



Universiteit
Leiden

The Netherlands

**Towards user-adapted training paradigms:
physiological responses to physical threat during
cognitive task performance**

Beurden, M. van; Brouwer, A.M.; Baardewijk, J.U. van; Binsch, O.;
Vermetten, E.; Roijendijk, L.

Citation

Beurden, M. van, Brouwer, A. M., Baardewijk, J. U. van, Binsch, O.,
Vermetten, E., & Roijendijk, L. (2020). Towards user-adapted
training paradigms: physiological responses to physical threat
during cognitive task performance. *Multimedia Tools And
Applications*, 79(47-48), 35867-35884.
doi:10.1007/s11042-020-09575-1

Version: Publisher's Version
License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)
Downloaded from: <https://hdl.handle.net/1887/3184540>

Note: To cite this publication please use the final published version
(if applicable).



Towards user-adapted training paradigms: Physiological responses to physical threat during cognitive task performance

Maurice van Beurden¹ · Anne-Marie Brouwer¹ · Jan Ubbo van Baardewijk¹ ·
Olaf Binsch¹ · Eric Vermetten^{2,3} · Linsey Roijendijk¹

Received: 3 December 2019 / Revised: 23 July 2020 / Accepted: 11 August 2020 /
Published online: 28 August 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Feedback of physiological responses have a great potential to support virtual training paradigms aimed to increase cognitive task performance under stressful threatening conditions. In the current study, we examined the sensitivity of a range of physiological indicators derived from electrodermal activity (EDA), blood pressure (BP) and heart rate (HR) to measure stress as induced by the threat of an electric shock (ES). In contrast to previous work that studied physiological stress responses compared to a rest condition, we compared conditions with high cognitive load combined with stress caused by threat of an ES, to conditions with high cognitive load without such stress. Twenty-five participants performed a cognitively demanding task in an experimental setup. At certain 10 s time intervals, indicated by a continuous tone, participants were either asked to do their best and increase cognitive task performance (non-threat condition), or they were told that they could receive an ES during this interval if cognitive task performance was not high enough (threat condition). Physiological measures, task performance and self-reported measures of stress and workload were analysed. Task performance and self-reported measures of stress and workload were roughly the same in both conditions. Especially EDA measures were affected by the threat of an ES. Threat and non-threat conditions could be distinguished with an across-participant classifier using EDA and BP features with an accuracy of 70%. These results suggest that EDA and BP can be used to evaluate stress coping training paradigms or to individually adapt the stress levels in virtual training environments.

Keywords Cognitive task performance · Stress · Physiological measures · Threat · Workload

✉ Maurice van Beurden
Maurice.vanbeurden@tno.nl

1 Introduction

The ability to cope with stress caused by physical threat is an important skill for military and law enforcement professionals. A potential promising tool to train professionals to cope with such stressors are virtual training programs, simulating professionals' working environments. Ideally, these programs adjust the intensity and frequency of presented stressors to individual's levels of stress [40, 21, 6, 17, 52]. For this, continuous information about an individual's stress level is valuable. In the present study, we explored the possible use of various physiological indicators in such technology by examining physiological responses under circumstances that are prerequisite for this area of application, but have not been studied together in previous literature. Firstly, we examined whether, and which, physiological responses differed between conditions presenting an increase in task demand (workload) alone, and conditions presenting physical threat in addition to an increase in workload. Secondly, we assessed response patterns against a background of performing a task, rather than to a resting baseline. Thirdly, we monitored and interpreted responses at the level of an individual person. Thus, besides examining signal characteristics aggregated across participants, we also trained classification models and assessed whether they could distinguish between threat and non-threat conditions at the level of a single individual. Finally, we designed our study such that any difference in physiology between conditions and task episodes can be attributed to mental processes rather than to differences in body movements. We focused on non-invasive and continuous physiological indicators, that can be applied in virtual training programs.

Previous literature has shown physiological effects of both workload and stress caused by physical threat. Workload has been manipulated by a range of controlled and naturalistic tasks, such as the n-back task that induces short term memory load in a controlled manner [11], mental arrhythmic [2], the Multi-Attribute Task Battery that simulates piloting tasks [22], and navigating aircrafts or vehicles under different workload conditions [5]. While specific results depend on task and other factors, physiological findings in these studies are consistent with increased activity of the sympathetic 'fight or flight' autonomous nervous system, such as increased pupil size [37], electrodermal activity or skin conductance [19], heart rate and blood pressure [30]. Threat also activates the sympathetic autonomous nervous system, at least when active coping mechanisms are possible [30]. Studies on stress caused by physical threat predominantly use the threat of electrical shocks (ES), with paradigms varying in ES predictability, whether ES can be evaded, and whether, or how often, ES are actually delivered [31, 44, 24, 43]. Previous studies reported that threats of (never administered) ES generally lead to heart rate acceleration [43, 20, 23]. Furthermore, Folkins [23] reported a rise in skin conductance after threat, and Smith et al. [43] a rise in blood pressure. For inducing threat in a virtual environment, threat of an ES without delivering it is a suitable tool since it is relatively easy to implement. In addition, the advantage of the threat of an ES is that compared to the situation an actual shock is administered, it will only affect the individual's mental state, without physiological effects of shock administration itself [18]. Finally, it mimics situations in which military professionals operate, where the threat of an event and the uncertainty if the feared event will occur contribute to their experienced stress [1, 13].

As outlined above, physiological responses to workload and physical threat are similar since both are associated with increased sympathetic nervous system activity [14, 28]. Importantly, in real-life situations, stress caused by (physical) threat will often co-occur with - or contribute to - increased experiences of workload [49, 39]. Workload is expected to increase, because in potentially stressful and threatening situations, individuals need to quickly

register and process information from their environment and decide on appropriate next moves, which has shown to be affected in stressful situations. [45]. Physiological responses following a ‘stressor’, such as physical threat, may then be attributed to stress while in fact, these responses are caused by increased workload. In our study we therefore examined whether we can differentiate between physiological responses in high workload conditions with and without concurrent experienced stress as caused by physical threat.

Furthermore, our study differs from most other studies in that we assessed effects of increases in workload, with or without threat, rather than compare it with a resting baseline. The majority of studies on physiological responses to threat view responses in a context without any (cognitively demanding) background task [43, 20, 23, 25]. In the application we have in mind, individuals are not resting but actively performing a task, both mentally and physically. Body movements and sensory input associated with the background task should be similar during the increased workload- and threat episodes, such that any physiological differences between these conditions can be ascribed to mental processes [12].

In this study we compared the responsiveness to stress caused by physical threat of several physiological indicators that can be measured continuously and relatively unobtrusively. We chose electrodermal activity (EDA) and several cardiovascular measures, namely heart rate (HR), systolic blood pressure (SBP) and diastolic blood pressure (DBP), following van der Vijgh et al. [53], who identified these as the most reliable psychophysiological indicators of stress responses in real time, and in a virtual context. In addition, we estimated cardiac output (CO) and total systemic peripheral vascular resistance (TPR, a measure of resistance of the arteries) since these measures hold promise to distinguish between experiencing stress as a threat or as a challenge [4]. These physiological variables were analysed at two different time scales. Firstly, at the overall level, comparing conditions with and without episodes of threat of ES. Secondly, at the level of seconds, we also provides insights in the response patterns of physiological indicators, specifically around the intervals when workload was increased either under threat of an ES or without such a threat. When the level of stress can be detected immediately after a stressor occurs, this could be directly used as feedback in a virtual learning environment.

2 Methods

2.1 Participants

Twenty-five males (age range = 18–35 years, mean age = 23,92 years) from the participant pool of TNO (Netherlands Institute of Applied Sciences, where the study was conducted) participated in the study. They all self-reported to be free of cardiovascular medical conditions. Participants were asked not to consume alcohol or other drugs 24 h prior to the experiment. In addition, they were asked to refrain from the consumption of caffeine and nicotine, and performing heavy physical exercise in the period three hours before the experiment. Participants received 25 euros for participating and compensation for their travel costs. The study was approved by the TNO ethics committee (approval number 2017-003).

2.2 Design

A within-subjects design was applied with a non-threat and a threat condition. Each condition was presented once to a participant in a task block of 11.5 min. The order of conditions was counterbalanced across participants. During both conditions, participants performed a continuous cognitive task (see below) for the whole task block period of 11.5 min. During the threat condition, participants were connected to a Transcutaneous Electrical Nerve Stimulation (TENS) device. They were told that, if they did not perform well, they would receive an electric shock (ES) at specific 10 s time periods that were indicated by the continuous sound of a tone. During the non-threat condition participants were disconnected from the TENS device and were instructed to increase cognitive effort during the indicated 10 s time periods.

2.3 Materials

2.3.1 Equipment

Figure 1 shows a schematic overview of the equipment and setup. Electrodermal activity (EDA), blood pressure (BP) and heart rate (HR) were continuously measured during the experiment using BIOPAC Systems at 500 Hz sampling rate. BP and HR were measured using a non-invasive blood pressure monitor (CNAP® Monitor 500, CN Systems, Graz, Austria) with a cuff on the upper arm and two electrodes: one on the index and one on the middle finger of the non-dominant hand. EDA was measured using two electrodes on the ring and little finger of the non-dominant hand (EL507, BIOPAC MP150, BIOPAC Systems). To increase the quality of the recordings, participants' non-dominant arm rested on a soft and comfortable armrest after the electrodes were placed for the duration of the measurements. All physiological data was recorded and synchronized using BIOPAC's software package *AcqKnowledge* on one pc, while on another pc (stimulus computer) the cognitive task software was run. To synchronize the physiological signals with the cognitive task, markers were sent by parallel port from the task pc to the acquisition pc. A Medisana TENS device was used to induce the threat of an ES. Two electrodes were placed between the shoulder blades and only in the threat condition the electrodes were connected to the TENS device to suggest the actual delivery of ES. However in this study ES were never delivered. Previous work applying the threat of an ES showed that the threat itself is an effective way to induce physiological responses indicative of stress [31, 44, 43, 20, 23].

2.3.2 Cognitive Task

Participants performed a cognitive visuomotor tracking task [48] in combination with an n -back task [9, 34, 29], as schematically depicted in Fig. 2. Participants used a joystick to continuously manoeuvre a virtual moving blue disc underneath a red target disc in the centre of the display. The blue disc was programmed to continuously float away from the red target. Simultaneously, participants had to perform a 2-back task. Letters were presented inside the red disc for 1 s each. Participants were asked to press a button on the joystick each time the current letter was the same as two letters earlier. This was the case for 20% of the letters. The outcome measures from this double task are the average Euclidian distance over time between the red and blue disc for the tracking task, the reaction time of pressing the response button and the accuracy (percentage correct) for the 2-back task.

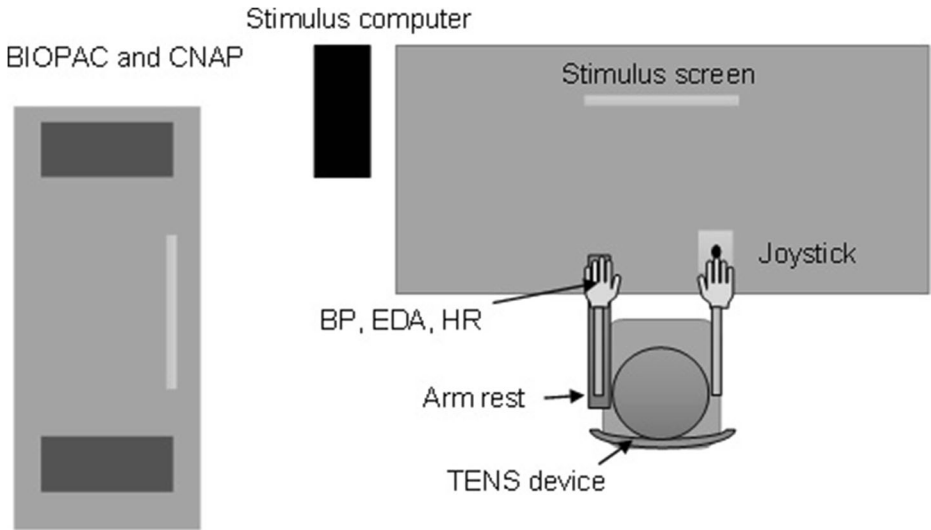


Fig. 1 Schematic representation of the setup. The participants were sitting in front of a desk with the stimulus screen in front of them. The electrodes to measure EDA, and the finger and handcuff to measure BP were attached to the participants' non-dominant hand resting on an armrest to avoid unwanted movements. The dominant hand was on the joystick to interact with the task. The two electrodes of the Medisana TENS were placed on the back between the shoulder blades. The CNAP blood pressure system and BIOPAC systems were located on the left side of the participant

During both 11.5-minute task blocks (threat and non-threat), a salient continuous tone was presented six times for a duration of ten seconds that indicated that they should increase their effort (with or without threat of the ES). The first continuous tone was presented thirty seconds after the start of the block, and each following tone was separated by a period varying between 90 and 100 s of uninterrupted task performance. Thus, for each participant 6 epochs of high workload, threat data and 6 epochs of high workload, non-threat data were recorded.

2.3.3 Subjective experience

The Perceived Stress Scale (PSS) was used to measure a participant's overall stress level during the past month [15]. Perceived anxiety was measured with a Dutch version of the short-STAI [50, 36, 51]. Cognitive workload was measured using the NASA-TLX [38]. Additionally, perceived stress during the experiment was measured by a single item question "How stressed did you feel?". For this, a 20-point scale was used that was modelled after the preceding NASA-TLX items, ranging from "very low" to "very high".

2.4 Procedure

On arrival participants were welcomed in a (de)briefing room and signed the consent form after the experimental procedure was explained. After that, the EDA electrodes were attached on the non-dominant hand, as well as the ES electrodes of the TENS between the shoulders. Next, the participants were escorted to the experimental room. Participants were seated behind a desk and filled out the PSS questionnaire. Before the start of the experiment, participants practiced the cognitive task in three practice blocks. These blocks lasted five minutes and were

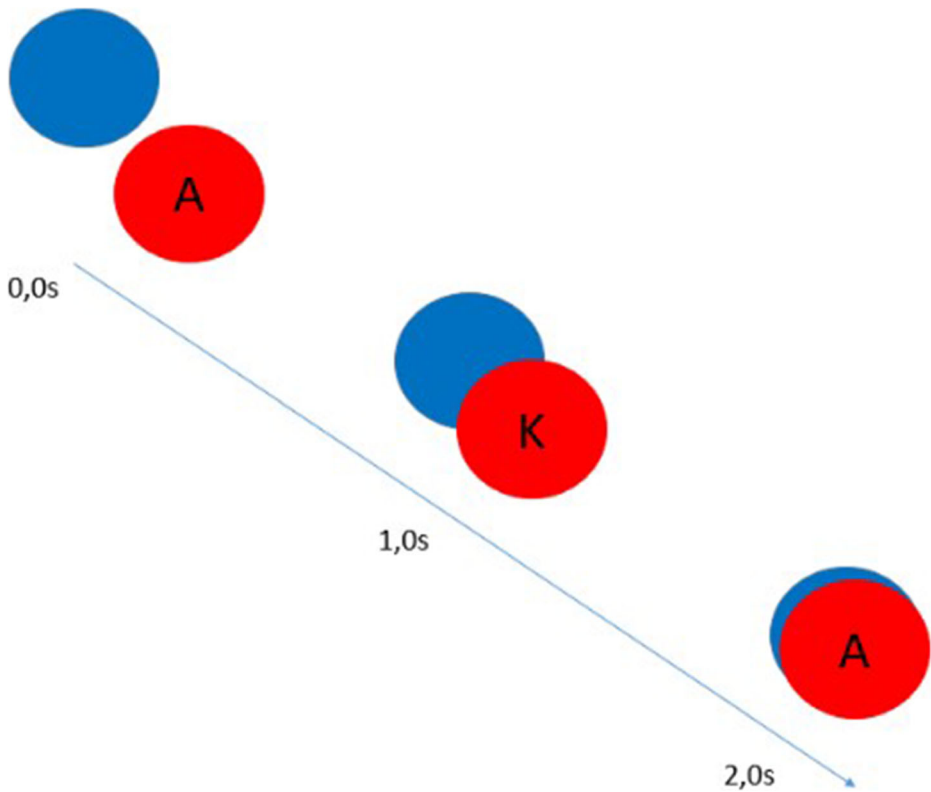


Fig. 2 A screenshot of the cognitive task. Participants had to keep the blue disc, that floats away from the red disc, underneath the red disc using a joystick. In the centre of the red disc, a new letter appeared every second. Participants had to press a button as soon as a letter was presented that was the same as two letters before

interleaved with two-minute breaks. During the first practice block, the experimenter stayed in the room to monitor whether the task was well understood and provided feedback on the task execution to increase learning speed. During the two remaining practice blocks, the experimenter left the room and only entered the room during the short break to answer questions, if necessary. After the practice blocks, BP cuffs were attached to the participant, the non-dominant arm was placed on a comfortable armrest and the measurement systems were calibrated.

Next, a three-minute rest measurement was conducted in which participants were instructed not to move. After that, the first experimental condition was conducted. The participant started either with the threat- or non-threat condition. In the non-threat condition participants were instructed to perform the task as good as possible and to do their best especially when they heard the tone. The TENS electrodes were not attached to the TENS device in the non-threat condition. In the threat condition, participants were told that an electric shock could be administered when their performance was below average during the tone, and the TENS electrodes were attached to the TENS device. After each experimental condition, participants filled out the NASA TLX, STAI and single-item stress surveys before they received a 5-minute break. When participants had performed the threat condition at first, the break between the two conditions was used to disconnect the TENS sensors from the device. After

completing both experimental conditions, a second rest measurement was conducted for three minutes. Next, the measurement equipment was disconnected, and participants were debriefed that we never intended to induce actual ES. The complete procedure, including (de)briefing, practice, filling out questionnaires, breaks, attaching and removing sensors took about 1 ½ hours.

2.5 Data analysis

Stress sensitivity of each physiological measure was investigated at the ‘overall level’ (i.e., a comparison between averages over complete threat- and non-threat task blocks as well as the rest measurement – at the level of multiple minutes) and at the ‘event level’ (i.e., a comparison between responses to the continuous tone in the threat and non-threat task blocks – at the level of seconds). Furthermore, to test for the possibility of detection of stress in a single individual, across participant classification models using the different physiological measures as features were trained to distinguish between threat and non-threat epochs.

2.5.1 Preprocessing physiological data

The arterial blood pressure data was processed using the Physio Toolkit MATLAB toolbox [27] available for free public use. First, data was down sampled from 500 to 128 Hz. Next, based on peak analysis of the blood pressure signal, the systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR) and cardiac output (CO) were calculated. CO was estimated using the Liljestrand method [32, 46]. Total peripheral resistance (TPR) was estimated as $80 \times MAP / CO$ [41], where MAP (mean arterial pressure) is defined as $\frac{2}{3} \times DBP + \frac{1}{3} \times SBP$. The signal quality index algorithm of the toolbox was used to detect abnormal arterial blood pressure beats. Intervals with abnormal beats were detected and data was linearly interpolated over these intervals. On average, the BP data contained 99.5% ($SD = 0.75\%$) of the data per participant.

EDA data was examined in two ways. Firstly, as epochs of raw measured data (EDA), and secondly, the number and mean amplitude of significant non-specific skin conductance response (nSCRs) per minute were extracted using the Continuous Decomposition Analysis method [3] as implemented in the Ledalab toolbox for Matlab. A skin conductance response was considered significant when the amplitude of the response was higher than $0.05 \mu S$ [7].

To analyse the physiological responses at the level of the first three-minute rest block and the 11-minute threat- and non-threat blocks, the mean of the physiological variables were calculated over these periods. As indicated above, we refer to this as the ‘overall level’.

To analyse the physiological response at the ‘event level’ in the threat- and non-threat conditions, data was first z-scored for each physiological measure and each participant using the overall mean and standard deviation of the data collected during both the threat and non-threat condition. Next, for each physiological measure the mean value in the period of -10 until 0 s before the tone was subtracted from the complete data trace, running from -10 to 50 s after the tone, followed by averaging this data per second.

For the statistical analysis, physiological data from the six epochs in the threat condition and the six epochs in the non-threat condition for each participant were averaged to reduce noise, whereas in the classification analysis every individual epoch was used.

2.5.2 Statistical analysis

R (version 4.0.0) and Matlab (version 2017a) were used for statistical analyses. For all statistical tests, the level of significance was set at 0.05. Normality of the distributions of the residuals were tested with Shapiro-Wilk tests. Though in the text below we refer to the parametric tests, non-parametric tests were used if normality was violated. Table 1 gives an overview of the used statistical tests.

Task performance At the overall level, to compare cognitive task performance between the non-threat and threat condition paired t-tests were used. At the event level, two-way repeated measures ANOVAs were used to test for effects of presence of tone and threat condition on task performance: average performance in the 10 s before the tone was compared to the average performance during the 10 s tone for the non-threat and threat condition. The effect of threat condition on n-back reaction time was not tested at the event level, as too few datapoints were available for a reliable estimate at the event level.

Subjective measures To compare the three subjective measures (Short-STAI, NASA-TLX, stress question) between the non-threat and threat conditions paired t-tests were used.

Physiological response To test for effects of threat condition on the physiological data at the overall level, one-way repeated measures ANOVAs were used comparing the first rest measurement, the non-threat condition and the threat condition. To compare the physiological responses between the non-threat and threat condition at the event level, a running t-test was used to compare all time windows in the range of 0 until 30 s after the tone. In order to deal with the issue of multiple testing, we followed the approach of Gladwin et al. [26] using a within-subject cluster-based nonparametric randomization test with 500 randomizations [35]. This approach indicated that the length of continuous time windows with a nominally

Table 1 Overview of statistical tests

Variables	Time scale	Comparisons	Parametric test	Non-Parametric test
Task: N-back accuracy, N-back response time, tracking distance Subjective: Short-STAI, NASA-TLX, stress question Physiological: HR, SBP, DBP, CO, TPR, EDA, number of nSCRs, amplitude of nSCRs	Overall	non-threat, threat	T-test	Wilcoxon signed-rank test
Task: N-back accuracy, racking distance Physiological: HR, SBP, DBP, CO, TPR, EDA, number of nSCRs, amplitude of nSCRs	Event	non-threat, threat; before tone, during tone	Two-way repeated measures ANOVA with post-hoc t-tests	Wilcoxon signed rank tests for main effects
Task: N-back accuracy, racking distance Physiological: HR, SBP, DBP, CO, TPR, EDA, number of nSCRs, amplitude of nSCRs	Event	non-threat, threat	Running t-tests	-

significant effect was 10 for EDA, 5 for SBP and TPR, and 4 for the remaining physiological measures.

2.5.3 Classification physiological response

A linear support vector machine (SVM) model [16] was used to classify threat and non-threat epochs (starting at the time of tone onset and ending 30 s later) over participants. Features for classification were obtained by averaging the pre-processed data of each epoch over 5 s for the period of 0 until 30 s after the tone, resulting in 6 features for each epoch and each physiological measure (EDA, HR, SBP, DBP, CO, TPR). Additionally, the mean values of each physiological measure in the period of -10 until 0 s before the tone were taken as additional features, resulting in a total of 42 features per epoch. Per participant 6 threat and 6 non-threat epochs were included in the analysis. The data set was randomly split into a training set with epochs from 18 out of 24 (75%) participants, and a test set of the remaining 6 participants (25%). Leave-one-subject-out cross-validation was performed on the training set. As including multiple types of signals does not necessarily increase classification performance in comparison to a single signal modality [33], additional models were trained using fewer signal modalities. One model was trained with only EDA features because EDA was the most significant measure from the overall and event level analyses (7 features). Another model was trained using two modalities (EDA and BP) (21 features).

3 Results

All data of one participant was removed from the dataset because the tone was absent in the threat of shock condition. Physiological data of another participant was removed because the physiological measurement failed. In total, the physiological data was analysed for 23 participants, and the task performance and survey data for 24 participants.

3.1 Task performance

3.1.1 Overall level

Performance in terms of 2-back accuracy was slightly but significantly higher, $t(23) = -2.09$, $p < .05$, $d = 0.43$, in the threat condition ($M = 89.1\%$, $SD = 4.1\%$) compared to the non-threat condition ($M = 88.6\%$, $SD = 4.2\%$). Tracking distance and 2-back reaction time were not significantly affected by threat condition, $Z = -0.58$, $p = .57$ and $Z = -1.03$, $p = .30$, respectively.

3.1.2 Event level

At the event level, a Wilcoxon signed-rank test on tracking distance showed a significant main effect of presence of tone, $Z = -3.47$, $p < .001$, $r = .71$, with a slight decrease in tracking distance during the tone ($Mdn = 0.071$) compared to before the tone ($Mdn = 0.073$). No main effect of condition was found, $Z = -0.03$, $p = .97$. This suggests that participants increased their effort during the tone in both conditions. The repeated measures ANOVA on 2-back accuracy

showed no significant main effect of presence of tone ($F(1,23) = 0.05, p = .83$), no main effect of non-threat/threat condition ($F(1,23) = 2.67, p = .12$) and no interaction effect ($F(1,23) = 3.04, p = .09$).

3.2 Subjective measures

The general perception of stress, as measured by the PSS, at the start of the experiment, was on average 12.67 ($SD = 4.44$). This indicates that our group of participants showed a normal level of stress as the norm for males is 12.1 ($SD = 5.9$) [15]. Paired t-tests showed no statistically significant differences in self-reported anxiety (short-STAI, $t(23) = 1.36, p = .19$), workload (NASA-TLX, $t(23) = 0.92, p = .37$), or stress ($t(23) = 0.74, p = .47$) between the non-threat and threat conditions. Averages and standard deviations are shown in Table 2.

3.3 Physiological response

3.3.1 Overall level

Figure 3 shows the average value for each physiological parameter in the rest, non-threat and threat task conditions. From the cardiovascular measures, HR, SBP, and CO showed a main effect of condition (rest, non-threat and threat) ($F(1,23) = 32.26, p < .0001, \eta_p^2 = 0.59$; $F(1,23) = 7.02, p < .01, \eta_p^2 = 0.24$; $\chi^2 = 16.7, p < .001$). Averaged EDA, number of nSCRs and amplitude of nSCRs all showed a significant main effect of condition ($\chi^2 = 7.3, p < .05, W = 0.16$; $\chi^2 = 10.5, p < .01, W = 0.23$; $Z = 15.9, p < .001, W = 0.35$). DBP and TPR were not significantly different among conditions ($F(1,22) = 1.43, p = .25$; $F(1,22) = 2.88, p = .08$).

Rest versus task Post-hoc comparisons showed that HR was higher in the non-threat and threat condition compared to the rest condition ($t(22,1) = -7.57, p < .001, D = 1.58$; $t(22,1) = -6.14, p < .001, D = 1.28$, respectively). CO and amplitude of nSCRs were also increased in both conditions compared to rest ($Z = -3.58, p < .001, r = .73$; $Z = -3.29, p < .01, r = .67$; $Z = -1.98, p < .05, r = .40$; $Z = -3.78, p < .001, r = .77$). SBP and number of nSCRs were only increased in the threat condition compared to rest ($t(22,1) = -3.58, p < .01, D = 0.75$; $Z = -4.11, p < .001, r = .84$). Other post-hoc tests were not statistically significant.

Non-threat versus threat All three EDA measures (averaged EDA, number of nSCRs and amplitude of nSCRs) were higher in the threat condition than in the non-threat condition ($Z = -2.81, p < .05, r = .57$; $Z = -2.31, p < .05, r = .47$; $F = -3.41, p < .01, r = .70$, respectively). SBP

Table 2 Mean and standard deviation of questionnaire scores for the non-threat and threat conditions

	Non-threat		Threat	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Short-STAI	17.42	5.91	16.54	5.91
NASA-TLX	11.26	2.07	10.91	2.77
Stress question	7.38	4.83	8.00	4.63

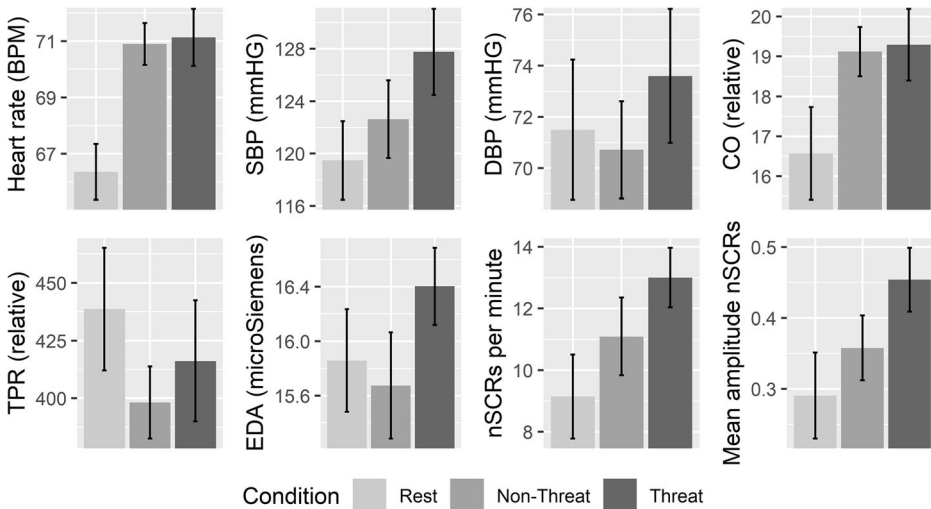


Fig. 3 Overall level averages for each physiological parameter as recorded during the first three minute rest block, the non-threat and threat task blocks. Error bars represent 95% confidence intervals for within-subject designs

was also significantly higher in the threat condition than in the non-threat condition ($t(22) = -2.239, p < .05, D = 0.47$). No significant differences were found for HR and CO.

3.3.2 Event level

Figure 4a-f shows respectively averaged EDA, HR, SBP, DBP, CO and TPR over time from 10 s before the tone onset until 50 s after. The onset of the tone, signalling to participants to perform the task extra well, induced a response in all variables, both in the threat and non-threat condition. Differences between the conditions are represented by asterisks that indicate moments at which the conditions significantly differed ($p < .05$) without correcting for multiple comparisons. Following the method by Gladwin et al. [26] for correcting for multiple comparisons, EDA, CO and TPR reached the sufficient number of successively significant differences between threat and non-threat conditions.

For a more continuous impression of when the differences occur for the different variables, Fig. 5 shows the p-values in one overview. During the period of the tone (0–10 s) p-values for EDA were low. For CO and TPR, differences between threat and non-threat emerged only after the tone had ended.

3.4 Classification

Table 3 shows the accuracy with which epochs can be correctly classified as originating from the threat or non-threat condition, based on combinations of physiological data. EDA in combination with SBP and DBP resulted in the highest accuracy, with 70% for the test set (i.e., for a participant that was not used in training the model).

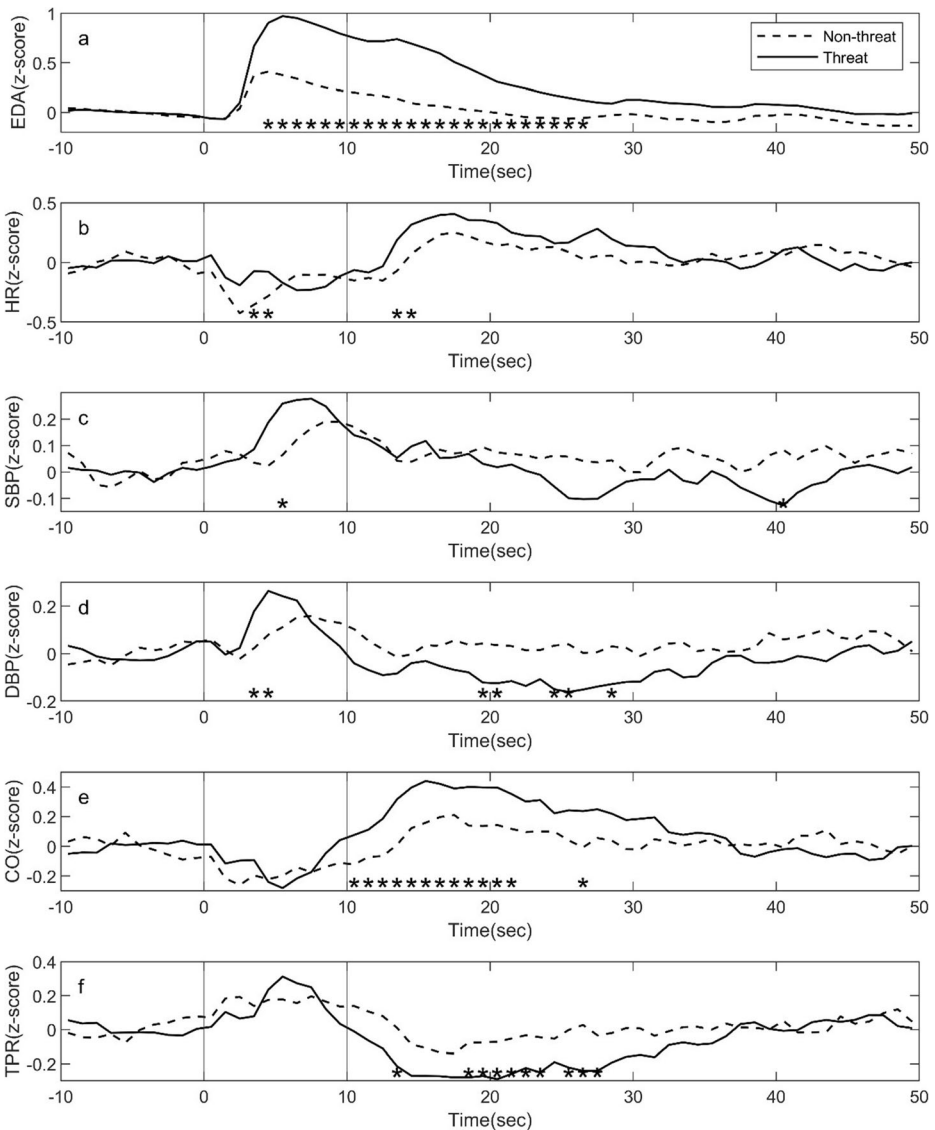


Fig. 4 a-f Response patterns of EDA, HR, SBP, DBP, CO and TPR over time from 10 s before tone onset until 50 s after. The vertical lines mark tone onset and offset. Individual traces were baselined by subtracting the average physiological value during the 10 s preceding the onset of the tone from the complete trace. They were then averaged across participants and condition. Asterisks mark the timepoints where the difference between non-threat and threat conditions were significant

4 Discussion

In the current study, we examined the response pattern of a range of physiological indicators derived from electrodermal activity (EDA), blood pressure (BP) and heart rate (HR) to increased workload with and without a physical threat. We prompted individuals to perform extra well on a cognitive demanding dual task during specific time periods, either accompanied by the threat of an ES, or

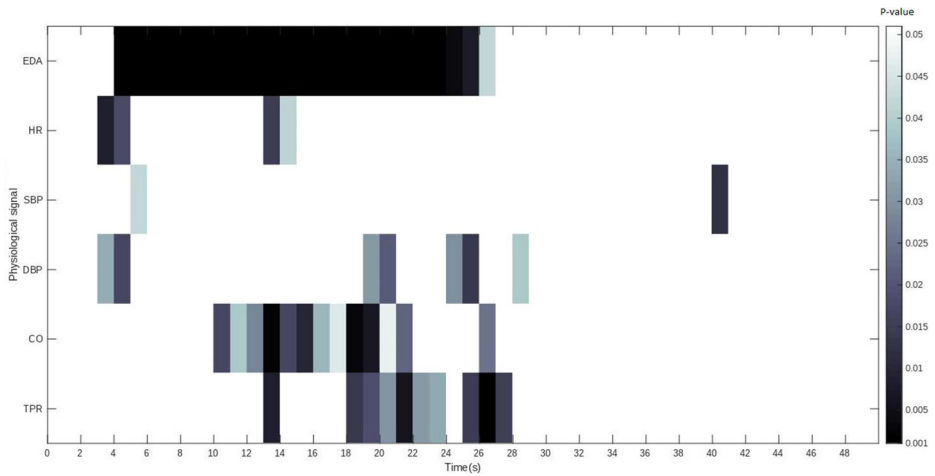


Fig. 5 P-values of running t-tests comparing the various physiological responses between threat and non-threat conditions. Time represents the time since the onset of the tone. Tone offset is at 10 s, marked by the vertical line

without such threat. This was done to examine, in the context of performing a cognitive task, the response to stress as induced by physical threat in several physiological indicators, self-reported indicators of stress and workload, and task performance.

During the threat condition, a higher level of EDA (reflected in the average skin conductance, as well as in the number and amplitude of phasic responses) and SBP was found compared to the non-threat condition. For HR, DBP, CO and TPR results revealed no statistically significant effect. When focussing on the response pattern at the event level, EDA showed a more pronounced rise in response to the tone in the threat condition compared to the non-threat condition. After the time period of the tone, CO increased and TPR decreased significantly in the threat compared to the non-threat condition. For HR, SBP and DBP no significant differences in response pattern between the experimental conditions was observed. This result shows that in terms of overall levels of physiological variables, SBP and EDA are good indicators to distinguish between periods of cognitive task performance with or without the threat of an ES. When focussing on the event-response pattern, EDA, CO and TPR markers can be used to indicate whether a short event is experienced as threatening or non-threatening. Classification analysis showed best results when using EDA, SBP and DBP to predict participants’ responsiveness to a threat vs. non-threat events. We think that this particular combination of variables worked well in the classification analysis, because they complement each other with respect to when, relative to the tone onset, they distinguished between threat and non-threat conditions. In future work it would be interesting to include a threat condition without a cognitive task and a rest condition with the 10 s beep, to further disentangle physiological effects of threat and sound [8] in different circumstances in one and the same experiment.

Table 3 Classification accuracy of different feature sets

Variables included	Cross-validation accuracy	Test set accuracy
EDA SBP DBP HR CO TPR	0.74	0.68
EDA SBP DBP	0.78	0.70
EDA	0.72	0.62

In contrast with previous literature on threat by ES [25, 53] both analyses (statistical and machine learning) showed that HR was not affected by threat, when comparing the result with the non-threat condition. A potential reason for this discrepancy is that we compared threat and non-threat conditions that both involve a continuous cognitive task, rather than comparing threat conditions with a rest condition. This is also reflected in the results that show that both threat and non-threat conditions revealed a significant increase in HR compared to the rest condition, suggesting that the increase is caused by increase in workload, rather than by stress caused by the threat of ES. Therefore, to study the potential of HR as marker for stress, future studies should compare stress manipulations to control conditions relevant within the stressor context rather than a resting baseline condition, especially when outcomes are applied in system support tools and multimedia applications used in real life settings.

While physiological effects of threat condition were found, participants did not report higher levels of stress or workload during the threat condition compared to the non-threat condition. We believe that towards the end of the condition participants learned that there will be no electrical shock. At the time of filling out the surveys at the end of each threat condition, this may have affected their memory of their earlier experience or their willingness to report stress.

We were able to correctly distinguish between threat and non-threat epochs using physiological data, even though the current models were trained using only 18 participants with each 6 instances of threat and 6 instances of non-threat data epochs. With more data, we expect higher classification accuracy. Also note that we focused on a model that classifies data from participants that were not included at all when training the model. Classification accuracies for within-participant models could be higher but would be unpractical since for the envisioned use (detecting stress in a virtual environment to train stress coping skills) we cannot record a large amount of threat and non-threat data of an individual before starting the training. Recording substantially more events from one participant using the current paradigm would also increase the likelihood that participants stop believing that ES are ever administered. Furthermore, it should be noted that our model distinguishes between episodes of increased workload and physical threat on the one hand and episodes of increased workload without physical threat on the other hand. While this more closely reflects the situation that we are interested in, this is likely more difficult than distinguishing between rest on the one hand, and increased task load and mental stress on the other hand. In addition, classification accuracies may improve when optimizing the pre-processing (e.g., selecting other features or using different outlier criteria) or when optimizing the classification approach, but this was not the focus of the current study.

In terms of task performance the results showed a lower tracking distance during the tone, suggesting that participants followed the instructions and increased their performance during the tone. Our study did not reveal detrimental effects of threat on performance. Indeed, stress as caused by physical threat does not necessarily decrease task performance and could even enhance it. However, (physical) stress can affect decision making, due to biased judgments or impaired cognitive functioning [45, 47, 54] and in the long run stress is considered to be detrimental for health [42].

We introduced our study with the long-term goal of adapting virtual training programs to ensure efficient tailoring of the training to the individual person. For example, a stressor can be made more intense when stress levels decrease. A similar application would be in tailoring stimuli for traumatic exposure to individuals with PTSD [40, 10]. Without the requirement of online adaptation, a well-trained stress detection model could be used to examine whether and

when professionals ‘keep their head cool’ under stressful conditions. This can support discussing their stress coping skills after the virtual training program. The model could also be used to evaluate the effectiveness of interventions to learn to cope with stress. Indeed, military professionals showed an attenuated physiological stress responses to an induced stressor compared to civilians [47]. However, it is unknown whether this reflects a positive effect of training or selection (i.e. only individuals who are relatively resilient to stress are likely to choose or be selected for a military profession). The discrepancy we found between physiological and self-reported data underlines the added advantage of integrating physiological indicators of stress over only self-report measures.

We do not know yet how specific the observed physiological responses are for the currently used stressor and task. The next step is to generalize a model trained on data following the current paradigm to situations with other events signalling stress and heightened workload. It may be required to collect more data using different events that co-occur with stress, followed by slightly different types of (mental) activity in order to be able to build a model that generalizes across situations varying in stressor intensity and type and the degree of increased workload.

In conclusion, the results showed that EDA and SBP were good indicators to distinguish between overall levels of threat and non-threat conditions while performing a cognitive task. When focussing on the response patterns, EDA, CO and TPR indicators can be used to indicate whether a short event is experienced as threatening or non-threatening. Classification analysis showed that the best indicators to differentiate between non-threat and threat were EDA, SBP and DBP. Another interesting result is that in this study HR, often used as indicator of stress or threat, was not a sensitive predictor to distinguish between threat and non-threat conditions.

Acknowledgements We would like to thank Lucas Heijdt and Emiel Stoelinga for assisting in data collection and implementing the cognitive task. This work was funded by the Dutch Ministry of Defense (V1532 program) and the TNO research ERP-Body Brain Interactions.

References

1. Bartone PT, Adler AB, Vaitkus MA (1998) Dimensions of Psychological Stress in Peacekeeping Operations. *Mil Med* 163:587–593. <https://doi.org/10.1093/milmed/163.9.587>
2. Beatty J (1982) Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol Bull* 91:276–292. <https://doi.org/10.1037/0033-2909.91.2.276>
3. Benedek M, Kaernbach C (2010) Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology* 47:647–658. <https://doi.org/10.1111/j.1469-8986.2009.00972.x>
4. Blascovich J, Mendes W (2000) Challenge and threat appraisals: The role of affective cues. In: Forgas J (ed) *Feeling and thinking: The role of affect in social cognition*. Cambridge University Press, Cambridge, pp 59–82
5. Borghini G, Astolfi L, Vecchiato G, Mattia D, Babiloni F (2014) Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci Biobehav Rev* 44:58–75. <https://doi.org/10.1016/j.neubiorev.2012.10.003>
6. Bosse T, Gerritsen C, de Man J, Treur J (2014) Towards virtual training of emotion regulation. *Brain Inform* 1:27–37. <https://doi.org/10.1007/s40708-014-0004-9>
7. Boucsein W, Fowles DC, Grimmes S, Ben-Shakhar G, Roth WT, Dawson ME, Filion DL (2012) Publication recommendations for electrodermal measurements. *Psychophysiology* 49:1017–1034. <https://doi.org/10.1111/j.1469-8986.2012.01384.x>
8. Bradley MM, Lang PJ (2000) Affective reactions to acoustic stimuli. *Psychophysiology* 37:204–215
9. Braver TS, Cohen JD, Nystrom LE, Jonides J, Smith EE, Noll DC (1997) A Parametric Study of Prefrontal Cortex Involvement in Human Working Memory. *Neuroimage* 5:49–62

10. Brouwer A-M, Neerinx MA, Kallen VL, van der Leer L, ten Brinke M (2011) EEG alpha asymmetry, heart rate variability and cortisol in response to virtual reality induced stress. *J Cybertherapy Rehabil* 4:21–34
11. Brouwer AM, Hogervorst MA, Holewijn M, van Erp JBF (2014) Evidence for effects of task difficulty but not learning on neurophysiological variables associated with effort. *Int J Psychophysiol* 93:242–252. <https://doi.org/10.1016/j.ijpsycho.2014.05.004>
12. Brouwer A-M, Zander TO, van Erp JB, Korteling JE, Bronkhorst AW (2015) Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Front Neurosci* 9:136. <https://doi.org/10.3389/fnins.2015.00136>
13. Campbell DJ, Nobel OB-Y (2009) Occupational stressors in military service: A review and framework. *Mil Psychol* 21:S47–S67. <https://doi.org/10.1080/0895600903249149>
14. Cinaz B, Arnrich B, La Marca R, Tröster G (2013) Monitoring of mental workload levels during an everyday life office-work scenario. *Pers Ubiquitous Comput* 17:229–239. <https://doi.org/10.1007/s00779-011-0466-1>
15. Cohen S, Kamarck T, Mermelstein R (1983) A global measure of perceived stress. *J Health Soc Behav* 24: 385–396
16. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297. <https://doi.org/10.1007/bf00994018>
17. Ćosić K, Popović S, Kukulja D, Horvat M, Dropuljić B (2010) Physiology-driven adaptive virtual reality stimulation for prevention and treatment of stress related disorders. *Cyberpsychology Behav Soc Netw* 13: 73–78. <https://doi.org/10.1089/cyber.2009.0260>
18. Davis M (1989) Sensitization of the acoustic startle reflex by footshock. *Behav Neurosci* 103:495–503
19. Dawson ME, Schell AM, Filion DL (2017) The electrodermal system. In: *Handbook of psychophysiology*, 4th edn. Cambridge University Press, New York, pp 217–243
20. Deane GE (1961) Human heart rate responses during experimentally induced anxiety. *J Exp Psychol* 61: 489–493. <https://doi.org/10.1037/h0049220>
21. Fairclough SH (2009) Fundamentals of physiological computing. *Interact Comput* 21:133–145. <https://doi.org/10.1016/J.INTCOM.2008.10.011>
22. Fairclough SH, Venables L (2006) Prediction of subjective states from psychophysiology: A multivariate approach. *Biol Psychol* 71:100–110. <https://doi.org/10.1016/j.biopsycho.2005.03.007>
23. Folkins C (1970) Temporal factors and the cognitive mediators of stress reaction. *J Pers Soc Psychol* 14: 173–184
24. Gehrman PR, Hall M, Barilla H, Buysse D, Perlis M, Gooneratne N, Ross RJ (2016) Stress Reactivity in Insomnia. *Behav Sleep Med* 14:23–33. <https://doi.org/10.1080/15402002.2014.940112>
25. Giannakakis G, Grigoriadis D, Giannakaki K, Simantiraki O, Roniotis A, Tsiknakis M (2019) Review on psychological stress detection using biosignals. *IEEE Trans Affect Comput* 1. <https://doi.org/10.1109/TAFFC.2019.2927337>
26. Gladwin TE, Hashemi MM, van Ast V, Roelofs K (2016) Ready and waiting: Freezing as active action preparation under threat. *Neurosci Lett* 619:182–188
27. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101. <https://doi.org/10.1161/01.cir.101.23.e215>
28. Goldstein DS (2010) Adrenal Responses to Stress. *Cell Mol Neurobiol* 30:1433–1440. <https://doi.org/10.1007/s10571-010-9606-9>
29. Kirchner WK (1958) Age differences in short-term retention of rapidly changing information. *J Exp Psychol* 55:. <https://doi.org/10.1037/h0043688>
30. Kreibitz SD (2011) Emotion, motivation, and cardiovascular response. In: *How motivation affects cardiovascular response: Mechanisms and applications*. American Psychological Association, Washington, D.C., pp 93–117
31. Lago TR, Hsiung A, Leitner BP, Duckworth CJ, Chen KY, Ernst M, Grillon C (2018) Exercise decreases defensive responses to unpredictable, but not predictable, threat. *Depress Anxiety* 35:868–875. <https://doi.org/10.1002/da.22748>
32. Liljestrand G, Zander E (1928) Vergleichende bestimmungen des minutenvolumens des herzens beim menschen mittels der stickoxydulmethode und durch blutdruckmessung. *Z Gesamte Exp Med* 59:105–122
33. Lingenfels F, Wagner J, Deng J, Brueckner R, Schuller B, André E (2018) Asynchronous and event-based fusion systems for affect recognition on naturalistic data in comparison to conventional approaches. *IEEE Trans Affect Comput* 9:410–423. <https://doi.org/10.1109/TAFFC.2016.2635124>
34. Manoach DS, Siewert B, Israel B, Edelman R (1997) Prefrontal cortex fMRI signal changes are correlated with working memory load *Vascular Imaging View project Emotional Responses to Music Perception View project*. <https://doi.org/10.1097/00001756-199701200-00033>

35. Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG-and MEG-data. *J Neurosci Methods* 164:177–190
36. Marteau TM, Bekker H (1992) The development of a six-item short-form of the state scale of the Spielberger State–Trait Anxiety Inventory (STAI). *Br J Clin Psychol* 31:301–306
37. McDougal DH, Gamlin PD (2015) Autonomic control of the eye. *Compr Physiol* 5:439–473. <https://doi.org/10.1002/cphy.c140014>
38. Moroney WF, Biers DW, Eggemeier FT (1995) Some measurement and methodological considerations in the application of subjective workload measurement techniques. *Int J Aviat Psychol* 5:87–106. https://doi.org/10.1207/s15327108ijap0501_6
39. Parent M, Pepsakhovich V, Mandrick K, Tremblay S, Causse M (2019) The diagnosticity of psychophysiological signatures: Can we disentangle mental workload from acute stress with ECG and fNIRS? *Int J Psychophysiol* 146:139–147. <https://doi.org/10.1016/j.ijpsycho.2019.09.005>
40. Repetto C, Gorini A, Vigna C, Algeri D, Pallavicini F, Riva G (2009) The use of biofeedback in clinical virtual reality: the INTREPID project. *J Vis Exp* 33:1554. <https://doi.org/10.3791/1554>
41. Sherwood A, Allen MT, Fahrenberg J, Kelsey RM, Lovallo WR, Van Doornen LJ (1990) Methodological guidelines for impedance cardiography. *Psychophysiology* 27:1–23
42. Slavich GM, Irwin MR (2014) From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. *Psychol Bull* 140:774–815. <https://doi.org/10.1037/a0035302>
43. Smith T, Houston B, Stucky R (1985) Effects of threat of shock and control over shock on finger pulse volume, pulse rate and systolic blood pressure. *Biol Psychol* 20:31–38
44. Soeter M, Kindt M (2012) Erasing fear for an imagined threat event. *Psychoneuroendocrinology* 37:1769–1779. <https://doi.org/10.1016/j.psyneuen.2012.03.011>
45. Starcke K, Brand M (2012) Decision making under stress: A selective review. *Neurosci Biobehav Rev* 36:1228–1248. <https://doi.org/10.1016/j.neubiorev.2012.02.003>
46. Sun JX, Reisner AT, Saeed M, Heldt T, Mark RG (2009) The cardiac output from blood pressure algorithms trial. *Crit Care Med* 37:72–80
47. Toet A, Bijlsma M, Brouwer A-M (2017) Stress response and facial trustworthiness judgments in civilians and military. *Sage open* 7:1–11. <https://doi.org/10.1177/2158244017725386>
48. Valk PJ, Simons RM, Struyvenberg PA, Kruit H, van Berge Henegouwen MT (1997) Effects of a single dose of loratadine on flying ability under conditions of simulated cabin pressure. *Am J Rhinol* 11:27–36
49. van Beurden M, Roijendijk L (2019) Towards a mixed reality platform for applied cognitive load evaluation. In: Harris D (ed) *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer International Publishing, Cham, pp 123–136
50. van der Bij AK, de Weerd S, Cikot RJ, Steegers EA, Braspenning JC (2003) Validation of the dutch short form of the state scale of the Spielberger State-Trait Anxiety Inventory: considerations for usage in screening outcomes. *Public Health Genomics* 6:84–87
51. van der Ploeg HM, Defares PB, Spielberger CD (1980) *Manual for the Self-assessment Questionnaire, STAI-DY, addendum. Handleid. bij Zelf-beoordelings Vragenlijst [Manual Dutch Adapt. STAI-Y]*
52. van der Vijgh B, Beun RJ, van Rood M, Werkhoven P (2014) GASICA: generic automated stress induction and control application design of an application for controlling the stress state. *Front Neurosci* 8:400
53. Van Der Vijgh B, Beun R-J, Van Rood M, Werkhoven P (2015) Meta-analysis of digital game and study characteristics eliciting physiological stress responses. *Psychophysiology* 52:1080–1098. <https://doi.org/10.1111/psyp.12431>
54. van Maanen L, van der Mijjn R, van Beurden MHPH, Roijendijk LMM, Kingma BRM, Miletić S, van Rijn H (2019) Core body temperature speeds up temporal processing and choice behavior under deadlines. *Sci Rep* 9. <https://doi.org/10.1038/s41598-019-46073-3>

Affiliations

Maurice van Beurden¹ · Anne-Marie Brouwer¹ · Jan Ubbo van Baardewijk¹ · Olaf Binsch¹ · Eric Vermetten^{2,3} · Linsey Roijendijk¹

¹ Department of Human Factors, Netherlands Organisation for Applied Scientific Research, Soesterberg, The Netherlands

² Central Military Hospital, Dutch Ministry of Defence, Utrecht, The Netherlands

³ University of Leiden, Leiden, The Netherlands