

Understanding the human performance envelope using electrophysiological measures from wearable technology

Jim Nixon¹ · Rebecca Charles¹

Received: 9 May 2017 / Accepted: 24 August 2017 / Published online: 13 September 2017
© The Author(s) 2017. This article is an open access publication

Abstract In this article, we capture electrophysiological measures from a new wearable technology to understand the human performance envelope. Using the NASA Multi-Attribute Task Battery (MATB II), participants completed tasks associated with flight control which included communication, tracking and system and resource monitoring. Electrophysiological measures relating to cardiac activity and respiration were taken using the new wearable technology. Our results show significant differences in both heart rate and respiration rate in response to different taskloads and that higher taskloads were associated with higher mental workload. Frequency measures of heart rate variability discriminated different task types but not taskloads. This finding may be related to differences in task complexity being more important than the number events which we have used to manipulate taskload. We suggest that this new generation of wearable sensors could be used to inform operator locus in a human performance envelope, indicating when assistance by the aircraft or another crew member may be necessary to maintain safe and efficient performance.

Keywords Human performance envelope · Mental workload · Wearable · Taskload · Aviation · MATB

1 Introduction

In this article, we validate the use of new wearable device that captures electrophysiological data in response to a variety of aviation tasks. We use these electrophysiological responses to characterise differences in the task types and taskloads that participants are exposed to in a suite of experimental tasks. We suggest that measures taken using wearable technologies could influence and specify user states in relation to the idea of a human performance envelope, analogous to the engineering envelopes specified for all aircraft.

Task performance is critical to the safety of flight operations. Our highly automated aircraft operates with extraordinary levels of safety. The UK Civil Aviation Authority's Global Fatal Accident Review reported that worldwide, the fatal accident rate between 2002 and 2011 was 0.6 per million flights flown (UK Civil Aviation Authority 2013). Civil aircraft operations are complex sociotechnical systems in which highly trained operators and high-integrity engineering work together to deliver safe flight every day (Cahill et al. 2014). All aircraft operate within a flight envelope (see Gratton 2015). A flight envelope is defined by the limits of the aircraft's structural and control capabilities. For example, there is an upper limit to the height at which aircraft can fly or the rate at which they can turn. Outside of these limits, structural damage to the aircraft or loss of control can occur. Recovery from highly unusual turn or climb rates can become increasingly difficult. The procedures we use in modern transport aircraft are designed to keep aircraft within this flight envelope. Indeed, in some highly automated aircraft, pilots are unable to make demands on the aircraft that exceed this manufacturer-specified flight envelope. As the flight envelope is approached or

✉ Jim Nixon
jim.nixon@cranfield.ac.uk

¹ Centre for Safety and Accident Investigation, Cranfield University, College Rd, Cranfield MK43 0AL, UK

exceeded, aircraft communicate this to the pilots through a range of alerts and alarms.

This article reports research conducted as part of the Future Sky Safety programme which looks to characterise the human performance envelope in a similar way in which we can specify and use the engineering performance envelope (Graziani et al. 2016; Silvagni et al. 2015). If we can reliably detect when a pilot is approaching or exceeding *their* performance envelope, we may be able to deploy automation that could itself recover control of the aircraft (for example, see Christensen and Estep 2013), or targeted guidance or warnings which would guide the pilot back to a safe zone within their individual human performance envelope.

The idea of a human performance envelope has been progressed in the air traffic management (ATM) domain (Edwards et al. 2012, 2014). Edwards and her co-authors characterise the human performance envelope using multifactorial human factors concepts which include situation awareness (SA), mental workload, stress, attention vigilance, teamwork, communication trust and fatigue. These concepts are proposed as performance shaping factors, which can differentially and interactively affect successful completion of a task. The model explicitly declares that boundaries exist where performance can degrade in line with the theoretical underpinnings for these concepts. Effective use of the human performance envelope to inform operations necessarily demands the measurement of the factors selected as being critical for the given task. A valid and reliable measure could be used to adapt a system to the state user or to inform users or their teams of their own states. Ideally, these measures would be predictive, having the ability to identify trends towards poorer performance—in essence a ‘yellow zone’. In this article, we build on the concept of the human performance envelope through examination of physiological measures that can give insight into mental workload in a set of representative tasks. Our research examines whether these physiological variables captured by the wearable technology can differentiate tasks and taskloads in the experimental setting. Broadly, we align the measures taken with the concept of mental workload as an explanatory concept. Mental workload is a key performance shaping variable in aviation operations (Dahlstrom and Nahlinder 2006) and more general in human factors as a whole (Young et al. 2014). Clearly, mental workload represents only part of the performance story, and we would expect interactions between different variables identified by Edwards et al. (2012). However, one way to look at the human performance envelope is to disassociate from the human factors meta-concepts such as SA, a view discussed by Dekker and Hollnagel (2004). The measure does not know what it means. If the measure is sensitive to differences in task

demand or task, then these differences could be used to indicate differences or exceedances in the human performance envelope. Physiological measures have been shown to be sensitive to differences in taskload and task demand across a range of domains. Measures relating to electrocardiac signals and respiration are represented in the recent peer-reviewed literature (for example, see Fallahi et al. 2016; Hsu et al. 2015; Matthews et al. 2015).

In this research, we capture physiological signals through exploitation of a newly developed wearable technology developed by the Swiss company CSEM SA (Chételat et al. 2015) shown in Fig. 1.

The research addresses three aims:

1. To identify which physiological measures captured using wearable technology can be used to understand the effects of taskload.
2. To identify which physiological measures captured using wearable technology can be used to understand the effects of task type.
3. To identify which physiological measures captured using wearable technology can be used to understand the effects of taskload gradient (from low to high vs from high to low).

The wearable technology acquires electrophysiological data, which is wirelessly communicated to a tablet computer. These data can then be displayed in real time and downloaded for further processing. These new wearable technologies do not require adhesive or wet electrodes in contrast to other methods of capturing electrophysiological data (Baig et al. 2013). The mobile technology means that the participant is also untethered from a computer further



Fig. 1 Wearable sensor technology developed by CSEM SA and used in this study

decreasing the gap between the laboratory and the eventual application (Guzik and Malik 2016). We capture cardiac measures that have been shown to differentiate taskloads (De Rivecourt et al. 2008; Fairclough et al. 2005; Splawn and Miller 2013) and respiration rates (Backs et al. 2000; Brookings et al. 1996; Wu et al. 2011; Yao et al. 2008).

Aviation-type tasks are generated using the NASA Multi-Attribute Task Battery 2 (MATB II) (Comstock and Arnegard 1992). This tool has been extensively used in the recent literature for the assessment of physiological variables in response to differential tasks and taskloads (Aricò et al. 2014; Dai et al. 2015; Hsu et al. 2015; Splawn and Miller 2013). Two taskload schedules, low and high, were developed, and three task groups were specified corresponding broadly to auditory, visual and psychomotor tracking skills.

2 Methods

2.1 Design

A mixed design was used. The within-subjects factors are task type (tracking task, communications task and system/resource monitoring task) and taskload (high or low). The between-subjects factor is gradient (low to high, or high to low taskload order presentation), which was randomised. Task order was also randomised. Each participant completed two five-minute sessions (one high and one low taskload) of each of the three tasks, and two five-minute baseline measurements pre- and post-task. During each block, physiological data from the participants were acquired by the smart harness. Each 5-min block was followed by a 2-min period during which subjective workload measurements were taken using the mental demand sub-scale of the NASA TLX. The single scale was used since the nature of the work is predominantly cognitive rather than physical. In addition, the single sub-scale reduced the time burden on participants in an already lengthy experimental procedure.

Physiological measures captured by the wearable technology and presented in this article include frequency domain measures (low, very low and high) and time domain measures (heart rate and the standard deviation of the heart rate). Breathing rate was also captured by the sensors. Selection of the heart rate variability (HRV) measure was guided by the guidelines offered by the European Society of Cardiology and the American Heart Association (AHA and ESC 1996). To constrain the scope of the study, the three standard frequency bands were adopted for the frequency-based measures. For the time-based measures, basic heart rate and one variation-based parameter (SDNN) were selected to provide sufficient

coverage of the key methods used to assess cardiac response.

2.2 Participants

Ethical approval for the study was granted through the university ethics board. Thirty-nine male participants took part in the experiment. Due to missing or incomplete data resulting from computer failure, eight participants were excluded from the final analysis. A further one participant was excluded after reporting health problems associated with cardiac function. These exclusions resulted in 30 male participants with a mean age of 34.3 years (SD 10.65). This number was not informed by a power calculation since no readily available data are available to reliably compute effect sizes.

All participants had normal or corrected-to-normal vision, and none reported consuming alcohol since waking prior to taking part in the experiment. Four participants also stated that they had some flying experience. However, these participants were not professional pilots.

2.3 Tasks

The MATB II was used to deliver tasks to participants. The MATB II can model a variety of tasks, which are central to aviation tasks but do not require a qualified pilot on which to assess performance. Four tasks from the MATB II were used in this study: the resource management task, the tracking task, the system monitoring task and the communication task (Table 1). The system monitoring and resource management tasks were combined creating three experimental tasks broadly representing emphasis on the psychomotor (tracking), auditory (communications) and visual attention/vigilance (system and resource monitoring) elements of cognition. These tasks are representative of the key elements of cognition required in the aircraft cockpit.

Two taskload levels were used: low and high. The event frequencies are listed in Table 2. The taskload was manipulated by adjusting the number of events in each task. Higher taskloads were specified by increasing the number of events, failures and increased input to maintain target.

2.4 Procedure

Participants were firstly given a voucher for participation in the study. Participants were then briefed and asked to provide informed consent. When informed consent was given, participants were asked to randomly select their participant number and select a piece of paper indicating the taskload condition they would be completing (low to high or high to

Table 1 MATB II interface elements and task descriptions

Task name	Task
System monitoring and resource management	In the system monitoring task, participants detect changes to the colour of the buttons and out-of-range movement of the scales. In the resource management task, participants switch pumps on and off to maintain flow. Participants detect failed pumps and adjust their plan
Tracking	Participants maintain moving target (circle) within inner square
Communications	Participants listen to auditory messages and dial in correct frequency when an ownship announcement is made. Participants are not required to speak in the communication task

Table 2 MATB II taskload stimulus

Task	Stimulus	Number of events in 5 min	
		Low taskload	High taskload
Systems monitoring and resource management tasks	Pump failure	6 × 5 s	10 × 15 s
	Gauge alert	6	30
	Green light off	4	10
	Red light on	4	10
Communications task (comms)	Communications	8	20
Tracking task (tracking)	Tracking	Default low target movement, low control gain	Default high target movement, high control gain

low). Each piece of paper was discarded after it was selected. Three cards were also presented to the participants stating the tasks on the back. Participants were asked to select the cards one at a time which determined the task order randomly.

Participants then provided brief biographical details and a stress and arousal checklist on a computerised questionnaire. Participants were then asked to change into the wearable technology and trained on the MATB II task groups for 25 min. After training, participants completed a 5-min baseline measurement sitting quietly looking at the MATB II screen. Participants then completed a total of six 5-min blocks of activity on the MATB II. During each block, physiological data from the participants were acquired by the wearable technology. Each 5-min block was followed by a 2-min period during which subjective workload measurements were taken. Participants were then instructed to change and given a full debrief.

3 Results

3.1 Data treatment

The physiological data were recorded in real time. The harness was worn from before the MATB II training to the end of the experimental procedure as one continuous

recording. The required recordings were the eight 5-min segments for each person. These were distinguished by markers throughout the recording made by the participant by ‘tapping’ one of the sensors. These markers were cross-checked with the time recorded for the start of each activity by the authors. Each recording was then split into eight individual recordings. The recordings were of 5-min duration. The quality of each of these recordings was checked. At this stage, six participants were excluded from the analysis due to poor quality recordings. In accordance with guidance provided by AHA and ESC (1996), the first 50 s and the last 10 s of each recording were then removed, resulting in eight four-minute segments per participant. Mean heart rate (HR), breathing rate (BR), the standard deviation of the N–N interval (SDNN) were derived. Frequency measures across the segments for very-low, low- and high-frequency spectral densities (VLF, LF, and HF) across each 4-min segment for each participant were captured using ensuring all zero measures were removed. Following application of Welch’s method to reduce noise, fast Fourier transforms were used to derive the different spectral densities. All data were subtracted from the mean of the post- and pretest baseline data. Physiological data represent the difference between a baseline signal and a signal stimulated by the experimental treatments.

3.2 Mental workload

A $2 \times 2 \times 3$ mixed ANOVA was conducted. The three factors are taskload (low taskload vs high taskload), gradient (low taskload to high taskload vs high taskload to low taskload) and task (communication vs tracking vs system and resource monitoring). Corrections for deviations in sphericity in the task-type condition used the Greenhouse–Geisser corrected df. Greenhouse–Geisser ϵ is reported to characterise the departure where significant. Gradient is a between-subjects factor. All other factors are within-subjects. Descriptive statistics are listed in Table 3. Since the main effect of gradient was not significant, this factor has been collapsed in order to simplify the table. Significant departure from sphericity was indicated for the task type \times taskload interaction ($\chi^2_2 = 23.76$, $p < 0.01$, $\epsilon = 0.63$). The main effect of task type was significant ($F_{2,56} = 48.68$, $p < 0.01$, $\eta^2_p = 0.64$). The main effect of taskload was also significant ($F_{1, 28} = 60.75$, $p = 0.01$, $\eta^2_p = 0.68$). No main effect of gradient was found ($F_{1, 28} = 0.36$, $p = 0.55$). No significant interactions were found between taskload and task type ($F_{1,26, 35,32} = 0.19$, $p = 0.72$), taskload and gradient ($F_{1, 28} = 1.81$, $p = 0.19$) or task type and gradient ($F_{2, 28} = 0.61$, $p = 0.40$). Bonferroni-corrected pairwise comparisons (Table 4) showed significant differences between all task types at both levels of taskload.

3.3 Physiological data

This section reports data associated with heart rate variability (HRV) including frequency- and time-based measures and respiration rate. For each physiological measure, a $2 \times 2 \times 3$ mixed ANOVA was conducted. The three factors are taskload (low taskload vs high taskload), gradient (low taskload to high taskload vs high taskload to low

taskload) and task (communication vs tracking vs system and resource monitoring). Corrections for deviations in sphericity in the task-type condition used the Greenhouse–Geisser corrected df. Greenhouse–Geisser ϵ is reported to characterise the departure where significant. Gradient is a between-subjects factor. All other factors are within-subjects.

Descriptive statistics for all physiological variables are shown in each section. Since the between-subjects factor of gradient is counterbalanced and no significant effects have been found in the analysis, the data have been collapsed across this factor to simplify the tables presented for each variable. It should also be remembered that the physiological variables were subtracted from baseline. In this way, a negative heart rate or respiration rate can be produced and interpreted.

3.3.1 Respiration rate

Descriptive statistics for the respiration rate variable are listed in Table 5. The main effect of task type was significant ($F_{2,56} = 8.04$, $p < 0.02$, $\eta^2_p = 0.22$). The main effect of taskload was significant ($F_{1, 28} = 10.84$, $p = 0.03$, $\eta^2_p = 0.28$). A significant interaction was found between taskload and task type ($F_{1,28} = 4.20$, $p = 0.02$, $\eta^2_p = 0.13$). The main effect of gradient was not significant ($F_{1, 28} = 2.70$, $p = 0.11$). No significant interactions were found between taskload and gradient ($F_{1, 28} = 1.98$, $p = 0.17$) or task type and gradient ($F_{1, 28} = 0.14$, $p = 0.87$). Bonferroni-corrected pairwise comparisons indicate that the source of the interaction effect is the absence of any differences in breath rate in the high taskload condition. In the low taskload condition, significant pairwise differences were found between the communications task and the tracking task (mean difference = 1.51 breaths, $SE = 0.33$, $p < 0.01$) and the communications task and the system monitoring task (mean difference = 1.46 breaths, $SE = 0.23$, $p < 0.01$). No other significant differences were found.

3.3.2 Heart rate variability (frequency domain)

HRV in the frequency domain was assessed in three power bands specified in accordance with the European Society of Cardiology and The North American Society of Pacing and Electrophysiology (AHA and ESC 1996). The very-low-frequency band is filtered at 0.0033–0.04 Hz, the low-frequency band at 0.04–0.15 Hz and the high-frequency band at 0.15–0.4 Hz. Power is measured in square milliseconds (ms^2) and offset from the baseline measurement in the frequency band under examination.

Table 3 Descriptive statistics for mental demand sub-scale of the NASA TLX

	Mean (SD)	Lower CI _{95%}	Upper CI _{95%}
<i>Communications</i>			
Low taskload	19.7 (18.0)	13.1	26.3
High taskload	35.2 (21.3)	27.2	43.1
<i>Tracking</i>			
Low taskload	41.8 (23.7)	32.9	50.7
High taskload	54.6 (26.0)	44.9	64.4
<i>System monitoring</i>			
Low taskload	53.7 (23.9)	44.8	62.7
High taskload	67.0 (22.8)	58.5	75.5

Table 4 Pairwise comparisons for mental demand at high and low taskload levels

Taskload	Comparison	Mean difference (SE)
Low	Communication versus tracking	22.1 (4.5)**
	Communication versus system monitoring	34.0 (3.7)**
	Tracking versus system monitoring	11.9 (4.3)*
High	Communication versus tracking	19.5 (4.6)**
	Communication versus system monitoring	31.9 (3.3)**
	Tracking versus system monitoring	12.49 (3.89)**

* $p < 0.05$
 ** $p < 0.01$

Table 5 Descriptive statistics for the respiration rate variable

	Mean (SD)	Lower CI _{95%}	Upper CI _{95%}	Range
<i>Communications</i>				
Low taskload	1.1 (2.2)	0.2	1.9	9.7
High taskload	2.3 (2.3)	1.5	3.2	9.6
<i>Tracking</i>				
Low taskload	2.6 (2.5)	1.6	3.5	9.6
High taskload	2.7 (2.2)	1.9	3.6	8.7
<i>System monitoring</i>				
Low taskload	2.5 (2.5)	1.6	3.5	10.3
High taskload	3.1 (2.5)	2.1	4.0	10.6

3.3.3 Very low frequency

Descriptive statistics for the very-low-frequency powers are listed in Table 6. Significant departure from sphericity was indicated for the task-type effect ($\chi^2 = 27.37, p < 0.01, \epsilon = 0.61$). The main effect of task type was significant ($F_{1.2, 34.2} = 7.2, p < 0.05, \eta_p^2 = 0.21$). Bonferroni-corrected pairwise comparisons showed one significant difference between the communications and system monitoring task (mean difference = 937.07 ms, SE = 359.36, $p < 0.05$) in the high taskload condition. The main effect of taskload did not reach significance

Table 6 Descriptive statistics for deviation in the very-low-frequency powers from baseline (ms²)

	Mean (SD)	Lower CI _{95%}	Upper CI _{95%}	Range
<i>Communications</i>				
Low taskload	316.8 (1797.1)	-354.2	987.9	10,868.0
High taskload	217.7 (1529.4)	-353.4	788.8	7022.5
<i>Tracking</i>				
Low taskload	-385.5 (1427.8)	-918.7	147.6	7022.5
High taskload	-619.0 (1005.5)	-994.4	-243.5	5018.0
<i>System monitoring</i>				
Low taskload	-516.1 (1079.0)	-919.0	-113.2	6117.5
High taskload	-719.4 (1048.6)	-1110.9	-327.8	6184.0

($F_{1,28} = 1.23, p = 0.28$). No main effect of gradient was found ($F_{1, 28} = 0.4, p = 0.85$). No significant interactions were found between taskload and task type ($F_{2, 56} = 0.06, p = 0.94$), taskload and gradient ($F_{1, 28} = 0.04, p = 0.53$) or task type and gradient ($F_{1, 28} = 0.37, p = 0.70$).

3.3.4 Low frequency

Descriptive statistics for the low-frequency variable are listed in Table 7. Significant departure from sphericity was indicated for the task-type effect ($\chi^2 = 8.60, p < 0.02, \epsilon = 0.79$) and the task type \times taskload interaction ($\chi^2 = 18.37, p < 0.01, \epsilon = 0.67$). The main effect of task type was significant ($F_{1.6, 44.0} = 8.56, p < 0.01, \eta_p^2 = 0.50$). Bonferroni-corrected pairwise comparisons (Table 8) showed significant differences between the communications and tracking task and the communication and system monitoring task. No significant pairwise differences between the tracking and system monitoring task were found.

The main effect of taskload did not reach significance ($F_{1, 28} = 3.41, p = 0.08$). No main effect of gradient was found ($F_{1, 28} = 1.03, p = 0.32$). No significant interactions were found between taskload and task type ($F_{1.4, 37.5} = 0.42, p = 0.58$), taskload and gradient ($F_{1, 28} = 0.40, p = 0.51$) or task type and gradient ($F_{2, 28} = 1.97, p = 0.15$).

Table 7 Descriptive statistics for deviation in the low-frequency powers from baseline (ms²)

	Mean (SD)	Lower CI _{95%}	Upper CI _{95%}	Range
<i>Communications</i>				
Low taskload	345.6 (1047.5)	−45.5	736.7	5127.0
High taskload	151.9 (1067.2)	−246.6	550.4	4483.5
<i>Tracking</i>				
Low taskload	−538.1 (856.3)	−857.8	−218.4	4483.5
High taskload	−630.5 (800.9)	−929.6	−331.5	3268.5
<i>System monitoring</i>				
Low taskload	−336.3 (870.1)	−661.2	−11.4	4124.5
High taskload	−579.2 (806.8)	−880.5	−278.0	3076.5

Table 8 Pairwise comparisons for the low-frequency power deviations at high and low taskload levels

Taskload	Comparison	Mean difference (SE)
Low	Communication versus tracking	883.7 (178.2)**
	Communication versus system monitoring	681.9 (190.8)*
	Tracking versus system monitoring	Not significant
High	Communication versus tracking	782.4 (154.1)**
	Communication versus system monitoring	731.1 (137.4)**
	Tracking versus system monitoring	Not significant

* $p < 0.05$
 ** $p < 0.01$

3.3.5 High frequency

Descriptive statistics for the high-frequency variable are listed in Table 9. The main effect of task type was significant ($F_{2,56} = 7.1, p = 0.02, \eta_p^2 < 0.20$). Bonferroni-corrected pairwise comparisons showed significant differences between the communications and tracking task (mean difference = 148.00, SE = 32.60, $p < 0.05$) in the low taskload condition and the communication and system monitoring task in the low (mean difference = 101.93 ms, SE = 38.39, $p < 0.01$) and high (mean difference = 121.53 ms, SE = 30.43, $p < 0.05$) taskload conditions.

The main effect of taskload did not reach significance ($F_{1,28} = 2.01, p = 0.17$). The main effect of gradient was not significant ($F_{1,28} = 0.87, p = 0.36$). No significant interactions were found between taskload and task type ($F_{2,56} = 0.72, p = 0.49$), taskload and gradient ($F_{1,28} = 1.20, p = 0.28$) or task type and gradient ($F_{2,33} = 0.38, p = 0.94$).

A consistent effect is the discrimination of the communications task and the other two tasks, regardless of the taskload effect by the HRV data. Figure 2 shows estimated marginal means for task type only clearly showing this pattern in each frequency band evaluated as part of the HRV analysis.

Table 9 Descriptive statistics for deviation in the high-frequency powers from baseline (ms²)

	Mean	Lower CI _{95%}	Upper CI _{95%}	Range
<i>Communications</i>				
Low taskload	−11.5 (227.5)	−96.4	73.5	1110.0
High taskload	−71.4 (332.1)	−195.4	52.6	1702.0
<i>Tracking</i>				
Low taskload	−159.5 (240.8)	−249.4	−69.5	1702.0
High taskload	−175.2 (237.4)	−263.8	−86.5	983.0
<i>System monitoring</i>				
Low taskload	−113.4 (258.2)	−209.8	−17.0	1207.0
High taskload	−193.0 (328.1)	−315.5	−70.4	1442.0

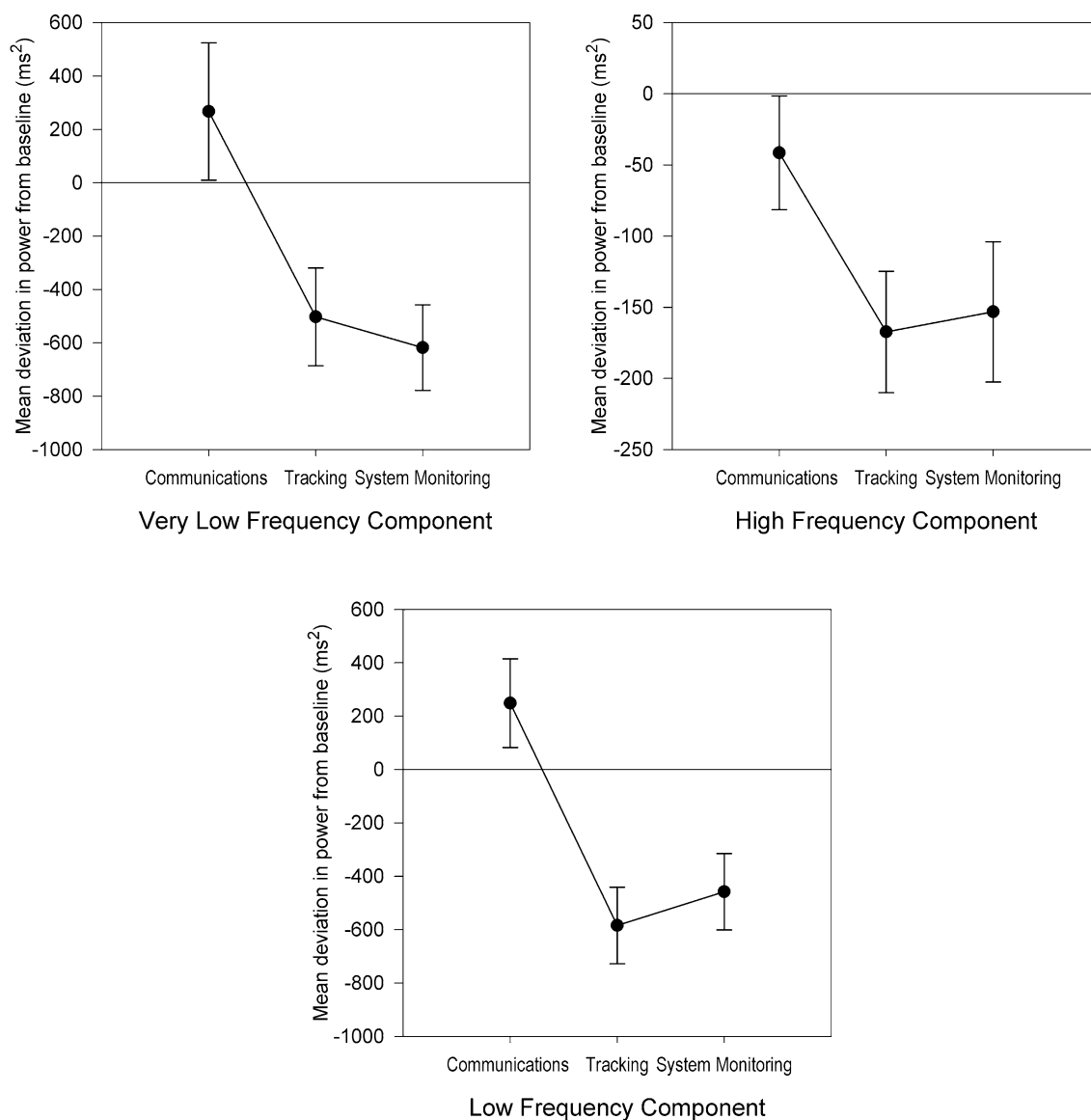


Fig. 2 Estimated marginal means for HRV deviation from the baseline for each task-type effect for very-low, low- and high-frequency domains. Error bars show one SE

Table 10 Descriptive statistics for mean deviation in heart rate from baseline (beats per min)

	Mean (SD)	Lower CI _{95%}	Upper CI _{95%}	Range
<i>Communications</i>				
Low taskload	0.5 (2.4)	-0.4	1.4	9.5
High taskload	0.8 (2.4)	-0.1	1.8	10.3
<i>Tracking</i>				
Low taskload	0.5 (3.6)	-0.9	1.8	10.3
High taskload	1.5 (3.6)	0.1	2.9	17.0
<i>System monitoring</i>				
Low taskload	0.7 (3.3)	-0.5	2.0	17.2
High taskload	1.0 (3.3)	-0.2	2.3	14.8

3.3.6 Mean heart rate

Descriptive statistics for the heart rate variable are listed in Table 10. The main effect of task type was not significant ($F_{2,56} = 0.14, p = 0.87$). The main effect of taskload was significant ($F_{1, 28} = 5.22, p = 0.03, \eta_p^2 = 0.16$). The main effect of gradient was not significant ($F_{1, 28} = 0.35, p = 0.85$). No significant interactions between taskload and task type ($F_{2, 56} = 1.16, p = 0.32$), taskload and gradient ($F_{2, 28} = 0.25, p = 0.62$) or task type and gradient ($F_{1, 28} = 0.62, p = 0.54$) were found.

Table 11 Descriptive statistics for the SDNN variable

	Mean	Lower CI _{95%}	Upper CI _{95%}	Range
<i>Communications</i>				
Low taskload	2.9 (17.6)	−4.2	10.1	93.5
High taskload	2.1 (14.4)	−3.7	7.9	59.5
<i>Tracking</i>				
Low taskload	−9.9 (15.0)	−15.9	−3.8	59.5
High taskload	−12.1 (13.0)	−17.4	−6.9	44.0
<i>System monitoring</i>				
Low taskload	−7.3 (14.3)	−13.1	−1.6	58.5
High taskload	−13.5 (13.1)	−18.8	−8.2	55.0

3.3.7 SDNN

Descriptive statistics for the SDNN variable are listed in Table 11. Significant departure from sphericity was indicated for task type ($\chi^2_2 = 10.66, p < 0.01, \epsilon = 0.73$). The main effect of task type was significant ($F_{1,46,35.02} = 15.93, p < 0.01, \eta^2_p = 0.40$). Bonferroni-corrected pairwise comparisons are listed in Table 12.

The main effect of taskload approached significance ($F_{1,28} = 3.87, p = 0.06, \eta^2_p = 0.14$). The main effect of gradient was not significant ($F_{1,28} = 1.56, p = 0.22$). No significant interactions between taskload and task type ($F_{2,56} = 0.93, p = 0.40$), taskload and gradient ($F_{1,28} = 0.28, p = 0.60$) or task type and gradient ($F_{2,28} = 0.49, p = 0.61$) were found.

3.4 Results summary

Table 13 summarises our findings in a concise way. In this table, findings are reported as effect sizes (partial η^2). Effect sizes of non-significant ($p < 0.05$) findings are not reported in this table. No significant main effect of gradient for any variable was found. Only one significant two-way interaction was found between task type and taskload. No other significant two-way interactions were found.

Patterns across the variables include the finding that mental workload was lowest in the communications task and

highest in the system and resource monitoring task. This pattern of differences was reflected across all physiological variables supportive of the general association between taskload, cognition and the physiological variables. Generally, the significant findings indicated that for higher taskloads and task types eliciting higher mental workload, heart and respiration rates were higher and power spectra across all frequency components reduced. The frequency domain measures of heart rate variability were not sufficiently sensitive to differences between the taskloads, but the heart rate and respiration rates differentiate the taskloads in a systematic way through the different types of task.

4 Discussion

In this article, we have validated the use of a new wearable technology to capture physiological signals. We have shown that the selected measures are sensitive to taskloads and task types generated in a systematic way using the NASA MATB II. Of the measures assessed, heart rate and respiration rate show significant differences between task types and taskloads. This is consistent with the wider literature and supports the use of the wearable technology in this way (for example, see Backs et al. 2000; Brookings et al. 1996; Fairclough and Venables 2006). A large effect size and significant differences between high and low taskloads for each task using the heart rate variables were found, indicating that this measure has promise in differentiating higher and lower taskloads. Overall, the patterns of differences across all variables reflect the general trend in the mental workload measured: lowest in the communications task and highest in the system and resource monitoring tasks. The SDNN and the frequency measures of heart rate showed differences between the tasks but not between the taskloads in the same task. One explanation for this pattern of results could be the differences in cognition demanded by the task itself. The communications task demands the use of auditory attention. Early cognitive psychology proposed that the echoic sensory store has greater capacity (between 2 and 4 s) than the iconic store

Table 12 Pairwise comparisons for SDNN at high and low taskload levels

Taskload	Comparison	Mean difference (SE)
Low	Communication versus tracking	12.6 (4.1)*
	Communication versus system monitoring	10.6 (3.3)*
	Tracking versus system monitoring	Not significant
High	Communication versus tracking	14.1 (3.3)**
	Communication versus system monitoring	15.4 (3.6)**
	Tracking versus system monitoring	Not significant

* $p < 0.05$

** $p < 0.01$

Table 13 Summary of significant effects across all variables

Variable type	Variable	Partial η^2		
		Taskload	Task type	Task type \times taskload
NASA TLX	Mental demand	0.68**	0.64**	–
Physiological variables	HRV (VLF power component)	–	0.21*	–
	HRV (LF power component)	–	0.50**	–
	HRV (HF power component)	–	0.20*	–
	Mean breath rate	0.28*	0.22*	0.13*
	Mean heart rate	0.16	–	–
	Mean SDNN	–	0.40**	–

– Not significant

* $p < 0.05$ ** $p < 0.01$

(Treisman 1964) allowing greater capacity to admit the auditory information to working memory. The highest levels of workload and the lowest spectral densities were found in the system and resource monitoring task. This finding is consistent with the overall trend in mental workload and is again consistent with the wider literature (for example, see Delaney and Brodie 2000; Splawn and Miller 2013). This finding may represent the effects of another variable: task complexity. This is in line with Gao et al. (2013) who cited task complexity as a limiting factor in explaining the relationships between task types and physiological measures. Although participants trained to stable performance on the NASA MATB II tasks, the system and resource monitoring tasks may make greater demands on endogenous attention, which must be allocated across the interface. The larger effect size in the low-frequency band reflects the findings, which indicate that this band is the most sensitive to changes in mental workload (Veltman and Gaillard 1998).

We did not find differences in any measures associated with the gradient of the taskload. Regardless of the direction of the taskload, the same patterns of differences were observed. This is important from an operational perspective since a measure that was sensitive to previous taskloads may under- or overestimate a new taskload (higher or lower), and this would radically increase the complexity of considering such a measure in examination of the human performance envelope.

Participant response to the wearable technology was positive. The new wearable technology developed by CSEM SA offers a step change in the quality of signals produced by a smaller number of electrodes; a clinical electrocardiogram can demand over 12 electrodes (Guzik and Malik 2016). These electrodes can be wet or sticky and look like a medical device. Embedding contact sensors in clothing is a realistic, and we suggest a more acceptable mechanism by which electrophysiological data could be captured and processed. Our results indicate that these physiological

measures captured in a wearable context could discriminate higher and lower taskloads across a range of tasks. These new wearable technologies could then be used in inform operator locus in a performance envelope.

Our study has a number of limitations. We of course acknowledge that mental workload is one part of the wider human performance envelope discussed in this special issue and the wider literature. However, we are of the view that this variable may be somewhat more amenable to measurement than other concepts such as SA which ignite stimulating, if at times fierce, debate (Klein 2015).

Although we have evidenced differences between different taskloads and task types, these differences have been established using specific task types. Clearly in the operational environment, tasks co-occur and the unit of success is most often the team rather than an individual operator. We very much doubt that task interaction is simply additive; however, we are encouraged that our data support a general task-type trend which follows the subjective reports of mental demand elicited by the NASA TLX. A limitation of the more general approach using the types of measures employed is the time within which the task is assessed. It is difficult to envisage a real-time one-to-one mapping of the signal–mental workload relationship that could inform a locus in a human performance envelope with the current measures used. As such, a time period must be sampled. This period could be continuously windowed, but then this decision would affect the types of task that could reasonably be monitored using the technique. The measures also have constraints in this regard. For example, the variance of the frequency measures increases with the length of the sample; the VLF band is less reliable at short recording times of less than 5 min (AHA and ESC 1996).

Overall, our results support the idea of using non-invasive electrophysiological sensors to give insight into workload in response to taskload. Using the MATB II, we have validated the signals captured by the wearable technology in support of this aim. Certainly, the simpler

measures of respiration and heart rate discriminate the taskloads effectively across the tasks in the case of heart rate in line with subjective reports of mental workload. The other measures may have discriminated task complexity rather than taskload; this would need testing in further research together with interaction effects between different tasks. However, measures that vary with taskload and are associated with mental workload have the potential to secure insight into the human performance envelope. Developing reliable and valid measures that can tell us when an operator is approaching the edge of their performance envelope has the potential to inform action on the part of another crew member or indeed the aircraft itself through changing automation (Christensen and Estep 2013). Indeed, this is a corollary of human-centred automation proposed by Billings: ‘The automated systems must also monitor the human operators’ (Billings 1997, p. 39). Understanding and specifying an individual’s performance envelope could permit the aircraft to understand the state of the crew, in addition to the crew understanding the state of the aircraft.

Acknowledgements The authors would like to thank Olivier Grossenbacher, Stephan Dasen and the wider team at CSEM SA for their superlative technical support and guidance associated with the wearable sensor technology and the Fourier transforms. This article is based on work performed in the programme: Future Sky Safety, which has received funding from the European Union’s Horizon 2020 research and innovation programme under Grant Agreement No. 640597. The views and opinions expressed in this paper are those of the authors and are not intended to represent the position or opinions of the Future Sky Safety consortium or any of the individual partner organisations.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- AHA and ESC (1996) Guidelines on heart rate variability measurement. *Eur Heart J* 17:354–381
- Aricò P, Borghini G, Graziani I, Taya F, Sun Y, Bezerianos A, et al (2014) Towards a multimodal bioelectrical framework for the online mental workload evaluation. In: Conference proceedings:… annual international conference of the IEEE engineering in medicine and biology society. IEEE engineering in medicine and biology society. annual conference, 2014 (Figure 1). pp 3001–3004. doi:10.1109/EMBC.2014.6944254
- Backs RW, Navidzadeh HT, Xu X (2000) Cardiorespiratory indices of mental workload during simulated air traffic control. In: Proceedings of the IEA 2000/HFES 2000 congress, 3. pp 89–92
- Baig MM, Gholamhosseini H, Connolly MJ (2013) A comprehensive survey of wearable and wireless ECG monitoring systems for older adults. *Med Biol Eng Comput* 51(5):485–495. doi:10.1007/s11517-012-1021-6
- Billings CE (1997) Aviation automation: the search for a human centered approach. LEA, Mahwah
- Brookings JB, Wilson GF, Swain CR (1996) Psychophysiological responses to changes in workload during simulated air traffic control. *Biol Psychol* 42(3):361–377. doi:10.1016/0301-0511(95)05167-8
- Cahill J, McDonald N, Losa CG (2014) A sociotechnical model of the flight crew task. *Hum Factors J Hum Factors Ergon Soc* 56(8):1337–1363. doi:10.1177/0018720814532684
- Chételat O, Ferrario D, Proença M, Porchet J, Falhi A, Grossenbacher O et al (2015) Clinical validation of LTMS-S: a wearable system for vital signs monitoring. *Embc*. doi:10.1109/EMBC.2015.7319054
- Christensen JC, Estep JR (2013) Coadaptive aiding and automation enhance operator performance. *Hum Factors* 55(5):965–975. doi:10.1177/0018720813476883
- Comstock JR, Arnegard RJ (1992) NASA-TM-104174, NAS 1.15:104174 the multi-attribute task battery for human operator workload and strategic behavior research. Hampton, VA, United States. <https://ntrs.nasa.gov/search.jsp?R=19920007912>
- Dahlstrom N, Nahlinder S (2006) A comparison of two recorders for obtaining in-flight heart rate data. *Appl Psychophysiol Biofeedback* 31(3):273–279. doi:10.1007/s10484-006-9021-7
- Dai Z, Príncipe JC, Bezerianos A, Thakor NV (2015) Cognitive workload discrimination in flight simulation task using a generalized measure of association. In: Arik S, Huang T, Lai W, Liu Q (eds) Neural information processing. Lecture notes in Computer Science, vol 9491. Springer, Cham
- De Rivecourt M, Kuperus MN, Post WJ, Mulder LJM (2008) Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics* 51(9):1295–1319. doi:10.1080/00140130802120267
- Dekker S, Hollnagel E (2004) Human factors and folk models. *Cogn Technol Work* 6(2):79–86. doi:10.1007/s10111-003-0136-9
- Delaney JP, Brodie DA (2000) Effects of short-term psychological stress on the time and frequency domains of heart-rate variability. *Percept Mot Skills* 91(2):515–524. doi:10.2466/pms.2000.91.2.515
- Edwards T, Sharples S, Wilson JR, Kirwan B (2012) Factor interaction influences on human performance in air traffic control: the need for a multifactorial model. *Work* 41(SUPPL.1):159–166. doi:10.3233/WOR-2012-0151-159
- Edwards T, Sharples S, Kirwan B, Wilson J, Balfe N (2014) Identifying markers of performance decline in air traffic controllers. In: Ahram T, Karwowski W, Marek T (eds) Proceedings of the 5th international conference on applied human factors and ergonomics AHFE 2014, Kraków, Poland. pp 19–23. doi:10.1201/9781315370460-27
- Fairclough SH, Venables L (2006) Prediction of subjective states from psychophysiology: a multivariate approach. *Biol Psychol* 71:100–110
- Fairclough SH, Venables L, Tattersall A (2005) The influence of task demand and learning on the psychophysiological response. *Int J Psychophysiol* 56(2):171–184. doi:10.1016/j.ijpsycho.2004.11.003
- Fallahi M, Motamedzade M, Heidarimoghdam R, Soltanian AR, Miyake S (2016) Effects of mental workload on physiological and subjective responses during traffic density monitoring: a field study. *Appl Ergon* 52:95–103. doi:10.1016/j.apergo.2015.07.009
- Gao Q, Wang Y, Song F, Li Z, Dong X (2013) Mental workload measurement for emergency operating procedures in digital nuclear power plants. *Ergonomics* 56(7):1070–1085. doi:10.1080/00140139.2013.790483

- Gratton G (2015) Initial airworthiness. Initial airworthiness: determining the acceptability of new airborne systems. Springer, London. doi:10.1007/978-3-319-11409-5
- Graziani I, Berberian B, Kirwan B, Blaye PL, Rognin L, Silvagni S, Kirwan B (2016) Development of the human performance envelope concept for cockpit HMI design development of the human performance envelope concept for cockpit HMI design. In: HCIAero 2016 international conference on human-computer interaction in Aerospace2016 international conference on human-computer interaction in aerospace. Paris, France
- Guzik P, Malik M (2016) ECG by mobile technologies. J Electrocardiol 49(6):894–901. doi:10.1016/j.jelectrocard.2016.07.030
- Hsu B-W, Wang M-JJ, Chen C-Y (2015) Effective Indices for monitoring mental workload while performing multiple tasks. Percept Mot Skills 121(1):94–117
- Klein G (2015) Whose fallacies? J Cognit Eng Decis Mak 9(1):55–58. doi:10.1177/1555343414551827
- Matthews G, Reinerman-Jones LE, Barber DJ, Abich J (2015) The psychometrics of mental workload: multiple measures are sensitive but divergent. Hum Factors J Hum Factors Ergon Soc 57(1):125–143. doi:10.1177/0018720814539505
- Silvagni S, Napoletano L, Graziani I, Le Blaye P, Rognin L (2015) Concept for human performance envelope. https://www.future-sky-safety.eu/wpcontent/uploads/2015/12/FSS_P6_DBL_D6.1-Concept-for-Human-Performance-Envelope_v2.0.pdf. Accessed 11 Sep 2017
- Splawn JM, Miller ME (2013). Prediction of perceived workload from task performance and heart rate measures. In: *Proceedings of the human factors and ergonomics society 57th annual meeting*, pp 778–782
- Treisman AM (1964) Verbal cues, language, and meaning in selective attention. Am J Psychol 77:206–219. doi:10.2307/1420127
- UK Civil Aviation Authority (2013) Global fatal accident review 2002 to 2011. The Stationery Office on behalf of the UK CAA, CAP1036, vol 1036. <http://publicapps.caa.co.uk/docs/33/CAP%201036%20Global%20Fatal%20Accident%20Review%202002%20to%202011.pdf>. Accessed 11 Sep 2017
- Veltman J A, Gaillard A W K (1998) Physiological workload reactions to increasing levels of task difficulty. Ergonomics 41(5):656–669. doi:10.1080/001401398186829
- Wu B, Hou F, Yao Z, Niu J, Huang W (2011) Using physiological parameters to evaluate operator's workload in manual controlled rendezvous and docking (RVD). In: Duffy VG (eds) Digital human modeling. ICDHM 2011. Lecture Notes in Computer Science, vol 6777. Springer, Berlin, Heidelberg
- Yao YJ, Chang YM, Xie XP, Cao XS, Sun XQ, Wu YH (2008) Heart rate and respiration responses to real traffic pattern flight. Appl Psychophysiol Biofeedback 33(4):203–209. doi:10.1007/s10484-008-9066-x
- Young MS, Brookhuis KA, Wickens CD, Hancock PA (2014) State of science: mental workload in ergonomics. Ergonomics 139(January 2015):1–17. doi:10.1080/00140139.2014.956151