www.nature.com/scientificdata

scientific data

OPEN

DATA DESCRIPTOR

# Open multi-session and multi-task EEG cognitive Dataset for passive brain-computer Interface Applications

Marcel F. Hinss[1]✉, Emilie S. Jahanpour[1], Bertille Somon[2], Lou Pluchon[1], Frédéric Dehais[1,3] & Raphaëlle N. Roy[1,3]

**Brain-Computer Interfaces and especially passive Brain-Computer interfaces (pBCI), with their ability to estimate and monitor user mental states, are receiving increasing attention from both the fundamental research and the applied research and development communities. Testing new pipelines and benchmarking classifiers and feature extraction algorithms is central to further research within this domain. Unfortunately, data sharing in pBCI research is still scarce. The COG-BCI database encompasses the recordings of 29 participants over 3 separate sessions with 4 different tasks (MATB, N-Back, PVT, Flanker) designed to elicit different mental states, for a total of over 100 hours of open EEG data. This dataset was validated on a subjective, behavioral and physiological level, to ensure its usefulness to the pBCI community. Furthermore, a proof of concept is given with an example of mental workload estimation pipeline and results, to ensure that the data can be used for the design and evaluation of pBCI pipelines. This body of work presents a large effort to promote the use of pBCIs in an open science framework.**

## Background & Summary

Since the industrial revolution, work environments have seen a shift away from physical work towards overseeing and controlling machines[1]. Even though accidents and incidents have become rarer, this shift in operator activity has led to an increase in cognitive demand through supervision, while simultaneously making errors more costly[2,3]. One emerging solution to optimize human-machine teaming and prevent human error is to implement passive Brain Compute Interfaces (pBCI[4]). PBCIs allow the estimation and monitoring of critical mental states during complex real-life tasks based on neuroimaging data collected from the user /operator. These mental state inference systems can be then used to dynamically drive human-machine interactions to overcome cognitive bottleneck[5,6]. Mental states that are commonly targeted include (but are not limited to) mental workload, vigilance and task-switching ability as they are known to precede a degradation of human performance[7]. In practice, researchers will use laboratory tasks (eg. the N-Back paradigm) to specifically target one of these mental states (working memory). Concomitantly, brain activity is recorded with electroencephalography (EEG) and/or functional near-infrared spectroscopy (fNIRS) to then be classified with machine-learning algorithms. As the ground truth about the task difficulty is known by the experimenter, the results of the classification can then be evaluated in terms of their accuracy[8]. However, BCI's further development and use in real life are unfortunately hindered by a decrease in performance due to strong variability that originates both between and within users factors[9]. In particular, EEG signals and BCI performance have been reported to significantly change across days, tasks, users' mental states and contexts[10–15]. A relevant solution to tackle this issue is to develop transfer learning methods. Transfer learning is an overarching term for a group of problems that refer to the decrease in effectiveness/accuracy of machine-learning algorithms and pipelines when applying the algorithm to slightly different but related data[16]. To that end, open datasets present not only a cost-effective alternative to collecting data individually for each experiment but also enables researchers to validate their algorithms and compare their performances on a similar set of data. Nonetheless, the number of freely available datasets is, with

[1]ISAE-SUPAERO, Université de Toulouse, Toulouse, France. [2]DTIS, ONERA, F-13661 Salon Cedex Air, France. [3]Artificial and Natural Intelligence Toulouse Institute – ANITI, Toulouse, France. ✉e-mail: marcel.hinss@isae-supaero.fr

exceptions[17], very limited see[10] for further details). The advantages of data sharing and the simultaneous lack of free data were the motivating factors for the creation of the COG-BCI database[18]. Hence, this database[18] was designed to propose a sufficient amount of fully informed data in a range of conditions representing various cognitive states, thus also allowing researchers to investigate transfer learning. Four tasks were carefully selected to allow the assessment of various mental states. The *Flanker task* is used for decision-making and conflict evaluation[19]. The task consists of deciding the direction of an arrow, while either having distracting (incongruent) or non-distracting (congruent) stimuli presented at the same time. Afterwards, participants receive feedback, allowing researchers to investigate the effects of errors as well as trial-based feedback. The *N-Back task* is known for taxing working memory, and has therefore been frequently used to elicit different levels of mental workload[20]. Here, participants have to retain numbers that appear on the screen and decide if the currently presented letter is the same as the Nth letter before (N equal to 0, 1 or 2). Another more ecological way to measure mental workload is the MATB-II, in which participants have to simultaneously perform four aviation-related subtasks. Lastly, the *Psychomotor Vigilance Task* (PVT) was selected as a simple and well-established measure for robust estimation of vigilance and fatigue[21]. In this task, participants have to react as fast as possible, by pressing a button, to a stimulus appearing pseudo-randomly on a computer screen. This article was meant as a description of a publicly available EEG database[18]. It details the methods used to create the tasks and database, the experimental choices and standards selected. Next, the analyses used for technical validation of the tasks are presented. The results (subjective, behavioral and physiological) are further interpreted to determine if the tasks were correctly implemented and elicited the expected mental states and their physiological markers. These results also include a proof of concept of their possible use for pBCI development with a three-class machine-learning estimation pipeline example. Lastly, usage notes are provided.

## Methods

### Participants.

Ethical Approval for the data collection and subsequent distribution was obtained from the Comité d'Éthique de la Recherche - CER at Université de Toulouse (CER number 2021-342). Based on the average number of participants presented in the pBCI literature (around 15 participants per experiment[5,22]), cognitive science literature (around 20 participants per experiment[23,24]) and similar databases (SEED[25] 15 participants and DEAP[26] 32 participants) we initially recruited 35 participants to perform this experiment. Due to participant dropout and technical issues with the data collection, a total of 29 participants were included in the final experimental setup. Participants (11 Female, 18 Male) were on average 23.9 (Std. 3.20) y.o. All but 4 participants were students, while the others were employees. 14 participants had obtained a bachelor's degree, while 13 had obtained a master's degree. Based on the Edinburgh Handedness scale, 2 participants were left-handed.

### Experimental paradigm.

The experiment was conducted over 3 sessions spaced one week apart during which participants had to perform 4 different tasks: the N-back task, the MATB-II, the PVT and the arrow-based Eriksen flanker task[19]. Upon their arrival for the first session, participants were informed about the study procedure and asked to read and sign the informed consent.Participants consented to the collection and subsequent distribution of their subjective, behavioral and physiological (cardiac and EEG) data for the duration of the experiment. Participants also consented to having a 3D picture taken. Then they were asked to fill in the demographic questionnaires, as well as the Edinburgh Handedness Inventory[27]. The EEG was set up while participants received task instructions. To ensure precise electrode location for within and between-subjects variability, a 3D picture of their head with the electrode cap was recorded at the beginning of each session. Then, participants completed a short training sequence for all 4 tasks. After training, they filled in the Karolinska Sleepiness Scale[28] (KSS) and performed a two-minute resting state (one-minute eyes open and one minute eyes closed) while their brain activity was recorded. Participants then started performing the tasks, which were presented twice in a pseudorandom order followed systematically by a Rating Scale Mental Effort[29] (RSME). An additional KSS was also completed following the PVT only. Participants were allowed to take short breaks in-between tasks. After all tasks were fully completed, participants performed another resting-state period, before filling in the KSS scale once more. The entire experimental protocol is detailed in Fig. 24. Completing the tasks took from 65 up to 80 minutes depending on the number of breaks the participants took. Together with setting up the electrodes and training, each session lasted for about two hours, with the first session often lasting a bit longer due to more detailed instructions. Due to the ongoing pandemic at the time of acquisition, a special sanitary protocol was adopted.

### Materials.

The database was created by acquiring subjective, behavioral and physiological data from participants using the experimental protocol presented above. Details on the questionnaires, tasks and acquisition devices used are given in this section.

*Subjective questionnaires.* All questionnaires were coded and administered with MATLAB v.2021a (The Mathworks Inc.), and presented on a 60 *Hz* LCD computer screen. The questionnaires can be found in Fig. 23.

Demographics. The demographics questionnaire includes questions assessing age, gender, level of education and occupation.

Edinburgh handedness inventory. The shortened version of the Edinburgh Handedness Inventory is administered to participants[30]. It encompasses four items (e.g., writing) scored on a 5-point scale ranging from: "Always right; Usually right; Both equally; Usually left; Always left". This shortened version is a faster measure while maintaining its reliability[27].

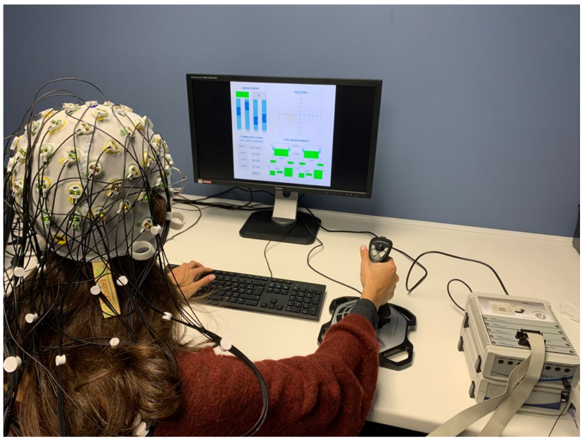


Fig. 1 Experimental setup with a participant equipped with the 64 electrode EEG system performing the MATB-II task. Consent for the publication of this picture was obtained from the participant.

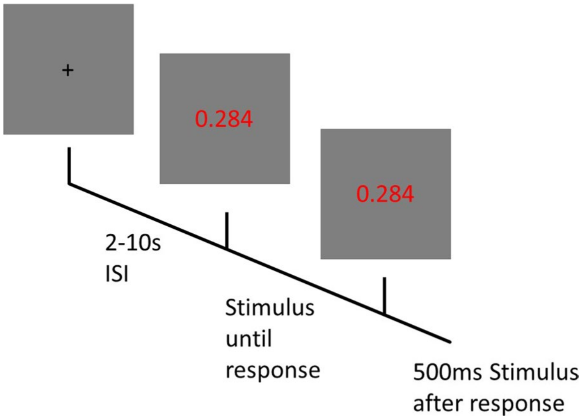


Fig. 2 The PVT task: Following an ISI of 2–10 seconds a red timer appears, which shows the time elapsed from trial onset until reaction. Once the participant reacts with a button push, the timer stops and shows the reaction time for 500 ms before the next trial starts.

Karolinska sleepiness scale (KSS). The KSS performs a measure of subjective sleepiness at any given time[31]. It is a simple 9-point scale ranging from "1 extremely alert" to "9 extremely sleepy - fighting sleep". The scale has been validated and used frequently thus providing a fast, straightforward and reliable measure of sleepiness[28].

Rating scale mental effort (RSME). The RSME is a scalar measure of Mental Effort. On a 150*mm*-long line, participants have to indicate their subjective level of workload. The scale is marked by 9 anchor points ranging from '0 = absolutely no effort' to '130 = extreme effort'. The instructions are written above the scale and read: "Please indicate, by marking the vertical axis below, how much effort it took for you to complete the task you have just finished". This scale has been validated and is favourable due to its short duration compared to other scales, such as the widely used NASA-Task Load Index[29].

*Tasks.* Psychomotor vigilance task. The PVT is a 10-minute task allowing to measure vigilance[21,32]. Participants are asked to react as fast as possible to the appearance of a timer on the computer screen by pressing the space bar of a keyboard. This task was designed to closely resemble the PC-PVT 2.0, an established computer version of the PVT[32]. Each trial starts with an interstimulus interval (ISI) lasting between 2 and 10 s. Then the timer is displayed on the screen until participants responded Fig 1. Then the timer stops and displays the participant's reaction time on the screen for another 500 *ms*. Participants have to perform a total of 90 trials per session, leading the task to last roughly 10 minutes. Participants' reaction times and responses are collected throughout the whole task. Figure 2 contains an illustration of the task.

N-Back task. The N-Back task is a widely used measure of both working memory and mental workload[20,33]. This task is selected to elicit different levels of mental workload across several conditions without changing the amount of visual information presented or the amount of motor responses required thus avoiding EEG data contamination.

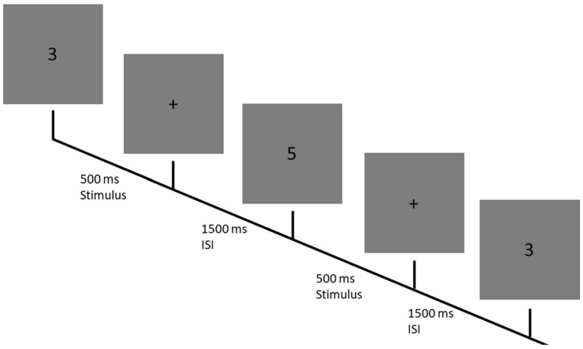


**Fig. 3** The N-Back task: Single digits are presented for 500 ms with a 1500 ms ISI. If in the 2-back condition the last 3 stimulus would be a hit trial and require a response from the participant, as the digit shown 2 trials before was the same.

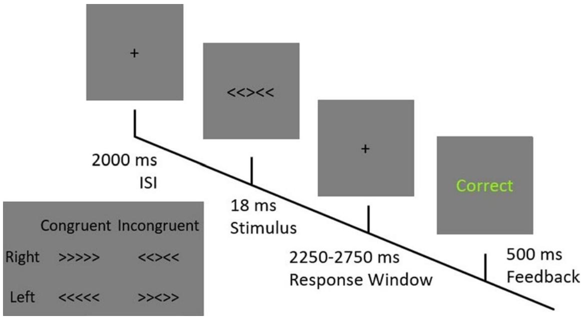


**Fig. 4** The Flanker task: Following a 2000 ms ISI one of the 4 possible stimuli (bottom left) is shown for 18 ms. Participants then have between 2250–2750 ms to react before they receive feedback for 500 ms.

On a computer screen, participants are presented with single letters appearing for a short period of time. Participants are instructed to remember the order in which the numbers appear and to react with a button press if the presented number is the same as the $N^{th}$ number presented before. The *N* here is the variable that determines the difficulty of a particular block. With the increasing size of *N*, the difficulty of this task increases, as more numbers need to be retained. Here 3 different conditions are chosen 0-Back, 1-Back and 2-Back corresponding to easy, medium and high workload levels. Trials begin with the presentation of a number (between 1 and 9) for 500 *ms*, followed by a blank screen for 1500 *ms* (Fig. 3). If the number presented is a hit number participants are instructed to respond by hitting the space bar. A hit number corresponds to the number 3 in the 0-back condition; the same number as the previous one in the 1-back condition; and the same number as 2 trials before in the 2-back condition. Each block consists of 48 trials and lasts approximately 2 minutes. The display frequency of hit numbers was fixed at $\frac{1}{3}$ (16 trials per block) in all three conditions. Participants completed 3 blocks (~6 minutes) of each condition for a total of 9 blocks.

Additionally, in the 2-back condition, 5 conflict trials are added per block. These trials are characterized by a number being followed immediately by the same number again. Participants are asked not to react to these trials, but this has been suggested to result in eliciting conflict[34]. As these trials would also occur during full randomization of the numbers, they have no adverse effect on the participant's performance. Before the onset of each block, the participant was informed about the condition of the block as well as given a short instruction on what to do. Participants' responses and reaction times are recorded throughout the entire task for each session.

Flanker task. The Flanker task is a simple choice reaction task to elicit errors and conflict during a binary decision[19]. In its arrowhead version, p articipants are presented with stimuli composed of 5 horizontal arrows. They are instructed to react to the middle arrow and ignore the flanking arrows on either side. These flanker stimuli can either point towards the same direction (congruent condition) or the other direction (incongruent condition) as the target, central arrow. A typical stimulus may therefore look like '<< ><<' or '<<<<<', where the respective targets are '>' and '<'. Each trial begins with an ISI of 2000 *m*, followed by the stimulus display for 16 *ms* (Fig. 4). This display time was determined in a pilot study based on changes in error rates. Each of the four possible stimuli ('>>< >>' '<<<<<' '<< ><<' '>>>>>') are presented equally frequently (25% of all trials) in a pseudorandom order. Following the stimulus presentation, participants are required to respond by stating the target direction with the "S" and "L" keys of the keyboard. During this time a blank screen with a fixation cross is shown for 2500 ± 250 *ms*. At the end of each trial, participants received feedback about the outcome (correct, incorrect, miss) of their trial for 500 *ms* In total, 120 trials were performed (30 for each type), with a complete run

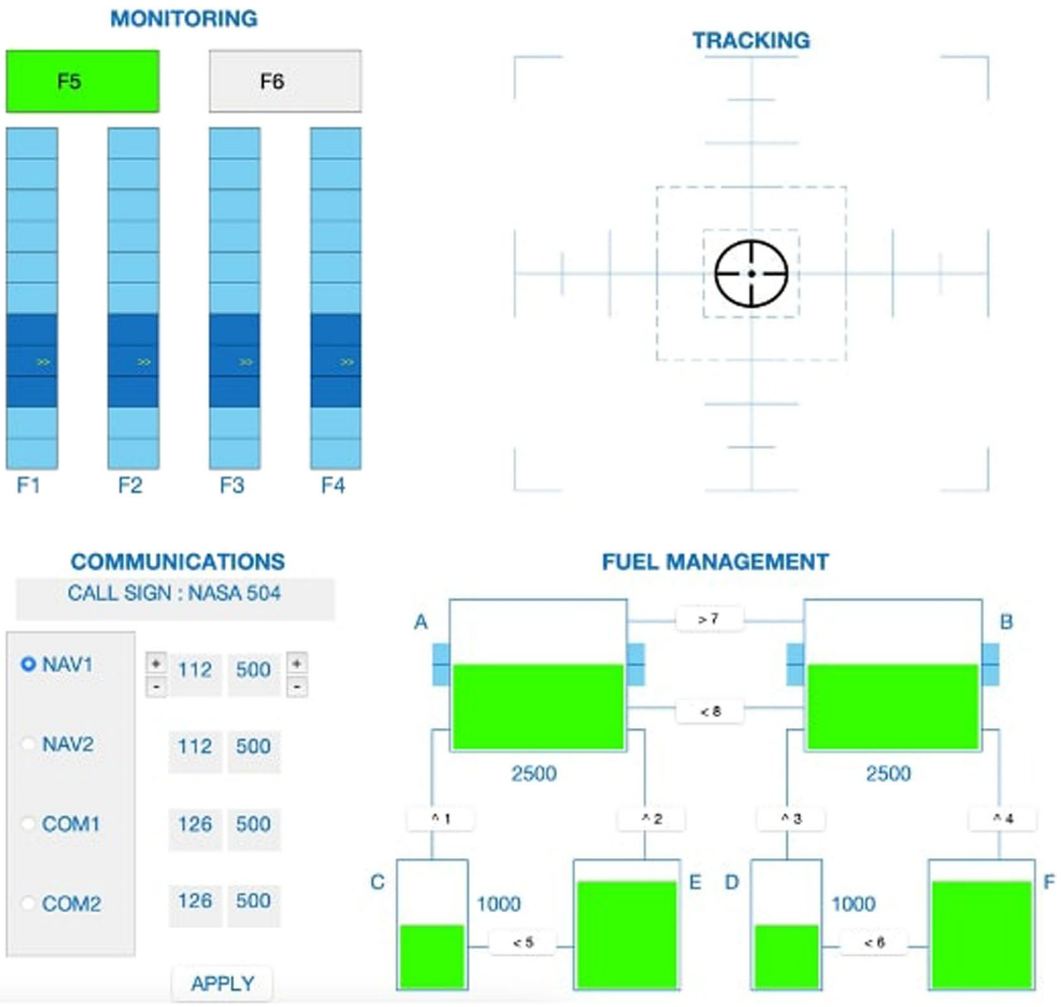


**Fig. 5** The MATB task: Top Left: Monitoring sub-task (SYSMON), where participants have to react to alarms with button presses, if any of the gauges indicate abnormal values (change in color for F5 and F6, extremely high or low values for F1 - F4). Top Right: Tracking task (TRACK), which requires the participants to keep the circle in the middle of the target by controlling a joystick. Bottom Left: Communications sub-task (COMM), requiring participants to react to radio messages and to change frequencies according to the message by using the Joystick. Bottom Right: The Fuel Management task (RESMAN), where participants need to keep the reservoir level of the two main tanks at a certain level, by activating and deactivating 8 pumps.

taking around 10 minutes. Before the onset of the run, participants received instructions. Participants' responses, error rates and reaction times to each type of stimulus are recorded throughout the entire task for each session.

MATB-II task. NASA developed the MATB-II task to assess task-switching and mental workload capacities in more realistic environments[35] 1. As the MATB-II also taxes mental workload the inclusion of this task not only allows for more ecologically valid taxation of mental workload, it also allows researchers to investigate transfer learning between the N-Back task and a more realistic measures of mental workload. Participants are presented with up to 4 different tasks that they have to complete simultaneously. This provides a highly realistic environment of operational systems that researchers can control to create different degrees of cognitive workload and difficulty. For this study, combinations of four of the available subtasks of the MATB-II were used and presented with an adapted version of the MATB-II software (Fig. 5). This version is coded in MATLAB v.2021a (The MathWorks Inc.) and provides the same measures as the original MATB-II task[22]. The four subtasks selected from the MTAB-II are tracking (TRACK), system monitoring (SYSMON), communication (COMM) and resource management (RESMAN); leaving task scheduling unused. In the tracking task (TRACK - top right corner), participants are presented with a moving target inside a window and have to keep the target within the central square area by controlling a joystick. The degree of difficulty can be adapted by modifying the degree and the speed at which the target moves. For the System monitoring task (SYSMON - top left corner), participants have to monitor gauges and warning lights. Action is required in the absence of green lights, the presence of red lights and deviations of four moving pointers dials from a midpoint. Participants have to press keys F5, F6 and F1 to F4 respectively in those cases. The degree of difficulty can be adapted by increasing the number of events to

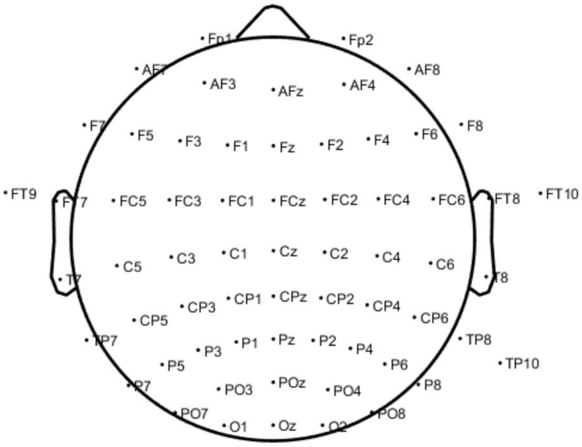


**Fig. 6** EEG channel locations following the extended 10–20 system[56]. Electrode TP9 was sacrificed to record peripheral ECG.

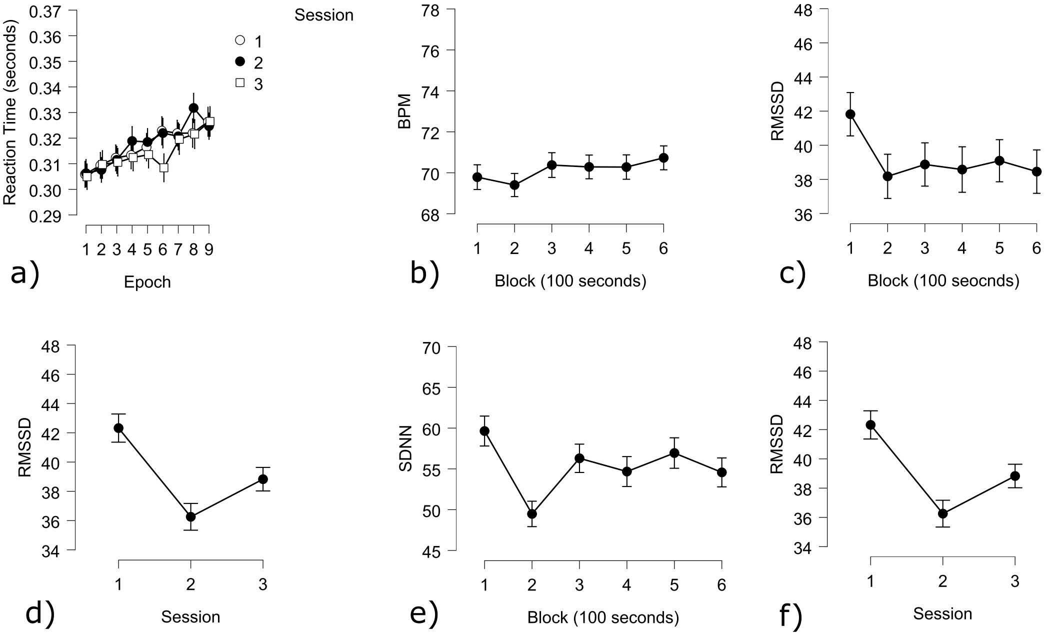


**Fig. 7** Behavioral and cardiac PVT results: (**a**) Average reaction time over trials (10 trials each) per session. A consistent increase in RT over sessions can be observed; (**b**) Average BPM over blocks of 100 seconds; (**c**) Average RMSSD (measure of HRV) over blocks; (**d**) Average SDNN (measure of HRV) over sessions.

which the participant has to react to. In the communication task (COMM - bottom left corner), participants are required to listen to radio messages and select specific radio channels and change frequencies accordingly when the messages are directed to him/her while ignoring messages not directed to them. Participants' responses are recorded with joystick-based button presses. Workload level can be adapted by including fewer or more messages directed or not to the participant. The last task used is the resource management task (RESMAN - bottom right corner). Participants are presented with an interface displaying two main fuel tanks (A and B) and four subsidiary tanks (C to F) interconnected via eight pumps (numbers 1 to 8, Type keys). The goal is to maintain a specific level of fuel in both of the main tanks. Participants can do this by activating or deactivating the pumps. To increase the difficulty of the task, events such as pump failures can be introduced. Pump activation/deactivation can be done by pressing buttons (F1-F6) Whenever a pump fails, it is inoperable for a short time window.

In the current study, participants are asked to perform three independent runs of 5 minutes each corresponding to a different degree of difficulty. For the easy condition, participants only engaged in the system monitoring and the tracking tasks. For the medium condition, participants engaged in both tasks as well as the resource management task. For the difficult condition, the communications task was added, as well as the tracking task was made more difficult. Before the start of each run, the participants also received a short instruction.

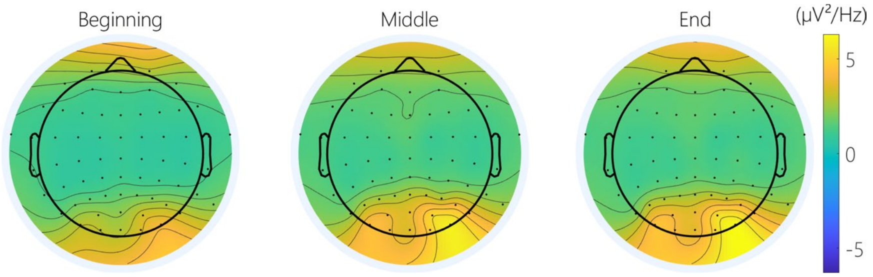


**Fig. 8** Topography of the average alpha power over three-time windows of 1 minute, in the beginning, the middle, and at the end of the PVT.

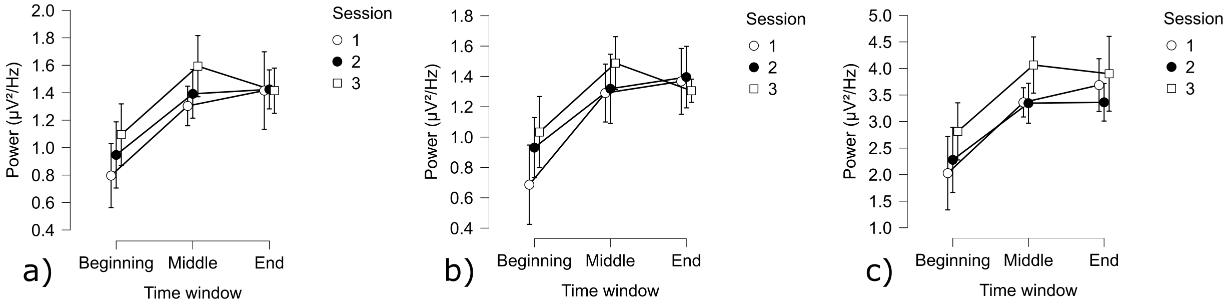


**Fig. 9** Cerebral PVT results: (**a**) Average alpha power in the frontal area; (**b**) Average alpha power in the central area; (**c**) Average alpha power in the posterior area.

### Data acquisition.

Subjective, behavioral and physiological data are recorded throughout the entire session for each session independently and with each participant. Accurate data synchronization and stimuli recording is performed through the LabRecorder of the LabStreaming Layer software and its related API (https://labstreaminglayer.readthedocs.io/info/intro.html).

*Stimulus presentation and response acquisition.* All tasks were coded in MATLAB v.2021a (The Mathworks Inc.), with the PsychToolBox-3 (http://psychtoolbox.org/). To display the stimuli, a desktop computer with a 60 Hz screen was used. Response acquisition occurred with a Keyboard and an Extreme 3D Pro Logitec Joystick for the MATB task. Participants were seated approximately 50 cm away from the screen Fig. 1.

*EEG system.* In this experimental campaign, we used an EEG system (electroencephalography) with 64 active Ag-AgCl electrodes (ActiCap, Brain Products Gmbh) and an ActiCHamp amplifier (Brain Products, Gmbh). Electrode locations followed the standard 10–20 system[36]. Electrode used here names can be found in Fig. 6. For participants 1–9 the electrode Cz was not recorded. In addition, one electrode dedicated to recording peripheral electrocardiographic (ECG) activity was placed on the left fifth intercostal. To obtain the precise location of the electrodes on the scalp of each participant at each session, a 3D scanning camera by STRUCTURE (https://structure.io/ and the get chanlocs plug-in developed specifically for electrode localisation purposes was used (GitHub.com/sccn/get chanlocs/wiki[37]). For this, a 3D camera was located on an IPad (Apple Inc.) and calibrated once for the entire experimental campaign. Then, after equipping the participant with the electrode cap, and right before the experiment, a 3D scanning of their head was realised as described on the Structure Sensor website and recommended specifically for EEG on the get_chanlocs wiki.

Data was recorded continuously during each session. Reference was located at Fpz (Fig. 6). Before acquisition impedances were improved such that acquisition started with all electrode impedances lower than 25 $k\Omega$. The signal was amplified, digitized at a 24-bit rate, and sampled at 500 Hz with a 0.05 $\mu V$ resolution. No filtering was applied during the acquisition.

### Data processing.

To allow the community to perform their pipeline design and tests, the data that we provide is raw. Hence, only the structure and format of these data is detailed below.

## Data Records

COG-BCI is available on Zenodo[18]. In order to improve standardization and ease of use, the data is presented in the BIDS format, intended to become the new standard for neuroimaging studies[38] (https://bids.neuroimaging.io/). An exemplary folder tree can be seen in Fig. 25. Participants are numbered from 1 to 29. For each session, the behavioral results, as well as the exact electrode locations and individual datasets for each task, are provided. The EEG data is saved in the .set and .fdt file format (two files per dataset). The resting state is divided into 4 different

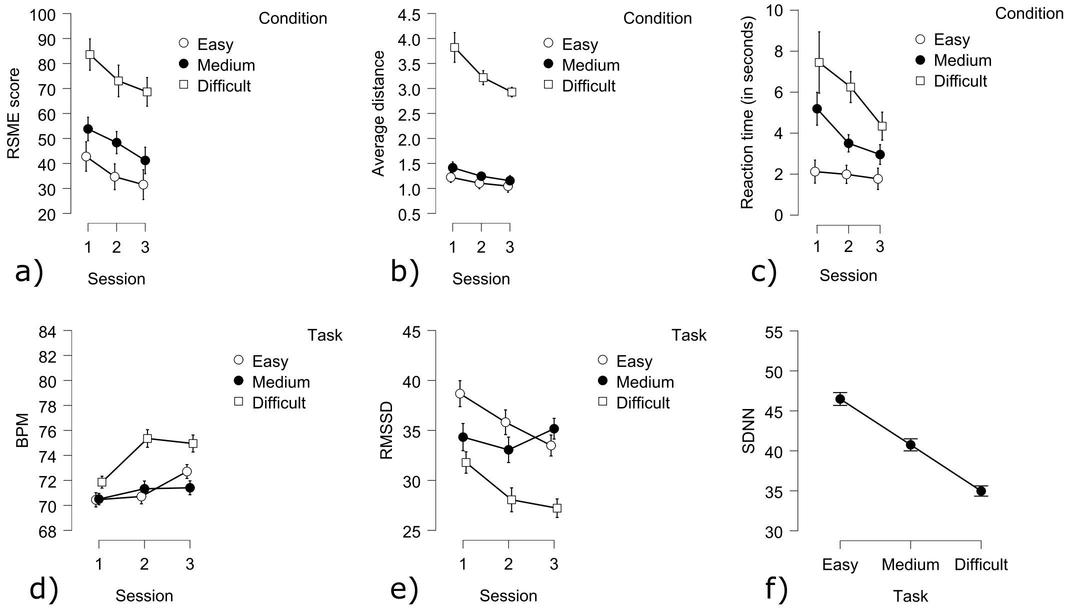


**Fig. 10** Behavioral and cardiac MATB results: (**a**) RMSE score divided by condition and by session; (**b**) Average absolute TRACK distance divided by condition and by session; (**c**) SYSMON reaction time divided by condition and by session (**d**) BPM divided by condition and by session; (**e**) RMSSD (measure of HRV) across conditions; (**f**) SDNN (measure of HRV) across conditions.

| Number in the Notebook | Task |
|---|---|
| 1 | PVT |
| 2 | Flanker |
| 3 | Two-Back |
| 4 | One-Back |
| 5 | Zero-Back |
| 6 | MATB-Easy |
| 7 | MATB-Medium |
| 8 | MATB-Difficult |

**Table 1.** Table with the numbers and the tasks they refer to within the Notebook. Allows for determining the order of the tasks of each session for each participant.

datasets: RS_Beg refers to the resting state at the beginning of the sessionand RS_End to the resting state at the end of the session; EC is the abbreviation for eyes closed and EO refers to eyes open. Furthermore, a notebook file is available within the database[18]. This file details the order in which the tasks - referenced with numbers available in Table 1 - were acquired, if the recording had to be interrupted at any point as well as other comments if applicable. The trigger list file contains all LSL triggers and what they refer to.

## Technical Validation

This section presents the analyses that were performed in order to ensure that the experimental protocol did elicit the mental states of interest to ensure the pBCI community of the relevance of the dataset. First, the processing applied to the data to enable statistical analysis of the results is presented. Next, the results obtained thanks to the statistical analyses are presented. Lastly, proof of usability for pBCI is provided with a simple pipeline and estimation results, followed by a conclusion on this technical validation. For all reported statistics a p-value of <0.05 was considered significant. For multiple correction, the Bonferroni method was applied and the reported p-values are multiplied by the factor of tests. Following the GLM contrast analyses (t-test) were used to report statistical significance[39].

**Data processing and analysis.** Data Processing and Analysis were performed using MATLAB v.2021a (The Mathworks Inc.) with the EEGlab toolbox v.2022.0 (https://sccn.ucsd.edu/eeglab/index.php) and JASP v. 0.16.3 (https://jasp-stats.org/).

*Cardiac activity.* The data from the cardiac electrode was first down-sampled to 250 Hz before a frequency band pass FIR filter was applied (1–40 Hz). The data were then epoched into 10-second segments. Using R

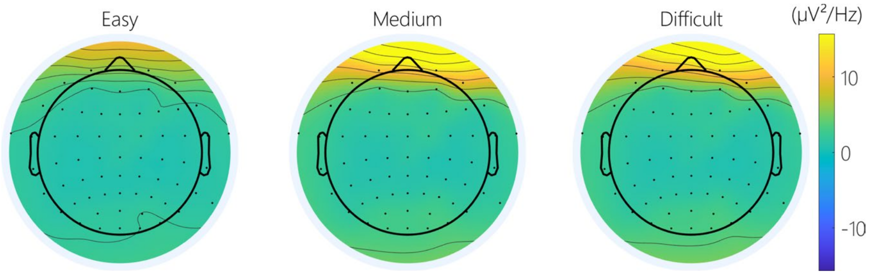


**Fig. 11** Topography of the average theta power across difficulty on the MATB.

peak-detection, the heart-rate (HR) as well as heart rate variability (HRV; in the time domain) were computed for each epoch. HRV was measured with the Standard deviation of NN (i.e., normal R-R) intervals (SDNN) and the root mean square of the successive differences (RMSSD). Both measures have been used in neuroergonomic experiments for mental workload and vigilance evaluation[40,41]. The RMSSD is a measure that is often used for shorter time-windows as it reflects beat-to-beat variations while the SDNN is a metric more suited for long-term analysis, reflecting overall cardiac variation[42].

*Cerebral activity.* The EEG data were processed with a usual pre-processing pipeline. Data were first down-sampled to 250 *Hz*. Next, a 1 *Hz* high-pass FIR filter was applied. An automatic channel rejection and interpolation, based on a two standard deviation criterion was then used. On average, 0.34 channels were interpolated per task. Using the cleanline[43] toolbox, the 50 *Hz* line noise was then filtered. The data were then epoched into 0.5-second segments and an automatic epoch rejection using a 2 standard deviation criterion was applied. On average, 16 epochs (8 seconds) were rejected per task. Finally, an independent component analysis (ICA) and subsequent component rejection were performed using the standard runica extended function in EEGLAB v. 2022.0 (https://sccn.ucsd.edu/eeglab/index.php) and the IClabel toolbox[44]. The thresholds for rejecting eye, heart and muscle components were all set to >90%. On average 7 components were rejected per participant and session. After the pre-processing, data was merged again into a continuous signal.

When the data was used for machine learning (i.e. proof of usability for pBCI, see section **Mental Workload Estimation**), an adapted version of this pipeline was applied. Here the data was first epoched before the pre-processing was applied to each epoch individually. The automatic epoch rejection was therefore not used.

For the EEG analysis, the power in the theta (4–8 *Hz*) and alpha (8–13 *Hz*) bands were extracted at each electrode of the continuous data using the spectopo EEGLAB function and the absolute mean power was computed for each frequency band. Electrode clusters were selected to calculate the power by brain areas, identical to those of Simon *et al.*[45]. For the frontal area, a cluster of 10 electrodes was averaged: F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4; for the central area, 10 electrodes were averaged: C3, C1, Cz, C2, C4, CP3, CP1, CPz, CP2, CP4; and for the parieto-occipital area, a cluster of 11 electrodes was averaged: P3, P1, Pz, P2, P4, PO3, POz, PO4, O1, Oz, O2.

*Specific per-task analyses.* Psychomotor vigilance task. The behavioral analysis of the PVT focused on the reaction times of participants. It was analyzed whether there was a significant effect of time-on-task (TOT; measured per trial), Session and ISI.

Cardiac analysis of the PVT focused on the normalized values of HR and HRV. A General Linear Model (GLM) with TOT (measured in the 10-second epochs) as a covariate and session as a fixed factor on the dependent variables of HRV (SDNN or RMSS) and HR (beats per minute; BPM) was performed.

For the EEG analysis, three 1-minute time windows were selected within the signal: one at the beginning, one in the middle and one at the end of the PVT[46]. A Greenhouse-Geisser correction was applied when the assumption of sphericity condition was violated. A GLM with Session (Session: 1, 2, 3) and Time window (Time window: beginning, middle, end) as fixed factors was performed on the alpha power on each Region of Interest (ROI: frontal, central and posterior).

MATB-II task. To analyze the subjective RSME scores, a 3 × 3 (Session: 1, 2, 3; and Difficulty: Easy, Medium, Difficult) repeated-measures analysis of variance (RMANOVA) was performed.

As with the subjective scores RMANOVAS were performed on the performance of the tracking and the system monitoring tasks. This was done as these were the only two tasks assessed across all conditions. For the tracking task, the average absolute distance from the center of the square (root mean squared distance) was used as a measure of performance. For the system monitoring task, the average reaction time to alarms was used.

For all analyses, a Greenhouse-Geisser correction was applied when the assumption of sphericity was violated. Cardiac data was analyzed using a GLM (Fixed factors: Session x Condition, TOT as a covariate) on the dependent variables of HRV (SDNN or RMSS) and HR (BPM).

For the EEG analysis, the data were analyzed according to the difficulty of the task (Easy; Medium; Difficult), the cortical area (frontal and posterior) and the theta and alpha bands. A GLM (Fixed factors: Session x Difficulty) was performed on each ROI (frontal, central and posterior).

N-Back task. A 3 × 3 × 3 RMANOVA (Session x Condition x Block) was performed on participants' error rates. A Greenhouse-Geisser correction was applied when the assumption of sphericity was violated.

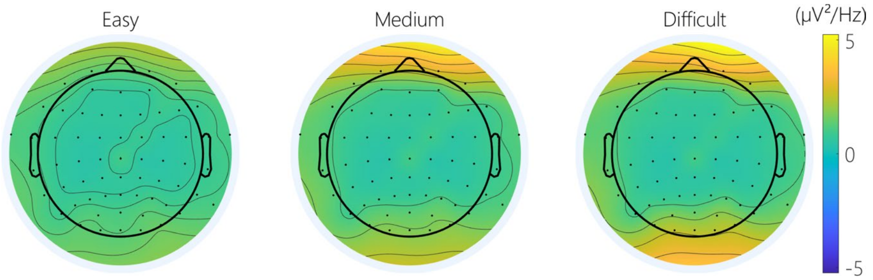


**Fig. 12** Topography of the average alpha power across difficulty on the MATB.

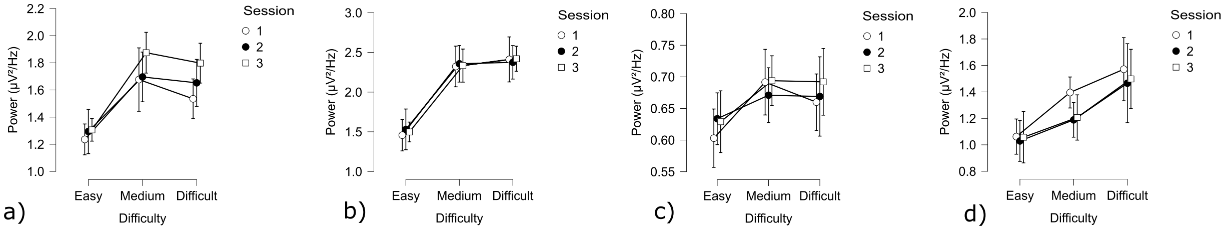


**Fig. 13** Cerebral MATB results: (**a**) Average theta power in the frontal area; (**b**) Average theta power in the posterior area; (**c**) Average alpha power in the frontal area (**d**) Average alpha power on the posterior area.

For reaction times, using the same factors the data was analyzed using a GLM (dependent variable: reaction time).

Cardiac data were analyzed in a similar way as was the cardiac data of the MATB task. A GLM (Fixed factors: Session x Condition x Block) on the dependent variables of HRV (SDNN or RMSS) and HR (BPM) was performed. Here Block was added as an additional factor. The EEG data were analyzed according to the difficulty (Easy; Medium; Difficult) of the N-back task, with a GLM (fixed factors: Session x Difficulty) on each ROI (frontal, central and posterior), and for both the alpha and theta band respectively.

Flanker task. The behavioral results of the Flanker task were analyzed using GLMs and RMANOVAs. The first analysis (GLM) focused on reaction times as a dependent variable with congruency (congruent vs incongruent trial), session (1, 2 or 3) and response accuracy (error vs correct response) as fixed factors. Furthermore, it was investigated if session or congruency affected accuracy (RMANOVA).

To analyze the EEG data, ERPs were computed for both the moment of response (when the participant reacted to the stimulus) and for the feedback (when participants received positive or negative feedback regarding the current trial). At the moment of the response, the analysis focused on the ERN amplitude, which is sensitive to accuracy[47]. In the feedback time window, the analysis focused on the FRN and P300, which have been shown to be sensitive to feedback valence[48]. To determine the peak amplitude of the three components, the same procedure was applied: the most negative (for the ERN and FRN) or positive (for the P300) point in the 0–150 *ms* post-response, and in the 175–325 *ms* and 250–450 *ms* post-feedback time-windows respectively for the ERN, FRN and P300 were identified for each epoch. Then the average around this point was computed (±2 points for a 5-point average). To analyze the results a GLM with the Fixed factors of congruency, correctness, and the session was performed on the dependent variables of peak amplitudes of the P300, FRN and ERN components.

Mental workload estimation. Mental Workload estimation was performed on both the N-Back and the MATB tasks, with three different mental workload levels each (0-back, 1-back, 2-back; and MATB Easy, MATB Medium, MATB Difficult). For both tasks, the same preprocessing, feature extraction and classification pipeline was used. To preprocess the data, the following steps were executed: First, 5-second non-overlapping epochs were created from the continuous data, which were then downsampled to 250 Hz. Next, FIR filters (eegfiltnew function from EEGLAB with default parameters) were applied to filter the data either in the theta or the alpha band (i.e.[4–8]*Hz* and [8–12]*Hz* respectively). Automatic channel rejection with subsequent interpolation of the rejected electrodes was performed. If the standard deviation of a channel was more than two standard deviations larger than that of the other channels, the channel was interpolated, using linear interpolation. To keep full rank for the following independent component analysis (ICA), a full-rank average referencing was performed. Finally, an ICA with IClabel component rejection was performed[44]. Eye, muscle and heart components with over 90% confidence were removed. Figure 26a) summarizes how many components on average, were removed per task. Figure 26b) details how many channels were interpolated per task. Differences between tasks with regard to the amounts of components removed channels interpolated are most likely due to differences in motor activity across conditions.

To classify the data, the covariance matrix of a subset of 10 channels (F3, Fz, F4, FCz, C3, C4, CPz, P3, Pz, P4) was calculated for each epoch. The data were then divided into training and testing sets using a 5-fold cross-validation procedure. The data were trained and tested on a Riemannian Minimum Distance to Mean

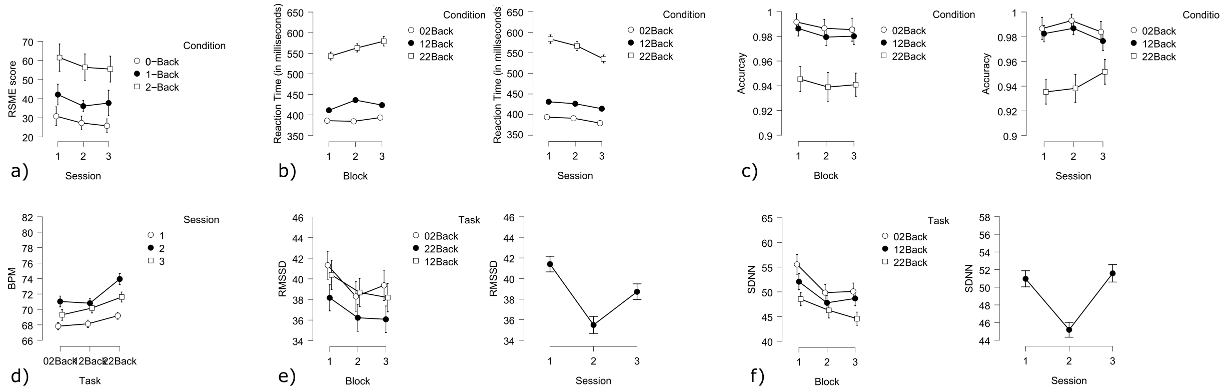


**Fig. 14** Behavioral and cardiac N-Back results: (**a**) RMSE score divided by condition and by session, showing an increase in subjective workload with increasing difficulty and a slight decrease in subjective workload over sessions (**b**) Reaction time across sessions divided by condition and Reaction time within sessions divided by condition. Over sessions participants improved in performance (lower reaction time). Within sessions the reaction time of participants increased (**c**) Accuracy (1 = 100%) across conditions; (**d**) BPM divided by condition and session; (**e**) RMSSD (measure of HRV) across conditions and sessions; (**f**) SDNN (measure of HRV) across conditions and sessions.

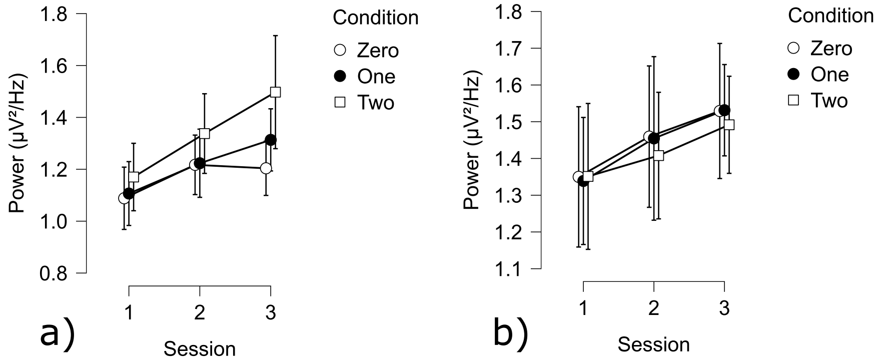


**Fig. 15** Cerebral N-Back results: Average theta power at (**a**) the frontal area; (**b**) the posterior area.

(MDM) classifier[49,50]. Based on at least 60 trials in each of the 3 classes, the upper limit for the 33% chance level is 41.7%[51]. Accuracy was then evaluated using a RMANOVA with the factors of session (1 vs. 2 vs. 3), difficulty, and bandwidth (theta vs. alpha).

*Validation results.* Here, we present the results obtained for each task independently, as well as the results for mental workload classification. RSME, as well as KSS scores, did not significantly change across sessions.

### Psychomotor vigilance task

- **Behavior**: The GLM showed that ISI, TOT and session all had significant effects on the reaction time (TOT: t(28) = 12.87 $p < 0.005$; ISI: t(28) = −19.32 $p < 0.005$; Session: t(28) = −2.31 $p < 0.05$; Fig. 7a). With increasing TOT, participants took significantly longer to respond to stimuli, while longer ISI resulted in quicker responses. Contrast analysis showed significantly lower reaction times during session 3 as compared to session 2 (t(28) = −2.31 $p < 0.05$).
- **Cardiac activity**: Both TOT and Session influenced participants' HR. It increased over sessions, as well as throughout each session (Epoch: t(4946) = 2.88 $p < 0.005$; Session 1 - Session 2: t(4946) = 9.86 $p < 0.001$; Session 1 - Session 3: t(4946) = 7.12 $p < 0.001$; Session 2 - Session 3: t(4946) = 2.67 $p < 0.01$; Fig. 7b).

  HRV however, significantly decreased as time increased (RMSSD: Epoch: t(4945) = −2.6 $p < 0.01$). The SDNN was not sensitive to changes within sessions. As observed with the N-Back task, the lowest HRV was observed during Session 2 (RMSSD: Session 1 - Session 2: t(4945) = −9.72 $p < 0.001$; Session 2 - Session 3: t(4945) = −5.50 $p < 0.001$; Session 1 - Session 3: t(4945) = −4.19 $p < 0.001$; Fig. 7; SDNN Session 1 - Session 2: t(4945) = −6.15 $p < 0.001$; Session 2 - Session 3: t(4945) = −3.45 $p < 0.001$; Session 1 - Session 3: t(4945) = −2.67 $p < 0.01$; Fig. 7).
- **Cerebral activity**: There was a significant effect of TOT, as indicated by the time windows, on the frontal, central and posterior alpha power (Fig. 8). A contrast analysis showed significantly lower alpha power at the

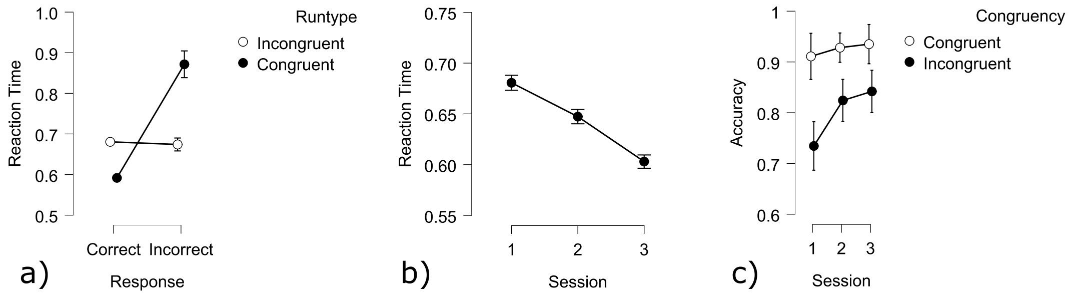


**Fig. 16** Behavioral and cardiac Flanker results: (**a**) Reaction time for Correct and Incorrect trials divided by Congruent and Incongruent trials; (**b**) Reaction time across sessions; (**c**) Accuracy across conditions and sessions.

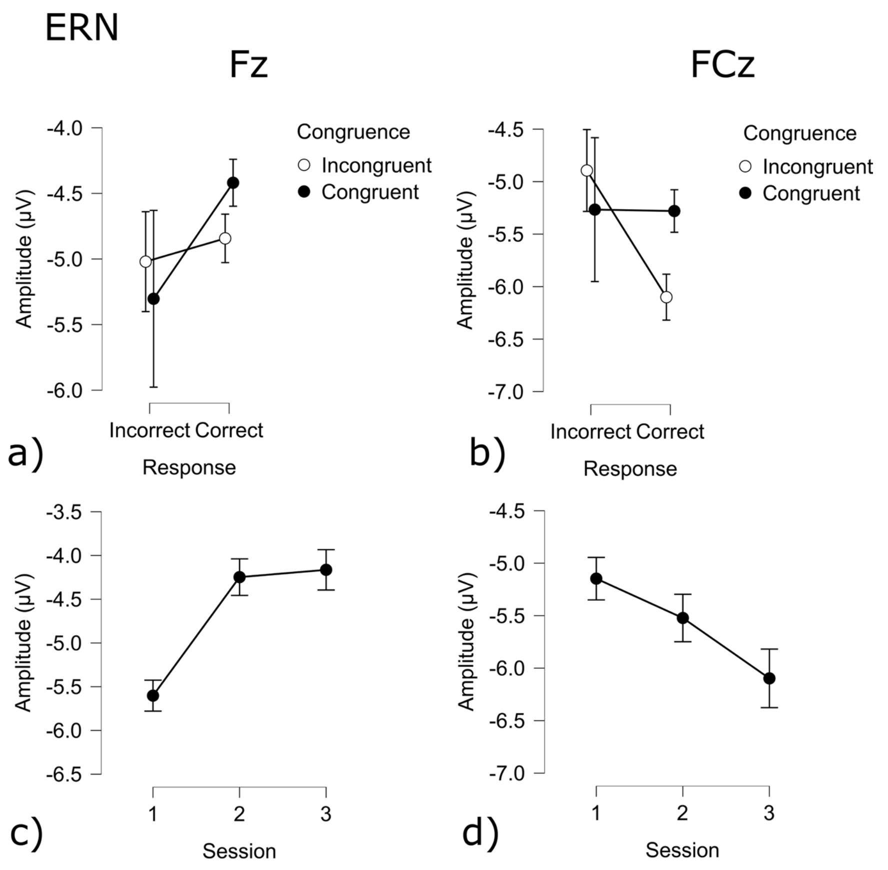


**Fig. 17** ERN results for the Flanker task: (**a**) Congruency and Response effect at the Fz electrode; (**b**) Congruency and Response effect at the Cz electrode; (**c**) Session effect at the Fz electrode; (**d**) Session effect at the FCz electrode.

beginning of the task, compared to the middle and the end of the task on all ROI (Beginning-Middle: frontal: t(256) = −2.51 $p < 0.05$; central: t(256) = −2.47, $p < 0.05$; posterior: t(256) = −2.47 $p < 0.05$; Beginning-End: frontal: t(256) = −2.51 $p < 0.05$; central: t(256) = −2.47, $p < 0.05$; posterior: t(256) = −2.47 $p < 0.05$; Fig. 9).

### MATB-II

- **RSME questionnaire**: RMSE scores were dependent on both session and difficulty (Session: F(2,104) = 17.15 $p < 0.005$; Difficulty: F(2,104) = 92.73 $p < 0.005$; Fig. 10a). Across sessions, RSME scores decreased, while increasing difficulty was reflected in higher scores.
- **Behavior**: The TRACK scores (average absolute distance from midpoint) showed significant effects of difficulty as well as session (Session: F(2,56) = 27.63 $p < 0.005$; Difficulty: F(2,56) = 421.99 $p < 0.005$). The interaction was also significant (Interaction F(4,112) = 23.12 $p < 0.005$; Fig. 10b). In general, the difficult condition

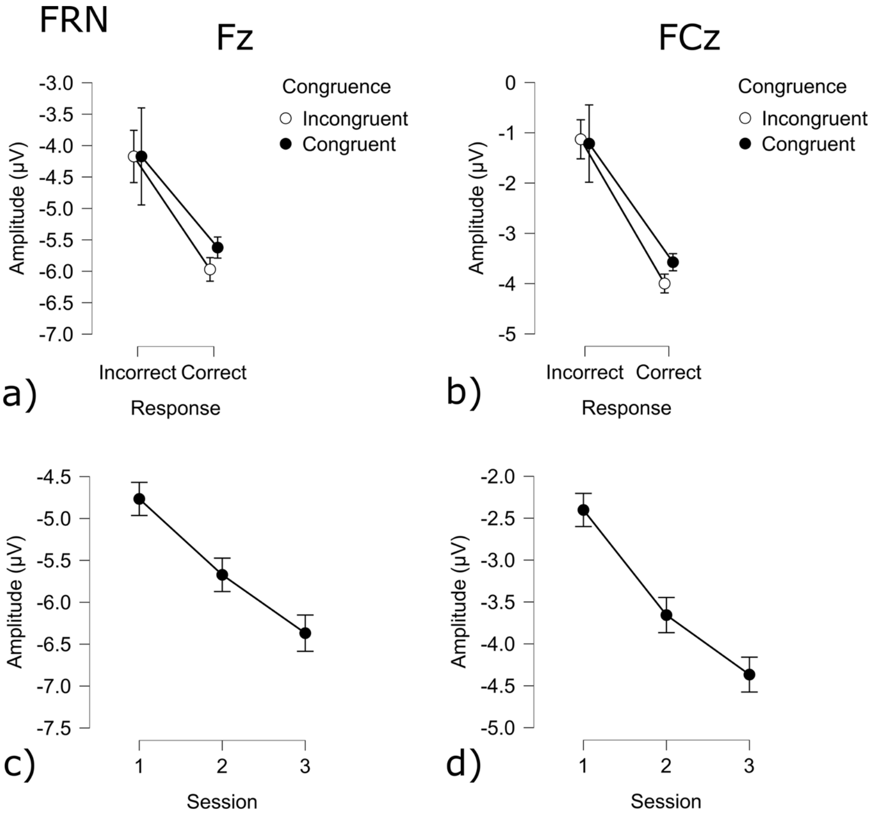


**Fig. 18** FRN results for the Flanker task: (**a**) Congruency and Response effect at the Fz electrode; (**b**) Congruency and Response effect at the Cz electrode; (**c**) Session effect at the Fz electrode; (**d**) Session effect at the FCz electrode.

showed much higher deviation scores as compared to the other conditions. However, an increase in performance was observed across sessions.

Analysis of the SYSMON scores (average reaction times to alarms) showed significant effects of difficulty as well as session (Session: F(2,56) = 15.17 $p < 0.005$; Difficulty: F(2,56) = 62.19 $p < 0.005$). The interaction was also significant (Interaction F(4,112) = 6.86 $p < 0.005$; Fig. 10c). As with the TRACK task, increase in difficulty resulted in lower performances as seen by higher reaction times. Across sessions, reaction times decreased in all sub-tasks.

- **Cardiac activity**: HR was found to be significantly affected by difficulty and TOT (Difficulty: Easy - Medium t(6860) = −4.11 $p < 0.001$; Easy - Difficult t(6860) = 7.44 $p < 0.001$; Medium - Difficult t(6860) = −11.51 $p < 0.001$; Fig. 10d). Contrast analysis showed that heart rate increased in the difficult condition, whereas the medium condition had the lowest heart rates. HR was shown to increase across sessions (Session 1 - Session 2 t(6860) = 6.27 $p < 0.001$; Session 1 - Session 3 t(6860) = 7.72 $p < 0.001$). However, the increase from session 2 to session 3 was not significant. Within sessions, an increase in heart rate was observed over time (TOT: t(6860) = 8.47 $p < 0.001$).

  HRV, as measured by RMSSD, was significantly influenced by difficulty, across and within sessions. In the difficult condition, significantly lower RMSSD values were observed as compared to the medium and easy conditions. Session 1 differed significantly from both sessions 2 and 3 (Difficulty: Easy - Difficult t(6860) = −9.89 $p < 0.001$; Medium - Difficult t(6860) = 9.75 $p < 0.001$; TOT: t(6860) = −4.72 $p < 0.001$; Session 1 - Session 2 t(6860) = −4.78 $p < 0.001$; Session 1 - Session 3 t(6860) = −4.85 $p < 0.001$; Fig. 10e). The heart rate variability was significantly lower in the difficult condition. During the first session, the variability was higher than during the other two sessions and within sessions, the HRV decreased. The results of the analysis of HRV as measured by SDNN showed similar results as the analysis using RMSSD. The difficult condition showed the lowest variability, while the easy condition had the highest SDNN values. SDNN values observed during session 1 were significantly higher as compared to the values observed during session 2(Difficulty: Easy - Difficult t(6860) = −16.29 p < 0.001; Easy - Medium t(6860) = −5.54 p < 0.001; Medium - Difficult t(6860) = 10.74 p < 0.001; Session 1 - Session 2 t(6860) = −2.24 p < 0.001; Fig. 10e). Within sessions a decrease in HRV was observed (TOT: t(6860) = −3.32 p < 0.001).
- **Cerebral activity**: There was a significant effect of difficulty on the frontal and posterior theta power (Fig. 11). A contrast analysis revealed that theta power during the easy condition was significantly lower compared to the medium and difficult conditions in both areas (Easy-Medium: frontal: t(256) = −4.35 $p < 0.001$; posterior: t(256) = −5.74 $p < 0.001$; Easy-Difficult: frontal: t(256) = −3.54 $p < 0.01$; posterior: t(256) = −6.18 $p < 0.001$; Fig. 13). Concerning the alpha power, there was a significant effect of difficulty in the posterior area

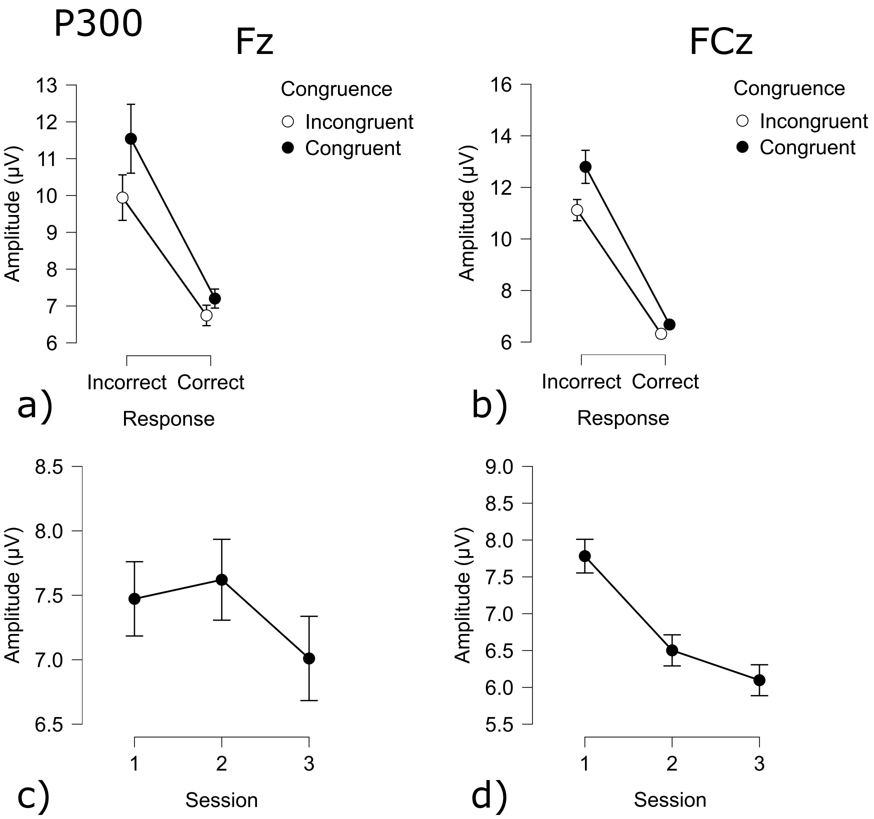


**Fig. 19** P300 results for the Flanker task: (**a**) Congruency and Response effect at the Fz electrode; (**b**) Congruency and Response effect at the Cz electrode; (**c**) Session effect at the Fz electrode; (**d**) Session effect at the FCz electrode.

(Fig. 12). A contrast analysis revealed that the alpha power was significantly lower in the easy than the difficult condition (Easy-Difficult: t(256) = −3.45 $p < 0.01$).

### N-Back task

- **RSME questionnaire**: RSME scores were significantly impacted by difficulty and session (Difficulty: F(2,104) = 55.56 $p < 0.005$; Session F(2,104) = 3.52 $p < 0.05$; Fig. 14a). An increase in difficulty was reflected by higher scores.
- **Behavior**: All GLM factors (session, condition, block and error) had significant main effects on reaction time (Session: t(28) = −8.89 $p < 0.005$; Condition: t(28) = 57.53 $p < 0.005$; Block t(28) = 6.29 $p < 0.005$; Error: t(28) = −10.29 $p < 0.005$; Fig. 14b). Reaction time decreased across sessions and increased within sessions. Errors were accompanied by longer reaction times. More difficult conditions (higher N) resulted in increases in reaction time. Regarding accuracy, a significant effect of condition on error rates was found (Condition: F(2,56) = 52.25 $p < 0.005$; Fig. 14c). More difficult conditions resulted in lower accuracy.
- **Cardiac activity**: Heart rate was found to be significantly higher during the 2-back condition (Difficulty: 0-back - 2-back t(6953) = 8.60 $p < 0.001$; 1-back - 2-back t(6953) = −7.52 $p < 0.001$; Fig. 14d). The lowest heart rate was observed during session 1, whereas the highest heart rate occurred during Session 2 (Session: Session 1 - Session 2: t(6953) = 13.96 $p < 0.001$; Session 1 - Session 3: t(6953) = 7.78 $p < 0.001$; Session 2 - Session 3: t(6953) = 6.12 $p < 0.001$). Block did not have any significant effect on heart rate.

  HRV as measured by SDNN was influenced by difficulty, session and block (Difficulty: 0-back - 1-back t(6953) = −4.02 $p < 0.001$; 1-back - 2-back t(6953) = 4.82 $p < 0.001$; 0-back - 2-back t(6953) = −8.75 $p < 0.001$; Session: Session 1 - Session 2: t(6953) = −9.45 $p < 0.001$; Session 2 - Session 3: t(6953) = −9.78 $p < 0.001$; Block: Block 1 - Block 2 t(6918) = −4.02 $p < 0.001$; Block 1 - Block 3 t(6918) = −6.67 $p < 0.001$; Fig. 14f). HRV was lower for more difficult tasks and decreased within each session. During session 2 a significantly lower HRV was observed. The first block had a significantly higher HRV as compared to both upcoming blocks.

  The results of the RMSSD metric largely confirmed these results. Difficulty, session and block showed significant effects on the RMSSD (Difficulty: 0-back - 2-back: t(6918) = −5.53 $p < 0.001$; 1-back - 2-back: t(6918) = 4.18 $p < 0.001$; Session: Session 1 - Session 2: t(6953) = −11.07 $p < 0.001$; Session 1 - Session 3: t(6953) = −5.32 $p < 0.001$; Session 2 - Session 3: t(6953) = −5.71 $p < 0.001$; Block: Block 1 - Block 2 t(6918) = −4.02 $p < 0.001$; Block 1 - Block 3 t(6918) = −3.62 $p < 0.001$; Fig. 14e). More difficult conditions resulted in a lower RMSSD value. As with the SDNN the lowest RMSSD value was observed during the

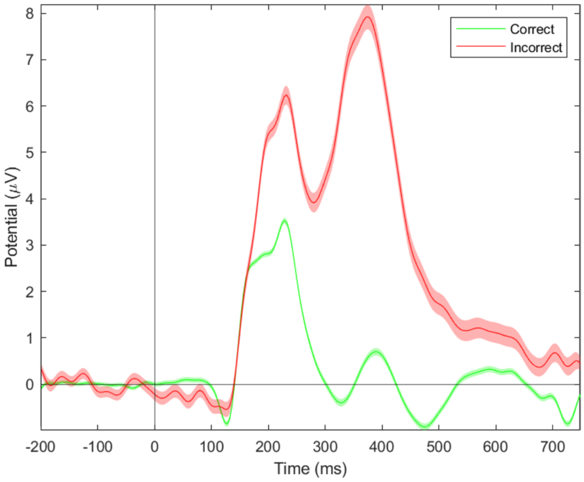


**Fig. 20** Event-related potentials elicited by feedback (correct/positive and incorrect/negative) at the FCz electrode (see ample P300 component in the incorrect condition).

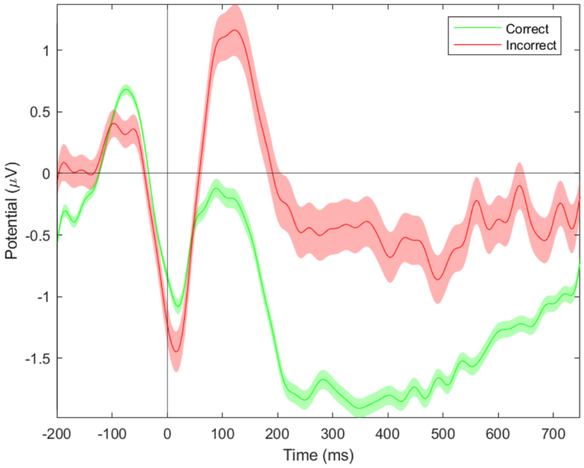


**Fig. 21** Error-related potentials time-locked to the participants' response at the FCz electrode during an arrow-based flanker task. Grand averages across participants show the ERP in response to correct (green) and incorrect (red) trials with the Error-Related Negativity peaking 30 ms post-response.

second session. In addition session 1 had a significantly higher RMSSD as compared to session 3. The HRV (as observed by the RMSSD) was significantly higher during the first block as compared to the other two blocks.

- **Cerebral activity**: Analysis of the EEG power showed that there was a significant effect of the session in the theta band on the frontal area (Fig. 15). A contrast analysis revealed that theta power during the third session was significantly higher compared to the first and second sessions in the frontal area (Session 1 - Session 3: t(256) = 2.12 $p < 0.05$). No effect of difficulty on theta power was found. Also, no significant effects of the session nor difficulty were found on alpha power for this task.

### Flanker task

- **RSME questionnaire**: RSME scores were significantly higher during session 1 as compared to the session 2 (t(54) = 3.38 $p < 0.005$).
- **Behavior**: All main factors (response, congruency and session as well as the covariate TOT) showed significant effects on the reaction time to the Flanker trials (Response: t(10433) = 46.84 $p < 0.001$; Congruency: t(10433) = −15.48 $p < 0.001$; Session: t(10433) = −14.83 $p < 0.001$; Fig. 16a,b). Contrast analysis revealed significant differences between: Correct Incongruent - Correct Congruent (t(10433) = 23.03 $p < 0.001$); Incorrect Incongruent - Correct Congruent (t(9980) = 11.29 $p < 0.001$); Correct Incongruent - Incorrect Congruent (t(9980) = 15.65 $p < 0.001$). Reaction times did not change due to correct responses for incongruent trials. However Congruent correct trials were significantly faster than responses to all other trials. Congruent Incorrect responses were significantly slower than all other trials. The analysis with

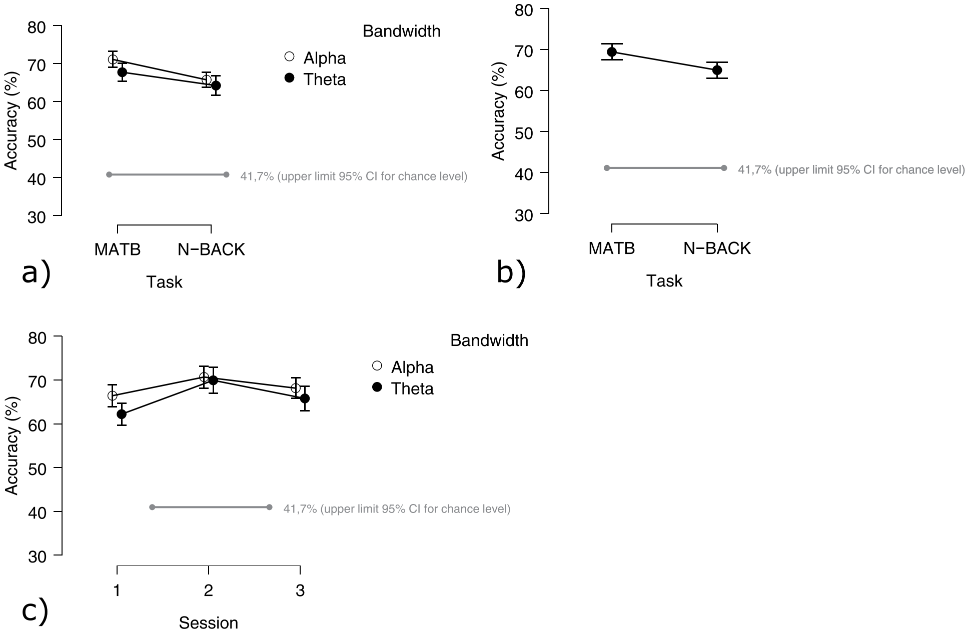


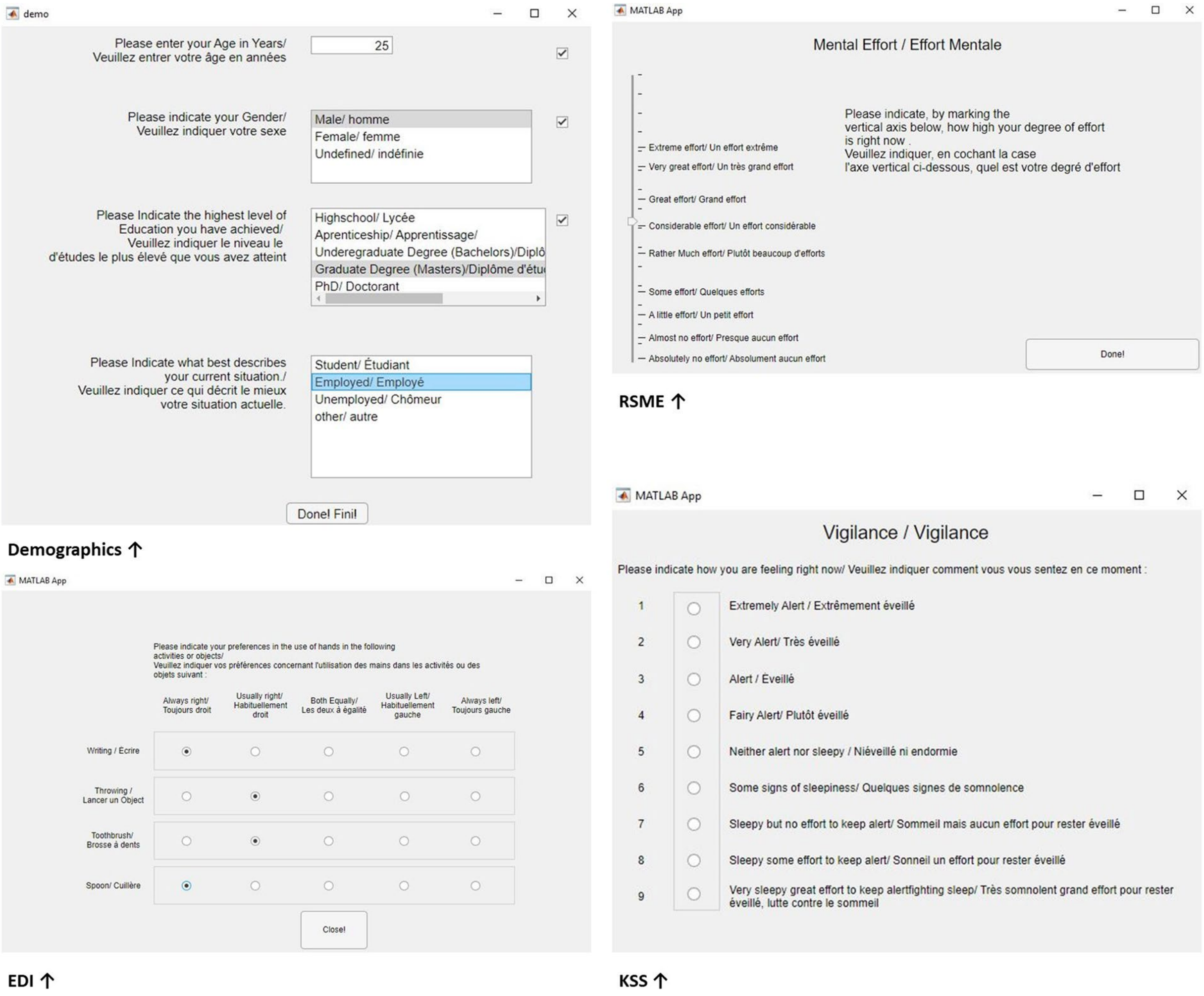
**Fig. 22** Results of the Mental Workload estimation: (**a**) Accuracy of estimation based on the alpha and theta band per task (i.e. MATB and N-back); (**b**) Accuracy of estimation based on the alpha and theta band per session (i.e. session 1, 2 or 3).: (**c**) Accuracy of the estimation based on the alpha and the theta band.

accuracy as a dependent variable showed significant main effects of congruency as well as session (Condition: F(2,28) = 20.79 $p < 0.005$; Session: F(2,56) = 10.33 $p < 0.005$; Fig. 14c). Participants' accuracy was significantly lower in incongruent trials. Their accuracy improved across sessions.

- **Cerebral activity**: Following the response of the participant, the event-related negativity ($ESN_{response}$) amplitude was sensitive to correctness, session and congruency (Correct: t(9941) = −3.51 $p = 0.001$; Session 1 - Session 2 t(9941) = −1.97 $p < 0.05$; Session 1 - Session 3 t(9941) = −5.24 $p < 0.001$; Session 2 - Session 3 t(9941) = 3.29 $p < 0.001$; Congruency t(9941) = 4.93 $p < 0.001$). Across sessions, the ERN value decreased. A contrast analysis of the interaction effects between Correct/Error and Incongruent/Congruent revealed several significant effects at the electrode site FCz: Incorrect incongruent - Correct congruent (t(9941) = 3.51 $p < 0.005$); Incorrect incongruent - Incorrect congruent (t(9941) = −4.9 $p < 0.001$); Incorrect Congruent - Correct Congruent t(9941) = −5.02 $p < 0.001$). Incongruent trials have a significantly lower amplitude during correct trials as compared to congruent correct trials.(Fig. 17a,c). There was also a significant effect of congruency and session at the Fz electrode (Session 1 - Session 2 t(9941) = 9.04 $p < 0.001$; Session 1 - Session 3 t(9941) = 9.55 $p < 0.001$; Congruency t(9980) = 3.11 $p < 0.005$). Contrast analysis showed the same effect as with electrode FCz: Incongruent trials had a significantly lower amplitude during correct trials as compared to congruent correct trials (Correct Congruent - Correct Incongruent t(9941) = −3.11 $p < 0.01$). The amplitude of session 1 was significantly lower as compared to the other sessions (Figs. 17b,d & 21).

  When participants received feedback for the preceding trial, the feedback-related negativity (FRN) showed to be affected by multiple factors. At electrode FCz Correctness, congruency as well as sessions had an effect on the amplitude of the FRN (Correct: t(9941) = −13.24 $p < 0.001$; Congruency: t(9941) = 2.95 $p < 0.005$; Session 1 - Session 2 t(9941) = −7.79 $p < 0.001$; Session 1 - Session 3 t(9941) = −12.42 $p < 0.001$; Session 2 - Session 3 t(9941) = 4.65 $p < 0.001$) Correct responses resulted in a lower amplitude, while congruency increased the amplitude. Across sessions, the amplitude decreased. (Fig. 18a,c). Fz also showed more negative amplitudes following a correct response (Correct - Error: t(9941) = −8.07 $p < 0.001$; Fig. 18a,b), while congruency increased the amplitude (Congruent - Incongruent: t(9941) = 2.46 $p < 0.05$). Across sessions the amplitude decreased (Session 1 - Session 2: t(9941) = −5.69 $p < 0.001$; Session 1 - Session 3 t(9941) = −10.29 $p < 0.001$; Session 2 - Session 3 t(9941) = 4.61 $p < 0.001$; Fig. 18b,d).

  The P300 amplitude during feedback was also seen to be affected by several conditions. The P300 amplitude at electrode Fz was influenced by session, congruency and accuracy(Correct - Error: t(9941) = −34.4 $p = 0.001$; Session 1 - Session 2: t(9941) = −6.72 $p < 0.001$; Session 1 - Session 3: t(9941) = −8.97 $p < 0.001$; Session 2 - Session 3: t(9941) = 2.26 $p < 0.05$; Congruent - Incongruent: t(9941) = 3.07 $p < 0.005$). The P300 component was significantly larger during incorrect trials and incongruent trials while decreasing across sessions (Figs. 19b,d & 20). The P300 component at electrode FCz was influenced by congruency, accuracy and session (Correct - Error: t(9941) = −12.0 $p = 0.001$; Session 2 - Session 3: t(9941) = −2.58 $p = 0.01$; Congruent - Incongruent: t(9989) = 3.02 $p < 0.005$). The amplitude was larger for incorrect trials, as well as for congruent trials. The amplitude during session 3 was significantly lower as compared to session 2 (Fig. 19a,c).



demo

Please enter your Age in Years/ Veuillez entrer votre âge en années: 25 ☑

Please indicate your Gender/ Veuillez indiquer votre sexe ☑
- Male/ homme (selected)
- Female/ femme
- Undefined/ indéfinie

Please Indicate the highest level of Education you have achieved/ Veuillez indiquer le niveau le d'études le plus élevé que vous avez atteint ☑
- Highschool/ Lycée
- Aprenticeship/ Apprentissage/
- Underegraduate Degree (Bachelors)/Diplô
- Graduate Degree (Masters)/Diplôme d'étu (selected)
- PhD/ Doctorant

Please Indicate what best describes your current situation./ Veuillez indiquer ce qui décrit le mieux votre situation actuelle.
- Student/ Étudiant
- Employed/ Employé (selected)
- Unemployed/ Chômeur
- other/ autre

Done! Fini!

**Demographics ↑**

MATLAB App

Mental Effort / Effort Mentale

Please indicate, by marking the
vertical axis below, how high your degree of effort
is right now .
Veuillez indiquer, en cochant la case
l'axe vertical ci-dessous, quel est votre degré d'effort

- Extreme effort/ Un effort extrême
- Very great effort/ Un très grand effort
- Great effort/ Grand effort
- Considerable effort/ Un effort considérable
- Rather Much effort/ Plutôt beaucoup d'efforts
- Some effort/ Quelques efforts
- A little effort/ Un petit effort
- Almost no effort/ Presque aucun effort
- Absolutely no effort/ Absolument aucun effort

Done!

**RSME ↑**

MATLAB App

Please indicate your preferences in the use of hands in the following activities or objects/
Veuillez indiquer vos préférences concernant l'utilisation des mains dans les activités ou des objets suivant :

| | Always right/ Toujours droit | Usually right/ Habituellement droit | Both Equally/ Les deux à égalité | Usually Left/ Habituellement gauche | Always left/ Toujours gauche |
|---|---|---|---|---|---|
| Writing / Ecrire | ● | ○ | ○ | ○ | ○ |
| Throwing / Lancer un Object | ○ | ● | ○ | ○ | ○ |
| Toothbrush/ Brosse à dents | ○ | ● | ○ | ○ | ○ |
| Spoon/ Cuillère | ● | ○ | ○ | ○ | ○ |

Close!

**EDI ↑**

MATLAB App

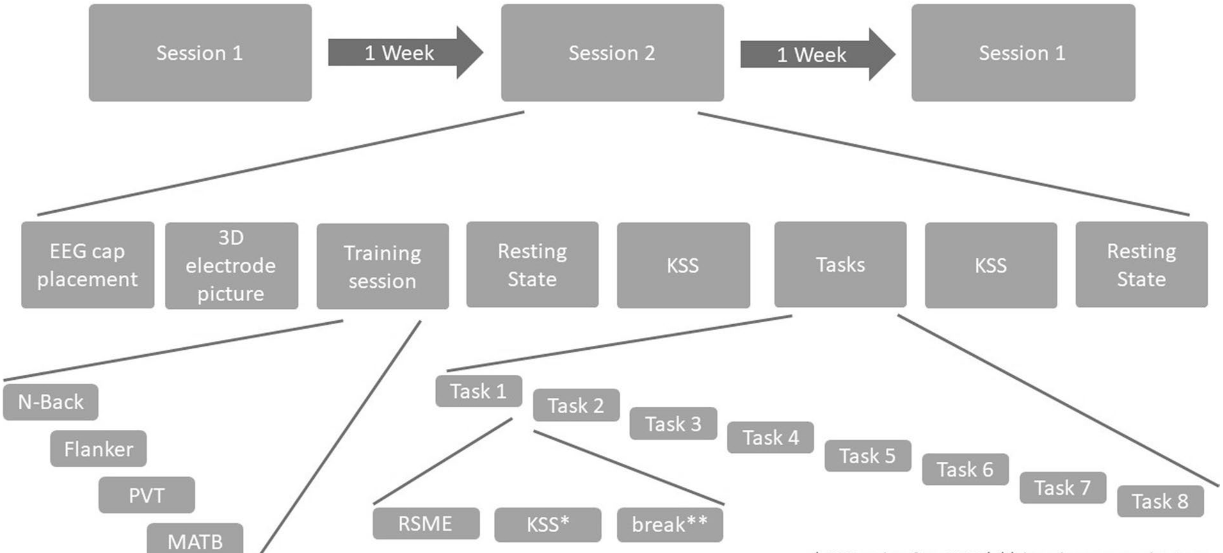
Vigilance / Vigilance

Please indicate how you are feeling right now/ Veuillez indiquer comment vous vous sentez en ce moment :

1 ○ Extremely Alert / Extrêmement éveillé
2 ○ Very Alert/ Très éveillé
3 ○ Alert / Éveillé
4 ○ Fairy Alert/ Plutôt éveillé
5 ○ Neither alert nor sleepy / Niéveillé ni endormie
6 ○ Some signs of sleepiness/ Quelques signes de somnolence
7 ○ Sleepy but no effort to keep alert/ Sommeil mais aucun effort pour rester éveillé
8 ○ Sleepy some effort to keep alert/ Sonneil un effort pour rester éveillé
9 ○ Very sleepy great effort to keep alertfighting sleep/ Très somnolent grand effort pour rester éveillé, lutte contre le sommeil

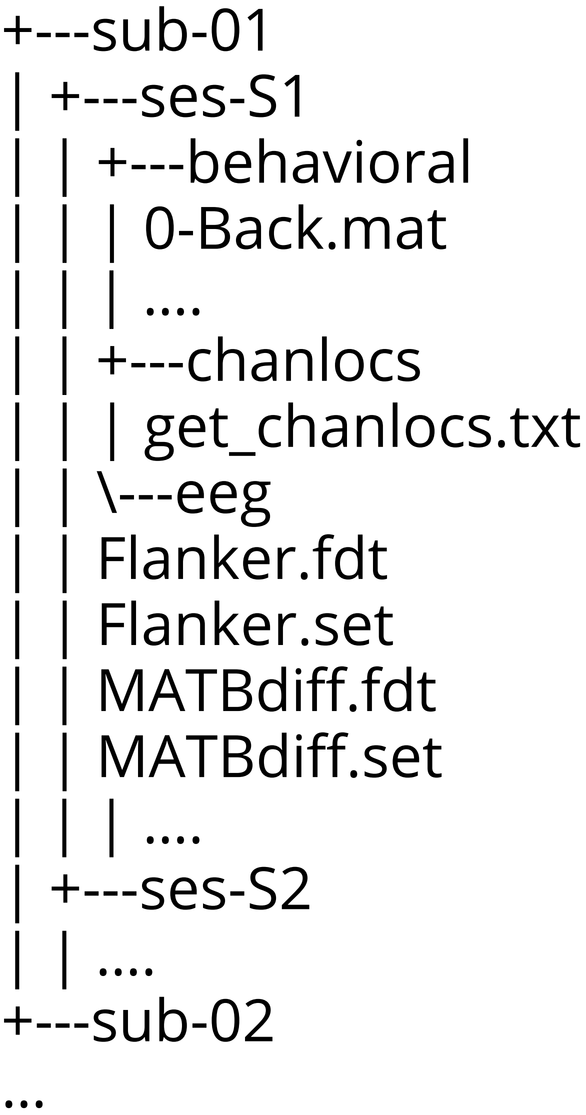
**KSS ↑**

**Fig. 23** Screenshots of the Matlab Questionnaires: Demographics, Karolinska Sleepiness Scale (KSS), Edinburgh Handedness Inventory (EDI), Rating Scale Mental Effort (RSME). The code is accessible in the **Code availability section**.



* KSS only after PVT | ** breaks were voluntary

**Fig. 24** Experimental Procedure, detailing each step for all three sessions.



```
+---sub-01
| +---ses-S1
| | +---behavioral
| | | 0-Back.mat
| | | ....
| | +---chanlocs
| | | get_chanlocs.txt
| | \---eeg
| | Flanker.fdt
| | Flanker.set
| | MATBdiff.fdt
| | MATBdiff.set
| | | ....
| +---ses-S2
| | ....
+---sub-02
...
```

**Fig. 25** Exemplary folder tree of the data in the BIDS structure.

Mental workload estimation. The estimation of mental workload using the Riemannian MDM classifier proved capable of detecting different levels of mental workload for both the MATB and the N-Back tasks. The RMANOVA with accuracy as a dependent variable and bandwidth (alpha or theta band), Session (1, 2, 3) and Task (MATB and N-BACK) showed several significant effects. With 69.40% (±12.50%), the accuracy for the MATB task was significantly higher than the accuracy of the N-Back (64.97 ± 12.99; F(1,28) = 13.00 $p < 0.001$; Fig. 22a). Using the data filtered in the alpha band resulted in significantly higher accuracies (68.41 ± 12.37%) as compared to the theta band (65.97 ± 13.37%; F(1,28) = 7.95 $p < 0.05$; Fig. 22a,c). The accuracies also changed across sessions (F(2,56) = 4.96 $p < 0.05$; Fig. 22b). Accuracy was highest during the second session (70.30 ± 12.73%) followed by the first session (64.30 ± 12.52%) with the lowest accuracy during the third session (66.96 ± 12.87). Multiple comparisons between the three tests showed a significant difference between session 1 and session 2 (t(28) = −3.14 $p < 0.01$). The interaction effect between Bandwidth and Session was also significant (F(2,56) = 3.50 $p < 0.05$; Fig. 22b). The interaction shows a more pronounced difference between classification performances of the alpha and theta band during the first session (Alpha Session 1 - Theta Session 1: t(28) = 3.67 $p < 0.001$). The strongest performance occurred with the alpha band during the second session (accuracy of 70.67%).

**Conclusion of the dataset validation.** In order to validate the acquired dataset, the tasks performed were assessed regarding their capability to elicit the well-known and expected effects on cognition. This validation occurred at three levels. Subjective levels of mental workload and vigilance decrement were assessed by the KSS and RSME subjective scales[28,29]. Behavioral performance for each task was obtained in the form of accuracies and reaction times. Physiological changes were assessed using cardiac (ECG) and cerebral (EEG) activity measurements.

Regarding the Psychomotor Vigilance Task, it was expected that over its 10-minute duration the vigilance of participants would decrease[21,32]. This was shown as reflected by the increase in reaction times as well as the

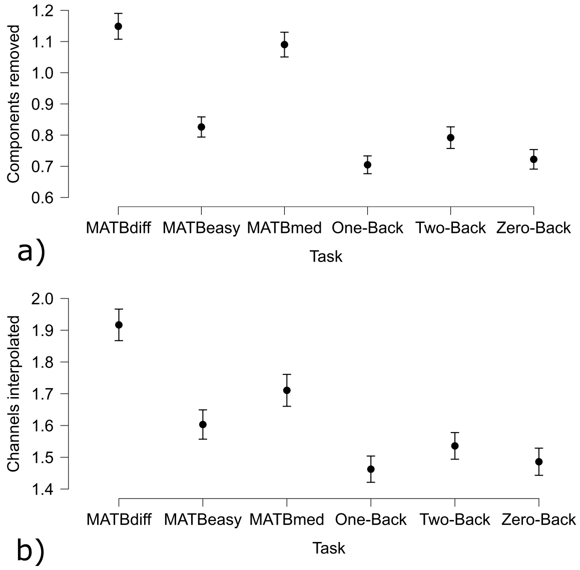


**Fig. 26** EEG data preprocessing for the mental state estimation example: (**a**) Average amount of removed components; (**b**) Average amount of interpolated channels.

increase in heart rate and decrease in HRV with increasing TOT. Concerning the cerebral activity, it was shown that slow EEG wave activity, including alpha, increases throughout the cortex as the person fatigues[52]. The results obtained here confirm this effect, namely the increase in alpha power with increasing TOT further confirms the vigilance decrement.

The N-Back task is a popular task for eliciting different levels of mental workload by taking working memory, while keeping visual stimuli as well as motor responses constant[20,33]. As shown here, differences in mental workload were observed at the subjective level with higher RSME scores (with increasing difficulty/mental workload), at the behavioral level with decreased performance with increasing difficulty, as well as at the physiological level with increased heart rates and decreased HRV during more difficult conditions. The current analyses revealed that there was an effect of the session on cerebral activity. Theta power increased with sessions in the frontal area. This result suggests that task engagement increased over the sessions[53] - in accordance with the increase in performance.

The MATB task represents an approach towards a more ecologically valid task for eliciting different levels of mental workload[35]. As with the N-Back task, the analysis showed increasing mental workload at the subjective, behavioral and physiological levels as difficulty increases. As with the N-Back, more difficult conditions resulted in a higher heart rate and a lower heart rate variability. The expected results concerning the modulation of brain activity during MATB was a peak of activity in the frontal theta band as the difficulty increased[54]. Regarding alpha power, a decrease in amplitude in the posterior sites was expected with increasing difficulty[54]. The cerebral analysis found an increase in theta power during more difficult conditions. While the alpha power was sensitive to changes in mental workload the results diverge from the traditionally seen decrease in alpha power with increasing difficulty. The opposite effect was observed here. However, behavioral and subjective analysis, as well as theta power, confirm different levels of mental workload across conditions of varying difficulty.

The Flanker task is a decision-making task that involves dealing with conflictual information[19]. Commonly it is observed, that incongruent trials result in less accurate and slower responses. On a cerebral level, it has been observed that errors result in stronger deflections of the ERN, FRN and P300 components. The behavioral results show a decrease in accuracy as well as an increase in reaction times for incongruent trials, as compared to congruent trials[47,55]. Furthermore, it was shown that errors resulted in significantly longer reaction times. The difference between erroneous and correct trials was also significant at the physiological level, as represented by the sensitivity of the ERN, FRN, and P300 components to errors sessions and congruency.

In addition to these measures of validation, the collected data was also tested in its ability to be used for mental workload estimation. On both the MATB and the N-Back task, an MDM Riemannian classifier was used to classify epochs of EEG data between three levels of mental workload. The obtained average accuracy of over 65% in this 3-class problem (more than 20% above the upper limit of a 95% CI for chance level) confirms that the COG-BCI database[18] may be of use for developing and testing classification pipelines. A higher accuracy was reached on the MATB-II data, which may be the result of an increased activity of the sensory-motor areas depending on task difficulty. In the N-back task, it is expected, that all conditions result in the same amount of motor activity.

Overall, the results presented here show that the COG-BCI database[18] is a useful asset for testing novel methods within passive BCI research or related fields. It should be noted that the analysis presented here does not in any form present an exhaustive analysis of all effects that may be present within the data as well as of all pipelines that may be designed to estimate users' cognitive state. The goal of the analyses performed here was to validate the usability of the collected dataset for working on passive BCIs and therefore to pave the way towards open science practices in pBCI research and development.

## Usage Notes

The behavioral output of the N-Back, Flanker and PVT tasks is saved in Tables. The behavioral output of the MATB is saved in a MATLAB structure, with individual substructures for each of the tasks.

- **TRACK:** 2 columns with the X and Y coordinates of the tracking tasks (2 Hz sampling rate).
- **SYSMON:** 2 columns referring to the onset of a specific alarm (column 1) and reaction time (column 2).
- **RESMAN:** 2 columns with the amount of fuel in the relevant reservoirs (1 Hz sampling rate).
- **COMM:** A table with 5 categories:
  - Target: Was the radio message a target (1 = yes)
  - TargetRadio: Which radio needed to be changed (1–4)
  - TargetFrequency: What was the target frequency
  - Reacted: Did the participant react to the radio message and change some radio frequency (1 = yes)
  - Correct Was the change in the radio frequency correct (1 = yes)

## Code availability

The dataset is only comprised of raw data, enabling researchers to work on their pipelines with their custom preprocessing steps for instance. Therefore, no code has been needed for generating these data. Regarding the technical validation, it was performed using Matlab and JASP software. The code for the Subjective Questionnaires is available on GitHub (https://github.com/Marcels-2-Neurons/Subjective_Questions/tree/main).

Received: 29 August 2022; Accepted: 14 December 2022;
Published online: 10 February 2023

## References

1. Hollnagel, E. (ed.) *Safer Complex Industrial Environments*, 0 edn (CRC Press, 2009).
2. Dekker, S. W. A. & Woods, D. D. MABA-MABA or Abracadabra? Progress on Human–Automation Co-ordination. *Cognition, Technology & Work* **4**, 240–244, https://doi.org/10.1007/s101110200022 (2002).
3. Berberian, B., Somon, B., Sahaï, A. & Gouraud, J. The out-of-the-loop Brain: A neuroergonomic approach of the human automation interaction. *Annual Reviews in Control* **44**, 303–315, https://doi.org/10.1016/j.arcontrol.2017.09.010 (2017).
4. Zander, T. O. & Kothe, C. Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *Journal of neural engineering* **8**, 025005, https://doi.org/10.1088/1741-2560/8/2/025005 (2011).
5. Aricò, P. *et al.* Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. *Frontiers in Human Neuroscience* **10**, 539, https://doi.org/10.3389/fnhum.2016.00539 (2016).
6. Dehais, F. *et al.* Dual passive reactive brain-computer interface: A novel approach to human-machine symbiosis. *Frontiers in Neuroergonomics* **7**, https://doi.org/10.1101/2021.12.23.473161 (2022).
7. Dehais, F., Lafont, A., Roy, R. & Fairclough, S. A neuroergonomics approach to mental workload, engagement and human performance. *Frontiers in neuroscience* **14**, 268, https://doi.org/10.3389/fnins.2020.00268 (2020).
8. Lotte, F. & Roy, R. N. Chapter 7 - Brain–Computer Interface Contributions to Neuroergonomics. In Ayaz, H. & Dehais, F. (eds.) *Neuroergonomics*, 43–48, https://doi.org/10.1016/B978-0-12-811926-6.00007-5 (Academic Press, 2019).
9. Fairclough, S. H. & Lotte, F. Grand challenges in neurotechnology and system neuroergonomics. *Frontiers in Neuroergonomics* 2, https://doi.org/10.3389/fnrgo.2020.602504 (2020).
10. Hinss, M. F., Somon, B., Dehais, F. & Roy, R. N. Open EEG Datasets for Passive Brain-Computer Interface Applications: Lacks and Perspectives. In *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, 686–689, https://doi.org/10.1109/NER49283.2021.9441214 (2021).
11. Hinss, M. F. *et al.* An eeg dataset for cross-session mental workload estimation: Passive bci competition of the neuroergonomics conference 2021 (version 2)[data set]. In Neuroergonomics Conference, Munich, Germany. *Zenodo*, https://doi.org/10.5281/zenodo.5055046 (2021).
12. Benaroch, C. *et al.* Long-term bci training of a tetraplegic user: Adaptive riemannian classifiers and user training. *Frontiers in Human Neuroscience* **15**, 635653, https://doi.org/10.3389/fnhum.2021.635653 (2021).
13. Roy, R. N., Bonnet, S., Charbonnier, S. & Campagne, A. Mental fatigue and working memory load estimation: Interaction and implications for EEG-based passive BCI. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6607–6610, https://doi.org/10.1109/EMBC.2013.6611070 (IEEE, Osaka, 2013).
14. Grosse-Wentrup, M. & Schölkopf, B. A review of performance variations in smr-based brain- computer interfaces (bcis). *Brain-Computer Interface Research* 39–51, https://doi.org/10.1007/978-3-642-36083-1_5 (2013).
15. Mühl, C., Jeunet, C. & Lotte, F. EEG-based workload estimation across affective contexts. *Frontiers in Neuroscience* **8**, https://doi.org/10.3389/fnins.2014.00114 (2014).
16. Pan, S. J. & Yang, Q. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* **22**, 1345–1359, https://doi.org/10.1109/TKDE.2009.191 (2010).
17. Shin, J. *et al.* Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset. *Scientific Data* **5**, 180003, https://doi.org/10.1038/sdata.2018.3 (2018).
18. Hinss, M. F. *et al. COG-BCI database:* A multi-session and multi-task EEG cognitive dataset for passive brain-computer interfaces. *Zenodo* https://doi.org/10.5281/zenodo.6874128 (2022).
19. Eriksen, B. A. & Eriksen, C. W. Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics* **16**, 143–149, https://doi.org/10.3758/BF03203267 (1974).
20. Brouwer, A.-M. *et al.* Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering* **9**, 045008, https://doi.org/10.1088/1741-2560/9/4/045008 (2012).
21. Lamond, N., Dawson, D. & Roach, G. D. Fatigue Assessment in the Field: Validation of a Hand-Held Electronic Psychomotor Vigilance Task. *Aviation, Space, and Environmental Medicine* **76**, 486–489 (2005).
22. Verdière, K. J., Roy, R. N. & Dehais, F. Detecting Pilot's Engagement Using fNIRS Connectivity Features in an Automated vs. Manual Landing Scenario. *Frontiers in Human Neuroscience* **12**, 6, https://doi.org/10.3389/fnhum.2018.00006 (2018).
23. Senoussi, M. *et al.* Pre-stimulus antero-posterior EEG connectivity predicts performance in a UAV monitoring task. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1167–1172, https://doi.org/10.1109/SMC.2017.8122770 (2017).

24. Zeng, H. *et al*. An EEG-Based Transfer Learning Method for Cross-Subject Fatigue Mental State Prediction. *Sensors* **21**, 2369, https://doi.org/10.3390/s21072369 (2021).
25. Zheng, W.-L. & Lu, B.-L. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Transactions on Autonomous Mental Development* **7**, 162–175, https://doi.org/10.1109/TAMD.2015.2431497 (2015).
26. Koelstra, S. *et al*. DEAP: A Database for Emotion Analysis;Using Physiological Signals. *IEEE Transactions on Affective Computing* **3**, 18–31, https://doi.org/10.1109/T-AFFC.2011.15 (2012).
27. Veale, J. F. Edinburgh Handedness Inventory - Short Form: a revised version based on confirmatory factor analysis. *Laterality* **19**, 164–177, https://doi.org/10.1080/1357650X.2013.783045 (2014).
28. Kaida, K. *et al*. Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology* **117**, 1574–1581, https://doi.org/10.1016/j.clinph.2006.03.011 (2006).
29. Ghanbary Sartang, A., Ashnagar, M., Habibi, E. & Sadeghi, S. Evaluation of Rating Scale Mental Effort (RSME) effectiveness for mental workload assessment in nurses. *Journal of Occupational Health and Epidemiology* **5**, 211–217, https://doi.org/10.18869/acadpub.johe.5.4.211 (2016).
30. Oldfield, R. C. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* **9**, 97–113, https://doi.org/10.1016/0028-3932(71)90067-4 (1971).
31. Åkerstedt, T. *et al*. Mental fatigue, work and sleep. *Journal of Psychosomatic Research* **57**, 427–433, https://doi.org/10.1016/j.jpsychores.2003.12.001 (2004).
32. Reifman, J., Kumar, K., Khitrov, M. Y., Liu, J. & Ramakrishnan, S. PC-PVT 2.0: An updated platform for psychomotor vigilance task testing, analysis, prediction, and visualization. *Journal of Neuroscience Methods* **304**, 39–45, https://doi.org/10.1016/j.jneumeth.2018.04.007 (2018).
33. Jaeggi, S. M., Buschkuehl, M., Perrig, W. J. & Meier, B. The concurrent validity of the N-back task as a working memory measure. *Memory* **18**, 394–412, https://doi.org/10.1080/09658211003702171 (2010).
34. Oberauer, K., Wilhelm, O., Schulze, R. & Süss, H.-M. Working memory and intelligence - their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin* **131**, 61–65, https://doi.org/10.1037/0033-2909.131.1.61 (2005).
35. Santiago-Espada, Y., Myer, R. R., Latorella, K. A. & Comstock, J. R. The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User's Guide. typeTech. Rep. numberL-20031 (2011). noteNTRS Affiliations: NASA Langley Research Center, Stinger Ghaffarian Technologies, Inc. (SGT, Inc.) NTRS Document ID: 20110014456 NTRS Research Center: Langley Research Center (LaRC).
36. Jasper, H. The ten-twenty electrode system of the international federation | CiNii Research. *Electroencephalogr. Clin. Neurophysio* **10**, 370–375, https://doi.org/10.1016/0013-4694(58)90053-1 (1958).
37. Homölle, S. & Oostenveld, R. Using a structured-light 3D scanner to improve EEG source modeling with more accurate electrode positions. *Journal of Neuroscience Methods* **326**, 108378, https://doi.org/10.1016/j.jneumeth.2019.108378 (2019).
38. Gorgolewski, K. J. *et al*. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data* **3**, 160044, https://doi.org/10.1038/sdata.2016.44 (2016).
39. Haans, A. Contrast Analysis: A Tutorial. *Practical Assessment, Research, and Evaluation* **23**, https://doi.org/10.7275/7dey-zd62 (2019).
40. Harbour, S., Christensen, J., Estepp, J. & Gray, T. An Integrated Neuroergonomic Assessment of In-Flight Pilot Workload. *17th International Symposium on Aviation Psychology* 621–626 (2013).
41. Labedan, P., Darodes-De-Tailly, N., Dehais, F. & Peysakhovich, V. Virtual reality for pilot training: study of cardiac activity. 81–88 (2021).
42. Shaffer, F. & Ginsberg, J. P. An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health* **5**, 258, https://doi.org/10.3389/fpubh.2017.00258 (2017).
43. Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M. & Robbins, K. A. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics* **9**, https://doi.org/10.3389/fninf.2015.00016 (2015).
44. Pion-Tonachini, L., Kreutz-Delgado, K. & Makeig, S. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage* **198**, 181–197, https://doi.org/10.1016/j.neuroimage.2019.05.026 (2019).
45. Simon, M. *et al*. EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions. *Clinical Neurophysiology* **122**, 1168–1178, https://doi.org/10.1016/j.clinph.2010.10.044 (2011).
46. Smith, M. R., Chai, R., Nguyen, H. T., Marcora, S. M. & Coutts, A. J. Comparing the Effects of Three Cognitive Tasks on Indicators of Mental Fatigue. *The Journal of Psychology* **153**, 759–783, https://doi.org/10.1080/00223980.2019.1611530 (2019).
47. Chavarriaga, R., Sobolewski, A. & Millán, J. D. R. Errare machinale est: the use of error-related potentials in brain-machine interfaces. *Frontiers in Neuroscience* **8**, https://doi.org/10.3389/fnins.2014.00208 (2014).
48. Nieuwenhuis, S., Holroyd, C. B., Mol, N. & Coles, M. G. H. Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience & Biobehavioral Reviews* **28**, 441–448, https://doi.org/10.1016/j.neubiorev.2004.05.003 (2004).
49. Barachant, A., Bonnet, S., Congedo, M. & Jutten, C. Multiclass Brain–Computer Interface Classification by Riemannian Geometry. *IEEE Transactions on Biomedical Engineering* **59**, 920–928, https://doi.org/10.1109/TBME.2011.2172210 (2012).
50. Barachant, A., Bonnet, S., Congedo, M. & Jutten, C. Riemannian Geometry Applied to BCI Classification. In Vigneron, V., Zarzoso, V., Moreau, E., Gribonval, R. & Vincent, E. (eds.) *Latent Variable Analysis and Signal Separation*, Lecture Notes in Computer Science, 629–636, https://doi.org/10.1007/978-3-642-15995-4_78 (Springer, Berlin, Heidelberg, 2010).
51. Mueller-Putz, G., Scherer, R., Brunner, C., Leeb, R. & Pfurtscheller, G. Better than random: A closer look on BCI results. *International Journal of Bioelectromagnetism* (2008).
52. Craig, A., Tran, Y., Wijesuriya, N. & Nguyen, H. Regional brain wave activity changes associated with fatigue. *Psychophysiology* **49**, 574–582, https://doi.org/10.1111/j.1469-8986.2011.01329.x (2012).
53. Berka, C. *et al*. EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks. *Aviation, space, and environmental medicine* **78**, 14 (2007).
54. Smith, M. E., Gevins, A., Brown, H., Karnik, A. & Du, R. Monitoring Task Loading with Multivariate EEG Measures during Complex Forms of Human-Computer Interaction. *Human Factors* **43**, 366–380, https://doi.org/10.1518/001872001775898287 (2001).
55. Chavarriaga, R. & Millan, Jd. R. Learning From EEG Error-Related Potentials in Noninvasive Brain-Computer Interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **18**, 381–388, https://doi.org/10.1109/TNSRE.2010.2053387 (2010).
56. Lotte, F., Bougrain, L. & Clerc, M. Electroencephalography (EEG)-Based Brain-Computer Interfaces. In *Wiley Encyclopedia of Electrical and Electronics Engineering*, 1–20, https://doi.org/10.1002/047134608X.W8278 (John Wiley & Sons, Inc., Hoboken, NJ, USA, 2015).

## Acknowledgements

This research was supported in part by the ERASMUS program (which funded Mr Hinss' internship), and by ANITI (Artificial and Natural Intelligence Toulouse Institute), Toulouse, France. The authors would also like to thank Dr S. Ladouce, Dr L. Darmet and Dr F. Lotte for their involvement in a previous related database and competition which fostered the energy for us to create the one presented here.

## Author contributions

Study design: M.H. and R.N.R.; Data acquisition: M.H., E.J., L.P. and B.S.; Data analysis: M.H. and E.J.; Drafting the article: M.H. and E.J.; Additions and critical revisions: B.S., F.D., R.N.R.; Supervision: R.N.R.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to M.F.H.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.