high as 20-30% unreliable). On the other hand, if the signal will be used

to automatically activate some kind of aid system, then it needs to be very, very good at rejecting unreliable signals (maybe 1-5% unreliable).

- This is very dependent on what application the signals are being used for, trade-off of type I vs. type II error rates, etc. If it's for detecting workload state it's relatively low sensitivity needed, but if it's detecting intent or commands (e.g., fire my weapon) then very high. So I guess I'll just give a middle-range of 0.8.
- Less than 5% of signal rejection.
- Unable to comment based on experimental work to date.
- 95%
- Same as above (in reference to the respiratory flow, mask pressure, gas composition, finger blood pressure)
- In ECG HRV, unreliable signals due to noise (rather than physiological abnormalities) are very easy to detect so the system does not need to be specific; instead, the analyst should have the expertise to apply standard data processing approaches to characterize ECG recordings (e.g., to exclude noisy periods).
- Significantly above chance
- Depends on the application. In general, I'd aim to capture at least 70% of the negatives.
- All need to reject unreliable signals.

Other measurements:

- EDA, SCL/SCR, medium specificity ($n = 1$)

Physical Characteristics - Participant Comments

Based on your personal experience in working with physiological data, please indicate the physical characteristics of pilots that may impact the signal quality and/or ability to secure the sensor to the body. Please include only those metrics you have direct experience in using.

Table D1. Physical Characteristics Respondent Comments

Physiological Measure	Respondent Comments on Physical Characteristic(s)
Actigraphy (<i>n</i> = 2)	▪ Movement disorders (<i>n</i> = 1) ▪ None really (<i>n</i> = 1)
Temperature (<i>n</i> = 1)	▪ Body hair (<i>n</i> = 1)
ECG HR (<i>n</i> = 6)	▪ Body hair (<i>n</i> = 5) ▪ High body fat (<i>n</i> = 1)
ECG HRV (<i>n</i> = 6)	▪ High body fat (<i>n</i> = 1) ▪ Body hair (<i>n</i> = 4) ▪ For Heart Rate Variability (ECG), electrodes can be integrated with physiological garments to ensure proper contact and location for collecting data (from my experience with these garments, it may eliminate the need for shaving chest hair). (<i>n</i> = 1)
EEG (<i>n</i> = 5)	▪ Scalp hair and skin (<i>n</i> = 1) ▪ Thick hair, oily or dry scalp (<i>n</i> = 1) ▪ Use of headphones and headrest. Difficult to integrate with everything. Headphones could interfere with EEG signals (shielded cables are needed), the contact between the head and headrest could cause several artefacts. (<i>n</i> = 1) ▪ Sweat can be a big issue for EEG (<i>n</i> = 1) ▪ Hair, scalp impedance (<i>n</i> = 1)
EMG (<i>n</i> = 1)	▪ Botox or other dermal injections (<i>n</i> = 1)
EOG (<i>n</i> = 2)	▪ Botox or other dermal injections (<i>n</i> = 1) ▪ None really (<i>n</i> = 1)
Eye Tracking (<i>n</i> = 7)	▪ Eye shape, color, presence of glasses or contacts (<i>n</i> = 1) ▪ Dark features, eye shape, nystagmus, or other oculomotor disorders (<i>n</i> = 1) ▪ Can be affected by glasses/contact lenses (<i>n</i> = 1) ▪ Some people are easier to track than others. Droppy eyeliner etc. can cause problems. Contact lenses are usually ok, but thick glasses can cause problems. (<i>n</i> = 1) ▪ Eye shape and vision correction (<i>n</i> = 1) ▪ Eye color, glasses, eye shape. Some systems have issues capturing light eyes. Some glasses will cause issues that cause the tracker to not pick up a participant's eye. Some trackers will not pick up an individual based on their eye shape. (<i>n</i> = 1) ▪ Eyeglasses. Glasses may obscure eyes (depending on angle of eye tracker) or may distort infrared light, or may interfere with calibration.

	Eye shape / eye lid. Some eye shapes make [sic] cause eye lid to overlap with the pupil. Movement. In desktop setups if a participant moves out of the calibration or view space of the eye tracker the data will be compromised. ($n = 1$)
Metabolites ($n = 1$)	High or low saliva volume ($n = 1$)
fNIRS ($n = 4$)	- Scalp hair and skin ($n = 1$) - High sweating, thick hair ($n = 1$) - Skin tone, hair color and thickness of the scalp. Different skin tones impact impedance. ($n = 1$) - Hair ($n = 1$)
PPG ($n = 3$)	- Interference with gestures and actions if placed on fingers ($n = 1$) - Hair at site ($n = 1$) - Hair ($n = 1$)
Pupillometry ($n = 3$)	- Older age, high expertise ($n = 1$) - Can be affected by large eyelashes, need for glasses/contact lenses ($n = 1$) - Eye shape and vision correction ($n = 1$)
Respiration Dynamics ($n = 1$)	None really ($n = 1$)

Other comments:

- In general, Aircrew Flight Equipment can impact sensor placement
- Electromagnetic (EM) radiation, head movement, sweat
- IMU: None
- Speech: None (ability to speak)
- All above (*IMU, pupillometry, eye tracking, speech*): Assumes a head-worn device (e.g., helmet, glasses)
- PPG and EDA are fully compatible if embedded in a wristband.
- Finger blood pressure is not an option in pilots who must be able to move their hands unimpeded and who also often must wear gloves.
- Gas sensors for exhaled gas and mask pressure sensors necessitate alterations in the mask. Inspiratory flow sensors can sit on the inlet tube. None of those touch the pilot.
- ECG HRV: There are a variety of methods to counter G forces that interact directly with autonomic function (intentionally). It would be necessary to acquire information on the application of these methods (e.g., use of Valsalva maneuver and pressure suits) in real-time in order to properly interpret changes in HRV and inferred autonomic function.
- EDA: Hair at site, ability to sweat.
- Hair - big interference with light-based sensors like PPG and fNIRS. Need to be clear of hair where the sensor is adhered. Also, hair can become an issue for adhesive-based sensors (pain)
- Skin color - sensors also need to be calibrated to account for a variety of skin tones.
- Muscle tone - may affect how tightly/securely a sensor is attached to a particular body part, like the bicep, or the chest area. individual differences may require individual

fittings of sensors and how tight or loose they need to be, and what is required to keep them in place.

- Sweat - sensors need to maintain position even under high sweat conditions.
- The flight suit and the helmet create constraints on mounting locations and they can lead to motion artifacts. The real flight (RF) environment of the cockpit can also create challenges in reliable, real-time wireless communication

Appendix E. Second Survey Results

Item 1

In the first round we asked questions regarding sensitivity. In this round we would like to use two use-case scenarios. Using the following definition of sensitivity, "proportion of truly positive cases that correspond to positive results," please respond to the following use-cases.

Use-case 1: Assuming a hypothetical physiological sensor system has been developed and integrated into the aircraft, within this use-case the information regarding operator state is being relayed to a mission commander. The mission commander is using this information to make determinations regarding the aviator's readiness level to continue mission. How sensitive would such a system need to be in determining operator state?

Responses

- "Operator state" is a little ambiguous and can include multiple sensors. Most human physiology is dynamic and constantly changing. In addition, most are relative to the individual. Measures would need to be time-stamped, synchronized, and normalized.
- I would presume to require greater than 90% sensitivity for estimating Soldier readiness level.
- It depends upon physiological sensors sensitivity and specificity. Technological and the human centered solutions must run simultaneously to identify any discrepancies between them.
- 95% sensitivity or near 0.95 area under the receiver operator characteristic (ROC) curve.
- With a mission commander overseeing the system and making decisions about the aviator's readiness or ability to continue the mission, I think that an 80-90% accuracy would be necessary for a commander to rely on the system. In addition, the commander would really have to understand what the information conveyed by the system means (including false positives and false negatives), and that would likely involve significant training.
- If it's affecting aviator's ability to continue the mission, I would think sensitivity would need to be 95%. I may not be qualified to answer this, however, being an engineer. You could also have a signal quality check to determine if the data being presented/recorded is "bad/poor" data and whether or not to make a decision based on it.
- In this case, in terms of safety, a low sensitivity would not have any dramatic consequence, the system will be particularly "cautious" and so it could assess the operator as no ready to continue even if it would have been still able. However, a too low sensitivity could induce skepticism and even frustration on the operator. Also, these psychological concerns, on a long term, could negatively affect operator's performance. 75% - 80% could be a proper minimum level of sensitivity.
- 80%
- Lack of false positives is important for a number of reasons in this scenario. The consequences of an unnecessary abort depend on the nature of the mission, as do the consequences of not aborting when there is a problem. However, the loss of confidence in the aircraft- or sensor systems could be critical.

Use-case 2: Assuming a hypothetical physiological sensor system has been developed and integrated into the aircraft, within this use-case the information regarding operator state is being used to trigger an adaptive automation system. The adaptive automation system is able to take over tasks (such as maintaining a specified flight route) for the aviator. How sensitive would such a system need to be in determining operator state?

Responses

- There are levels of autonomy, and until a system is proven trustworthy, I would recommend the human being able to override the automation. Further research would need to be conducted given the exact timing, but just because we can measure operator state to the second, doesn't mean we should intervene at that time. The data may need to be aggregated over minutes before automation takes over tasks. I would also recommend a priority automated system, as well. For example, loss of life, loss of aircraft, loss of mission.
- I would also require greater than 90% sensitivity for offloading tasks automatically.
- The adaptive automation system must be very sensitive to react to over taking aviator's decision. Suppose that a time lag between autonomous system and the aviator incapacitation is 5 seconds. In this example, if an aircraft starts flying at 250 knots true airspeed (KTAS) it can cover a distance of around 2460 feet. This distance can be a forward, backward, ascending or descending.
- 90% sensitivity or near 0.90 area under the ROC curve.
- Under routine operating conditions, an 80-90% accurate system would probably be acceptable for routine tasks (like maintaining a specified flight path). Under combat or non-routine conditions, I'd be hesitant to recommend a system that wasn't close to 100% accurate (and even then, I'd think very hard about the tasks the automation would take over).
- I would say in this case, similar to the above, but it would require the same or a greater sensitivity. So 95% or higher. And an assurance of quality data.
- Also in this case, in terms of safety, a low sensitivity would not have any dramatic consequence, sometimes the system could decide to replace operator's role when not necessary. However, in this case a too low sensitivity could have immediate negative effects on operator's performance, since inaccurate triggers will cause operator's mistrust in the system, resulting in a poor human-machine interaction. In this case values close to 90 % would be preferable.
- 90%
- If the situation is straight and level flight, a false positive that leads to a temporary "self-fly" situation is nuisance level only, but if the false positive arose because the pilot was maneuvering, etc., this could be mission- and life-threatening. Adaptive automation triggering needs high sensitivity, for example, triggers from multiple independent sensors.

Item 2

Likelihood of Each Operator State as a Candidate for Detection in Real-time

We received varying responses regarding whether some of the operator states are candidates for detection in real-time. These are indicated below. We invite you to provide any justification for why or why not a specific state (or states) may **not** be feasible for detection in real-time.

Alternatively, provide justification for why you may be **unsure** of whether or not a state (or states) is feasible for detection in real-time. *Note, you may choose to skip this item.*

Table E1. Summary of Operator State Responses

Operator State	Very Unlikely	Unlikely	Neutral	Likely	Very Likely	Unsure
Autonomic Tone		12.50 %	4.20 %	20.80 %	16.70 %	33.33 %
Mental Workload			29.20 %	25.00 %	29.20 %	8.30 %
Fatigue / Sleepiness			4.20 %	33.33 %	54.20 %	
Distraction		4.20 %	12.50 %	41.70 %	16.70 %	12.50 %
Mental Fatigue		8.30 %	20.80 %	41.70 %	12.50 %	
Hypoxia			8.30 %	16.70 %	33.33 %	33.33 %
Spatial Disorientation		4.20 %	25.00 %	20.80 %	8.30 %	29.20 %

Responses

- Cognitive workload is measurable in real time using the Cognitive Assessment Tool Set (CATS). CATS transforms raw electrocardiogram waveforms to classify cognitive state in real time. I have been using CATS for several years for instant, valid, and reliable measures.
- Unsure about spatial disorientation and its relationships to physiologic sensors for real-time. This could be a combination of inattention, low state of mental arousal, fatigue, drug effects, or innate individual different than like those which might have occurred before a mishap. Also, the mental workload (defining low and high boundaries).
- Workload is a very broad construct with no agreed objective and universal assessment criteria. Workload quantification is always linked to a specific task - so may be detectable within the particular task for which the algorithm was trained/designed. If you created an operationalized definition of distraction linked to overt attention (e.g., 30% of fixations must be related to the primary task) may be possible to detect with high enough accuracy.
- I will argue that autonomic tone, mental workload, and distraction are too uncoupled from the possible sources/drivers in the environment to be able to make judgments solely on operator state about the correct intervention/aid that should be deployed. Without being able to couple the cause with the distracted state, any intervention/aid could alleviate the problem, do nothing, or make things worse. Unless/until state detection algorithms are also able to be coupled with environmental information about potential causes, aid cannot be narrowly targeted to provide relief to the operator, and so aid must be general (e.g., reduce task load), and in its generality, potentially unhelpful and perhaps counterproductive.

With regard to fatigue, mental fatigue, and spatial disorientation, I'm not sure there are sufficient, observable indicators of each state to say whether or not they could be deployed in real-time. I know at present there are not robust indicators of each to deploy.

- I am quite aligned with the distribution of scores, I am only surprised by the amount of "unsure" scores about autonomic tone, since there are established biomarkers of such parameter, intended as the overall level of operator's activation/arousal, for example the skin conductance level or specific features of heart rate variability, that can be easily computed online by using wearable devices for data recording.
- It's interesting to me that autonomic tone is rated as less feasibly measured than most of the other constructs (mental workload, fatigue/sleepiness, distraction, and mental fatigue), even though autonomic tone has clearly defined behavioral and physiological manifestations (GSR/EDA, various cardiac signals like HRV and baroreflex sensitivity, stimulus-related cardiac deceleration, and pupil dilation) and uncontroversial anatomical underpinnings (vagal cardiac efferents, locus coeruleus-noradrenergic system). Autonomic tone exists more clearly as a useful functional concept than the higher level cognitive constructs which are often circularly defined in terms of behavioral dispositions.

I think that if autonomic tone were more clearly spelled out in the survey (in terms of its constituent signals) it would have received higher feasibility ratings like hypoxia. As a cognitive psychologist I'm no behaviorist-- I'm merely commenting on the surprisingly low rating for a promising measure.

- Autonomic tone is often detected based on heart-rate variability. That is itself a function of heart rate, and not only because high HR generally implies parasympathetic withdrawal and sympathetic activation. (There is also less time for variability). It also requires high-fidelity ECG sampled at 500 hertz (Hz) or more. I swing to "unlikely" on this for interpretable results.

Hypoxia is easy to detect reliably in real time if one means the condition of inhaled gas. This is what I think should be meant. This translates well to arterial blood unless breathing is impossible or the g-load is intolerable. Tissue hypoxia would not be the goal of monitoring. -- I say "very likely."

Introduction to Set of Survey Items

Similar to the item regarding likelihood of different operator states being candidates for detection in real-time, we also received varying responses on **which physiological measures** may be capable of detecting the various operator states. We report the results of the operator states that received mixed responses.

We invite you to provide any justification for why or why not a specific measure (or measures) may **not** be feasible for use in detecting the operator state. Alternatively, provide justification for why you may be **unsure** of whether or not a measure (or measures) is feasible for detecting the operator state. *Note, you may choose to skip items.*

Item 3

Likelihood of Each Physiological Measure Capable of Detecting Mental Workload

Table E2. Summary of Physiological Measures Responses

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	33.3 %	12.5 %	8.3 %	12.5 %	12.5 %	4.2 %
Body/face Temperature	25.0 %	4.2 %	20.8 %	16.7 %	20.8 %	
ECG HR	16.7 %	4.2 %	16.7 %	29.2 %	16.7 %	8.3 %
ECG HRV	12.5 %		12.5 %	16.7 %	37.5 %	12.5 %
Electroencephalography	8.3 %			8.3 %	29.2 %	41.7 %
Electromyography	25.0 %		25.0 %	16.7 %	20.8 %	
Electrooculogram	12.5 %		8.3 %	25.0 %	29.2 %	12.5 %
Excreted Metabolites	16.7 %	12.5 %	25.0 %	12.5 %	16.7 %	4.2 %
Eye Tracking			4.2 %	16.7 %	50.0 %	16.7 %
fNIRS	29.2 %		4.2 %	12.5 %	20.8 %	20.8 %
Photoplethysmography	29.2 %	4.2 %	12.5 %	25.0 %	16.7 %	
Pupillometry	4.2 %		8.3 %	12.5 %	45.8 %	20.8 %
Respiration Dynamics	25.0 %		29.2 %	25.0 %	4.2 %	4.2 %

Other responses received included head movement (very likely); electrodermal activity (neutral)

Responses

- ECG Heart Rate Variability is the measure CATS uses - and classifies cognitive state by analyzing the differences between each beat. The one current limitation of CATS would involve physical activity, such as running that may also impact the ECG waveform.
- Pupillometry has promise, in principle, for detecting mental workload, but the current state of the art of technology and modeling does not yet allow us to interpret pupil size signals in complex and/or uncontrolled visual environment as would be encountered in a real-world mission scenario. While I think it is somewhat likely to do so eventually, I do not see pupillometry working well to detect mental workload in a real-world mission in the very near term (much more basic research and modelling work is needed to enable

this capability in my opinion). The same applies to using pupillometry for detecting any other state below in real-time in a real-world context - there is promise but also basic research gaps to overcome and realize the full potential of pupillometry.

- Actigraphy, temperature, ECG, EMG, PPG, and respiration are not specific enough to provide a generalized indicator of mental workload - but may be useful in specific tasks for which baseline distributions are well parameterized.
- Actigraphy, body temperature, ECG HRV, electromyography, electrooculogram, photoplethysmography, pupillometry, and respiration dynamics are all too insensitive/nonspecific and uncoupled with environmental sources.
 - * EEG, fNIRS, and eye tracking are too susceptible to artifacts for real deployment.
 - * All/most measures on your list are too non-specific to ensure high enough detection accuracy for deployment.
 - * I'm not aware of any research directly tying excreted metabolites to mental workload (maybe good for physical workload, though).
- I think that ECG Heart rate is a good measure to input to a greater algorithm that takes into account other measures to detect mental workload. In flight tests that I've had, I've seen pilots heart rate's spike at times when there was relatively quiet environmental conditions but they were doing a mentally demanding task. I even talked to them about their data and they told me why their heart rate spiked at that moment, due to the task they were doing.

I believe Photoplethysmography is also a good indicator, because we've seen that with hypoxia, your mental fatigue is increased and your ability to concentrate and perform tasks correctly decreases.
- I am almost aligned with the overall scores, I have only some doubts about the applicability of pupillometry since its robustness has been demonstrated only in laboratory environments. In real applications its validity is debated, since it will be sensitive to a lot of confound factors, such as light, distances of objects, etc.
- Motor and gross physiological features (actigraphy, ECG HR, EMG, EOG, eye tracking, PPG, pupillometry, respiration) can likely be used in conjunction with appropriate cognitive task analyses to diagnose mental workload induced by task switching. This would be very specific to mission context but could be useful in very routine scenarios.

Item 4

Likelihood of Each Physiological Measure Capable of Detecting Fatigue / Sleepiness

Table E3. Summary of Physiological Measures for Fatigue / Sleepiness Responses

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	16.7 %			16.7 %	33.3 %	16.7 %
Body/face Temperature	29.2%		4.2%	20.8%	25.0%	8.3%
ECG HR	8.3%		20.8%	20.8%	41.7%	
ECG HRV	12.5%		16.7%	12.5%	37.5%	12.5%
Electroencephalography	12.5%			8.3%	25.0%	37.5%
Electromyography	16.7%		25.0%	20.8%	20.8%	4.2%
Electrooculogram	20.8%		4.2%	8.3%	29.2%	25.0%
Excreted Metabolites	25.0%	8.3%	25.0%	20.8%	12.5%	
Eye Tracking	2.4%			16.7%	20.8%	37.5%
fNIRS	25.0%		4.2%	16.7%	29.2%	12.5%
Photoplethysmography	29.2%	4.2%	8.3%	29.2%	12.5%	
Pupillometry	8.3%		8.3%	12.5%	33.3%	25.0%
Respiration Dynamics	12.5%		25.0%	33.3%	16.7%	

Other responses included head-worn infrared cameras for eye area-based computer vision (very likely); electrodermal activity (neutral)

Responses

- I believe there are correlations between cognitive state and fatigue.
- Unlike mental workload, the appropriate aid for fatigue is fairly well specified - sleep or a (strong) stimulant. Therefore, unspecific measures that assess large changes over time, such as actigraphy, respiration dynamics, HR & HRV, etc., that have been previously demonstrated to change with fatigue, could be usefully deployed for monitoring. However, are such systems necessary? Time-on-task is a much cheaper, and probably reasonably accurate, assessment approach.
- Same previous doubts about the use of pupillometry. I am also surprised by the not high value of EOG, actually by means of a vertical EOG channel it is possible to get information about eye blink rate, that is one of the physiological parameters mostly sensitive to sleepiness.
- Alpha power and sleep spindles in EEG are pretty well-established measures of sleepiness alongside eye tracking/EOG measures of drift and blink rate and speed. EMG frequency domain features are used to assess muscular fatigue, but may be less applicable to flight deck settings where the fatigue may be more general in nature rather than specific to muscle groups.

Item 5

Likelihood of Each Physiological Measure Capable of Detecting Distraction

Table E4. Summary of Physiological Measures for Distraction Responses

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	29.2 %	12.5 %	8.3 %	20.8 %	12.5 %	
Body/face Temperature	37.5%	12.5%	12.5%	20.8%		
ECG HR	16.7%	12.5%	29.2%	29.2%	4.2%	
ECG HRV	20.8%	12.5%	20.8%	25.0%	12.5%	
Electroencephalography	20.8%	8.3%		16.7%	20.8%	20.8%
Electromyography	25.0%	8.3%	29.2%	20.8%	4.2%	
Electrooculogram	20.8%	4.2%		12.5%	41.7%	8.3%
Excreted Metabolites	37.5%	25.0%	16.7%	8.3%	4.2%	
Eye Tracking	4.2%			8.3%	33.3%	41.7%
fNIRS	33.3%	8.3%	8.3%	20.8%	4.2%	12.5%
Photoplethysmography	45.8%	12.5%	16.7%	12.5%		
Pupillometry	12.5%	4.2%		29.2%	41.7%	
Respiration Dynamics	33.3%	8.3%	25.0%	16.7%	4.2%	

Other responses included electrodermal activity (very unlikely)

Responses

- Actigraphy, temperature, ECG, EMG, PPG, and respiration are not specific enough to provide a generalized indicator of distraction - but may be useful in specific tasks for which baseline distributions are well parameterized.
- Actigraphy, body temperature, ECG HRV, electromyography, electrooculogram, photoplethysmography, pupillometry, and respiration dynamics are all too insensitive/nonspecific and uncoupled with environmental sources.
 - * EEG, fNIRS, and eye tracking are too susceptible to artifacts for real deployment.
 - * Eye tracking may be "best" in the sense that you can assess whether someone has inspected (in the sense that they foveated it) a specific part of the visual scene within some specified temporal window.
 - * I'm not aware of any research directly tying excreted metabolites to distraction.
- No remarks on that.
- Aside from more direct brain measures (EEG, fNIRS) and measures that would track the motor -behavioral characteristics of distraction (actigraphy, EMG, EOG), I think the most promising approaches here would be very similar to those used for characterizing mental workload.

Item 6

Likelihood of Each Physiological Measure Capable of Detecting Mental Fatigue

Table E5. Summary of Physiological Measures for Mental Fatigue Responses

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	33.3 %	4.2 %	4.2 %	20.8 %	16.7 %	4.2 %
Body/face Temperature	33.3%		29.2%	8.3%	12.5%	
ECG HR	12.5%		37.5%	25.0%	16.7%	
ECG HRV	16.7%		20.8%	25.0%	25.0%	4.2%
Electroencephalography	8.3%			16.7%	25.0%	37.5%
Electromyography	16.7%	4.2%	20.8%	41.7%	4.2%	
Electrooculogram	16.7%		8.3%	33.3%	16.7%	12.5%
Excreted Metabolites	25.0%	12.5%	20.8%	20.8%	12.5%	
Eye Tracking	8.3%			29.2%	29.2%	20.8%
fNIRS	25.0%		8.3%	12.5%	29.2%	12.5%
Photoplethysmography	37.5%	4.2%	16.7%	20.8%	8.3%	
Pupillometry	8.3%			29.2%	25.0%	25.0%
Respiration Dynamics	25.0%	4.2%	20.8%	33.3%	8.3%	

Other responses included electrodermal activity (unlikely)

Responses

- I believe there are correlations between cognitive state and mental fatigue.
- (Same response as Mental Workload).
- Detection of mental fatigue is not likely to be accompanied by strong, specific physiological reactions (as is the case with fatigue/sleepiness). As such, most measures on your list are too non-specific to ensure high enough detection accuracy for deployment.
- No remarks on that.
- I'm slightly surprised that EDA wasn't viewed as capable of detecting mental fatigue, though I've forgotten how I categorized it myself.

Item 7

Likelihood of Each Physiological Measure Capable of Detecting Hypoxia

Table E6. Summary of Physiological Measures for Hypoxia Responses

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	45.8 %	12.5 %	8.3 %	16.7 %		4.2 %
Body/face Temperature	45.8%	8.3%	4.2%	12.5%	12.5%	
ECG HR	37.5%	4.2%	8.3%	12.5%	12.5%	12.5%
ECG HRV	45.8%	8.3%	4.2%	8.3%	16.7%	8.3%
Electroencephalography	41.7%	8.3%	8.3%	16.7%	12.5%	4.2%
Electromyography	41.7%	8.3%	12.5%	16.7%	8.3%	
Electrooculogram	45.8%	4.2%	4.2%	8.3%	25.0%	
Excreted Metabolites	45.8%	12.5%	20.8%		12.5%	
Eye Tracking	45.8%	4.2%		16.7%	20.8%	
fNIRS	41.7%		4.2%	4.2%	12.5%	25.0%
Photoplethysmography	41.7%		8.3%		8.3%	29.2%
Pupillometry	45.8%	4.2%	4.2%	12.5%	16.7%	
Respiration Dynamics	33.3%			4.2%	25.0%	20.8%

Other responses included electrodermal activity (unlikely); inhaled/exhaled gas analysis (likely); Oximetry – SpO₂ % (dual wavelength variant of PPG – note, not all oximetry systems are equal) (very likely)

Responses

- The coupling of available oxygen and hypoxia is a strong and well understood. Measures such as inhaled/exhaled gas analysis and oximetry are capable of assessing oxygen availability, making them well suited to detecting hypoxia. All the other measures seem poor for this purpose, though some of them, like actigraphy could be used to detect onset/full unconsciousness (not really useful if you want to prevent this state, though).
- HR is a good indicator of POSSIBLE hypoxia. Typically, as your SpO₂ drops, your heart rate increases to compensate for the decrease in oxygen.
- I do not feel qualified enough to provide reliable feedback on that.
- I would rely entirely on inhaled gas and "good" SpO₂, which may include photoplethysmography to ensure a proper signal. All the others are irrelevant, secondary, unnecessarily complicated (fNIRS)

Item 8

Likelihood of Each Physiological Measure Capable of Detecting Spatial Disorientation

Table E7. Summary of Physiological Measures for Spatial Disorientation Responses

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	45.8 %	8.3 %	8.3 %	4.2 %	16.7 %	
Body/face Temperature	41.7%	4.2%	16.7%	12.5%	8.3%	
ECG HR	41.7%		12.5%	25.0%	8.3%	
ECG HRV	41.7%		16.7%	16.7%	12.5%	
Electroencephalography	33.3%		4.2%	12.5%	25.0%	8.3%
Electromyography	37.5%		20.8%	16.7%	8.3%	
Electrooculogram	41.7%		4.2%	4.2%	33.3%	
Excreted Metabolites	37.5%	16.7%	8.3%	20.8%	4.2%	
Eye Tracking	25.0%			4.2%	37.5%	16.7%
fNIRS	37.5%		4.2%	16.7%	20.8%	4.2%
Photoplethysmography	41.7%	4.2%	16.7%	20.8%		
Pupillometry	37.5%		8.3%	12.5%	20.8%	4.2%
Respiration Dynamics	37.5%	4.2%	12.5%	29.2%	4.2%	

Other responses included pupillary light response (very likely)

Responses

- To my knowledge, spatial disorientation is not accompanied by strong, specific physiological reactions that can be assessed using the methods specified in the list. As such, most measures on your list are too non-specific to ensure high enough detection accuracy for deployment.
- I do not feel qualified enough to provide reliable feedback on that.
- Makes sense that the downstream eye behavioral measures (eye tracking, EOG, pupil measures) are viewed as likely feasible. I'd imagine that there would be head movement-related measures that would be just as good, neck EMG etc.

Item 9

Likelihood of Each Physiological Measure Capable of Detecting Autonomic Tone

Table E8. Summary of Physiological Measures for Autonomic Tone Responses

Measure	Unsure	Very Unlikely	Unlikely	Neutral	Likely	Very Likely
Actigraphy	50.0 %	4.2 %	12.5 %	12.5 %	4.2 %	4.2 %
Body/face Temperature	37.5%		8.3%	12.5%	29.2%	
ECG HR	25.0%		4.2%	1.25%	37.5%	12.5%
ECG HRV	25.0%			4.2%	29.2%	33.3%
Electroencephalography	45.8%		12.5%	12.5%	12.5%	4.2%
Electromyography	41.7%		20.8%	4.2%	16.7%	4.2%
Electrooculogram	45.8%		12.5%	12.5%	16.7%	
Excreted Metabolites	33.3%	4.2%	8.3%	4.2%	20.8%	20.8%
Eye Tracking	37.5%		12.5%	12.5%	25.0%	
fNIRS	41.7%	4.2%	12.5%	16.7%	12.5%	
Photoplethysmography	41.7%		8.3%	12.5%	12.5%	12.5%
Pupillometry	41.7%		12.5%	12.5%	12.5%	8.3%
Respiration Dynamics	37.5%		12.5%	12.5%	20.8%	8.3%

Other responses included electrodermal activity (very likely)

Responses

- Combining several measures of physiological arousal should be sufficient to get a fairly good idea of autonomic tone. Combining actigraphy, temperature, HR and HRV, EMG, and respiration dynamics (or some subset of them), should give a pretty accurate assessment of state.
I fully agree with the distribution of scores, I see only some more potential also in terms of electroencephalography, since there is some evidence about EEG correlates of human level of activation and engagement.
- The ECG measures of HR variability require high fidelity measurements and artifact rejection. Further, they are valid only at low heart rate.

Item 10

In the past year, what would you consider the most important finding or paper related to operator state monitoring? Please indicate why.

- Reddix, M.D., et al. (2020). Cognitive and physiological state monitoring for reduced crew and single pilot operations in civil aviation. Naval Medical Research Unit, Dayton, Wright-Patterson AFB, OH.
Although this is from 2020, it is a valuable white paper providing statistics and discussion around human factors and physiological causes of operator error.
- I am not sure.
- A plethora of literature exists supporting the use of physiological metrics to measure operator state, how reliable and valid these matrices are not fully established (e.g., Matthews et al., 2015). We therefore must evaluate promising physiologic sensors in the simulator study before it is further integrated into the aircraft.
- My lab is no longer working on single-operator state monitoring, and as such, I've somewhat lost track of pubs from this year on the topic. However, for monitoring teams (current research focus), I highly recommend:
Kazi, S., Khaleghzadegan, S., Dinh, J. V., Shelhamer, M. J., Sapirstein, A., Goeddel, L. A., ... & Rosen, M. A. (2021). Team physiological dynamics: A critical review. *Human Factors*, 63(1), 32–65. <https://doi.org/10.1177/0018720819874160>
- It is hard to provide just one answer to this question, it would need more discussion. Monitor the operator state is a matter of compromises to accept, i.e., to wear devices, to properly calibrate these devices, minimum levels of sensitivity and accuracy, time resolution needed, etc. Each one of the huge number of findings in scientific literature favours only some of these concerns to the detriment of others. Therefore, before considering relevant findings, it is crucial to have a clear understanding of user's needs. Maybe the work at <https://doi.org/10.1088/1361-6579/aad57e> could be a good starting point to have an overview of the current scenario.

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**