ORIGINAL ARTICLE

Investigating the relationship between mental state (workload and affect) and physiology in a control room setting (ship bridge simulator)

Andreas Simskar Wulvik[1] · Henrikke Dybvik[1] · Martin Steinert[1]

Abstract

This paper discusses how to investigate the human element in a control room setting in terms of situational settings (monitoring and active control) and mental state (workload and affect). We show an explorative experiment in a ship bridge simulator context to investigate measurement practices and uncover correlations between mental state and changes in physiology. 31 participants from an engineering student population participated in the experiment. Data were collected from two scenarios through surveys (workload and affect) and physiology sensors (electrocardiography and electrodermal activity). We highlight the following findings from our experiment: One, there is a significant difference in variables measuring mental and physiological states between two regularly occurring scenarios in the context of large ship navigation. With changes in mental and physiological states, the capacity and reaction pattern of users change, so there are different demands of the user interface and system behavior. Two, elements of mental state are correlated with changes in physiological state. Most prominently, stress and workload covary with electrodermal activity and elements of heart rate variability. This finding can support designers in evaluating different solutions by enabling them to assess changes in the mental state of users working in control rooms through physiology sensor data.

Keywords Control room · Interaction design evaluation · Workload · Affect · Heart rate variability · Electrodermal activity

1 Introduction: including users' mental state

In modern industry, many aspects of operation have moved into the control room. When combined with increasing degrees of automation, users must keep track of more information and are responsible for an increasing number of systems. Woods et al. (2002) highlight this issue of data overload, describing the issue of users being presented with an abundance of data without being able to process and act on the data in an efficient manner. In addition, control room activities are shifting toward monitoring systems,

ensuring that operation is within normal limits. In the event of anomalies, the user must be able to perform appropriate actions to ensure safe and efficient operation. Bainbridge (1983) highlights a challenge of automation—namely, that of users mainly monitoring systems instead of actively controlling them. This can be problematic when action must be taken, because users are less familiar with active control and might not have a feel for the process that they are controlling. When designing such systems, the user is usually modeled as having stable, rational behavior—i.e., responding to external events in a predictable manner (Balters and Steinert 2015). However, it has been shown that humans do not behave in stable, rational patterns (Kahneman and Tversky 1979, 1984). Behavior can be influenced in part by the mental state of users. We define mental state as the combination of affective state and experienced workload. A positive affect has been shown to influence decision making by increasing risk averseness (Isen 2001; Isen and Reeve 2005). Hart and Staveland (1988) claim that experienced workload influences problem-solving strategies. If a user

✉ Andreas Simskar Wulvik
andreas.wulvik@ntnu.no

1 Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

experiences a high workload, he might start shedding tasks or adopt a lower criterion for performance.

This knowledge—namely, the notion of behavior being influenced by mental state—has been adopted by the fields of engineering design and human–computer interaction. Affective computing (Picard 1997) aims to improve the interaction between users and computers by including knowledge of the user's mental state in the design and behavior of the system. Jiao et al. (2017) highlight the importance of integrating affective and cognitive needs when designing the user experience. A promising method for measuring mental state is through physiology. Many studies explore this correlation in strictly controlled experiments (Hjortskov et al. 2004; McDuff et al. 2014; Nasoz et al. 2004; Zhai and Barreto 2006; Zhou et al. 2011). In an engineering context, users are more prone to noise, which might influence the correlation between physiology and mental state (Balters and Steinert 2015). There are several studies that investigate more complex tasks, such as driving (Healey and Picard 2005), aviation (Nixon and Charles 2017; Wilson 2002), nuclear power plants (Gao et al. 2013), and ship navigation (Cohen et al. 2015).

Ship navigation is of special interest for the authors, especially because there is increasing activity in the field of remote and autonomous shipping. The YARA Birkeland will be one of the first autonomous ships in operation and will be operating on the Norwegian coast as of 2019 with a small crew for supervision, and then autonomously as of 2020. We believe that this is only the beginning of a shift in the maritime industry toward remote and autonomous vessels. This shift means that there will be people operating and monitoring ships from onshore control rooms. There is little research on how this type of operation will affect the people performing the tasks, which in turn may have unknown consequences on safety and efficiency of operation.

The aim of this paper is to investigate the mental state aspect of people who remotely operate and monitor ships. More specifically, we set out to understand how mental state and physiology change between different operational situations, and if mental state can be estimated from physiological responses. This can then form a basis for understanding how mental state is related to performance by monitoring physiological state in future experiments, which again can help designers when developing and evaluating new concepts.

We pursue this goal through an experiment investigating how physiological changes are related to changes in mental state in the context of ship bridges. The experiment was conducted with consumer ship simulation software. 31 participants from a student population were tasked to navigate a large ship in two scenarios: on open sea and in a narrow harbor. Stimuli were designed and verified in cooperation with an industry partner, one of the world's largest suppliers of ship bridge systems. Our contacts were professional ship simulator instructors with extensive experience operating large ships. The stimuli were created to be as realistic as possible—i.e., to replicate demands and actions encountered in a real context under normal operation. Tasks were consciously selected to represent the majority of activities on large ships, not extreme events that occur seldom, if ever. This will most likely result in smaller effect sizes in changes of mental state and physiology compared to extreme situations. Data were collected through surveys and physiology sensors. Participants rated their own affective state and workload through survey questions. Physiological state was evaluated through electrocardiography (ECG) and electrodermal activity (EDA). The results showed significant changes in mental and physiological state between the two scenarios. Concepts of stress and workload were correlated with EDA and elements of heart rate variability (HRV). These results indicate that users may have changing demands in their interfaces and system behavior as their affective state changes and that the experienced stress and workload of users are related to EDA and HRV.

2 Background: Mental state

In this work, we are interested in the mental state of users. In our definition of mental state, we include the constructs of affective state and workload. Both are of interest to the human–computer interaction community, because these concepts might influence users' behavior and how they perceive their situations. Below we present the constructs of affective state and workload, along with common subjective (Table 1) and physiological (Table 2) measurement tools for the respective constructs.

2.1 Affective state

We use the definition of Balters and Steinert (2015), which describes affect, or emotions, as a set of variables that might moderate behavior. We define affective state as the manifestation of the concept affect—i.e., the value of these variables at a specific time for a specific individual.

There are two main schools on how to describe affect in the field of psychology. The first considers affect as a set of discrete categories (Tomkins 1962; Ekman and Friesen 1969, 1971; Ekman 1992). Tomkins and McCarter (1964) describe eight categories of emotions, naming them at medium and high levels. They are interest–excitement, enjoyment–joy, surprise–startle, distress–anguish, fear–terror, shame–humiliation, contempt–disgust, and anger–rage. Ekman et al. (1972) define the six basic emotions as happiness, surprise, anger, disgust, and fear. Later, Ekman expanded the number of emotions to amusement, anger,

Table 1 Subjective evaluation of mental state

	Model	Dimensions	Method	Authors
Affective state	Circumplex model of affect	Arousal–sleepiness, Pleasure–displeasure	Affect grid, rate along separate dimensions	Russell (1980)
	PANAS	Positive affect, Negative affect	Two ten-item mood scales	Watson et al. (1988)
	AD-ACL	Energetic arousal, Tense arousal	20-25 activation-descriptive adjectives	Thayer (1986)
Workload	NASA TLX	Mental-, Physical-, Temporal demand, Performance, Effort, Frustration	Weighted sum of dimensions	Hart and Staveland (1988)
	SWAT	Time load, Mental effort load, Stress load	Weighted sum of dimensions	Reid and Nygren (1988)
	OW	Overall workload	Single scale	Vidulich and Tsang (1987)
	MCH	Overall workload	Decision tree	Wierwille and Casali (1983)

Table 2 Studies using physiological measurements to evaluate mental state

	Authors	Measurement
Affective state	Healey and Picard (2005)	Heart rate variability, electrodermal activity, respiration, muscle tension
	Pedrotti et al. (2014)	Pupil dilation, electrodermal activity
	Zhou et al. (2011)	Electrodermal activity, respiration, muscle tension, brain activity
	Katsis et al. (2008)	Heart rate variability, electrodermal activity, respiration, muscle tension
	Kim et al. (2004)	Heart rate variability, electrodermal activity, skin temperature,
	Schmidt et al. (2016)	Heart rate variability, electrodermal activity, skin temperature, respiration
	Baltaci and Gokcay (2016)	Pupil dilation, skin temperature
	Horlings et al. (2008)	Brain activity
	Mandryk and Atkins (2007)	Heart rate variability, electrodermal activity, muscle tension
Workload	Nixon and Charles (2017)	Heart rate variability, respiration
	McDuff et al. (2014)	Heart rate variability
	Gao et al. (2013)	Pupil dilation, Heart rate variability, eye blink
	Hjortskov et al. (2004)	Heart rate variability, blood pressure
	Nourbakhsh et al. (2012)	Electrodermal activity
	Brookings et al. (1996)	Heart rate variability, respiration, brain activity

contempt, disgust, embarrassment, excitement, fear, guilt, pride in achievement, relief, sadness/distress, satisfaction, sensory pleasure, and shame (Ekman Ekman 1999). The second school of thought considers affect as a combination of multiple dimensions (Russell 1980; Russell and Barrett 1999; Thayer 1967; Watson and Tellegen 1985). Watson and Tellegen describe affect as a combination of the dimensions of positive affect and negative affect (Watson et al. 1988; Watson and Tellegen 1985). Positive affect is related to the extent a person feels enthusiastic, active, and alert. Negative affect is related to subjective feelings of distress and unpleasurable engagement. Positive affect and negative affect are measured through the Positive and Negative Affect Schedule (PANAS) scales, asking people to rate two ten-item mood scales assessing positive and negative affect on a five-point scale. Thayer (1967, 1978, 1986) explains affect as the two dimensions of energetic arousal and tense arousal, consisting of the four factors general activation (energy), deactivation-sleep (tiredness), high activation (tension), and general deactivation (calmness). This is measured through the activation–deactivation adjective checklist by rating 20–25 activation-descriptive adjectives on a four-point scale. Russell et al. (Russel et al. 1989; Russell 1979, 1980) propose a model of affect with the main dimensions arousal and pleasure–displeasure in which different affective states can be described as combinations of these two dimensions—e.g., stress being the combination of displeasure and high arousal. Levels of arousal and pleasure can be assessed either through the single-item affect grid (Russel et al. 1989) or through rating arousal and pleasure along separate dimensions. Russell (1979) shows that Thayer's dimensions of energetic arousal and tense arousal can be seen as an approximate rotation of arousal and pleasure. Watson and Tellegen's dimensions can—according to Russel et al. (1989)—be seen as a 45° rotation of the same dimensions.

2.2 Workload

The notion of workload or cognitive load has been used in human factor research in relation to performance. Parasuraman et al. (2008) argue that workload is one of the few constructs that is predictive of both performance in complex human–machine interactions and of the mental state of the operator. Cooper and Harper (1969) define workload as the sum of physical and mental effort and attention required to maintain a given level of performance. When viewing workload as a function of effort, one should consider both the capabilities of operators and their state. This could be how the skill level of operators might influence their effort, as well as their physical and mental state—i.e., tired or stressed. Parasuraman et al. (2008) describe workload as a function of the demand on mental resources in relation to the resources available from the human operator. Hart and Staveland (1988) describe workload as a multidimensional construct describing the cost incurred by a human operator to achieve a particular level of performance. Workload is not an inherent property but emerges from the interaction among task requirements, context, operator skills, behavior, perceptions, and affective state (Hart and Staveland 1988; Sheridan and Stassen 1979; Xie and Salvendy 2000). These definitions of workload are human centered, focusing on the subjective perception of workload. The notion of workload as a subjective experience is supported by Johanssen et al. (1979) and Sheridan (1980). The reason why the subjective experience of workload is important, according to Hart and Staveland (1988), is that this might alter behavior. Should an operator experience a situation as a high workload, they might adopt strategies to mitigate workload and experience distress.

Subjective workload is usually measured through surveys. Commonly used tools are the NASA Task Load Index (TLX) (Hart and Staveland 1988), Subjective Workload Assessment Technique (SWAT) (Reid and Nygren 1988), Modified Cooper–Harper scale (MCH) (Wierwille and Casali 1983), and Overall Workload (OW) (Vidulich and Tsang 1987). TLX and SWAT use multiple dimensions to assess workload, offering better diagnostic properties than one dimension when trying to assess the underlying mechanisms of workload. MCH and OW are unidimensional, providing less detail, but with the advantage of being faster to fill out. In addition, there is evidence that univariate methods have greater sensitivity than multivariate methods when estimating OW (Hendy et al. 1993; Vidulich and Tsang 1987).

2.3 Measuring mental state through physiology

Affective state and workload have been shown to be reflected in physiological responses (Andreassi 2010; Boucsein 2012; Ekman et al. 1983; Levenson 2003; Wilson and Eggemeier

1991). These physiological responses are controlled by the autonomic nervous system (ANS). The ANS is divided into two branches, the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). The PNS and SNS work antagonistically to regulate physiological arousal (Appelhans and Luecken 2006). The SNS can be described as responsible for the body's fight-or-flight reactions, whereas the PNS has been said to be in charge of rest and digest. Sympathetic and parasympathetic activities are expressed in various physiological phenomena, such as heart rate, respiration, EDA, brain activity, muscle tension, pupil dilation, skin temperature, and blood pressure. For an exhaustive overview of the use of physiological measurements in an engineering context, we refer to Balters and Steinert (2015). In Table 2, we show examples of previous work on correlating mental state with physiological changes.

3 Method: experiment investigating mental state in a ship simulator

An experiment was created to investigate how self-reported mental state relates to physiology in large ship navigation settings. Our focus has been to develop an experiment that addresses regularly occurring situations when operating a large ship. This means that stimuli have been designed to replicate reality rather than to elicit specific mental states or reactions. The experiment was conducted in commercial ship simulation software with student participants. For an exhaustive description of the experiment setup and execution, we refer to Dybvik et al. (2018).

4 Stimuli: open sea and harbor

Participants were asked to steer a 200-m-long cruise ship using a commercial ship simulation software in two scenarios. These scenarios were created in cooperation with experienced ship simulator instructors from one of the world's largest suppliers of ship bridge systems. The stimuli were created to be as realistic as possible—i.e., to replicate demands and actions encountered in a real context under normal operation. It was our goal to create scenarios that represent the majority of activities on large ships, not extreme events that occur seldom, if ever. This decision will likely result in smaller effect sizes in changes of mental state and physiology compared to extreme situations.

The first scenario was designed to replicate the task of sailing on an open sea. This can be described as long periods of time with little activity, mostly spent monitoring systems. Participants were tasked to sail the ship out from port, across an empty body of water, and into a new port. The only additional stimuli given during this scenario was a low-frequency

(LF) engine rumble to create a realistic backdrop. The scenario lasted 15 min, but the duration was not communicated to participants in advance to reduce any expectation effects towards the end. When the scenario ended, the ship would be approximately halfway between the two ports.

The second scenario was supposed to recreate tasks associated with harboring. This includes navigating narrow channels and performing secondary tasks related to going to harbor under a time constraint. Participants were instructed to navigate to a berth marked on their map. When the ship started moving, a 10-min visible timer could be seen on the screen, instructing participants to reach their destination before time ran out. The secondary tasks consisted of asking prerecorded questions regarding crew and cargo over the radio at regular intervals. Crew and cargo lists were printed and placed face-down on the table in front of the participants. The lists were turned over by the participants at the beginning of the second scenario after the participants were given instructions to do so on screen. Participants were instructed in advance to give their answers using a hand-held walkie-talkie. Questions were repeated after 90 s if no reply was given or upon request from participants. Should participants reach their designated berth, a new destination would be given. This was intended to make it nearly impossible to finish the primary task within the allotted 10 min. Throughout the second scenario, radio chatter and LF engine noise were added to create a realistic backdrop.

5 Data collection: subjective and physiological

42 participants were sampled from an engineering student population. Eleven were excluded because of technical issues or failure to follow instructions. 31 participants were included in the analysis—13 females and 18 males. Ages of the participants ranged from 19 to 33 years (24.0 ± 2.74). To address our research questions, we collected both physiological data and subjectively assessed affective state and workload.

5.1 Subjective measurements

Affective state and workload were assessed through survey questions. We adopted the circumplex model of affect (Russell 1980), evaluating arousal and pleasure, as a framework for affective state. Pilot studies showed that the participants had trouble understanding the concept of arousal. Thus, alertness and awakeness were added to arousal and pleasure in an attempt to triangulate the concept of arousal described by the circumplex model of affect. Because stress is a known term for most people, it was included to see how a subjective rating of arousal and pleasure would relate to perceived levels of stress. According to the circumplex model of affect, stress should show up as a combination of high arousal and displeasure. Workload was assessed in two ways: as a single dimension of OW (Vidulich and Tsang 1987) and through the multidimensional TLX (Hart and Staveland 1988) scheme. These were selected because TLX and OW have been shown to be superior in terms of sensitivity and user acceptance (Hill et al. 1992). OW can provide an overview of subjective workload, whereas TLX gives a more detailed view of which dimensions influence subjective workload in our context. Affective state and overall workload were rated on eleven-point Likert scales. TLX was rated on seven-point Likert scales plus pairwise comparisons of the six dimensions.

5.2 Physiological measurements

Selection of the physiological data type and related sensors to use in the experiment was guided by the feasibility of integration in future products. By this, we mean sensors that could be worn by users without interfering with normal operation—e.g., wireless and comfortable to wear. In this experiment, physiological data were collected through ECG and EDA. Heart rate and HRV can be calculated from ECG. Heart rate and HRV are associated with both sympathetic and parasympathetic nervous system activity. Sympathetic activity tends to increase heart rate and decrease HRV, and vice versa for parasympathetic activity (Appelhans and Luecken 2006; Berntson et al. 1997; Camm et al. 1996). High-frequency (HF) variations (0.04–0.15 Hz) in heart rate are believed to be parasympathetically mediated, whereas LF variations (0.15–0.4 Hz) are considered a product of both parasympathetic and sympathetic activities (Berntson et al. 1997; Camm et al. 1996; Malliani et al. 1991). The normalized frequency components of LF and HF HRV are supposed to assess sympathetic and parasympathetic activity, respectively (Furlan et al. 2000; Pagani et al. 1997).

EDA is influenced only by sympathetic nervous system activity (Boucsein 2012; Dawson et al. 2007). EDA can be divided into phasic and tonic activity. Phasic and tonic activities are, according to Dawson et al. (2007), related to attention and activation, respectively. Phasic activity, or skin conductance response (SCR), is elicited by almost any stimulus that is novel, unexpected, or potentially important (Siddle 1991). Tonic activity, or skin conductance level (SCL), is related to continuous stimuli—e.g., performing a task (Bohlin 1976). ECG data were collected using the Shimmer3 ECG device (Shimmer3 ECG/EMG Unit, 2017) with a sampling rate of 512 Hz. EDA data were collected with the Shimmer3 GSR + device (Shimmer3 GSR + Unit, 2017) with a sampling rate of 128 Hz. Both devices transmitted data wirelessly via Bluetooth to a central computer

running iMotions 6.4 (iMotions 2017) for synchronization and storage.

5.3 Procedure

The IMotions (iMotions 2017) software platform was used to present stimuli and collect and synchronize subjective and physiological data. After the participants expressly consented to the experiment, physiology sensors were attached to them, and they were seated in front of a computer screen. The ECG sensor was attached with five leads on the chests of participants per the instructions provided by Shimmer, with the V_x lead in position six. The EDA sensor was attached to the middle part of the index and middle fingers on the left hand. After the sensors were attached, the participants were seated in front of a computer screen in the simulator environment (Fig. 1). The participants were instructed on how the experiment would proceed and told that they should follow instructions given either onscreen or via audio.

Figure 2 shows the experimental procedure as a timeline. First, the participants were asked to fill out a survey on their affective state to provide a baseline. This was followed by information about the ship they were supposed to steer and a video demonstrating the controls of the ship. After the instructions were completed, the first scenario was presented, and the participants were supposed to sail the ship out of harbor and across an open expanse of sea. At the end of the first scenario, surveys on affective state and workload were filled out. This was repeated for the second scenario. After both scenarios were completed and surveys on mental state were filled out, an additional survey on demographics was filled out. This concluded the experiment, and the participants were debriefed and sensors were disconnected.

## 5.4 Analysis: classical statistics and multivariate analysis

Subjective measurements used for analysis were collected after each scenario. Measurements used for analyzing the physiological state were sampled from the last 5 min of each scenario (see Fig. 2). This was intended to avoid carryover effects from previous stimuli.

Heart rate and HRV in the time and frequency domains were calculated from ECG data using Kubios HRV

Fig. 1 Experiment environment (Harbor scenario), both physical and virtual. Physiological sensors are highlighted in the red dashed rectangle, ECG (top) and GSR (bottom) (Dybvik et al. 2018)

Fig. 2 Experimental procedure
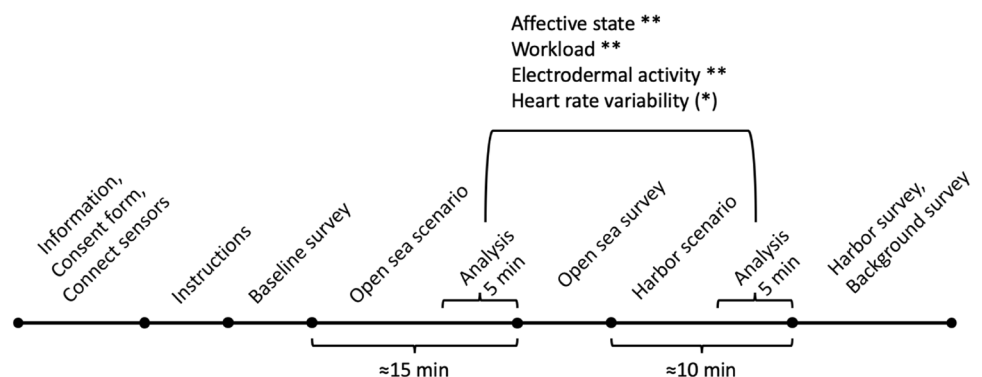
(Tarvainen et al. 2014). The frequency domain of HRV is typically calculated using either fast Fourier transformation or autoregressive modeling (AR). We have selected AR because of its increased robustness and accuracy for shorter periods (Malliani et al. 1991; Montano et al. 2009). EDA data were processed in Ledalab using continuous decomposition analysis (CDA) (Benedek and Kaernbach 2010). CDA has been shown to be more sensitive to peak detection and estimation of tonic activity as opposed to through-to-peak algorithms (Benedek and Kaernbach 2010).

Table 3 provides an overview of all physiological variables included in the analysis. Assumptions of normal distribution, significant outliers, and skewness were evaluated, and statistical tests selected accordingly. The paired-sample *t* test investigates whether there is a difference in mean values between populations. The Wilcoxon signed-rank test and sign test investigate whether there is a difference in median values between populations. In addition to statistical tests comparing populations, the results were analyzed through correlation tests and multivariate analysis (i.e., principal component analysis (PCA) and partial least-squares regression (PLSR)) to investigate the relationship between different variables. For PCA and PLSR, each variable was mean centered and standardized (1/std.dev. of each variable). Multivariate analysis was performed using the software program (The Unscrambler X 2018).

6 Results

Three outliers were removed from the arousal values owing to inconsistencies in subjective reporting of arousal compared to awakeness and alertness. In these cases, arousal levels were reported much lower than awakeness and alertness, and we assume the nonnative English-speaking participants misinterpreted arousal as sexual arousal. This was a known misunderstanding from pilot studies. The data used for both classical statistics and multivariate analysis can be found in Online Resource 1.

7 Subjective variables: affective state and workload

Table 4 shows the results from statistical tests of subjective variables. Data used for the analysis are based on surveys filled out after each scenario. Participants show significant ($p < 0.01$) changes in self-reported affective state and workload between the two scenarios, open sea and harbor conditions, for paired samples with the exception of subjective experience of performance ($p = 0.694$), as previously shown by Dybvik et al. (2018).

Table 3 Physiology variables

	Variable	Description
Heart rate variability	Mean RR (ms)	Average RR interval
	SDNN (ms)	Standard deviation of RR interval
	Mean HR (bpm)	Average heart rate
	SD HR (bpm)	Standard deviation of heart rate
	RMSSD (ms)	Root mean square of successive differences. Measure of short-term variability
	NN50 (n.u.)	Number of successive intervals that differ more than 50 ms
	pNN50 (%)	Relative number of successive intervals that differ more than 50 ms
	LFpeak, HFpeak (Hz)	Peak frequencies for LF and HF bands
	LFpow, HFpow (ms^2)	Absolute power in LF and HF bands
	LFpow, HFpow (n.u.)	Powers of LF and HF bands in normalized units $LF[n.u.] = \frac{LF[ms^2]}{LF[ms^2]+HF[ms^2]}$ $HF[n.u.] = \frac{HF[ms^2]}{LF[ms^2]+HF[ms^2]}$
	TOTpow (ms^2)	Total spectral power
	LF_HF_ratio (n.u.)	Ratio between LF and HF powers
Electrodermal activity	nSCR (n.u.)	Number of significant skin conductivity responses (SCR)
	AmpSum (μS)	Sum of SCR amplitudes
	SCR (μS)	Average phasic driver
	PhasicMax (μS)	Maximum value of phasic activity
	Tonic (μS)	Mean tonic activity
	RawMean (μS)	Mean skin conductivity (SC) value

Table 4 Subjective variable results

Variable	$\bar{X}_{sea}$	σ_{sea}	$\bar{X}_{harbour}$	$\sigma_{harbour}$	$\bar{\Delta}$	σ_{Δ}	Paired samples	Effect size (d)
Arousal 0–10[d]	6.07	1.98	7.32	1.89	1.25	1.71	0.001[b]**	0.73
Awake 0–10	6.48	2.05	7.61	2.04	1.13	1.09	< 0.001[c]**	1.04
Alert 0–10	6.26	1.95	7.42	1.86	1.16	1.61	< 0.001[c]**	0.72
Pleasant 0–10	6.35	1.62	4.68	1.80	− 1.68	1.54	< 0.001[a]**	1.09
Stress 0–10	3.23	1.94	6.39	2.14	3.16	1.88	< 0.001[a]**	1.68
Overall Workload 0–10	2.03	1.70	8.03	1.78	6.00	2.93	< 0.001[c]**	2.05
Mental demand 1–7	2.55	1.41	6.13	0.85	3.58	1.43	< 0.001[a]**	2.50
Physical demand 1–7	1.58	0.85	3.74	1.91	2.16	1.83	< 0.001[c]**	1.18
Temporal demand 1–7	1.94	0.93	6.26	0.96	4.32	1.28	< 0.001[c]**	3.39
Performance 1–7	3.90	2.07	4.13	1.88	0.23	3.17	0.694[a]	0.07
Effort 1–7	2.71	1.47	5.90	1.14	3.19	1.64	< 0.001[b]**	1.95
Frustration 1–7	2.74	1.69	4.42	1.82	1.68	2.09	< 0.001[a]**	0.80
TLX 1–7	2.86	1.23	5.60	0.74	2.73	1.37	< 0.001[a]**	1.99

[a]Paired samples t test, [b]Wilcoxon signed-rank test, [c]Sign test, *$p < 0.05$, **$p < 0.01$

[d]Three outliers were removed due to inconsistencies in subjective reporting of arousal compared to awakeness and alertness. 28 data points used for this specific analysis

8 Physiological variables: heart rate variability and electrodermal activity

The results from the statistical tests can be seen in Table 5. Heart rate variables such as mean and standard deviation of heart rate show significant ($p < 0.05$) changes. Significant changes are found in the frequency domain for normalized LF and HF powers, whereas nonnormalized powers do not exhibit this change. The LF to HF ratio of HRV, which is a common metric used to describe physiological arousal, has a positive change from the open sea scenario

Table 5 Physiological variable results

Variable	$\bar{X}_{sea}$	σ_{sea}	$\bar{X}_{harbour}$	$\sigma_{harbour}$	$\bar{\Delta}$	σ_{Δ}	Paired samples	Effect size (d)
Mean RR (ms)	802.83	134.31	726.55	116.73	− 76.29	66.05	< 0.001[a]**	1.16
SDNN (ms)	41.86	18.16	42.52	17.14	0.66	12.49	0.782[a]	0.05
Mean HR (bpm)	76.63	11.87	84.58	13.00	7.95	6.79	< 0.001[a]**	1.17
SD HR (bpm)	4.63	1.65	5.57	1.67	0.94	1.17	< 0.001[a]**	0.80
RMSSD (ms)	34.34	19.95	30.91	16.77	− 3.43	9.67	0.072[a]	0.35
NN50 (beats)	46.14	43.48	43.71	42.33	− 2.43	29.47	0.400[b]	0.08
pNN50 (%)	13.75	14.55	11.67	12.12	− 2.08	8.69	0.215[a]	0.24
LFpeak (Hz)	0.093	0.014	0.097	0.018	0.004	0.020	0.036[c]*	0.18
HFpeak (Hz)	0.214	0.070	0.155	0.024	− 0.059	0.072	0.001[b]**	0.82
LFpow (ms^2)	1273.34	1030.42	1308.92	877.50	35.58	929.32	0.855[b]	0.04
HFpow (ms^2)	549.38	851.69	497.28	655.89	− 52.10	309.36	0.185[c]	0.17
LFpow (n.u.)	71.81	13.47	76.60	10.17	4.79	9.57	0.013[a]*	0.50
HFpow (n.u.)	28.13	13.45	23.34	10.15	− 4.79	9.55	0.013[a]*	0.50
TOTpow (ms^2)	1989.34	1836.75	2017.95	1558.11	28.60	1225.82	0.964[b]	0.02
LF_HF_ratio	3.78	3.15	4.45	3.50	0.68	2.44	0.053[b]	0.28
nSCR (n.u.)	62.10	44.60	111.45	35.96	49.35	44.79	< 0.001[a]**	1.10
AmpSum (µS)	5.99	8.04	12.96	10.27	6.97	6.84	< 0.001[c]**	1.02
SCR (µS)	0.006	0.006	0.011	0.008	0.005	0.005	< 0.001[a]**	1.00
PhasicMax (µS)	2.03	2.05	2.77	2.28	0.73	1.91	0.041[a]*	0.38
Tonic (µS)	2.47	1.39	3.75	1.70	1.28	0.86	< 0.001[c]**	1.49
RawMean (µS)	2.56	1.44	3.93	1.76	1.37	0.90	< 0.001[a]**	1.53

[a]Paired samples t test, [b]Wilcoxon signed-rank test, [c]Sign test, *$p < 0.05$, **$p < 0.01$

to the harbor scenario ($p = 0.053$). Data on electrodermal activity show highly significant changes between the two scenarios for all variables except for the maximum phasic driver ($p = 0.041$).

8.1 Relation within subjective variables

In the following two sections, we highlight interesting correlations from the analysis. The full correlation table can be found in Online Resource 2. Stress is defined according to Russell (1980) as a combination of high arousal and displeasure—i.e., the opposite of feeling pleasant. Our results show that arousal has a stronger correlation to stress ($r = 0.47$, $n = 62$, $p < 0.001$) than displeasure ($r = 0.25$, $n = 62$, $p = 0.01$). For this context, we interpret arousal as the main factor of stress. In a professional context such as large ship navigation, this makes sense, because we expect affective changes to be more linked to arousal—i.e., energy level—than feelings of displeasure. We find that workload, both OW and TLX, is associated with levels of stress, displeasure, and arousal, in that order.

The subjective variables were also analyzed using principal component analysis (PCA). Values are given as cross-validated results followed by calibration in parentheses. Cross-validation was performed by leaving one participant out at a time. The results show that the first component accounted for 46% (52%) of the variance in the subjective variables. This component was mainly dominated by workload and stress, as seen in Fig. 3a. The second component, accounting for 16% (18%) of variance, consisted mainly of arousal and displeasure (Fig. 3b). These first two components, explaining 62% (70%) of the total variance in the subjective data, align with Watson and Tellegen's (1985) model of affect consisting of the dimensions positive and negative affect.

What we see from both the statistical tests and the PCA is that the largest changes in mental state are in workload and stress. This could mean that in an environment where participants are supposed to perform professional tasks, emotions—i.e., arousal and displeasure—are less influenced than the more task-related concepts stress and workload. One other explanation for the smaller changes in self-reported affective state could be the semantic understanding of the different concepts, given that our experience reveals that it is more common to discuss concepts of stress and workload in a professional context as opposed to emotions. Evaluating one's own state of arousal and displeasure could be more challenging than stress and workload for nonnative English speakers owing to unfamiliarity of the concepts.
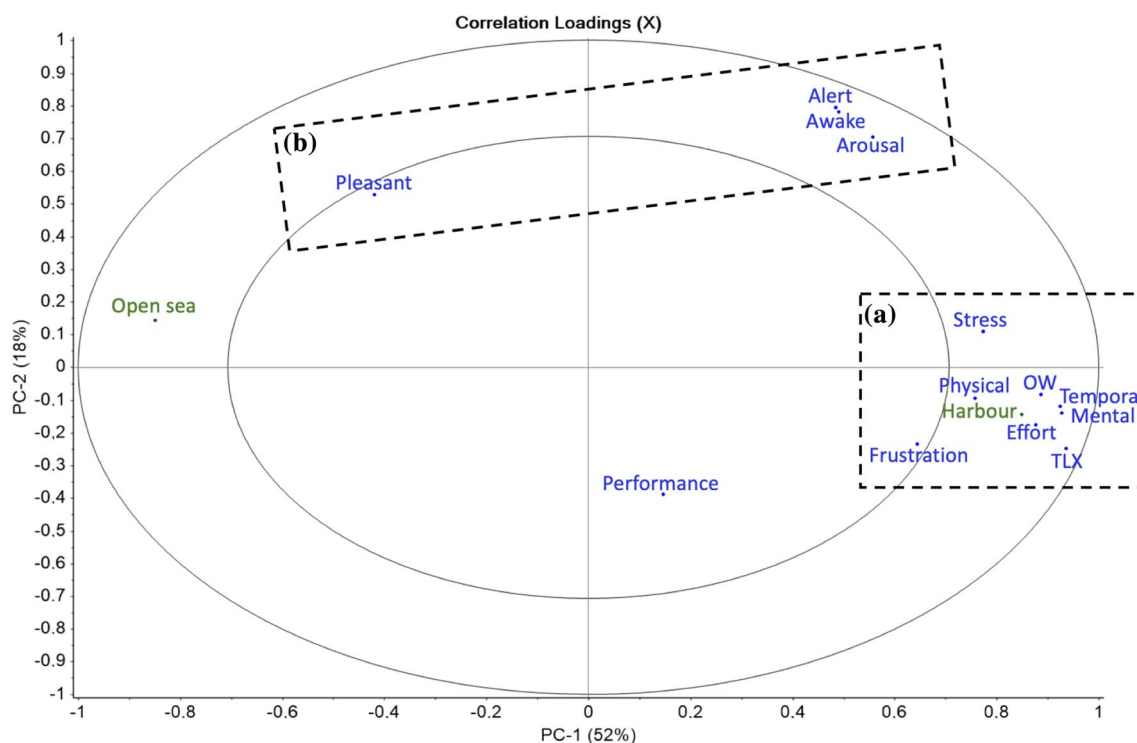

Fig. 3 PCA—correlation loadings for subjective variables

8.2 Relation between subjective and physiological variables

We find that the variables workload and stress have the strongest correlation to both HRV and EDA variables. Among the HRV variables, the HF peak frequency has the highest correlation to both stress and workload (stress: $r = -0.33$, $n = 57$, $p < 0.001$; OW: $r = -0.43$, $n = 57$, $p < 0.001$; TLX: $r = -0.61$, $n = 57$, $p < 0.001$). The number of skin conductivity responses (nSCRs), tonic level, and raw mean SC signal have the strongest correlation to stress and workload of the EDA variables. Displeasure was found to be correlated with nSCRs ($r = 0.30$, $n = 62$, $p = 0.02$). Awakeness shows correlations to the LF peak of HRV ($r = 0.37$, $n = 57$, $p = 0.01$) and HR standard deviation ($r = 0.36$, $n = 57$, $p = 0.01$). Arousal and alertness were not correlated with any of the physiological variables ($r < 0.30$). The strongest correlations to mental state were peak frequency in HF HRV and phasic and tonic EDA.

The relation between physiological and subjective variables was also investigated through PLSR (Fig. 4). The values are given as cross-validated results followed by calibration in parentheses. Cross-validation was performed by leaving one participant out at a time. We found that the first component accounts for 13% (19%) of variance in the subjective variables (marked in red) and 20% (29%) in

the physiological variables (marked in blue). The second component accounted for 30% (30%) of the variance in the physiological variables but did not contribute to explaining the variance in the subjective variables. Here, the explained variance was -2% (2%), indicating a small amount of overfitting.

The first component was dominated by HF peak, EDA variables, mean RR/HR, and to some extent normalized LF and HF components in terms of physiological variables (Fig. 4a, b). The mental state variables in the first component mainly consisted of workload and stress variables (Fig. 4a). The second component consisted of absolute power and time-domain HRV variables (Fig. 4c). The two scenarios (marked in green), cruising on open sea and docking in a narrow harbor, were down-weighted when calculating the factors. By including these variables in the model, we see the tendencies of how the variables changed between the two scenarios. Although the explained variance was relatively low, we found an emerging pattern of how the variables covaried. We observed a decrease in peak frequency HF HRV when workload and stress increased—i.e., HRV moved toward lower frequencies, which are associated with lowered parasympathetic activity, and could be a sign of increased sympathetic activity. This corresponds well to the increase in EDA for increasing levels of workload and stress, which is mediated solely by the SNS. The increase in both
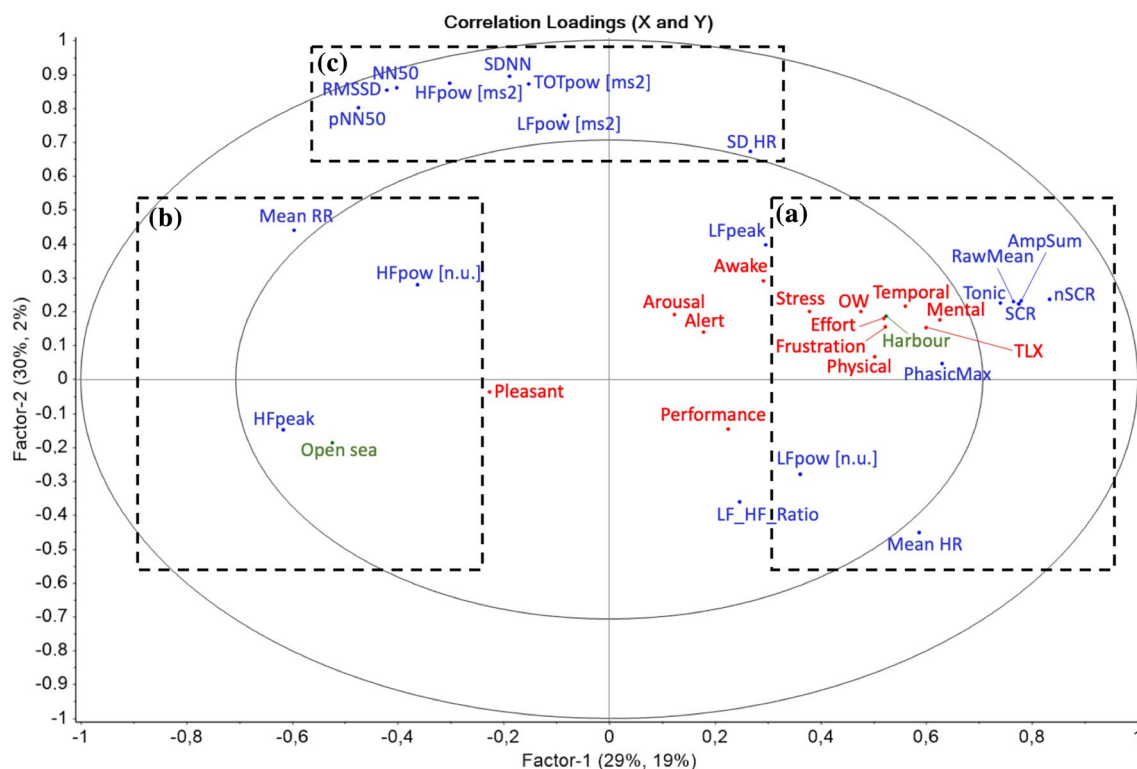

Fig. 4 PLSR—correlation loadings for physiological to subjective variables

phasic and tonic EDA could be explained by the increased number of external stimuli and task demands, which should influence phasic and tonic levels, respectively.

9 Discussion

Our results show significant changes in both mental and physiological states between the two scenarios in the experiment. There is a general increase in arousal, displeasure, stress, and workload in the scenario where participants navigated a large ship in a narrow harbor compared to the scenario where they navigated on open sea. The same increase is found for EDA and partially for HRV. Through PLSR, we find that workload and stress make up the majority of variations in the mental state of participants. Most variation in physiological data along the same component can be found in EDA and the HRV HF band peak frequency. When calculating correlations between mental and physiological state variables, we find that the strongest relations are between the dominant variables found in the PLSR analysis—namely, stress and workload for mental state and EDA and the HF HRV peak for physiological state. These correlation scores range from approximately 0.4 to 0.6. The relation between mental and physiological state variables might be influenced by the following factors: First, physiological responses are highly individual. If given the same stimuli, one participant might show strong changes in EDA, whereas another has little to no change. Second, participants might have different reactions to the stimuli. What is considered stressful for one person might be routine for the other. Third, there might be differences in how individuals rate concepts of affective state and workload, owing either to language barriers or their own understanding of the concept in question. Finally, the analysis in this paper is based on average results from participants, meaning that individual differences are not considered. Although we did not consider individual differences, we found correlations between mental and physiological states. Our interpretation is that when these results can be found despite the aforementioned factors, analysis of multiple data points from individual participants should provide even stronger results.

Changes in HRV LF/HF ratio are similar to the findings of Gao et al. (2013) and Hjortskov et al. (2004), although our LF/HF ratio is higher. McDuff et al. (2014) found a much larger change in the LF/HF ratio, doubling the mean value from approx. 0.5 to 1.0 between the resting and stress conditions. We believe that our large ratio values are due to a lower HF power in our data compared to that of Hjortskov et al. This could indicate parasympathetic withdrawal, which could be an indicator of high baseline physiological arousal. Total power is comparable to the studies of Hjortskov et al. and Gao et al., although the latter report a much

lower total power for their low-complexity task. Healey and Picard (2005) show that driver stress level was correlated most strongly to mean skin conductivity and HRV, when compared to respiration rate and electromyography in addition to the aforementioned variables. Out of these, mean skin conductivity had the strongest correlation to their stress metric. Wilson (2002) found EDA and heart rate to be more sensitive to changes in cognitive demand than HRV, which is similar to our results when comparing physiological data to the subjective ratings.

From the experiment setup, we know that the harbor scenario, where higher levels of perceived stress and workload were measured, contained more stimuli. EDA, or more specifically SCR, has been shown to be highly reactive to external stimuli (Siddle 1991). We observe a relative increase in number of SCRs in our data between the open sea and harbor scenarios, corresponding to the known reactivity to external stimuli. At the same time, we observe an increase in the tonic level of EDA as well as stress and workload. We show that there is a relation among the number of stimuli, subjectively assessed mental state, and electrodermal activity in the context of large ship navigation. However, we do not claim causality between the variables, only that they covary.

By designing tasks that replicate normal working conditions onboard a ship, it is our belief that the results of this paper are representative of the real context. However, the results are limited to the relevance of the participants partaking in the experiment. When interpreting the results of this experiment, it is important to keep in mind that the data were collected from a student population performing tasks in an unfamiliar environment, steering a large ship in a simulator. Even if the results are consistent for a student population, we do not know how this relates to the responses of professionals if they were to participate in the same experiment. We believe that students' reactions to stimuli might be stronger than would be the case for professionals owing to their training and experience in handling the given tasks. Our assumption is that the direction of responses would be similar for students and professionals, although the magnitude might differ. This should be tested in a future experiment.

10 Conclusion: mental state and physiology

In this paper, we have presented an experiment aiming to investigate changes in mental state and physiology between different scenarios and how mental state and physiology are related in the context of large ship navigation. The motivation behind the experiment has been to create a foundation for future work on continuously monitoring users' mental state in the context of remote operation and monitoring of ships.

The experiment tests the feasibility of using changes in electrodermal activity and heart rate variability to act as a proxy for users' self-reported mental state. 31 participants from a student population were tasked with navigating a large ship in two scenarios: on open sea and in a narrow harbor. The results show that there are significant changes in the variables used to measure mental state between the two scenarios. We found significant changes in EDA and several variables representing HRV. We found that self-reported stress and workload were correlated with EDA and the peak frequency in the HF band of HRV. Awakeness was found to be positively correlated with the LF band of HRV and the standard deviation of heart rate. No correlation was found between arousal or alertness and physiological variables. Multivariate analysis showed that one component could explain 13% of the variance in mental state and 20% in physiological data. The second component did not contribute to explaining any variance in the mental state of users but did explain 30% of the variability of physiological data. The first component was dominated by workload, stress, EDA, and elements of HRV. The second component was dominated by absolute power- and time-domain heart rate variability.

We draw the following conclusions from our experiment: One, there is a significant difference in variables used to measure mental and physiological states between two regularly occurring scenarios in the context of large ship navigation. As these changes occur, the capacity and reaction pattern of the user change, and the user could have different or changing demands of the user interface and system behavior. Two, elements of mental state are correlated with changes in physiological state. Most prominently, stress and workload covary with EDA and elements of HRV. This finding can serve as a foundation for how to assess changes in mental state of users remotely operating and monitoring ships through measuring changes in physiological state.

We believe our findings to be representative of similar control rooms. To verify this, experiments should be conducted in situ by collecting data on professionals working in their normal environments, such as ship bridges, power plants, airplanes, or air traffic control. Such experiments would be more prone to noise and uncontrollable variables that may influence results. Despite the challenges of noisy data and uncontrolled influencing factors, this would be the real situation we are interested in. Results from such an experiment would have the highest ecological validity, because it is the real situation. For research on how mental and physiological states are related in a control room context, this is the necessary step to find an answer. Should a connection between mental and physiological states be found from in situ experiments, it would be possible to evaluate the mental state of users unobtrusively—i.e., not interfering with their tasks by having to fill out surveys or answer questions. With this information, designers could compare concepts on how they influence the mental state of users, which again may influence task performance. We hope this can be a tool in the designers' toolkit on how to evaluate designs when working on remote operation of ships.

References

Andreassi JL (2010) Psychophysiology: human behavior and physiological response. Psychology Press, London

Appelhans BM, Luecken LJ (2006) Heart rate variability as an index of regulated emotional responding. Rev General Psychol 10(3):229. https://doi.org/10.1037/1089-2680.10.3.229

Bainbridge L (1983) Ironies of automation. Automatica 19(6):775–779. https://doi.org/10.1016/0005-1098(83)90046-8

Baltaci S, Gokcay D (2016) Stress detection in human–computer interaction: fusion of pupil dilation and facial temperature features. Int J Hum–Comput Stud 32(12):956–966. https://doi.org/10.1080/10447318.2016.1220069

Balters S, Steinert M (2015) Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices. J Intell Manuf. https://doi.org/10.1007/s10845-015-1145-2

Benedek M, Kaernbach C (2010) A continuous measure of phasic electrodermal activity. J Neurosci Methods 190(1):80–91. https://doi.org/10.1016/j.jneumeth.2010.04.028

Berntson GG, Thomas Bigger J, Eckberg DL, Grossman P, Kaufmann PG, Malik M et al (1997) Heart rate variability: origins, methods, and interpretive caveats. Psychophysiology 34(6):623–648. https://doi.org/10.1111/j.1469-8986.1997.tb02140.x

Bohlin G (1976) Delayed habituation of the electrodermal orienting response as a function of increased level of arousal. Psychophysiology 13(4):345–351. https://doi.org/10.1111/j.1469-8986.1976.tb03088.x

Boucsein W (2012) Electrodermal activity. Springer Science & Business Media, New York

Brookings JB, Wilson GF, Swain CR (1996) Psychophysiological responses to changes in workload during simulated air traffic control. Biol Psychol 42(3):361–377. https://doi.org/10.1016/0301-0511(95)05167-8

Camm AJ, Malik M, Bigger JT, Breithardt G, Cerutti S, Cohen RJ et al (1996) Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation 93(5):1043–1065. https://doi.org/10.1161/01.CIR.93.5.1043

Cohen I, Brinkman W-P, Neerincx MA (2015) Modelling environmental and cognitive factors to predict performance in a stressful training scenario on a naval ship simulator. Cogn Technol Work 17(4):503–519. https://doi.org/10.1007/s10111-015-0325-3

Cooper GE, Harper RP (1969) The use of pilot rating in the evaluation of aircraft handling qualities (No. AGARD-567). Advisory Group for Aerospace Research and Development Neuilly-Sur-Seine (France)

Dawson ME, Schell AM, Filion DL (2007) The electrodermal system. Handb Psychophysiol 2:200–223

Dybvik H, Wulvik A, Steinert M (2018) Steering a ship-investigating affective state and workload in ship simulations. In: DS92: Proceedings of the DESIGN 2018 15th international design conference (pp. 2003–2014). https://doi.org/10.21278/idc.2018.0459

Ekman Paul (1992) An argument for basic emotions. Cogn Emot 6(3–4):169–200. https://doi.org/10.1080/02699939208411068

Ekman Paul, Friesen WV (1969) The repertoire of nonverbal behavior: categories, origins, usage, and coding. Semiotica 1(1):49–98. https://doi.org/10.1515/semi.1969.1.1.49

Ekman Paul, Friesen WV (1971) Constants across cultures in the face and emotion. J Pers Soc Psychol 17(2):124. https://doi.org/10.1037/h0030377

Ekman P, Levenson RW, Friesen WV (1983) Autonomic nervous system activity distinguishes among emotions. Science 221(4616):1208–1210. https://doi.org/10.1126/science.6612338

Ekman P (1999) Basic emotions. In: Scientist TDR, MJP of C. Psychology (eds) Handbook of cognition and emotion. Wiley, New York, pp 45–60. https://doi.org/10.1002/0470013494.ch3

Ekman P, Friesen WV, Ellsworth P (1972) Emotion in the human face: guide-lines for research and an integration of findings: guidelines for research and an integration of findings. Pergamon

Furlan R, Porta A, Costa F, Tank J, Baker L, Schiavi R et al (2000) Oscillatory patterns in sympathetic neural discharge and cardiovascular variables during orthostatic stimulus. Circulation 101(8):886–892. https://doi.org/10.1161/01.CIR.101.8.886

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. https://doi.org/10.1080/00140139.2013.790483

Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol 52:139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Healey JA, Picard RW (2005) Detecting stress during real-world driving tasks using physiological sensors. IEEE Trans Intell Transp Syst 6(2):156–166. https://doi.org/10.1109/TITS.2005.848368

Hendy KC, Hamilton KM, Landry LN (1993) Measuring subjective workload: when is one scale better than many? Hum Factors 35(4):579–601. https://doi.org/10.1177/001872089303500401

Hill SG, Iavecchia HP, Byers JC, Bittner AC, Zaklade AL, Christ RE (1992) Comparison of four subjective workload rating scales. Hum Factors 34(4):429–439. https://doi.org/10.1177/001872089203400405

Hjortskov N, Rissén D, Blangsted AK, Fallentin N, Lundberg U, Søgaard K (2004) The effect of mental stress on heart rate variability and blood pressure during computer work. Eur J Appl Physiol 92(1–2):84–89. https://doi.org/10.1007/s00421-004-1055-z

Horlings R, Datcu D, Rothkrantz LJM (2008) Emotion recognition using brain activity. In: Proceedings of the 9th international conference on computer systems and technologies and workshop for PhD students in computing, CompSysTech '08. ACM, New York, NY, pp 6:II.1–6:1. https://doi.org/10.1145/1500879.1500888

Isen AM (2001) An influence of positive affect on decision making in complex situations: theoretical issues with practical implications. J Consum Psychol 11(2):75–85. https://doi.org/10.1207/S15327663JCP1102_01

Isen AM, Reeve J (2005) The influence of positive affect on intrinsic and extrinsic motivation: facilitating enjoyment of play, responsible work behavior, and self-control. Motiv Emotion 29(4):295–323. https://doi.org/10.1007/s11031-006-9019-8

Jiao RJ, Zhou F, Chu C-H (2017) Decision theoretic modeling of affective and cognitive needs for product experience engineering: key issues and a conceptual framework. J Intell Manuf 28(7):1755–1767. https://doi.org/10.1007/s10845-016-1240-z

Johanssen G, Moray N, Pew R, Rasmussen J, Sanders A, Wickens C (1979) Final report of experimental psychology group.

Mental workload. Springer, Boston, pp 101–114. https://doi.org/10.1007/978-1-4757-0884-4_7

Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. Econometrica 47(2):263–292

Kahneman D, Tversky A (1984) Choices, values, and frames. Am Psychol 39(4):341–350

Katsis CD, Katertsidis N, Ganiatsas G, Fotiadis DI (2008) Toward emotion recognition in car-racing drivers: a biosignal processing approach. IEEE Trans Syst Man Cyber Part A: Syst Human 38(3):502–512. https://doi.org/10.1109/TSMCA.2008.918624

Kim KH, Bang SW, Kim SR (2004) Emotion recognition system using short-term monitoring of physiological signals. Med Biol Eng Comput 42(3):419–427

Levenson RW (2003) Blood, Sweat, and Fears. Ann N Y Acad Sci 1000(1):348–366. https://doi.org/10.1196/annals.1280.016

Malliani A, Pagani M, Lombardi F, Cerutti S (1991) Cardiovascular neural regulation explored in the frequency domain. Circulation 84(2):482–492. https://doi.org/10.1161/01.CIR.84.2.482

Mandryk RL, Stella Atkins M (2007) A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. Int J Hum-Comput Stud 65(4):329–347

McDuff D, Gontarek S, Picard R (2014) Remote measurement of cognitive stress via heart rate variability. In: 2014 36th annual international conference of the IEEE engineering in medicine and biology society, pp. 2957–2960. https://doi.org/10.1109/EMBC.2014.6944243

Montano N, Porta A, Cogliati C, Costantino G, Tobaldini E, Casali KR, Iellamo F (2009) Heart rate variability explored in the frequency domain: a tool to investigate the link between heart and behavior. Neurosci Biobehav Rev 33(2):71–80. https://doi.org/10.1016/j.neubiorev.2008.07.006

Nasoz F, Alvarez K, Lisetti CL, Finkelstein N (2004) Emotion recognition from physiological signals using wireless sensors for presence technologies. Cogn Technol Work 6(1):4–14. https://doi.org/10.1007/s10111-003-0143-x

Nixon J, Charles R (2017) Understanding the human performance envelope using electrophysiological measures from wearable technology. Cogn Technol Work 19(4):655–666. https://doi.org/10.1007/s10111-017-0431-5

Nourbakhsh N, Wang Y, Chen F, Calvo RA (2012) Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In: Proceedings of the 24th Australian computer-human interaction conference. New York, NY, USA: ACM, pp 420–423. https://doi.org/10.1145/2414536.2414602

Pagani M, Montano N, Porta A, Malliani A, Abboud FM, Birkett C, Somers VK (1997) Relationship between spectral components of cardiovascular variabilities and direct measures of muscle sympathetic nerve activity in humans. Circulation 95(6):1441–1448. https://doi.org/10.1161/01.CIR.95.6.1441

Parasuraman R, Sheridan TB, Wickens CD (2008) Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. J Cognit Eng Decis Mak 2(2):140–160. https://doi.org/10.1518/155534308X284417

Pedrotti M, Mirzaei MA, Tedesco A, Chardonnet J-R, Mérienne F, Benedetto S, Baccino T (2014) Automatic stress classification with pupil diameter analysis. Int J Hum Comput Stud 30(3):220–236

Picard RW (1997) Affective computing. The MIT Press, Cambridge, p 167, 170

Reid GB, Nygren TE (1988) The subjective workload assessment technique: a scaling procedure for measuring mental workload. In: Hancock PA, Meshkati N (eds) Advances in psychology, vol 52. Elsevier, North-Holland, pp 185–218

Russel JA, Weiss A, Mendelsohn GA (1989) Affect grid: a single-item scale of pleasure and arousal. J Pers Soc Psychol 57(3):493–502

Russell JA (1979) Affective space is bipolar. J Pers Soc Psychol 37(3):345. https://doi.org/10.1037/0022-3514.37.3.345

Russell JA (1980) A circumplex model of affect. J Pers Soc Psychol 39(6):1161–1178. https://doi.org/10.1037/h0077714

Russell JA, Barrett LF (1999) Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. J Pers Soc Psychol 76(5):805. https://doi.org/10.1037/0022-3514.76.5.805

Schmidt E, Decke R, Rasshofer R (2016) Correlation between subjective driver state measures andpsychophysiological and vehicular data in simulated driving. In: 2016 IEEE intelligent vehicles symposium (IV), Gothenburg, Sweden, 19–22 June 2016, pp 1380–1385. https://doi.org/10.1109/IVS.2016.7535570

Sheridan T (1980) Mental workload: what is it? Why bother with it. Human Factors Soc Bull 23(2):1–2

Sheridan TB, Stassen HG (1979) Definitions, models and measures of human workload. Mental workload. Springer, Boston, pp 219–233. https://doi.org/10.1007/978-1-4757-0884-4_12

Shimmer3 ECG/EMG Unit (2017) Dublin, Ireland: Shimmersense. http://www.shimmersensing.com/products/shimmer3-ecg-sensor. Accessed 12 Nov 2017

Shimmer3 GSR + Unit (2017) Dublin, Ireland: Shimmersense. Retrieved from http://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor. Accessed 12 Nov 2017

Siddle DAT (1991) Orienting, habituation, and resource allocation: an associative analysis. Psychophysiology 28(3):245–259. https://doi.org/10.1111/j.1469-8986.1991.tb02190.x

Tarvainen MP, Niskanen J-P, Lipponen JA, Ranta-aho PO, Karjalainen PA (2014) Kubios HRV—Heart rate variability analysis software. Comput Methods Programs Biomed 113(1):210–220. https://doi.org/10.1016/j.cmpb.2013.07.024

Thayer RE (1967) Measurement of activation through self-report. Psychol Rep 20(2):663–678. https://doi.org/10.2466/pr0.1967.20.2.663

Thayer RE (1978) Toward a psychological theory of multidimensional activation (arousal). Motiv Emotion 2(1):1–34. https://doi.org/10.1007/BF00992729

Thayer RE (1986) Activation-deactivation adjective check list: current overview and structural analysis. Psychol Rep 58(2):607–614. https://doi.org/10.2466/pr0.1986.58.2.607

The Unscrambler X (2018) (Version 10.5). Camo Software, Inc., Woodbridge

Tomkins S (1962) Affect imagery consciousness: volume I: the positive affects. Springer, New York

Tomkins SS, McCarter R (1964) What and where are the primary affects? Some evidence for a theory. Percept Mot Skills 18(1):119–158. https://doi.org/10.2466/pms.1964.18.1.119

Vidulich MA, Tsang PS (1987) Absolute magnitude estimation and relative judgement approaches to subjective workload assessment. Proc Hum Factors Soc Ann Meet 31(9):1057–1061. https://doi.org/10.1177/154193128703100930

Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 54(6):1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Watson D, Tellegen A (1985) Toward a consensual structure of mood. Psychol Bull 98(2):219. https://doi.org/10.1037/0033-2909.98.2.219

Wierwille WW, Casali JG (1983) A validated rating scale for global mental workload measurement applications. Proc Hum Factors Soc Ann Meet 27(2):129–133. https://doi.org/10.1177/154193128302700203

Wilson GF (2002) An analysis of mental workload in pilots during flight using multiple psychophysiological measures. Int J Aviat Psychol 12(1):3–18. https://doi.org/10.1207/S15327108IJAP1201_2

Wilson GF, Eggemeier FT (1991) Psychophysiological assessment of workload in multi-task environments. In: Multiple-task performance. Taylor & Francis, London, pp 329–360

Woods DD, Patterson ES, Roth EM (2002) Can we ever escape from data overload? A cognitive systems diagnosis. Cogn Technol Work 4(1):22–36. https://doi.org/10.1007/s101110200002

Xie B, Salvendy G (2000) Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. Work Stress 14(1):74–99. https://doi.org/10.1080/026783700417249

Zhai J, Barreto A (2006) Stress detection in computer users based on digital signal processing of noninvasive physiological variables. In: 2006 International conference of the IEEE engineering in medicine and biology society, pp 1355–1358. https://doi.org/10.1109/IEMBS.2006.259421

Zhou F, Qu X, Helander MG, Jiao JR (2011) Affect prediction from physiological measures via visual stimuli. Int J Hum Comput Stud 69(12):801–819. https://doi.org/10.1016/j.ijhcs.2011.07.005

iMotions (2017) (Version 6.4). iMotions, Boston