



HAL
open science

The diagnosticity of psychophysiological signatures: Can we disentangle mental workload from acute stress with ECG and fNIRS?

Mark Parent, Vsevolod Peysakhovich, Kevin Mandrick, Sébastien Tremblay,
Mickaël Causse

► To cite this version:

Mark Parent, Vsevolod Peysakhovich, Kevin Mandrick, Sébastien Tremblay, Mickaël Causse. The diagnosticity of psychophysiological signatures: Can we disentangle mental workload from acute stress with ECG and fNIRS?. *International Journal of Psychophysiology*, 2019, 146, pp.139-147. 10.1016/j.ijpsycho.2019.09.005 . hal-03977435

HAL Id: hal-03977435

<https://hal.science/hal-03977435>

Submitted on 7 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The diagnosticity of psychophysiological signatures: Can we disentangle mental workload from acute stress with ECG and fNIRS?

Mark Parent^{a,*}, Vsevolod Peysakhovich^b, Kevin Mandrick^b, Sébastien Tremblay^a, Mickaël Causse^b

^a Co-DOT Laboratory, École de psychologie, Pavillon Félix Antoine Savard, bureau 1144, Université Laval, Québec, QC G1K 7P4, Canada

^b ISAE-SUPAERO, 10 avenue Edouard Belin, 31055 Toulouse, France

ABSTRACT

Keywords:

Mental workload
Stress
Functional near infrared spectroscopy
Electrocardiography
Machine learning
Neuroergonomics

The ability to identify reliable and sensitive physiological signatures of psychological dimensions is key to developing intelligent adaptive systems that may in turn help to mitigate human error in complex operations. The challenge of this endeavor lies with diagnosticity. Despite different underlying causes, the physiological correlates of workload and acute psychological stress manifest in rather similar ways and can be easily confounded. The current work aimed to build a diagnostic model of mental state through the simultaneous classification of mental workload (varied through three levels of the n-back task) and acute stress (the presence/absence of aversive sounds) with machine learning. Using functional near infrared spectroscopy (fNIRS) and electrocardiography (ECG), the model's classifiers was above-chance to disentangle variations of mental workload from variations of acute stress. Both ECG and fNIRS could predict mental workload level, the best accuracy resulted from the two measures in combination. Stress level could not be accurately diagnosed through ECG alone, only with fNIRS or ECG and fNIRS combined. Individual calibration may be important since stress classification was more accurate for those with higher subjective state anxiety, perhaps due to a greater sensitivity to stress. Mental workload and stress were both better classified with activity in lateral prefrontal regions of the cortex than the medial areas, and the HbO2 signal generally lead to better classification than HbB. The current model represents a step forward to finely discriminate different mental states despite their rather analog physiological correlates.

1. Introduction

The high mental workload imposed by complex activities such as piloting an airplane or operating a nuclear power plant can reduce performance and increase human error, with sometimes serious financial or even life-threatening consequences (Byrne and Parasuraman, 1996; Causse et al., 2016; Durantin et al., 2014; Parasuraman et al., 2008). Conversely, a very low level of mental workload can lead to boredom (Sawin and Scerbo, 1995) and a vigilance decrement that equally risks a cost to human life (Pattyn et al., 2008). Furthermore, complex activities are often also associated with acute psychological stress, the effects of which – an impairment to attention, memory, or decision making (Bryce and Floresco, 2016; LeBlanc, 2009) – can be difficult to disentangle from that of excessive mental workload. To ascertain the relative impact of mental workload and psychological stress in parallel during complex tasks is challenging, particularly as traditional methods such as questionnaires fail to capture moment-to-moment fluctuations in cognitive state. A reliable and sensitive set of measures to discriminate between different levels of mental workload

and psychological stress in real time is key to developing intelligent adaptive systems, which in turn can help to prevent the deleterious consequences of human error in the performance of complex tasks.

The literature on adaptive automation suggests that both under- and over-load are detrimental to human performance (Parasuraman et al., 2000): under-load can lead to vigilance decrement (Pattyn et al., 2008), while over-load promotes errors, cognitive tunneling, or difficulties in adapting to situational changes (Durantin et al., 2014; Plessow et al., 2011; da Silva, 2014). Some authors have proposed an optimal workload level that is neither too low nor too high (Hou and Fidopiastis, 2014; Parasuraman et al., 2000). Acute stress is also often viewed as detrimental to performance (e.g., Elzinga and Roelofs, 2005; Hembree, 1988), with extended exposure to stress leading to adverse health issues (Murphy, 1996). Neuroimaging evidence supports the idea that acute psychological stress can temporarily alter the functioning of the prefrontal lobes (e.g., Arnsten, 2009; Qin et al., 2009; Schoofs et al., 2008), thus contributing to the occurrence of human error.

Mental workload and psychological stress are often used indistinctively (e.g., with the term “mental stress”, see Hjortskov et al.,

* Corresponding author at: Co-DOT Laboratory, École de psychologie, Pavillon Félix Antoine Savard, bureau 1144, Université Laval, Québec, QC G1K 7P4, Canada.
E-mail address: mark.parent@psy.ulaval.ca (M. Parent).

2004), perhaps because the source and their effects are often very similar. For example, it is well-documented that a task generating a very high mental workload can also lead to an increase in psychological stress when an individual feels overwhelmed by difficulty (Warm et al., 2008). Furthermore, even if mental workload and psychological stress are thought to be triggered by different mechanisms (e.g., cognitive effort vs. emotion), and underpinned by distinct brain centers (e.g., prefrontal cortex vs. limbic system; Gaillard and Wientjes, 1994), the probability of observing similar physiological effects during a task is high as they both impact upon the autonomic nervous system (e.g., Causse et al., 2010; Critchley et al., 2013; Mandrick et al., 2016). However, the two should be considered as distinct phenomena (Chen, 2006; Hidalgo-Muñoz et al., 2018); a high level of mental workload does not necessarily elicit a high psychological stress level, and a high stress level may also occur when mental workload is low.

Both mental workload and psychological stress are common in naturalistic work settings. There is evidence that the two concepts may mutually impact upon each other. For instance, coping with acute stress requires mental resources that will be no longer available for the performance of the cognitive task (Mandrick et al., 2016; Stawski et al., 2006; Williges and Wierwille, 1979). Similarly, Matthews and Campbell (2010) revealed that high levels of distress may be related to lower working memory performance, due to higher distress lowering the threshold for an acceptable mental workload. Such a performance decrease under stress is flexible since the recruitment of additional cognitive resources seems to protect from its negative impacts (Eysenck and Calvo, 1992). A moderate level of stress can also enhance motivation (Anderson, 1976), which might increase the actual amount of effort invested in the task. In this sense, maintaining a certain level of stress might be desirable during particular situations. For example, Wolf (2009) has shown that during learning, stress can help to consolidate information in memory. Thus, it seems that different situations come with different optimal levels of workload and stress, and if one can identify these 'sweet spots' it may be possible to maximize desirable effects (i.e., performance, learning). In attempting to distinguish between the two concepts, Gaillard and Wientjes (1994) proposed that mental workload (mental effort in their study) is characterized by efficient energy expenditure, increasing or sustaining performance, while stress (distress) is seen as an inefficient expenditure of energy that does not sustain performance, instead provoking unpleasant feelings such as anxiety.

Quantifying in real-time both mental workload and psychological stress in operators has been a research endeavor for decades. Physiological measures are often viewed as well-suited to quantifying and even predicting mental workload in dynamic settings (Durkee et al., 2013), since physiological assessment of mental workload can be performed in real time without interrupting individuals in their ongoing activity and tasks (Borghini et al., 2017; Borghini et al., 2012; Hincks et al., 2016). Furthermore, physiological measurements are objective and not prone to desirability and other biases that affect questionnaires and self-reports. One approach is to study the peripheral manifestations of a high mental stress such as cardiovascular responses (Hjortskov et al., 2004), while it is also possible to explore more direct measures of brain activity such as electroencephalography (EEG) or functional near infrared spectroscopy (fNIRS). For example, using fNIRS, Herff et al. (2014) were able to classify workload levels during an n back task with an accuracy level of 50% (higher than a chance level of 33% with three n back levels). A promising way to improve classification performance is to combine several physiological modalities (Hirshfield et al., 2009; Liu et al., 2017a), for example, Aghajani et al. (2017) demonstrate that a hybrid system (EEG + fNIRS) allows higher classification accuracy for mental workload than using EEG or fNIRS alone. One recent study achieved accuracy of over 90% in distinguishing between stress and non-stress conditions during a mental arithmetic task using EEG and fNIRS combined (Al-Shargie et al., 2016). However, in this study stress was induced using time pressure and negative feedback which arguably

may have forced the participant to increase effort, thus blurring the line between a stress and a mental workload classifier.

Many studies investigating physiological measures of mental workload have focused on a unidimensional construct (e.g., Liu et al., 2017a; Wang et al., 2012), yet workload is considered by many researchers as multidimensional (Hart, 2006; Wickens, 2002; Wilson and Eggemeier, 1991). A unitary view would constrict the ability to respond to fluctuations in workload, for example, time load (or time pressure) might negatively impact drivers' performance (Rendon-Velez et al., 2016) while other forms of cognitive load might actually increase performance (Victor, 2014). In other cases, it could be beneficial to consider the specific dimensions of multiple resource theory (i.e., stage, modality, codes; Wickens, 2002) to avoid interference between tasks, and for this reason, some authors have called for further investigation into multidimensional physiological workload (Matthews et al., 2015). The ability to physiologically distinguish and predict up to six sources of workload (see NASA-TLX, Hart, 2006), is appealing but, we believe, unrealistic based on the current level of knowledge. Rather, as a first step, we suggest trying to dissociate between cognitive and affective aspects.

As mentioned previously, one of the challenges in dissociating mental workload and psychological stress is their overlap in terms of physiological markers. For example, both mental workload and acute stress are known to impact heart rate and heart rate variability (Bernardi et al., 2000; Fairclough et al., 2005). One study reported a cross-participant classification accuracy of 73% using pressure sensors embedded in a chair to distinguish between cognitive load and stress (Arnrich et al., 2010). However, their study did not include sub-levels of workload and stress, nor did it include a situation in which cognitive load and stress were simultaneously administered. Finally Mühl et al., 2014 used EEG to predict two workload levels in an n-back task (similar to that used in the current study) across two affective conditions (relaxation and stress), with around 70% accuracy. However, this was achieved using within subject classifiers, thus limiting the ability to generalize to new participants without a training/calibration session. Furthermore, they did not attempt to perform a classification of stress level.

Another challenge that comes with classifying mental workload and psychological stress is related to inter-individual variability. Individuals might react differently, both psychologically and physiologically in response to workload or stress, making physiological metrics person-specific when predicting mental states (Macaš et al., 2009). Prefrontal cortex activations are not only correlated with task difficulty, but also with individual levels of mental effort (Causse et al., 2017). Age is another factor to consider and has been shown to differentially affect the hemodynamic response of the brain during mental tasks (Laguë-Beauvais et al., 2013). Furthermore, individuals with behavioral disorders have been shown to have lower cardiac and electrodermal responses when facing stressors compared to the rest of the population (Popma et al., 2006), as do unfit individuals (de Rooij and Roseboom, 2010). It has been investigated whether such under-reactivity to stress, coined blunted stress reaction, could be explained by factors like self-appraisal of stress/difficulty, or the amount of invested mental effort (Brindle et al., 2017). However, results indicated that individuals displaying blunted stress reaction were no different when appraising stress/difficulty and invested a similar mental effort, thus other factors might be responsible for blunted stress reactions, increasing the variance among individual's physiological responses.

The present study sought to build a diagnostic model of mental state by detecting simultaneously, instead of concurrently or independently, mental workload and stress. Particular care was invested in the experimental design to minimize overlap between the two concepts. Machine learning techniques able to handle a large number of features were used in order to combine fNIRS and electrocardiographic (ECG) measures. Finally, the study took into account individual characteristics (gender, age, state anxiety, and perceived difficulty) in the

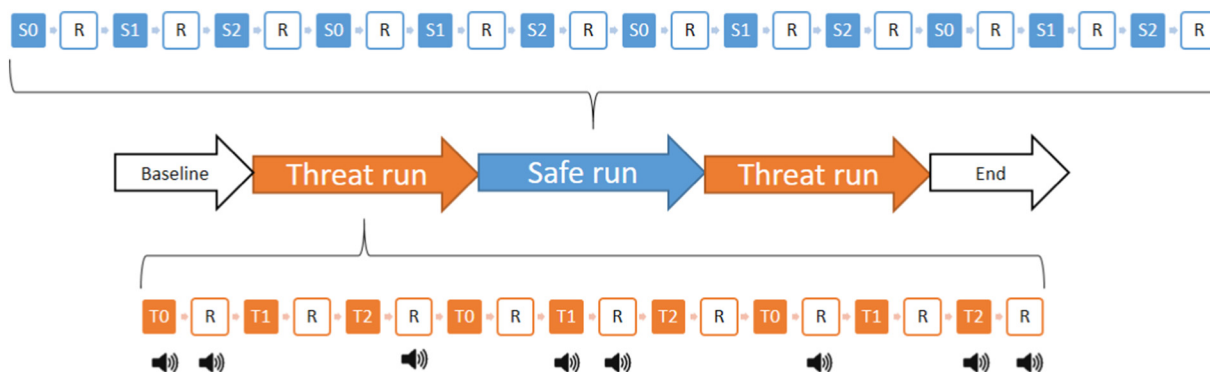


Fig. 1. The experiment was composed of resting blocks (R), safe blocks with 3 levels of n-back (S0, S1 and S2) and threat block, also with 3 levels of n-back (T0, T1 and T2). Unpredictable aversive loud sounds (volume icon) were played at random during threat runs. Run and trial orders were counterbalanced and pseudo-randomized.

interpretation of the physiological measures. All results presented in this paper derive from models that predict participant responses having been trained on the responses of other participants only.

2. Materials and method

2.1. Participants

Eighteen participants (4 females), aged between 20 and 35 (mean = 24.6, SD = 5.0), were recruited from National Civil Aviation School (ENAC) and Higher Institute of Aeronautics and Space Engineering School (ISAE-SUPAERO) in Toulouse, France. Participants reporting psychological, neurological conditions or cardiovascular disease, or taking medication affecting the brain or autonomic functions could not participate in the study. Participants gave written informed consent. The study complied with the Declaration of Helsinki for human experimentation and was approved by medical committee (CPP du Sud-Ouest et Outre-Mer IV, no CPP15-010b/2015-A00458-41).

2.2. Toulouse n-back task

This study used a modified version of the n-back task called Toulouse n-back Task (TNT). The TNT is a combination of a classical n-back task with a mental arithmetic task (Mandrick et al., 2016; Peysakhovich et al., 2016). This variation allows one to add a processing load to the working memory load already elicited during the classical version of the task. During the TNT, participants are required to memorize and compare results of arithmetic operations instead of just single characters. Operations were either additions or subtractions of dual-digit multiples of 5 (e.g., 75–20, 15 + 70) displayed in the center of a gray background. Just like the classical version of the task, difficulty was modulated by changing the “n” parameter. In this study, this parameter was either 0 (low load), 1 (medium load) or 2 (high load). During the 0-back, operations of which the result was equal to 50 were the target stimuli. During 1- and 2-back, operations of which the result was the same as the last (1-back) or the one before-last (2-back) were the target stimuli. For example during a 1-back condition, “60 – 15” is a target stimulus if it was preceded by “35 + 10”, since the two computations share the same result (45). Participants were instructed to press a green button on a Cedrus response pad (RB-740, Cedrus Corporation, San Pedro, CA) to report target stimuli. To report non-target stimuli, participants pressed a red button on the same pad. Participants were instructed to respond as quickly and as accurately as possible.

2.3. Experimental design

The experiment was divided into 2 run types, safe run and threat

run, according to whether or not loud aversive background sound was played during the trial. Participants completed one safe run and two threat runs. The safe run contained 12 task blocks (4 of each difficulty) and 12 resting blocks. The two threat runs contained 9 task blocks (3 of each difficulty) and 9 resting blocks each. Order of run type and difficulty levels were counterbalanced across participants.

Experimental blocks were composed of 12 trials (12 arithmetic operations), each displayed on screen for 2 s with an inter-stimulus interval of 1 s, for a total length of 36 s per task block. Each experimental block contained 4 target stimuli (33%) in random positions. A resting block was presented after each experimental block. During the resting blocks, the screen displayed “00 + 00” in the same fashion as during experimental blocks. Participants were not required to report any target during this resting blocks. The resting blocks lasted 18 s (6 trials of 3 s each). The blocks were pseudo-randomized for each participant so that there were no two consecutive blocks of the same difficulty. In threat runs, aversive sounds were administered during six blocks (one block of each level of difficulty plus three resting periods). These blocks were chosen pseudo-randomly so that the aversive sound distribution was uniform during the entire run duration (to avoid expectations from the participants). Fig. 1 summarizes the task design.

2.3.1. The auditory stressors

A set of 34 sounds was used to induce participant stress. This set was created based on the literature (Grillon et al., 2009; Hirano et al., 2006; Kumar et al., 2008; Zald and Pardo, 2002) and was validated by subjective ratings from 34 separate participants (see Mandrick et al., 2016) and by a spectral frequency-temporal modulation analysis (Chi et al., 2005). Sounds were modified to equalize loudness and duration. Each sound was played only once for each participant to avoid habituation. Participants were told that unpredictable aversive sounds could be played during threat runs, even during resting trials. The sounds were played for 7 s via AKG K171 MkII monitor headphones in stereo mode and at a 95-dB sound pressure level, as controlled using a noise meter. Unbeknownst to the participants, half of the threat trials did not contain any sound. For analysis of the threat runs, those blocks containing sounds (6 blocks in total for the two threat runs) were discarded to exclude any potential effect of sound distraction, and to focus on the anxiety related to the expectancy of the unpredictable sounds (Clarke and Johnstone, 2013). Thus, the period during which the sounds were played was not analyzed. After having excluded these 6 blocks, the number of blocks analyzed in the safe and threat conditions were identical (12 blocks). During safe runs, participants were reminded that no sounds could be played. A 5-minute break was introduced between runs to diminish effects of fatigue. Before the experiment, participants performed a short practice session. They were submitted to one block of each level of TNT difficulty (without aversive sounds). The order of administration was fixed, namely 0-back, 1-back, and 2-back, to allow

an easier familiarization with the task. Then, they were also presented with all examples of the aversive sounds that could occur during the aversive runs.

2.3.2. Subjective measures and performance

After the practice session, participants were invited to evaluate the TNT task difficulty levels using a DP15 scale (Delignières et al., 1994). Anxiety was also measured before (i.e. just after signing the consent form) and after the whole experimental protocol using the State-Trait Anxiety Inventory (STAI form Y-A, French translation; Gauthier and Bouchard, 1993). Task performance was measured using the percentage of correct responses (both true positives and true negatives).

2.4. fNIRS and ECG measurements

fNIRS measurements were performed using a fNIRS 16-channel headband model fNIR 100 (BIOPAC). Raw light intensities at 730 and 850 nm were recorded at 2 Hz for all 16 channels. The fNIR 100 has a fixed 2.5 cm source-detector separation. The differential pathlength factor (DPF), which accounts for the increased distance travelled by light due to scattering, was set at 5.76. This value is in the recommended range for an adult head (Van der Van der Zee et al., 1992; Gong et al., 2016). Participants were asked to relax for approximately 2 min, and a ten-second baseline measurement was performed at rest. Data acquisition was performed using COBI Studio, and processed with fnirSoft 4.0. The fNIRS signal was recorded through the entire experiment without interruption. The modified Beer-Lambert law was used to convert light signal to concentration changes of oxygenated and deoxygenated hemoglobin (HbO₂ and HHb, respectively). We removed higher-frequency cardiac or respiratory activity and other noise with other frequencies than the target signal (Roche-Labarbe et al., 2008; Lu et al., 2010; White et al., 2009, Sasai et al. 2011) applying a band-pass FIR filter with an order of 20 (0.02–0.40 Hz) on this raw time series of HbO₂ and HHb signal changes. Signal drifts and global trends were further removed using normalization to zero mean (Z-normalization), applied to the whole raw time series of HbO₂ and HHb signal changes. The data was then visually inspected. Data exclusion criteria were channels without visible heart rate oscillations; Channels with coefficient of variation > 15%; Non-measuring channels (e.g. flat lines), see Pinti et al. (2018). This visual inspection led to the exclusion of one participant (nearly all channels were impacted). For the remaining participants, < 2% of the samples were removed from the analysis. To dissociate effects of TNT difficulty (0-back vs. 1-back vs. 2-back) and threat (safe vs. threat), we extracted the fNIRS response from each trial. More precisely, all analyzes were performed on changes in HbO₂ and HHb concentrations from the average of the initial ten-second rest period baseline. We calculated the average signal change for all trials of each condition, the standard procedure with the BIOPAC system (Ayaz et al., 2012; Durantin et al., 2014; Foy et al., 2016; Causse et al., 2019; Andéol et al., 2017). We then also computed the average slope for each experimental condition. Slope was calculated on the entire block durations (see Mandrick et al., 2013). Average changes and slopes were calculated for both HbO₂ and HHb for the 16 optodes, yielding a total of 64 fNIRS features. Based on previous literature using the same fNIRS device (Causse et al., 2017; Gateau et al., 2018; Kreplin and Fairclough, 2015), we created 3 ROIs: left lateral PFC (optodes 1–6), medial PFC (optodes 7–10), and right lateral PFC (optodes 11–16). ECG activity was measured using a single lead ECG, recorded at 500 Hz. From the ECG signal, we computed the RR intervals and used Kubios HRV software 2.2 (University of Eastern Finland, <http://kubios.uef.fi>) to obtain 22 ECG features. Nine features were obtained using time domain analysis (e.g., mean heart rate, RMSSD) and 13 were obtained using frequency domain analysis (e.g., LF and HF power).

2.5. Machine learning modeling

Before training the models, baseline and resting blocks were removed. The remaining 382 samples were then folded using a Leave-One-Participant-Out scheme, meaning that results presented are always predictions on new unseen participants. We attempted prediction on three variables of interest: 1) mental workload level (classification of the 3 n-back levels), 2) stress level (classification as either threat or safe), and 3) joint mental workload/stress level (classification of the 6 possible mental workload/stress combinations). In order to predict these three variables of interest, three different feature subsets were tested: 1) fNIRS only subset, 2) ECG only subset, and 3) joint fNIRS-ECG subset. We used a MATLAB implementation of a Naive Bayes classifier to perform machine learning modeling (using fitcnb function, MATLAB R2016b). This classifier was used because it has a high capacity for generalization and can perform well with highly dimensional datasets such as the one used in this study (Fan et al., 2011). Naive Bayes classifiers use the assumption that each feature is independent of the others. They are fairly simple and efficient (Rish, 2001). Naive Bayes classifiers were configured to use a kernel distribution, whereby kernel width optimization was achieved using MATLAB's Bayesian Optimization. Training data were 5-folded and 100 optimization iterations were performed for validation. The subjective, behavioral, fNIRS, and ECG data of the current study were part of a previously published dataset (Mandrick et al., 2016).

3. Results

3.1. Subjective and behavioral data

Self-reported subjective ratings of difficulty (DP15) increased with the n-back level: the 0-back task was rated as “Very easy” (average DP15 = 4.8 ± 1.0), the 1-back task as “Somewhat difficult” (average DP15 = 8.2 ± 1.1), and the 2 back task as “Very difficult” (average DP15 = 12.0 ± 0.8). A one-way repeated measures ANOVA revealed that this increase was significant, $F(2, 32) = 110.9, p < .001, \eta_p^2 = 0.87$, with each of the three levels rated differently from each other ($p < .001$ for all three comparisons). Analysis of the state anxiety ratings (STAI Y-A) revealed that participants were more anxious after the experimental protocol (average STAI Y-A = 32.9 ± 3.4; $F(1, 16) = 10.7, p = .005, \eta_p^2 = 0.40$) than before (average STAI Y-A = 27.8 ± 2.2). Table 1 shows the average accuracy for all six conditions during the n-back task. The main effect of mental workload was significant, performance was lower as the n-level increased, $F(2, 32) = 27.8, p < .001, \eta_p^2 = 0.64$. Post hoc comparison analysis (Tukey-Kramer method) confirmed significant differences between all three levels ($p < .01$ for all three comparisons). Repeated measures ANOVA revealed neither main effect of stress nor mental workload x stress interaction on accuracy.

3.2. Classification performance

Table 2 shows average classification performance per participant for each combination of variable and feature subset. These feature subsets encompassed all combinations: 1) fNIRS features, 2) ECG features, and 3) combined fNIRS and ECG features. Chance rate was about 33% (1 out of 3) for n-back classification, 50% (1 out of 2) for threat

Table 1
Performance (accuracy) by mental workload (n-back level) and stress levels (safe vs. threat).

	0-back	1-back	2-back
Safe	94.5 ± 7.1	83.7 ± 16.6	74.7 ± 18.4
Threat	94.4 ± 6.8	84.5 ± 15.7	72.1 ± 18.1

Table 2
Classification accuracy of classifiers.

Feature subset	Workload (0 vs. 1 vs. 2-back)	Stress (safe vs. threat)	Workload + stress (all 6 conditions)
fNIRS	0.43 ± 0.07*	0.63 ± 0.06*	0.22 ± 0.06*
ECG	0.42 ± 0.07*	0.53 ± 0.07	0.24 ± 0.04*
fNIRS + ECG	0.47 ± 0.07*	0.62 ± 0.07*	0.24 ± 0.04*

* Significantly better than chance level ($p < .001$).

classification and 16% (1 out of 6) for joint n-back/stress classification. Results of binomial test showed that all but one classifier achieved accuracy above chance level ($p < .001$). Workload classification performance was between 43% and 47% ($p < .001$ for all feature subsets), and for stress classification, two combinations of parameters were significantly better than chance (fNIRS, $p < .001$; fNIRS + ECG, $p < .001$). By itself, ECG was not sufficient to classify stress. Regarding condition classification, accuracy was between 22% and 24% ($p < .001$ for all feature subsets).

3.3. Deeper analysis of classification performance with fNIRS

Besides generating models, we examined more carefully which part of the fNIRS signal contributed the most to classification. To this aim, we trained extra classifiers using only specific feature subsets. In contrast to the previous analysis that integrated the 16 prefrontal optodes indistinctly, fNIRS features were separated into three areas of interest (AOI). The left and right AOIs grouped the 6 lateral optodes (1 to 6 and 11 to 16, respectively), while the medial AOI grouped optodes 7 to 10. We also separated oxygenated and deoxygenated features, thus leading to six specific feature subsets (i.e., left-HbO2, left-HHb, medial-HbO2, medial-HHb, right-HbO2, right-HHb). These six classifiers were once again trained for the three variables of interest (i.e., 3 mental workload levels, 2 stress levels and all 6 conditions). No ECG features were used to train these classifiers. As illustrated in Fig. 2, classification accuracy using these feature subsets showed that all variables were best predicted using activity of the lateral prefrontal cortices rather than the medial prefrontal cortex. Regarding oxygenated and deoxygenated features, stress level was best predicted with HHb while classification of mental workload level (or of the six conditions separately) was always more accurate with the HbO2 signal (see Fig. 2).

3.4. Analysis of classification performance obtained with fNIRS, demographic information, and subjective ratings

Finally, we used a generalized linear regression model following a binomial distribution to analyze which factors contributed the most to classification performance. While some of this information is conveyed in the previous figures, a generalized linear regression model allows statistically rigorous verification of the previously obtained results and the inclusion of several other independent variables. We used classification success of each 2292 trials (382 trials × 6, since six classifiers were compared) as dependent variable (0 = wrong classification, 1 = correct classification), while independent variables included in the model were AOIs (i.e., left, medial, right), fNIRS signal type (HbO2 or HHb), mental workload (0-back, 1-back, 2-back), and stress levels (safe, threat). We also included demographic information (gender and age), anxiety ratings (from the STAI Y-A questionnaire, before and after the experiment), and difficulty ratings (from the DP-15 questionnaire, for 0-back, 1-back, and 2-back). Multicollinearity was assessed using the Variance Inflation Factor (VIF; Belsley et al., 1980). The highest VIF obtained was 3.9, which is lower than the threshold commonly considered for multicollinearity issues (Kutner, Nachsheim & Neter, 2004; Sheather, 2009). Thus, all variables were kept in the regression analysis.



Fig. 2. Classification performance of the 3 variables (mental workload, stress, mental workload+stress) for the 3 prefrontal AOIs (left lateral, medial, and right lateral) and the 2 fNIRS signal (HbO2 and HHb). Dotted line shows chance level.

3.4.1. Mental workload classification

The analysis confirmed that the left and right PFC AOIs led to better classification compared to the medial PFC AOI (left AOI: $\beta = 0.28$, $p = .008$; right AOI: $\beta = 0.31$, $p = .003$). HbO2 features also led to better classification compared to those of HHb ($\beta = 0.23$, $p = .005$). Mental workload, stress level, gender, and age did not affect classification performance. As for the subjective ratings, scores of the DP15 during 0-back was shown to be associated with better classification ($\beta = 0.59$, $p = .03$).

3.4.2. Stress classification

Classification was better using the left AOI ($\beta = 0.22$, $p = .03$) and even better using the right AOI ($\beta = 0.27$, $p = .007$; compared to the medial AOI in both cases). In this analysis, fNIRS signal type, mental workload and stress levels, gender and age did not affect classification accuracy for stress. Scores of the STAI Y-A, recorded before the experiment, were shown to be associated with better classification accuracy ($\beta = 0.63$, $p < .001$). Again, scores of the DP15 (for 0-back and 2-back) were also associated with better accuracy (respectively: $\beta = 0.95$, $p < .001$; $\beta = 0.43$, $p = .005$).

3.4.3. Mental workload + stress classification

The right PFC AOI yielded better classification accuracy compared to the medial AOI ($\beta = 0.40$, $p = .002$). This time, the left AOI was no better than the medial AOI. HbO2 features were once again associated

with better classification ($\beta = 0.22, p = .03$). Ratings of the STAI Y-A before the experiment and ratings of the DP15 (0 back) were also associated with better classification (respectively: $\beta = 0.66, p = .005$; $\beta = 1.10, p = .001$).

4. Discussion

The current work aimed to disentangle the separate contributions of mental workload and stress to task performance and physiological activity. The monotonic decline in accuracy in line with each increase in load on the n-back task confirmed that we had elicited marked variations of mental workload. The threat condition had minimal effect on task performance. Noise, as a stressor, can impair task performance (e.g. Szalma and Hancock, 2011), but this is not always the case. Personality traits (Belojević et al., 1992), noise properties (Smith, 1989) and task modalities (Driskell et al., 1992) can mitigate the effects on the performance. In some cases, noise has been reported to increase performance (Alimohammadi et al., 2013; Saeki et al., 2004). In the current work, despite using a stressor which is known to increase subjective anxiety, no apparent behavioral effects were measurable. This shows the benefits of physiologically recognizing stress independently from the observable behavior.

Our models achieved an average three-category (n-back level) classification accuracy of 43%, 42%, and 47% for the fNIRS-alone, ECG-alone, and fNIRS + ECG combined approaches, respectively. This is consistent with the existing literature, for example, Liu et al. (2017a) classified three mental workload levels (during a classical n-back task) using fNIRS-alone, EEG-alone, and combined EEG and fNIRS, achieving 42%, 43%, and 49% accuracy, respectively. Our models achieved levels of classification accuracy on mental workload comparable to that of Liu et al. (2017a) despite having to deal with concurrent variations in participants' emotional state (stress level).

Our results showed that mental workload classification was possible using only ECG. To our knowledge, no other study has used a 3-level n-back classifier using only ECG features, although significant differences have previously been found between baseline, 0-back, 1-back and 2-back using only a measure of heart rate (Mehler et al., 2009). Furthermore, Cinaz et al. (2013) computed common ECG features for 3 levels of a dual n-back task; while they did not train classifiers, they showed that some ECG features (e.g., RMSSD) were sensitive enough to present significant differences between levels. Such differences in RMSSD were also observed in the current study (see Mandrick et al., 2016), suggesting that a simple, and more affordable ECG measurement system could perform just as well as a more complex fNIRS setup with regard to mental workload classification.

In terms of stress classification, in the current study, this was possible using fNIRS and a fNIRS + ECG combination, but not using ECG alone. At first glance, this result might seem surprising given that the effects of stress on cardiovascular activity are well documented (e.g., Schubert et al., 2009); however, we argue that our results do not oppose this common knowledge. While investigating the physiological response pattern of this study (see Mandrick et al., 2016), we find that stress does indeed affect cardiovascular metrics but in order to make correct classifications, it is necessary to have sufficient sensitivity and specificity performances. As such, it is possible to surmise that cardiovascular changes induced by the threat condition were not specific enough to that stressor. Stress is often associated with the fight-or-flight response, in which the body increases its sympathetic activity in order to prepare the body to respond to a threat. However, an increase of mental workload is also associated with an increase in sympathetic activity and therefore, our classifiers might have been trained to discount the ECG response (which might have been captured by the n-back classifier), and trained instead to classify according to the fNIRS pattern associated with anxiety. Also, according to Liu et al. (2017b), it is possible that integrating ECG measures with fNIRS does not significantly improve classification (the workload in their study) if the

physiological measurements do not provide additional information to the brain signal measurements (for the stress in our case).

Finally, we also postulate that the nature of the stressor, which was only of moderate emotional intensity (i.e., far from a life-threatening stressor that pilots might encounter) and induced relatively little variation in ECG activity (i.e., around +1.5 bpm in the threat vs. safe condition, see Mandrick et al., 2016), could have limited the classifying performance. A more intense stress, jeopardizing task performance and generating a more marked physiological response (in particular cardiovascular), might have improved classification accuracy (safe vs. threat conditions and mental workload factor vs. stress factor). We believe that the lack of effects of the auditory stressors on task performance was mainly due to their too low emotional intensity. However, it can also be partly due to coping mechanisms. The processing efficiency theory (Eysenck and Calvo, 1992) proposes that adverse effects of anxiety are not always visible on performance outcome. Stressful situations can generate an increased motivation to minimize the aversive anxiety state. Motivation promotes enhanced effort, thus, potential performance impairments caused by stress can be compensated if auxiliary processing resources are available. As said previously, this is one of the advantages of the psychophysiological measurements. They are able to detect a moderate increase of the stress level, without "visible" effects on the behavior, but that could be intolerable on the long term.

Beside independent mental workload and stress classifications, we were able to disentangle variations in physiological activity resulting from combined mental workload and stress level changes. Although the six-category classification accuracy (the 6 combinations of mental workload and stress levels) was moderate (24% at best), it was significant and able to outperform chance level (16%). These results support the notion that fNIRS + ECG in combination could be used to disentangle the two concepts. Furthermore, given the leave-one-participant-out scheme used in the training of classifiers, we suggest that it is possible to generalize these predictions to new individuals without having to calibrate models on them first. We cannot exclude that increasing n-back levels may have also increased the level of psychological stress. For example, the cognitive appraisal processes refer to situational evaluations in terms of their significance for one's well-being (Lazarus, 1991). Acute stress can increase when situational demands are perceived as exceeding coping resources or abilities (Penley and Tomaka, 2002). Some participants may have felt overwhelmed by task difficulty. In addition, coping with stress can consume cognitive resources (Eysenck and Calvo, 1992). In this sense, high n-back level can elicit a mixture of mental workload and psychological stress. Purely generating mental workload remains complex, and depends on individual characteristics and personality. Thus, a possible limitation of the study is that the comparison of safe vs. threat conditions in the high level of mental workload (in particular 2-back) can have finally resulted in a mixture of workload and acute stress (safe condition) vs. a mixture of workload and higher acute stress, with an additional stress due to the aversive sounds (threat condition). A future study should assess the level of subjective anxiety or the level of subjective acute stress after each n-back level, to control for this possible effect, especially in the higher level of difficulty. Even if it is difficult to purely and separately manipulate mental workload and stress, it does not impact the relevance of our results, in particular the ability of the model to disentangle combinations of mental workload and stress level changes.

A more detailed analysis separating three prefrontal AOIs and distinguishing between the two fNIRS signals (HbO2 and Hb) revealed that mental workload levels were best classified with the activity in the lateral prefrontal regions rather than medial regions. Classifiers using either the left or right prefrontal cortices were able to achieve a similar level of accuracy using the full set of optodes. This result reinforces previous studies that propose a prominent role for these regions in working memory (Curtis and D'Esposito, 2003; De Pisapia et al., 2007; Owen et al., 1996) and mental arithmetic (Gruber et al., 2001). Regarding the fNIRS signal, except for the two-category classification of

stress that was slightly more accurate using HHB, classification accuracy was always better using HbO₂, for both the three-category mental workload classifier and the six-category (all possible mental workload/stress combinations) classifier.

The additional analysis of classification performance using a generalized linear regression model showed that objective factors (mental workload and stress) were associated with no better performance, suggesting that our classifiers were equally good in every experimental condition. There were some subjective ratings that were associated with higher classification accuracy, for example, increased ratings of perceived difficulty during the 0-back (the easiest condition) were associated with better classification accuracy for all classifiers. In other words, individuals demonstrating higher perceived difficulty during the 0-back condition were more likely to be correctly classified. We can assume that individuals who felt more difficulty during the 0-back condition were more likely to be overwhelmed and overloaded during the much-harder 2-back condition. As some authors suggest, overload might be characterized by its own physiological signature, in particular, a disengaging from the task associated with a decline of fNIRS activity (Durantin et al., 2014). Thus, our classifiers might have detected high variance between normal load and overload for high 0-back raters. Following this hypothesis, individuals who perceived 0 back as an easy task might have been unfazed by 2-back, resulting in less physiological variation between task difficulty and, ultimately, less opportunity for classifiers to distinguish correctly. Finally, classification of stress level was better if the participant rated a high level of pre-experiment anxiety. We suggest that participants showing a higher level of anxiety were probably more inclined to react strongly to the threat condition, and this reaction might have provoked a more marked physiological reaction that was more easily detected by the classifiers. On the other hand, participants rating as low anxious might have been much less influenced by the threat condition.

5. Conclusion

In this study, we showed that a combination of fNIRS + ECG achieved the best accuracy for predicting variations in mental workload. Nevertheless, ECG alone was able to classify variations of mental workload quite efficiently, which confirmed that this simple measure can be sufficient for field studies or assessment in an operational setting where more cumbersome sensors are difficult to employ. The stress level was well predicted with fNIRS alone or a combination of fNIRS and ECG, however, ECG alone was no greater than chance level. Most likely, the level of stress generated in the study was too moderate to elicit a marked cardiovascular activity; as such, participants with a higher subjective level of anxiety were classified more accurately. Finally, and most importantly, we were able to disentangle variations of mental workload from variations of stress with all features (fNIRS-alone, ECG-alone, a combination of fNIRS + ECG). Classification accuracy was moderate, but focusing on the lateral regions of the prefrontal cortex can improve classification performance. To the best of our knowledge, very few studies have successfully separated the two concepts on the basis of their physiological signatures. Given the lack of effect of our stressor on task performance, future studies should attempt to replicate the current design using a stressor with higher intensity, for example by adding an induction inspired by the Trier Social Stress Test (Kirschbaum et al., 1993). In addition, in future works, more psychological constructs could be added to further increase diagnosticity of classification models. Such constructs could include mental fatigue, but also physical activity, which is often disregarded in laboratory experiments despite being present in many work settings.

Acknowledgments

This work was supported by the French Research National Agency and the French Defence Procurement Agency via the Accompagnement

Spécifique des travaux de Recherches et d'Innovation Défense (ASTRID).

References

- Aghajani, H., Garbey, M., Omurtag, A., 2017. Measuring mental workload with EEG + fNIRS. *Front. Hum. Neurosci.* 11, 359. <https://doi.org/10.3389/fnhum.2017.00359>.
- Alimohammadi, I., Soltani, R., Sandrock, S., Azkosh, M., Gohari, M.R., 2013. The effects of road traffic noise on mental performance. *Iranian journal of environmental health science & engineering* 10 (1), 18.
- Al-Shargie, F., Kiguchi, M., Badruddin, N., Dass, S.C., Hani, A.F.M., Tang, T.B., 2016. Mental stress assessment using simultaneous measurement of EEG and fNIRS. *Biomedical Optics Express* 7 (10), 3882. <https://doi.org/10.1364/BOE.7.003882>.
- Andéol, G., Suiéd, C., Scannella, S., Dehais, F., 2017. The spatial release of cognitive load in cocktail party is determined by the relative levels of the talkers. *J. Assoc. Res. Otolaryngol.* 18 (3), 457–464.
- Anderson, C.R., 1976. Coping behaviors as intervening mechanisms in the inverted-U stress-performance relationship. *J. Appl. Psychol.* 61 (1), 30–34. <https://doi.org/10.1037/0021-9010.61.1.30>.
- Arnrich, B., Setz, C., La Marca, R., Tröster, G., Ehlert, U., 2010. What does your chair know about your stress level? *IEEE Trans. Inf. Technol. Biomed.* 14 (2), 207–214. <https://doi.org/10.1109/TITB.2009.2035498>.
- Arnsten, A.F.T., 2009. Stress signalling pathways that impair prefrontal cortex structure and function. *Nat. Rev. Neurosci.* 10 (6), 410–422. <https://doi.org/10.1038/nrn2648>.
- Ayaz, H., Shewokis, P.A., Bunce, S., Izzetoglu, K., Willems, B., Onaral, B., 2012. Optical brain monitoring for operator training and mental workload assessment. *Neuroimage* 59 (1), 36–47.
- Beloević, G., Öhrström, E., Rylander, R., 1992. Effects of noise on mental performance with regard to subjective noise sensitivity. *Int. Arch. Occup. Environ. Health* 64 (4), 293–301.
- Belsley, D.A., Kuh, E., Welsch, R.E., 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* (Wiley Series in Probability and Statistics).
- Bernardi, L., Wdowczyk-Szulc, J., Valenti, C., Castoldi, S., Passino, C., Spadacini, G., Sleight, P., 2000. Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. *J. Am. Coll. Cardiol.* 35 (6), 1462–1469. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10807448>.
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F., 2012. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.* 1–18. <https://doi.org/10.1016/j.neubiorev.2012.10.003>.
- Borghini, G., Aricò, P., Di Flumeri, G., Cartocci, G., Colosimo, A., Bonelli, S., ... Benhacene, R., 2017. EEG-based cognitive control behaviour assessment: an ecological study with professional air traffic controllers. *Sci. Rep.* 7 (1), 547. <https://doi.org/10.1038/s41598-017-00633-7>.
- Brindle, R.C., Whittaker, A.C., Bibbey, A., Carroll, D., Ginty, A.T., 2017. Exploring the possible mechanisms of blunted cardiac reactivity to acute psychological stress. *Int. J. Psychophysiol.* 113, 1–7. <https://doi.org/10.1016/j.ijpsycho.2016.12.011>.
- Bryce, C.A., Floresco, S.B., 2016. Perturbations in effort-related decision-making driven by acute stress and corticotropin-releasing factor. *Neuropsychopharmacology* 41 (8), 2147. <https://doi.org/10.1038/npp.2016.15>.
- Byrne, E.A., Parasuraman, R., 1996. Psychophysiology and adaptive automation. In: *Biological Psychology*. vol. 42. pp. 249–268. [https://doi.org/10.1016/0301-0511\(95\)05161-9](https://doi.org/10.1016/0301-0511(95)05161-9).
- Causse, M., Sénard, J.M., Démonet, J.F., Pastor, J., 2010. Monitoring cognitive and emotional processes through pupil and cardiac response during dynamic versus logical task. *Applied Psychophysiology Biofeedback* 35 (2), 115–123. <https://doi.org/10.1007/s10484-009-9115-0>.
- Causse, M., Peysakhovich, V., Fabre, E.F., 2016. High working memory load impairs language processing during a simulated piloting task: an ERP and pupillometry study. *Front. Hum. Neurosci.* 10 (May), 1–14. <https://doi.org/10.3389/fnhum.2016.00240>.
- Causse, M., Chua, Z., Peysakhovich, V., Del Campo, N., Matton, N., 2017. Mental workload and neural efficiency quantified in the prefrontal cortex using fNIRS. *Sci. Rep.* 7 (1). <https://doi.org/10.1038/s41598-017-05378-x>.
- Causse, M., Chua, Z.K., Rémy, F., 2019. Influences of age, mental workload, and flight experience on cognitive performance and prefrontal activity in private pilots: a fNIRS study. *Sci. Rep.* 9 (1), 7688.
- Chen, F., 2006. Designing human interface in speech technology. *Designing Human Interface in Speech Technology*. <https://doi.org/10.1007/b104643>.
- Chi, T., Ru, P., Shamma, S.A., 2005. Multiresolution spectrotemporal analysis of complex sounds. *J. Acoustical Soc. Am.* 118 (2), 887–906. <https://doi.org/10.1121/1.1945807>.
- Cinaz, B., Arnrich, B., La Marca, R., Tröster, G., 2013. Monitoring of mental workload levels during an everyday life office-work scenario. *Pers. Ubiquit. Comput.* 17 (2), 229–239. <https://doi.org/10.1007/s00779-011-0466-1>.
- Clarke, R., Johnstone, T., 2013. Prefrontal inhibition of threat processing reduces working memory interference. *Front. Hum. Neurosci.* 7. <https://doi.org/10.3389/fnhum.2013.00228>.
- Critchley, H.D., Eccles, J., Garfinkel, S.N., 2013. Interaction between cognition, emotion, and the autonomic nervous system. *Handb. Clin. Neurol.* 117, 59–77. <https://doi.org/10.1016/B978-0-444-53491-0.00006-7>.
- Curtis, C.E., D'Esposito, M., 2003. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* [https://doi.org/10.1016/S1364-6613\(03](https://doi.org/10.1016/S1364-6613(03)

- 00197-9.
- da Silva, F.P., 2014. Mental workload, task demand and driving performance: what relation? *Procedia Soc. Behav. Sci.* 162 (Panam), 310–319. <https://doi.org/10.1016/j.sbspro.2014.12.212>.
- De Pisapia, N., Slomski, J.A., Braver, T.S., 2007. Functional specializations in lateral prefrontal cortex associated with the integration and segregation of information in working memory. *Cereb. Cortex* 17 (5), 993–1006. <https://doi.org/10.1093/cercor/bhl010>.
- de Rooij, S.R., Roseboom, T.J., 2010. Further evidence for an association between self-reported health and cardiovascular as well as cortisol reactions to acute psychological stress. *Psychophysiology* 47 (6), 1172–1175. <https://doi.org/10.1111/j.1469-8986.2010.01023.x>.
- Delignières, D., Famose, J., Genty, J., 1994. Validation of a scale for the assessment of perceived task difficulty. *Staps* 34, 77–88.
- Driskell, J.E., Mullen, B., Johnson, C., Hughes, S., Batchelor, C.L., 1992. Development of Quantitative Specifications for Simulating the Stress Environment. (Armstrong Lab Wright-Patterson Afb OH).
- Durantini, G., Gagnon, J.-F., Tremblay, S., Dehais, F., 2014. Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behav. Brain Res.* 259, 16–23. <https://doi.org/10.1016/j.bbr.2013.10.042>.
- Durkee, K., Geyer, A., Pappada, S., Ortiz, A., Galster, S., 2013. Real-time workload assessment as a foundation for human performance augmentation. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8027 LNAI, pp. 279–288. https://doi.org/10.1007/978-3-642-39454-6_29.
- Elzinga, B.M., Roelofs, K., 2005. Cortisol-induced impairments of working memory require acute sympathetic activation. *Behav. Neurosci.* 119 (1), 98–103. <https://doi.org/10.1037/0735-7044.119.1.98>.
- Eysenck, M.W., Calvo, M.G., 1992. Anxiety and performance: the processing efficiency theory. *Cognit. Emot.* 6 (March 2015), 409–434. <https://doi.org/10.1080/02699939208409696>.
- Fairclough, S.H., Venables, L., Tattersall, A., 2005. The influence of task demand and learning on the psychophysiological response. *Int. J. Psychophysiol.* 56 (2), 171–184. <https://doi.org/10.1016/j.ijpsycho.2004.11.003>.
- Fan, J., Fan, Y., Wu, Y., 2011. High-dimensional classification. *High-Dimensional Data Analysis* 3–37.
- Foy, H.J., Runham, P., Chapman, P., 2016. Prefrontal cortex activation and young driver behaviour: a fNIRS study. *PLoS One* 11 (5), e0156512.
- Gaillard, A.W.K., Wientjes, C.J.E., 1994. Mental load and work stress as two types of energy mobilization. *Work Stress.* 8 (2), 141–152. <https://doi.org/10.1080/02678379408259986>.
- Gateau, T., Ayaz, H., Dehais, F., 2018. In silico versus over the clouds: on-the-fly mental state estimation of aircraft pilots, using a functional near infrared spectroscopy based passive-BCI. *Front. Hum. Neurosci.* 12, 187.
- Gauthier, J., Bouchard, S., 1993. Adaptation canadienne-française de la forme révisée du State-Trait Anxiety Inventory de Spielberger./A French-Canadian adaptation of the revised version of Spielberger's state-trait anxiety inventory. *Canadian Journal of Behavioural Science/Revue Canadienne Des Sciences Du Comportement* 25 (4), 559–578. <https://doi.org/10.1037/h0078881>.
- Gong, H., Luo, Q., Zeng, S., Chance, B., Nioka, S., Kuroda, Y., 2016. Monitoring of Brain Activity With Near-infrared Spectroscopy.
- Grillon, C., Pine, D.S., Lissek, S., Rabin, S., Bonne, O., Vythilingam, M., 2009. Increased anxiety during anticipation of unpredictable aversive stimuli in posttraumatic stress disorder but not in generalized anxiety disorder. *Biol. Psychiatry* 66 (1), 47–53. <https://doi.org/10.1016/j.biopsych.2008.12.028>.
- Gruber, O., Indefrey, P., Rabin, S., Vythilingam, M., 2001. Dissociating neural correlates of cognitive components in mental calculation. *Cereb. Cortex* 11 (4), 350–359. <https://doi.org/10.1093/cercor/11.4.350>.
- Hart, S.G., 2006. NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50 (9), 904–908. <https://doi.org/10.1037/e577632012-009>.
- Hembree, R., 1988. Correlates, causes, effects, and treatment of test anxiety. *Rev. Educ. Res.* 58 (1), 47. <https://doi.org/10.2307/1170348>.
- Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., Schultz, T., 2014. Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS. *Front. Hum. Neurosci.* 7 (January), 1–9. <https://doi.org/10.3389/fnhum.2013.00935>.
- Hidalgo-Muñoz, A.R., Mouratille, D., Matton, N., Causse, M., Rouillard, Y., El-Yagoubi, R., 2018. Cardiovascular correlates of emotional state, cognitive workload and time-on-task effect during a realistic flight simulation. *Int. J. Psychophysiol.* 128, 62–69. <https://doi.org/10.1016/j.ijpsycho.2018.04.002>.
- Hincks, S.W., Afergan, D., Jacob, R.J., 2016. Using Fnrirs for Real-time Cognitive Workload Assessment. Springer, pp. 198–208. https://doi.org/10.1007/978-3-319-39955-3_19.
- Hirano, Y., Fujita, M., Watanabe, K., Niwa, M., Takahashi, T., Kanematsu, M., ... Onozuka, M., 2006. Effect of unpleasant loud noise on hippocampal activities during picture encoding: an fMRI study. *Brain Cogn.* 61 (3), 280–285. <https://doi.org/10.1016/j.bandc.2006.02.003>.
- Hirshfield, L.M., Chauncey, K., Golotta, R., Girouard, A., Solovey, E.T., Jacob, R.J., ... Fantini, S., 2009. Combining Electroencephalograph and Functional Near Infrared Spectroscopy to Explore Users' Mental Workload. Springer, pp. 239–247. https://doi.org/10.1007/978-3-642-02812-0_28.
- Hjortskov, N., Rissén, D., Bångsted, A.K., 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>.
- Hou, M., Fidopiastis, C.M., 2014. When Designing Intelligent Adaptive Systems. pp. 26–34.
- Kirschbaum, C., Pirke, K.M., Hellhammer, D.H., 1993. The 'Trier Social Stress Test'—a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychologia* 28 (1–2), 76–81.
- Kreplin, U., Fairclough, S.H., 2015. Effects of self-directed and other-directed introspection and emotional valence on activation of the rostral prefrontal cortex during aesthetic experience. *Neuropsychologia* 71, 38–45.
- Kumar, S., Forster, H.M., Bailey, P., Griffiths, T.D., 2008. Mapping unpleasantness of sounds to their auditory representation. *J. Acoustical Soc. Am.* 124 (6), 3810–3817. <https://doi.org/10.1121/1.3006380>.
- Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 1992. *c. Applied Linear Statistical Models.* Mc Graw-Hill Companies, Inc., New York, NY, pp. 214–508.
- Laguë-Beauvais, M., Brunet, J., Gagnon, L., Lesage, F., Bherer, L., 2013. A fNIRS investigation of switching and inhibition during the modified Stroop task in younger and older adults. *NeuroImage* 64 (1), 485–495. <https://doi.org/10.1016/j.neuroimage.2012.09.042>.
- Lazarus, R.S., 1991. *Emotion and Adaptation.* Oxford University Press, New York.
- LeBlanc, V.R., 2009. The effects of acute stress on performance: implications for health professions education. *Acad. Med. J. Assoc. Am. Med. Colleges* 84 (10 Suppl), S25–S33. <https://doi.org/10.1097/ACM.0b013e3181b37b8f>.
- Liu, Y., Ayaz, H., Shewokis, P.A., 2017a. Mental workload classification with concurrent electroencephalography and functional near-infrared spectroscopy. *Brain-Computer Interfaces* 4 (3), 175–185. <https://doi.org/10.1080/2326263X.2017.1304020>.
- Liu, Y., Ayaz, H., Shewokis, P.A., 2017b. Multisubject "learning" for mental workload classification using concurrent EEG, fNIRS, and physiological measures. *Front. Hum. Neurosci.* 11, 389.
- Lu, C.M., Zhang, Y.J., Biswal, B.B., Zang, Y.F., Peng, D.L., Zhu, C.Z., 2010. Use of fNIRS to assess resting state functional connectivity. *J. Neurosci. Methods* 186 (2), 242–249.
- Macaš, M., Vavrecka, M., Gerla, V., Lhotská, L., 2009. Classification of the Emotional States Based on the EEG Signal Processing. *IEEE*, pp. 1–4.
- Mandrick, K., Derosiere, G., Dray, G., Coulon, D., Micallef, J.P., Perrey, S., 2013. Utilizing slope method as an alternative data analysis for functional near-infrared spectroscopy-derived cerebral hemodynamic responses. *Int. J. Ind. Ergon.* 43 (4), 335–341.
- Mandrick, K., Peysakhovich, V., Rémy, F., Lepren, E., Causse, M., 2016. Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biol. Psychol.* 121, 62–73. <https://doi.org/10.1016/j.biopsycho.2016.10.002>.
- Matthews, G., Campbell, S.E., 2010. Dynamic relationships between stress states and working memory. *Cognit. Emot.* 24 (2), 357–373. <https://doi.org/10.1080/02699930903378719>.
- Matthews, G., Reinerman-jones, L., Wohleber, R., Lin, J., Mercado, J., Iv, J.A., 2015. Workload Is Multidimensional, Not Unitary: What Now? 9183. pp. 44–55. <https://doi.org/10.1007/978-3-319-20816-9>.
- Mehler, B., Reimer, B., Coughlin, J., Dusek, J., 2009. Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transport. Res. Record J. Transport. Res. Board* 2138, 6–12. <https://doi.org/10.3141/2138-02>.
- Mühl, C., Jeunet, C., Lotte, F., 2014. EEG-based workload estimation across affective contexts. *Front. Neurosci.* 8 (8 JUN), 1–15. <https://doi.org/10.3389/fnins.2014.00114>.
- Murphy, L.R., 1996. Stress management in work settings: a critical review of the health effects. *Am. J. Health Promot.* <https://doi.org/10.4278/0890-1171-11.2.112>.
- Owen, A.M., Evans, A.C., Petrides, M., 1996. Evidence for a two-stage model of working memory processing within the lateral frontal cortex: a positron emission tomography study. *Cereb. Cortex* 6, 31–38.
- Parasuraman, R., Sheridan, T.B., Wickens, C.D., 2000. A model for types and levels of human interaction. *IEEE Transactions on Systems Man and Cybernetics Part A Systems and Humans* 30 (3), 286–297.
- Parasuraman, R., Sheridan, T.B., Wickens, C.D., 2008. Situation awareness, mental workload, and trust in automation: viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making* 2 (2), 140–160. <https://doi.org/10.1518/155534308X284417>.
- Pattyn, N., Neyt, X., Henderickx, D., Soetens, E., 2008. Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue? *Physiol. Behav.* 93 (1–2), 369–378. <https://doi.org/10.1016/j.physbeh.2007.09.016>.
- Penley, J.A., Tomaka, J., 2002. Associations among the big five, emotional responses, and coping with acute stress. *Personal. Individ. Differ.* 32 (7), 1215–1228.
- Peysakhovich, V., Vachon, F., Dehais, F., 2016. The impact of luminance on tonic and phasic pupillary responses to sustained cognitive load. *Int. J. Psychophysiol.* 112, 40–45. <https://doi.org/10.1016/j.ijpsycho.2016.12.003>.
- Pinti, P., Aichelburg, C., Gilbert, S., Hamilton, A., Hirsch, J., Burgess, P., Tachtsidis, I., 2018. A review on the use of wearable functional near-infrared spectroscopy in naturalistic environments. *Jpn. Psychol. Res.* 60 (4), 341–373.
- Plessow, F., Fischer, R., Kirschbaum, C., Goschke, T., 2011. Inflexibly focused under stress: acute psychosocial stress increases shielding of action goals at the expense of reduced cognitive flexibility with increasing time lag to the stressor. *J. Cogn. Neurosci.* 23 (11), 3218–3227. https://doi.org/10.1162/jocn_a.00024.
- Popma, A., Jansen, L.M.C., Vermeiren, R., Steiner, H., Raine, A., Van Goozen, S.H.M., ... Doreleijers, T.A.H., 2006. Hypothalamic pituitary adrenal axis and autonomic activity during stress in delinquent male adolescents and controls. *Psychoneuroendocrinology* 31 (8), 948–957. <https://doi.org/10.1016/j.psyneuen.2006.05.005>.
- Qin, S., Hermans, E.J., van Marle, H.J.F., Luo, J., Fernández, G., 2009. Acute psychological stress reduces working memory-related activity in the dorsolateral prefrontal cortex. *Biol. Psychiatry* 66 (1), 25–32. <https://doi.org/10.1016/j.biopsycho.2009.03.006>.

- Rendon-Velez, E., van Leeuwen, P.M., Happee, R., Horváth, I., van der Vegte, W.F., de Winter, J.C.F., 2016. The effects of time pressure on driver performance and physiological activity: a driving simulator study. *Transport. Res. F: Traffic Psychol. Behav.* 41, 150–169. <https://doi.org/10.1016/j.trf.2016.06.013>.
- Rish, I., 2001. An empirical study of the naive Bayes classifier. *Empirical Methods in Artificial Intelligence Workshop. IJCAI 22230 (JANUARY 2001)*, 41–46. <https://doi.org/10.1039/b104835j>.
- Roche-Labarbe, N., Zaaimi, B., Berquin, P., Nehlig, A., Grebe, R., Wallois, F., 2008. NIRS-measured oxy-and deoxyhemoglobin changes associated with EEG spike-and-wave discharges in children. *Epilepsia* 49 (11), 1871–1880.
- Saeki, T., Fujii, T., Yamaguchi, S., Harima, S., 2004. Effects of acoustical noise on annoyance, performance and fatigue during mental memory task. *Appl. Acoust.* 65 (9), 913–921.
- Sasai, S., Homae, F., Watanabe, H., Taga, G., 2011. Frequency-specific functional connectivity in the brain during resting state revealed by NIRS. *Neuroimage* 56 (1), 252–257. <https://doi.org/10.1518/001872095778995616>.
- Sawin, D.A., Scerbo, M.W., 1995. Effects of instruction type and boredom proneness in vigilance: implications for boredom and workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37 (4), 752–765. <https://doi.org/10.1518/001872095778995616>.
- Schoofs, D., Preuß, D., Wolf, O.T., 2008. Psychosocial stress induces working memory impairments in an n-back paradigm. *Psychoneuroendocrinology* 33, 643–653. <https://doi.org/10.1016/j.psyneuen.2008.02.004>.
- Schubert, C., Lambertz, M., Nelesen, R. a, Bardwell, W., Choi, J.-B., Dimsdale, J.E., 2009. Effects of stress on heart rate complexity—a comparison between short-term and chronic stress. *Biol. Psychol.* 80 (3), 325–332. <https://doi.org/10.1016/j.biopsycho.2008.11.005>.
- Sheather, S., 2009. *A Modern Approach to Regression With R*. Springer Science & Business Media.
- Smith, A., 1989. A review of the effects of noise on human performance. *Scand. J. Psychol.* 30 (3), 185–206.
- Stawski, R.S., Sliwinski, M.J., Smyth, J.M., 2006. Stress-related cognitive interference predicts cognitive function in old age. *Psychol. Aging* 21 (3), 535. <https://doi.org/10.1037/0882-7974.21.3.535>.
- Szalma, J.L., Hancock, P.A., 2011. Noise effects on human performance: a meta-analytic synthesis. *Psychol. Bull.* 137 (4), 682.
- Van der Zee, P., Cope, M., Arridge, S.R., Essenpreis, M., Potter, L.A., Edwards, A.D., ... Delpy, D.T., 1992. Experimentally measured optical pathlengths for the adult head, calf and forearm and the head of the newborn infant as a function of inter optode spacing. In: *Oxygen Transport to Tissue XIII*. Springer, Boston, MA, pp. 143–153.
- Victor, T., 2014. Analysis of Naturalistic Driving Study Data. *Driver Inattention, and Crash Risk, Safer Glances*. <https://doi.org/10.17226/22297>.
- Wang, Z., Hope, R.M., Wang, Z., Ji, Q., Gray, W.D., 2012. Cross-subject workload classification with a hierarchical Bayes model. *NeuroImage* 59 (1), 64–69. <https://doi.org/10.1016/j.neuroimage.2011.07.094>.
- Warm, J.S., Parasuraman, R., Matthews, G., 2008. Vigilance requires hard mental work and is stressful. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50 (3), 433–441. <https://doi.org/10.1518/001872008X312152>.
- White, B.R., Snyder, A.Z., Cohen, A.L., Petersen, S.E., Raichle, M.E., Schlaggar, B.L., Culver, J.P., 2009. Resting-state functional connectivity in the human brain revealed with diffuse optical tomography. *Neuroimage* 47 (1), 148–156.
- Wickens, C.D., 2002. Multiple resources and performance prediction. *Theor. Issues Ergon. Sci.* 3 (2), 159–177. <https://doi.org/10.1080/14639220210123806>.
- Williges, R.C., Wierwille, W.W., 1979. Behavioral measures of aircrew mental workload. *Human Factors: The Journal of Human Factors and Ergonomics Society* 21 (5), 549–574. <https://doi.org/10.1177/001872087902100503>.
- Wilson, G. F., & Eggemeier, T. F. (1991). Psychophysiological assessment of workload in multi-task environments. In *Multiple-task performance* (pp. 329–360).
- Wolf, O.T., 2009. Stress and memory in humans: twelve years of progress? *Brain Res.* 1293, 142–154. <https://doi.org/10.1016/j.brainres.2009.04.013>.
- Zald, D.H., Pardo, J.V., 2002. The neural correlates of aversive auditory stimulation. *NeuroImage* 16 (3 1), 746–753. <https://doi.org/10.1006/nimg.2002.1115>.