Brain Informatics

RESEARCH Open Access

Beyond the lab: real-world benchmarking of wearable EEGs for passive brain-computer interfaces

Ronca Vincenzo[1*], Cecchetti Marianna[1], Capotorto Rossella[2], Di Flumeri Gianluca[3], Giorgi Andrea[2], Germano Daniele[1], Borghini Gianluca[3], Babiloni Fabio[4,5] and Aricò Pietro[1]

Abstract

Purpose Wearable EEG systems are increasingly used for brain–computer interface (BCI) applications beyond controlled laboratory environments. However, there is still limited evidence on their reliability in real-world cognitive monitoring, especially for deriving robust mental-state indicators. This study investigates the signal quality, computational stability, and neurometric consistency of two widely used consumer-grade EEG devices (Emotiv EPOC X and Muse S) compared to a validated research-grade system (Mindtooth Touch) during naturalistic tasks relevant to passive BCIs and brain-machine intelligence.

Method Twenty-four participants completed a multimodal protocol including video observation, multitasking under varying cognitive loads, and a simulated driving task. Each participant used all three EEG systems in a counterbalanced order to avoid any bias induced by the order. Signal quality was assessed through artefact analysis and Power Spectral Density (PSD) stability. Neurometrics, i.e., metrics related to specific mental and emotional states that can be extracted from EEG signal processing (workload, attention, vigilance, and approach–withdrawal) were extracted and compared across devices, conditions, and subjective reports of effort and comfort.

Finding The research grade system demonstrated higher signal stability, fewer artefacts, and more consistent neurometric responses to cognitive variations, with high significant correlation with subjective measures. Post-processing improved data continuity in consumer devices, but neurometrics remained less sensitive to task demands and less aligned with subjective ratings. Each device reflected different trade-offs between data quality, usability, and cost.

Conclusion Research-grade systems remain more reliable for passive BCI applications requiring high-resolution cognitive state monitoring. Nevertheless, consumer-grade headsets may still be appropriate for exploratory studies or non-critical applications. This work highlights key trade-offs between signal quality, usability, and application goals, contributing to the broader integration of wearable neurotechnologies into brain–machine intelligence frameworks.

Keywords Biosignal processing, Human factors, Electroencephalography, Wearable EEG

*Correspondence:
Ronca Vincenzo
vincenzo.ronca@uniroma1.it

Full list of author information is available at the end of the article

1 Introduction

1.1 Background

Brain–Computer Interfaces (BCIs) have emerged as a transformative technology enabling novel forms of interaction between users and external systems, with applications in neuroergonomics, training, adaptive automation, assistive technologies, and beyond. Among BCI paradigms, passive BCIs are particularly promising, as they allow the implicit monitoring of the user's mental and emotional state without requiring intentional commands [1, 2]. These systems rely primarily on EEG signals to estimate key cognitive states such as mental workload, attention, vigilance, and affective engagement, providing actionable information to adapt systems, even in real time.

In this context, wearable EEG devices represent a crucial innovation. They offer improved usability, portability, and comfort compared to traditional laboratory EEG setups, making it possible to extend passive BCIs beyond the lab into realistic operational settings [3–6]. The availability of commercial wearable systems, ranging from research-grade to consumer-grade, has grown rapidly, with each offering different trade-offs between cost, usability, and signal quality.

1.2 Limitations of current technologies

Despite the increasing availability of wearable EEG devices, their effectiveness in naturalistic contexts remains poorly characterized. Wearable systems span a range of quality levels, from research-grade devices—typically more expensive but offering high performance—to consumer-grade alternatives that are more affordable and accessible. Research-grade systems often deliver performance comparable to conventional gel-based EEG setups in validation studies, incorporating dry or semi-dry electrodes, active signal amplification, and optimized processing pipelines, which are essential for extracting robust neurophysiological markers (i.e. neurometrics [7–12]) in real time.

In contrast, consumer-grade devices tend to prioritize ease of use, portability, and affordability, often featuring fewer electrodes, lower sampling rates, and simplified signal processing. These limitations raise concerns regarding their ability to reliably capture subtle variations in mental states. Most comparative studies in the literature focus on basic signal characteristics under resting-state conditions, without assessing how these systems perform in ecologically valid, cognitively demanding tasks—an essential consideration for the deployment of brain–machine intelligence systems in real-world environments.

Furthermore, most existing comparisons do not integrate subjective data or assess how signal fidelity affects the extraction of neurometrics used in adaptive cognitive monitoring systems. This gap limits the translational potential of consumer EEG for intelligent brain-aware applications, where accurate mental state tracking is crucial for machine adaptation. In this regard, recent comparative studies on wearable EEG devices have mainly focused on signal-centric metrics such as signal-to-noise ratio, power spectral density (PSD) stability, or event-related potential (ERP) robustness under tightly controlled laboratory conditions [13–15]. While these works demonstrated that consumer-grade systems can capture the main spectral components of the EEG, they also reported higher variability, reduced amplitude in ERP responses, and lower reliability across sessions when compared to medical- or research-grade systems. A more recent review [16] highlighted that motion- and electrode-related artefacts remain one of the main limitations of wearable EEG, especially when dry or saline sensors are used in dynamic or ecological contexts. In addition, studies assessing connectivity or higher-order neurometric indices are still scarce and limited to highly controlled resting-state paradigms [17]. Consequently, the reliability of wearable EEG in real-world and task-oriented scenarios—particularly concerning its ability to reproduce meaningful neurometric indicators (e.g., workload, engagement, attention) and their convergence with subjective experience—remains largely unexplored.

1.3 Motivation for EEG systems selection

To provide a meaningful comparison representative of both research- and consumer-grade EEG systems, we selected Emotiv EPOC X, Muse S, and Mindtooth Touch. The choice of EPOC X and Muse S was guided by their extensive adoption in the scientific literature and their complementary design approaches. A bibliometric review conducted over 916 studies found that approximately 68% of consumer-grade EEG applications employed Emotiv systems, whereas Muse-based devices accounted for around 8%, ranking second among commercial wearables [18–21]. A separate systematic review of wireless EEG studies reported similar findings, with Emotiv devices representing 24.6% of the total sample, followed by Muse among the most frequently used headsets [20, 22, 23]. Conversely, Mindtooth Touch was included as a research-grade reference, as it integrates active saline sponge electrodes and high-resolution front-end electronics, previously validated against gel-based laboratory systems for applications in neuroergonomics and human factors research [1, 24–26].

Together, these three devices capture the current landscape of wearable EEG—from low-cost consumer to high-fidelity research systems—offering a technically and practically representative benchmark for real-world passive-BCI applications.

To address the above-described gaps, the present study adopts an ecological benchmarking approach that simultaneously evaluates three representative EEG devices across four complementary axes: signal quality, temporal stability, neurometric discriminability, and alignment with subjective reports. A structured summary of previous comparative investigations and their methodological gaps is provided in Supplementary Table S1.

1.4 Objective of the study

This study aims to provide a comprehensive and systematic comparison of two wearable and widely used consumer-grade EEG systems—Emotiv EPOC X and Muse S—within the context of passive BCI applications, with respect to a research-grade EEG system (i.e., the Mindtooth Touch) chosen as reference. The Mindtooth Touch was selected as a reference device due to prior validation studies that demonstrated its ability to provide EEG signal quality comparable to traditional gel-based systems in passive BCI contexts [26]. The three systems considered in this work—Mindtooth Touch, Emotiv EPOC X, and Muse S—differ in design philosophy, technical specifications, and price, with each device optimized for different trade-offs between cost, usability, and signal fidelity. The evaluation focuses on the following key dimensions:

1. *Signal Quality* Evaluation of artefact incidence and the temporal stability of EEG signals during both resting and task-engaged states.
2. *Neurometric Reliability* Assessment of the robustness and consistency of cognitive indices (e.g., workload, attention, vigilance, approach–withdrawal) derived from each device across time and conditions.
3. *Responsiveness to Cognitive Demand* Examination of each system's ability to distinguish between experimental conditions involving different levels of task difficulty.
4. *Correlation with Subjective rating* Analysis of the alignment between EEG-derived metrics and participants' self-reported measures of effort, comfort, and perceived demand.

To address these goals, participants engaged in a multimodal experimental protocol including emotionally evocative video viewing, multitasking under varying difficulty levels, and immersive driving simulations. EEG data were analysed using validated neurometrics and complemented by subjective assessments to benchmark each system's performance in a realistic operational context. By combining objective signal quality assessments with neurometric validation and subjective self-reports, this study contributes to the field of brain informatics by informing the development and deployment of wearable EEG systems for passive BCIs and computational models of cognitive state in real-world scenarios.

Table 1 Summary of the technical specifications associated to each of the three wearable EEG systems considered in the study

	Emotiv EPOC X	Mindtooth Touch	Muse S
Weight	170 g	158 g	41 g
Sample rate	2048 internal downsampled to 128 SPS or 256 SPS (user configured)	125, 250 or 500 Hz	256 Hz
Sample Resolution	14 bits or 16 bits	24 bits	14 bits
Battery Life	Up to 12 h using USB receiver, up to 6 h using Bluetooth Low Energy	Up to 4 h	10 h
Electrodes	Saline soaked felt pads	Salt-water sponge-based	Silver thread fabric (20% silver content, 80% polyamide fabric (nylon))
Num. Of Channel	14	8	4
Electrode sites	AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4	AFz,AF3, AF4, AF7, AF8, PZ, P3, and P4	AF7, AF8, TP9, and TP10
Ref	CMS/DRL references at P3/P4; left/right mastoid process alternative	Mastoid	FPz
Starting price (Official website)	~999,00 $	~8.500,00 €	~379,99 €

2 Material and methods

2.1 EEG devices included in the study

The three devices under analysis—EPOC X (*Emotiv*), Mindtooth Touch (*Brain Products GmbH*), and Muse S (*InteraXon*)—are described below in relation to the objectives of the present study. Pairwise comparisons were conducted between the Mindtooth Touch–Emotiv EPOC X and Mindtooth Touch–Muse S configurations, with the Mindtooth Touch serving as the reference device. This choice is motivated by prior validation studies demonstrating that the Mindtooth Touch achieves research-grade signal quality, comparable to traditional gel-based EEG systems [26].

A summary of the main technical specifications of the devices is provided in Table 1.

The Emotiv EPOC X (*Emotiv Inc., San Francisco, California, USA*) is equipped with 14 acquisition channels, with electrodes positioned on specific points of the cerebral cortex. The electrodes are located on F3, F4, AF3, AF4, F7 and F8 to monitor the neural activity of the frontal lobe of the subject's brain; T7, T8, FC5 and FC6 for the temporal lobe, while the signals from the parietal lobe are acquired through the electrodes positioned on P8 and P7. The EEG signals are recorded using saline-soaked felt sensors, which ensure good quality of contact with the skin. The system includes CMS/DRL references located on P3/P4, with the possibility of also using the left and right mastoid processes as alternative references, depending on the experimental needs. As for data acquisition, the device records data with an internal sampling rate of 2048 Hz, which is then reduced to 128 or 256 samples per second (SPS), depending on the user-configurable settings. The signal resolution is 14 bit (0.51 μV) or 16 bit (0.1275 μV), depending on the selected mode. The system has a bandwidth between 0.16 and 43 Hz, with digital notch filters at 50 Hz and 60 Hz to reduce network noise. The device is powered by an internal 595 mAh lithium polymer battery, which allows up to 12 h of autonomy when used with the USB receiver and up to 6 h in Bluetooth Low Energy mode. The device measures 9 × 15 × 15 cm and weighs 170 g, making it compact.

The Mindtooth Touch headset (*Brain Products GmbH, Gilching, Germany*) is an advanced wearable EEG system designed to monitor brain activity in real time in a comfortable and precise way. The device consists of two main parts, anterior and posterior, connected to each other by an adjustable elastic band that allows an optimal adaptation to the shape of the user's head. The frontal part integrates five EEG sensors (AFz, AF3, AF4, AF7, AF8), positioned on the forehead area. All electrodes and bases for the sponge electrode holders are incorporated into the structure of the device, minimizing invasiveness. The posterior section houses three parietal EEG sensors (P3, Pz, P4) and the housing for the amplifier module. The reference and mass sensors are placed on the mastoid processes. The electrodes are mounted on flexible arms with extended rotating segments, designed to ensure stable contact with the scalp, adapting to different cranial conformations. The system supports up to 8 active EEG channels, in addition to the reference and ground electrodes. EEG data can be sampled at 125, 250 or 500 Hz, with a resolution of 24 bits per sample. The bandwidth varies according to the sampling frequency: from 0.1 to 33 Hz (at 125 Hz), up to a maximum of 131 Hz (at 500 Hz). Data transmission occurs via Bluetooth® Low Energy (BLE) version 5.0. The device is powered by two lithium button batteries (3.7 V, 240 mAh), which ensure lightness and autonomy for prolonged sessions. The physical interface with the amplifier module is ensured by a 10-contact magnetic connector and 3 magnets. The compact size (53 × 53 × 15 mm) and low weight (158 g) facilitate its use even in field research contexts.

The Muse S (*InteraXon, Ontario, Canada*) features 4 EEG channels and 4 amplified auxiliary channels, with a sampling rate of 256 Hz and a recently improved resolution of 14 bits per sample. The EEG electrodes, made of conductive fabric with 20% silver content and 80% polyamide (nylon), are located at TP9, AF7, AF8, and TP10, with the reference electrode at FPz (CMS/DRL). The electrodes are dry and use silver fibre technology, which allows for signal collection without the need for conductive gel, facilitating daily and extended use. The central module is made of polycarbonate plastic with a soft-touch finish, equipped with integrated magnets and glued using specially selected glues for assembly. The device is fixed to a soft and adjustable band. The compact size (6 × 3 × 1.5 cm) and low weight (approximately 41 g) make this EEG system particularly suitable for night-time use.

2.2 Experimental protocol design

The experimental sample consisted of 24 healthy participants (11 males and 13 females), aged between 24 and 28 years, with no prior diagnosis of psychiatric or neurological disorders, in order to avoid interference related to physiological or pathological conditions. Recruited on a voluntary basis, all participants held a category B driving licence, considered essential to ensure familiarity with real-world driving scenarios. Prior to the experiment, each participant received detailed information about the study's objectives and procedures and signed an informed consent form authorising the recording of electroencephalographic (EEG) activity during the execution of simulated tasks in a controlled operational environment. All participants authorised the use of their data exclusively for scientific purposes and the publication of results in anonymised form. The experiments were conducted following the principles outlined in the 1975 Declaration of Helsinki, as revised in 2000. Experiments were approved by the Ethical Committee of the Sapienza University of Rome (1807/2020).

The experimental protocol (Fig. 1) consists of a sequence of tasks specifically designed to evaluate the quality of the devices in realistic contexts (Case Study 1, Case Study 2 and Case Study 3). Upon arrival, participants were asked to complete an informed consent form, authorizing the anonymous use of their data for research purposes. This was followed by a training phase lasting approximately 7 min, during which they had the opportunity to familiarize themselves with the realistic tasks. This phase was essential to ensure a proper understanding of the subsequent experimental activities. In the initial phase of the protocol, two acquisitions lasting one minute each are planned, during which the subject first keeps his eyes open (EO) and then his eyes closed (EC). The EC condition was included for the calculation of the individual alpha frequency (IAF) [30]. The EO condition

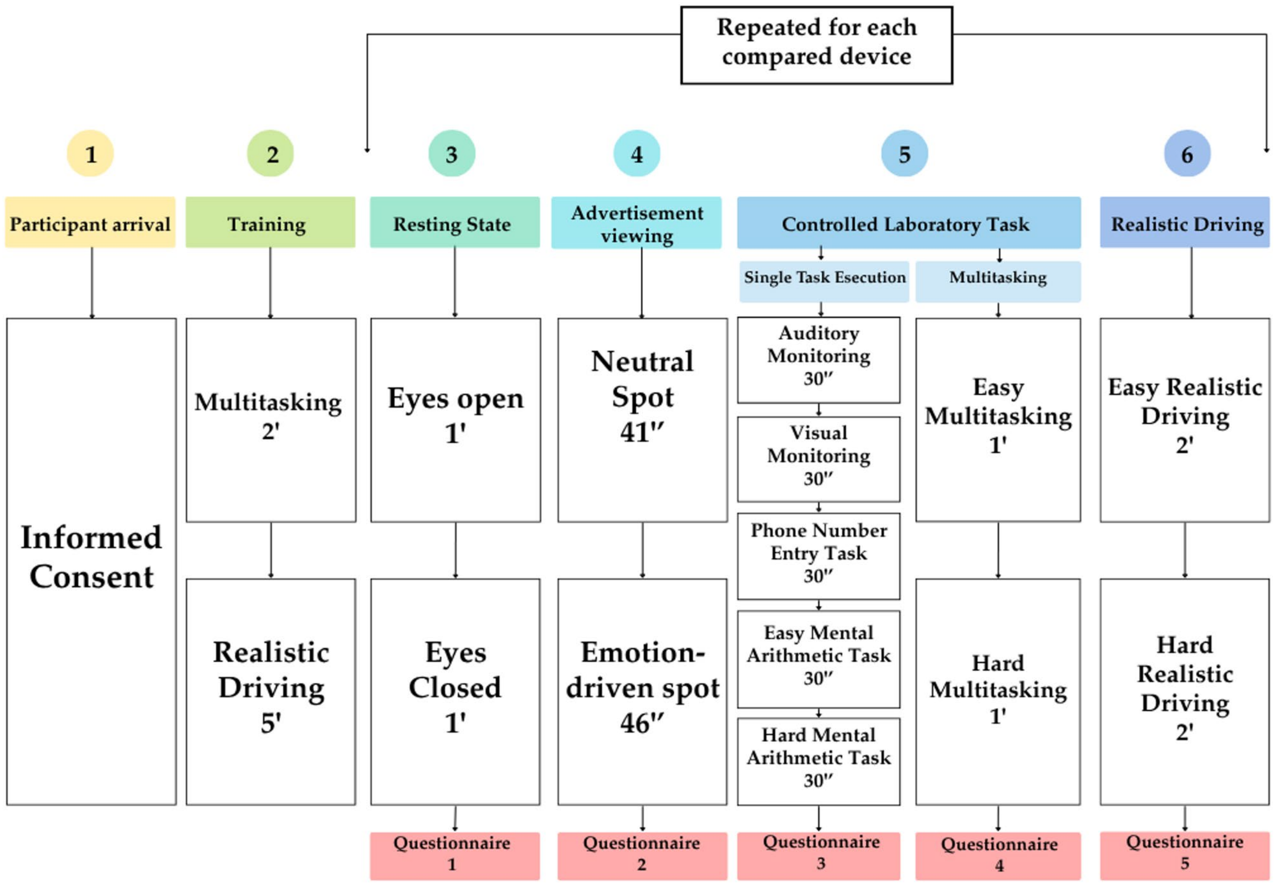


Fig. 1 Summary of the experimental protocol design

is used to compare the three devices in terms of spectral characteristics [31].

2.2.1 Case study 1: advertisement viewing

Wearable EEG devices are widely used in the neuromarketing sector [32–34], as they allow analysing users' neurophysiological responses to emotional stimuli in real time. In line with this perspective, the experimental protocol adopted in this study included a section dedicated to the viewing of video content designed to evoke different emotions, with the aim of investigating the variations in brain activity associated with different affective conditions. Regarding the audiovisual materials employed, the neutral video used consisted of a 41-s excerpt from a documentary about the city of Pietraperzia (Italy). This content, which had previously been validated in the E-Movie study, was classified as neutral, exhibiting a low level of arousal and valence values close to the mean. The emotional video, on the other hand, was a 46-s excerpt from the music video Firestone by Kygo. It was selected because it presented similar characteristics to those of the high valence and high arousal videos included in the DEAP dataset (Dataset for Emotion Analysis Using Physiological Signals) [35], widely recognized in literature for the classification of audiovisual stimuli based on the emotional responses evoked. It was preferred not to use the DEAP videos directly because, being dated material, they offered a visual resolution that is too low, potentially inadequate for current perceptual standards and therefore less effective in emotional stimulation. The videos were viewed using a computer, with the participant sitting on a chair in a comfortable position, in order to ensure maximum relaxation and stability during the exposure phase to the visual stimuli.

2.2.2 Case study 2: controlled laboratory task

This phase of the experimental protocol initially involved the separate execution of cognitive tasks, followed by their simultaneous execution in two multitasking conditions, which differed in terms of difficulty level. The tasks selected for this phase was as follows:

1. Mental calculation (top left): Participants had to solve addition operations by entering the results using a numeric keypad. As the difficulty increases, the number of digits and the presence of carryover in the operations increases.

2. Auditory monitoring (top right): In this phase, subjects had to identify a target tone emitted between two tons of different frequencies, presented at regular intervals. Increasing difficulty manifests itself in a greater similarity between the target tone and the distractor tone.
3. Visual monitoring (bottom left): Participants had to reset a horizontal fill bar as soon as it reaches the end. As the difficulty increased, the speed of filling the bar increased.
4. Telephone number entry task (bottom right): Subjects were asked to enter a sequence of numbers on a numeric keypad. Increasing difficulty was associated with an increase in the number of digits to be entered.

Participants were initially asked to complete each of the four tasks separately, with a duration of 30 s for each task (auditory monitoring, visual monitoring, entering a phone number, and mental calculation in the easy and difficult versions). Subsequently, the subjects performed the tasks simultaneously during a multitasking phase, divided into two levels of increasing difficulty, with a duration of 1 min for each level.

It has been demonstrated in previous studies that the simultaneous execution of multiple cognitive tasks involves a significant increase in mental workload compared to managing a single task, requiring greater activation of cognitive resources.

2.2.3 Case study 3: realistic driving

The objective of this phase of the experimental protocol was to evaluate the performance of the three wearable EEG devices during a simulated driving task in a realistic and immersive context. The simulator used replicated the typical configuration of a car, including a steering wheel, gear shift, and a pedal set with clutch, brake, and accelerator. The participant was seated on a non-rolling chair positioned in front of a desk. On the desk, the steering wheel and three 27-inch monitors were installed in a panoramic arrangement, recreating a highly immersive virtual driving environment. This type of task was well-suited for analysing variations in cognitive load, attention, and stress levels—key parameters for assessing the reliability and sensitivity of EEG devices in complex operational scenarios. The driving phase was divided into two sessions, each lasting two minutes. The first one, of lower difficulty, included the route called "Fiera di Roma", during which the subject was asked to respect the road signs and the speed limits imposed (about 40 km/h), simulating ordinary driving in an urban environment. The second session, characterized by higher complexity, used the "Fiera di Roma – Circuit" route and did not include speed limits, thus increasing the cognitive demand and time pressure. This distinction allowed us to analyse the impact of the difficulty of the driving context on the recorded neurophysiological activity.

2.3 Subjective data

At the end of each block of the protocol, participants were asked to complete a questionnaire—written in Italian—aimed at collecting subjective data related to their experience. The questionnaire consisted of three questions, each of which required a response on a VAS (Visual Analogue Scale) scale from 1 to 10, where 1 indicated the minimum value and 10 the maximum value. The first question was aimed at assessing the level of comfort perceived during the use of the EEG device, and it was formulated as follows: "*What is your level of comfort regarding the device you are wearing?*". The next two questions were instead aimed at measuring the psychophysical state of the subject and the effort required by the task performed: "*How tired do you feel right now?*" and "*How much effort did you put into completing the last task?*". It has to be underlined that, even if reported in English for reading purposes, all the questionnaires were provided in Italian (i.e., the native language of the involved participants). These subjective measurements were systematically collected at the end of each of the following phases: recording in resting state conditions, viewing of the two commercials, single task execution, multitasking session, and realistic driving with increasing difficulty. Their analysis allowed us to evaluate, in a complementary manner to the EEG data, the cumulative effect of the experimental protocol and the perceived usability of the tested wearable devices.

2.4 EEG data recording and processing

Each participant repeated the experimental protocol three times, wearing each of the analysed devices. To avoid a possible bias deriving from the sequence of assembly of the devices, the order of use was randomized among the participants. The EEG signal was acquired using the three abovementioned EEG headset: Emotiv EPOC X, Mindtooth Touch and Muse S, with sampling rates of 256 Hz, 250 Hz and 256 Hz, respectively. To standardize the data and allow direct comparison between devices, all signals were resampled to 512 Hz. Before starting the EEG signal analysis, a pre-processing phase was performed to identify and correct eventual physiological and non-physiological content unrelated to the cerebral activity of interest (i.e., ocular, muscular, and movement-based artifacts). In this regard, the EEG signal was band-pass filtered with a 5th-order Butterworth filter in the interval 2–30 (Hz). The eye blink artefacts were detected and corrected online by employing the o-CLEAN method [36, 37]. For further sources of artefacts, such as the ones deriving by muscular activity and movements, ad-hoc algorithms based the EEGLAB toolbox [38] were applied. More specifically, two statistical criteria were applied to the 1-s-long pre-processed EEG signal. Firstly, EEG epochs with the signal amplitude exceeding ± 120 μV were marked as "artefacts". Secondly, such EEG epochs marked as "artefacts" were removed from the EEG dataset to obtain an artefacts-free EEG dataset. Once artifact-free signals were obtained, the Global Field Power (GFP) [39, 40] was calculated for each frequency band. The GFP represents an index of the degree of spatial synchronization of cortical activity and constitutes a preliminary step for the calculation of specific neurophysiological indices of mental state, such as Stress, Workload, Attention, Vigilance and Approach Withdrawal.

The frequency bands definition was performed individually, on the basis of the Individual Alpha Frequency (IAF), computed for each participant during the recording in the closed-eye condition, the phase in which the alpha peak typically occurs. In accordance with the IAF value, the bands were defined as follows:

- Theta: [IAF – 6 ÷ IAF – 2] Hz
- Alpha Low: [IAF – 2 ÷ IAF] Hz
- Alpha High: [IAF ÷ IAF + 2] Hz
- Beta Low: [IAF + 2 ÷ IAF + 11] Hz
- Beta High: [IAF + 11 ÷ IAF + 16] Hz.

2.5 Performed analyses

In order to best evaluate the potential application of the devices under examination in real-world passive BCI contexts, the analysis was structured within four main sections: *EEG Signal Comparisons*; *EEG Spectral Features Comparisons: Neurometrics*; *Passive BCI Monitoring*; and *Subjective Perceptions*. For all the performed statistical analyses, to reduce the probability of false positives due to multiple comparisons, the Benjamini–Hochberg false-discovery-rate (FDR) correction (q = 0.05) was applied within logically defined families of analyses: i) artefact-rate comparisons, ii) PSD correlations, iii) condition-discrimination tests on neurometrics, and iv) repeated-measures correlations between physiological and subjective variables. Therefore, the adjusted p-values were reported alongside effect sizes. To control for potential effects related to the sequential use of the three EEG headsets (Mindtooth Touch, Emotiv EPOC X, Muse S), the order of device use was randomized across participants. In addition, a post-hoc analysis was performed to verify whether learning, fatigue, or sequence-related factors could have influenced the results. Specifically, for each neurometric (Mental Workload, Attention, Vigilance, and Approach–Withdrawal), a linear mixed-effects model (LME) [41] was fitted including *Device*, *Condition*, and *Order* (i.e., first, second, or third device used) as fixed effects, and Subject as a random intercept. The *Device* × *Order* interaction was examined to assess whether device performance varied as a function of testing order. No significant main or interaction effects involving *Order* were found (all $p > 0.20$), indicating that neither learning nor fatigue biased the observed results. To further verify this outcome, a sensitivity check restricted to each participant's first session confirmed the same pattern of significance.

2.5.1 EEG signal comparisons

The initial comparison involved a technical analysis between the Mindtooth Touch device (reference) and the two consumer grade devices, Emotiv Epoc X and Muse S. This analysis was conducted exclusively for technical purposes, with the aim of evaluating the instrumental performance of the devices under examination.

In the first phase, for each device the percentages of artifacts detected on all available channels were calculated, with the aim of estimating the overall quality of the measurements. Subsequently, a paired-sample t-test was performed, aimed at comparing the percentages of error recorded on the electrodes shared between the reference device and each of the other two devices analysed.

In the second phase, the percentage of artifacts detected across all available channels was calculated for each device, in order to estimate the overall quality of the recorded signals. For each participant and each experimental condition, artifact percentages were averaged across electrodes. Subsequently, a statistical analysis was performed using repeated-measures ANOVA, to compare the mean artifact percentages, computed as channel averages, for each device.

Furthermore, the eyes-open (EO) resting-state condition was analysed for each subject and each electrode shared across the Mindtooth Touch–Emotiv EPOC X and Mindtooth Touch–Muse S pairs. In this condition, the EEG signal was segmented into two equal temporal halves. For each subject and each shared electrode, the Power Spectral Density (PSD) was computed separately for the first and second halves of the EO signal.

As an initial step, intra-device correlations were computed between the power spectral densities (PSDs) of the first and second halves of the EEG recording obtained with the Mindtooth Touch. This analysis was performed separately for each participant and electrode, providing a reference measure of temporal spectral stability under a controlled condition. The eyes-open (EO) resting state, used as the baseline condition, is characterized by minimal cognitive and behavioral fluctuations; thus, the EEG signal—and its corresponding spectral profile—is expected to remain stable over time. High correlation values between the two halves would therefore indicate a consistent signal, whereas notable deviations from this pattern may suggest the presence of temporal instabilities or artefactual noise in the recording.

Subsequently, inter-device correlations were computed by comparing the PSDs from the first half of the EO resting-state signal recorded with the Mindtooth Touch and the second half recorded with either the Emotiv EPOC X or the Muse S, for the same participant and electrode. This approach enabled the assessment of signal similarity and temporal coherence across devices, under comparable experimental conditions, while accounting for intersession variability.

To statistically compare the intra-device (reference) and inter-device correlations, paired-sample t-tests were conducted for each shared electrode, testing whether the correlations between Mindtooth Touch and each consumer device significantly differed from the Mindtooth Touch's own internal consistency.

Additionally, to assess the intrinsic temporal stability of the consumer-grade devices, intra-device PSD correlations were computed for both the Emotiv EPOC X and Muse S, using the same split-half procedure previously applied to the Mindtooth Touch. These values were then statistically compared to the Mindtooth-derived reference correlations using paired t-tests. This analysis aimed to evaluate whether each device, independently of inter-device comparisons, was capable of maintaining consistent spectral features over time during a condition—namely, the eyes-open resting state—in which EEG dynamics are expected to remain relatively stable.

2.5.2 EEG spectral features comparisons: neurometrics

In many applications outside the laboratory context passive Brain-Computer Interfaces (BCIs) represent a promising technology for the continuous monitoring of the user's psychophysiological state. Such systems allow to extract information related to different human factors, enabling the development of solutions aimed at improving operational performance and safety, especially in real and dynamic scenarios.

In this study, with the aim of comparing the performance of three wearable EEG devices in the context of passive BCI, four neurometrics, widely used in research fields on pBCIs were considered: Vigilance, Mental Workload, Attention and Approach Withdrawal (AW). The term neurometric is used to indicate the quantitative measure of each human factor, obtained through a specific combination of the Global Field Power (GFP) values, calculated in the different EEG frequency bands.

Vigilance Such a mental state can be defined as the ability to maintain a constant level of attention and reactivity over time. A physiological decrease in Vigilance is commonly associated with a progressive degradation of performance, evidenced by an increase in reaction times and a reduction in situational awareness. On the contrary, maintaining an adequate level of activation during the entire execution of the task is essential to guarantee stable and optimal performance [42, 43]. The processes related to Vigilance mainly involve the right inferior frontal brain regions.

$$Vigilance = FrontalBeta_{GFP}^{right}$$

Mental Workload (MWL) It quantifies the mental resources employed during a task and is correlated with performance [44–46]. It is defined as the ratio between frontal activity in the theta band and parietal activity in the alpha band [47, 48].

$$MentalWorkload = \frac{FrontalTheta_{GFP}}{ParietalAlpha_{GFP}}$$

Approach Withdrawal (AW) Since wearable devices in realistic contexts are widely used in advertising research, the Approach Withdrawal (AW) index was calculated to quantify the type of motivation (degree of approach/avoidance) towards the two spots included in the protocol. This index, which was initially developed by Davidson et. al in 1990 [49], is based on hemispheric asymmetries in the alpha frequency range on prefrontal electrodes and represents a widely accepted correlation between motivation related to approach and avoidance [50–53].

$$AW = FrontalAlpha_{GFP}^{right} - FrontalAlpha_{GFP}^{left}$$

Attention is defined as the set of cognitive processes responsible for the selection, filtering and processing of relevant stimuli, in order to produce behavioural responses appropriate to the context.

$$Attention = -FrontalAlpha_{GFP}$$

However, considering that the aim of the analysis was to compare the measurements obtained by the two devices under examination with respect to the reference device (i.e., Mindtooth Touch), the formulas relating to the calculation of the neurometrics were appropriately adapted by taking into account only the electrodes that were common between the different configurations. Specifically, the AF3 and AF4 electrodes were used for the comparison between Emotiv EPOC X and Mindtooth Touch, while the AF7 and AF8 electrodes for the comparison between Muse S and Mindtooth Touch. In particular, as regards the neurometrics relating to the Mental Workload, the formula was modified by excluding the denominator, as some electrodes necessary for the complete calculation were not available on both devices.

After calculating the metrics in the temporal domain for each subject, the data were normalized with respect to the baseline, corresponding to the eyes-open condition. All analyses and comparisons involving the use of neurometrics were therefore performed on the normalized data. Subsequently, a moving average filter was applied with a 5, 10, 15, and 30 s-length window to the neurometrics. The application of moving averages with progressively larger windows was employed for a reduction of missing values along the neurometrics time series because of the artifacts, as removed epochs are compensated within the averaging window. In other words, by including a greater number of samples, the averaging operation 'absorbs' the artifacts, preventing the formation of 'gaps' in the neurometric time series. Given these considerations, the use of a moving average can be regarded as a potentially effective strategy to reduce the impact of artefacts. It therefore becomes important to determine the minimum temporal resolution—i.e., the smallest averaging window—needed to achieve an artefact rejection rate comparable to that of a research-grade system in the resulting neurometric time series. To further characterize the effect of the smoothing-window length on signal robustness and neurometric consistency, we systematically quantified how different averaging windows (0, 5, 10, 15, and 30 s) affected both the residual artefact rate and the correlation between each consumer-grade device and the reference Mindtooth Touch. This analysis was conducted using the same datasets employed for the main comparisons and included all participants.

Furthermore, the percentages of artifacts associated with neurometrics in each experimental condition were calculated for each device, with the aim of ensuring not only the reliability of the results, but also of evaluating the quality of the metrics themselves. Given that the Neurometric has a temporal resolution of one second, all missing values resulting from the removal of invalid data within the analysed one-second interval—such as those caused by noise or signal loss—were defined as artifacts. To this end, statistical comparisons were carried out using paired-sample t-tests, comparing the percentages of artifacts detected by using the Mindtooth Touch device with those obtained, respectively, by using the Emotiv EPOC X and the Muse S, in each condition.

Taking into account the conditions in which the metrics could be better appreciated, comparisons were performed through correlation analyses between the metrics obtained from the reference device, the Mindtooth Touch, and those recorded using the Emotiv EPOC X. Similarly, correlations were also carried out between the metrics recorded with the Mindtooth and those obtained with the Muse S.

2.5.3 Passive BCI monitoring

In this case, for each subject and for each normalized metric, the mean relative to the different experimental conditions was calculated. The comparisons were conducted separately for each pair of devices: Mindtooth Touch–Emotiv EPOC X and Mindtooth Touch–Muse S. For each comparison, a Friedman ANOVA was performed, considering the following pairings of conditions: neutral spot vs emotional spot, math easy vs math hard, multitasking easy vs multitasking hard and driving easy vs driving hard. The 'device' factor was inserted as a between-subjects variable, in order to evaluate, for each pair, the ability of wearable EEG systems to discriminate between the different experimental conditions.

2.5.4 Subjective perceptions

At the end of the various experimental conditions (viewing videos, performing a single task, multitasking, and simulated driving), as well as following the initial resting phase (with eyes open and eyes closed), participants were asked to fill a questionnaire aimed at collecting a subjective evaluation of perceived comfort, the effort required to perform the task, and the level of fatigue experienced at the end of each condition. Statistical analyses were conducted on the data related to perceived comfort. Since the collected data did not meet the assumptions of normality, a non-parametric Friedman ANOVA was performed, which is appropriate for ordinal or non-normally distributed data. This analysis allowed for the assessment of whether statistically significant differences in perceived comfort existed across the different experimental conditions, as well as, consequently, over time and according to the device used. With regard to subjective measures of effort, the values of mental workload neurometrics obtained from each device were subsequently compared with the corresponding scores reported in the questionnaires. For each device, the level of effort was normalized with respect to the values recorded during the baseline phase, i.e., those reported by participants at the end of the initial EO and EC conditions. The average Workload value was then calculated for each experimental condition (commercials, single task, multitasking, driving) with the aim of verifying its consistency with the subjective assessments provided through the questionnaire. The initial comparison between the neurometric values of Mental Workload and the subjective evaluations of effort was conducted using a Friedman ANOVA, aiming to identify any significant differences across the various experimental conditions and between devices.

A further analysis was conducted by calculating the repeated measures correlations, comparing the subjective scores of efforts with the Mental Workload values recorded by the three devices considered: Mindtooth Touch, Emotiv EPOC X and Muse S. This technique allows to estimate the common association within each individual between two variables measured on two or more occasions, avoiding the violation of the assumption of independence of observations, a typical violation of simple correlations or traditional regressions. In the graphs resulting from this analysis, each line represented the estimated correlation for each participant, while the coloured dots indicated the distribution of the data relative to the single subject [54]. Through this analysis, the relationship between subjective indices and neurophysiological measures was examined, both at the individual and group level. For each participant, four values were considered, corresponding to the four experimental conditions, excluding the baseline, used exclusively for normalization. The analysis focused on the MWL neurometric, as it is the most coherent and sensitive with respect to the subjective perceptions of effort.

3 Results

The results have been organized in four different subsections, according to the main sections constituting the performed analysis described in the previous paragraph, to facilitate the reader. All the results are here shortly described, pointing out the relevant findings, while they will be commented in a more comprehensive view along the *Discussion* section.

3.1 EEG signal comparisons

The results presented in this subsection were obtained following an analysis aimed at comparing the devices mainly from a signal quality point of view, through the direct examination of the EEG signals recorded by each device.

The comparison of artefact percentages on shared electrodes was carried out for each pair analysed. For the Mindtooth Touch–Emotiv EPOC X pair, the comparison of artefact percentages, calculated for each condition, on the shared electrodes was conducted using the paired-samples statistical test. The analysis showed that, in most cases, no statistically significant differences were found between the two devices with regard to the percentage of artefacts recorded on electrodes AF3 and AF4. However, while performing the driving tasks the percentage of artifact identified for Emotiv EPOC X on AF3 and AF4 channels was significantly higher compared to the one computed for Mindtooth Touch (Easy driving condition, AF3: $p = 0.014$,; Easy driving condition, AF4: $p = 0.048$; Hard driving condition, AF3: $p = 0.001$; Hard driving condition, AF4: $p = 0.025$). Furthermore, a significant difference was also found in the single-task Easy Math condition, where the anterofrontal EEG channels of the Emotiv EPOC X were affected by a higher artifact percentage (AF3: $p = 0.036$; AF4: $p = 0.021$).

A similar statistical analysis was performed for the Mindtooth Touch-Muse S pair. In that case, no statistically significant differences were observed between the artifact percentages calculated for each experimental condition on shared electrodes between Mindtooth Touch and Muse, i.e. AF7 and AF8. However, the Muse S showed extremely high artefact percentage values on channels TP9 and TP10. The statistical comparison with the Mindtooth Touch was nonetheless carried out exclusively on the shared electrodes.

An additional analysis was also conducted on the percentage of artefacts detected on the electrodes. The values were averaged across electrodes for each participant and for each experimental condition. The result of the statistical analysis, performed using Friedman ANOVA and aimed at comparing the mean artefact percentages among the three devices, showed that the higher averaged percentage of artifacts was observed for Muse S device, while the averaged percentage of artifacts identified for Emotiv EPOC X device was statistically higher than the one identified for the Mindtooth Touch ($F = 3.97$, $p < 0.001$, $\eta^2 = 0.26$) (Fig. 2).

As can be observed from the above representation, the statistical test revealed a significant difference between the Muse S and the Emotiv EPOC X ($p < 0.001$), as well as between the Muse S and the Mindtooth Touch ($p < 0.001$).

Finally, a further statistical analysis was performed by considering the reference correlations, i.e., PSD of the first half of the Mindtooth Touch EEG data – PSD of the second half of the Mindtooth Touch EEG data, both recorded during the eyes open condition, and the correlations between the PSD of the first half of the EEG data recorded with the Mindtooth Touch and the PSD of the second half of the EEG data recorded with the Emotiv EPOC X on OA data. In this specific case, it was observed a significant difference for the AF4 channel ($p < 0.001$). Similarly, the statistical comparison between the Mindtooth Touch and Muse S in terms of spectral temporal consistency during the eye open condition was performed. Statistically significant differences emerged for both considered channels (AF7: $p < 0.001$; AF8: $p < 0.001$) (Fig. 3).

Finally, to verify whether such temporal consistency was also present within the EEG signal portion collected through the same device, a statistical comparison was carried out between the correlation referred to the reference device (i.e., the PSD computed on the Mindtooth Touch EEG signal) and the respective correlation associated with the PSD computed on the EEG signal collected through the Emotiv EPOC X and Muse S devices. According to the previous analysis, in this case the first and second half of the eyes open condition was considered. The results revealed that, for both pairs of shared

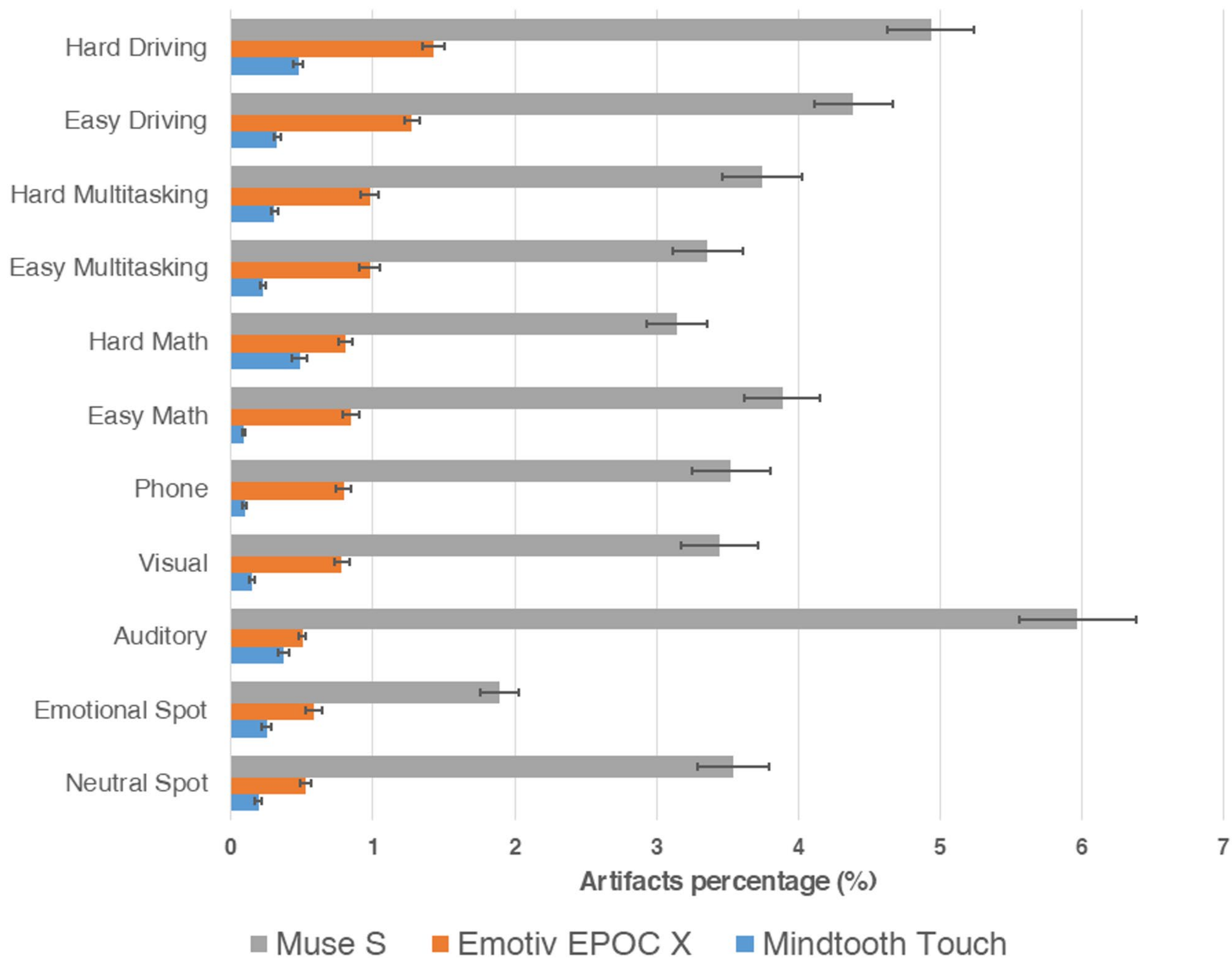


Fig. 2 Artifacts percentage computed for each considered EEG device and for each experimental condition. The values were averaged across all the EEG channels

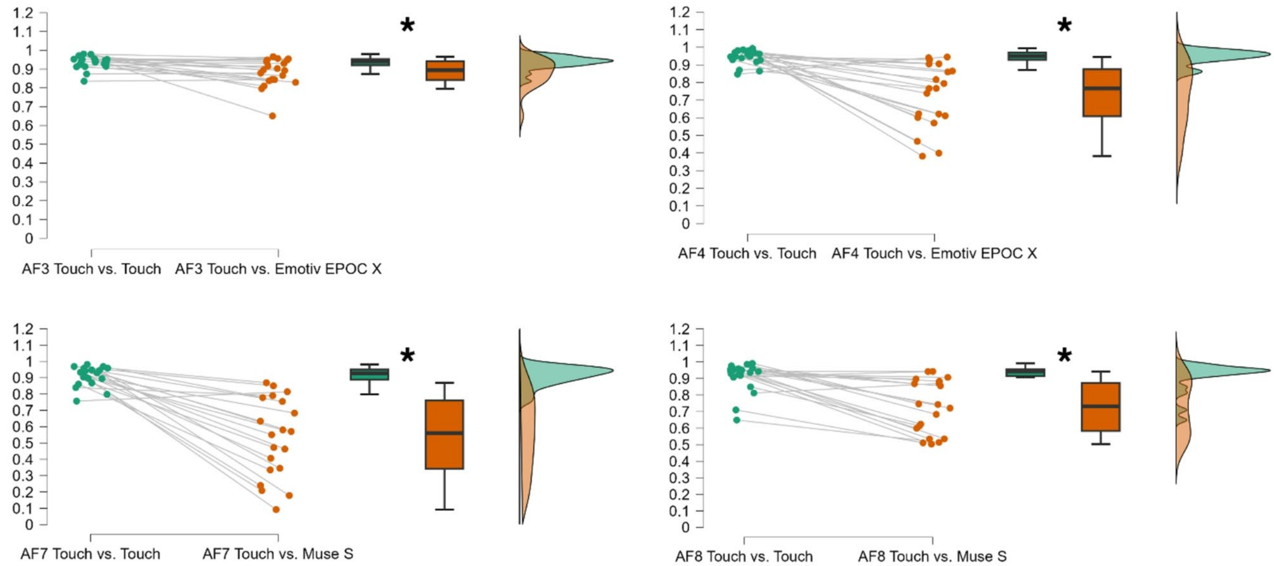
electrodes (i.e., AF3 and AF4 when considering the Mindtooth Touch-Emotiv EPOC X, and AF7 and AF8 when considering Mindtooth Touch-Muse S), the Mindtooth Touch exhibited a statistically higher temporal consistency (all $p < 0.01$) (Fig. 4).

3.2 Spectral features comparison: neurometrics

The percentages of artefacts identified along the pre-processed EEG signal are directly affecting the neurometrics time series. In this analysis, the missing values percentages occurring in the neurometrics time series (i.e., deriving from the artifacts occurrences along the pre-processed EEG signal) were considered to compare the three EEG systems included in the study. The first statistical analysis revealed that the artifact percentages was significantly higher for both the Emotiv EPOC X and the Muse S compared to the Mindtooth Touch (all $p < 0.01$). In relation with the moving average application described in *Material and Methods* section, a second statistical analysis was performed to estimate the minimum moving average window for obtaining a non-significant p-value when comparing the considered gold-standard (i.e., the Mindtooth Touch) and each of the Emotiv EPOC X and Muse S.

From Table 2 it can be observed that, when considering the comparison between the Mindtooth Touch and the Emotiv EPOC X, the minimum temporal resolution for which the p-value was non-significant was, for most of the experimental conditions, 10 s. A similar result emerged when comparing the Mindtooth Touch and the Muse S, where the minimum temporal resolution beyond which the artefact percentages were statistically non-different was 10 s. However, larger moving average window size (i.e., 15 s) was required during more dynamical experimental tasks, as the driving ones.

An additional analysis was performed to compute the Pearson's correlations between the Mindtooth Touch-Emotiv EPOC X and Mindtooth Touch-Muse S pairs in terms of neurometrics. The results showed, for most of the cases, low significant correlation values (Table 3).



I)

S)

Fig. 3 Intra-device and inter-device Pearson's correlations computed by comparing the PSD from the: i) first and second half of the EO signal recorded with the Mindtooth Touch, and ii) first half of the EO signal recorded with the Mindtooth Touch with that from the second half recorded with either the Emotiv EPOC X or the Muse S

Finally, a complementary analysis was conducted to explicitly quantify how smoothing affected both artefact rate and inter-device neurometric correlation across all smoothing windows. Supplementary Figure F1 provides an overview of these effects for the three EEG systems. Panel A shows that the Mindtooth Touch retained consistently low artefact levels across all windows, while the Emotiv EPOC X and Muse S exhibited pronounced improvements in signal robustness as the window increased to 10–15 s. Panel B illustrates that inter-device neurometrics correlations with the Mindtooth Touch followed a similar trend, with the Emotiv EPOC X showing higher convergence than the Muse S. Together, these results highlight the device-specific trade-off between temporal responsiveness and robustness, establishing an empirical basis for selecting smoothing parameters in consumer-grade EEG applications.

3.3 Passive BCI monitoring

The results obtained from the Friedman ANOVA, aimed at evaluating the ability of the devices under examination to discriminate between experimental conditions, showed that, in the absence of the application of the moving average, the Mindtooth Touch was the only device to statistically significantly discriminate among experimental conditions coherently with the experimental hypotheses. In contrast, the other two devices, without moving averaging, exhibited unsatisfactory performance. Overall, the Emotiv EPOC X showed better performance than the Muse S, particularly after the application of the moving average operator. All the details associated with the presented statistical analysis are reported in Table 4.

3.4 Subjective perceptions

The statistical analysis performed on the subjective ratings along all the experimental conditions demonstrated that the comfort associated with the Emotiv EPOC X was statistically lower than the one associated with the Muse S, but not with respect to the Mindtooth device, that is positioned in the middle. No significant drops in terms of perceived comfort were found during the experiment progression for all three devices (Fig. 5).

Furthermore, a repeated measure statistical analysis was performed in order to evaluate the coherence between the subjective and physiological assessments. More specifically, the effort subjective ratings and its respective MWL neurometrics were compared across all the experimental conditions.

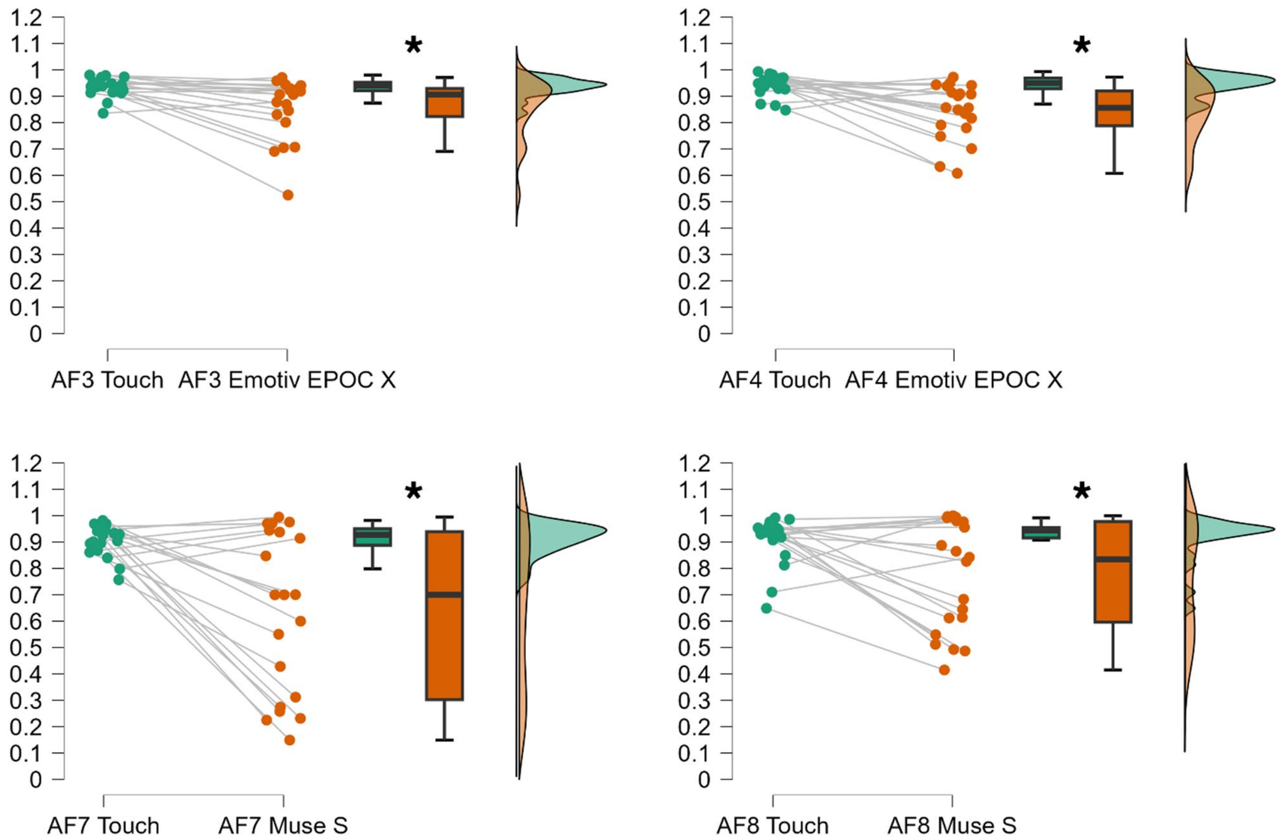


Fig. 4 Intra-device PSD Pearson's correlations computed within the first and second half of the EO signal recorded with the Mindtooth Touch, Emotiv EPOC X and Muse S

As illustrated in Fig. 6a, the subjective ratings indicated a significantly greater effort perception during the multitasking phase, compared to the other conditions. This outcome was perfectly consistent with what was found when assessing the MWL neurometric evaluated through the Mindtooth Touch (Fig. 6d) (all $p < 0.02$). With regard to the Emotiv EPOC X and Muse S devices, the statistical increase in terms of subjective effort perception was not confirmed by the MWL neurometrics computed through such devices, i.e., no statistical differences during the multitasking experimental phase in terms of EEG-based MWL (Fig. 6e and 6f).

Additionally, the repeated measure correlation analyses were performed between the EEG-based MWL index evaluated through the three considered devices and the effort subjective ratings. The results revealed that the MWL measurements performed through Mindtooth Touch showed a greater temporal coherence with the subjective perception compared to the ones evaluated through the Emotiv EPOC X and Muse S devices (Mindtooth Touch: $R = 0.84$, $p < 0.001$; Emotiv EPOC X: $R = 0.57$, $p = 0.002$; Muse S: $R = 0.33$, $p = 0.12$) (Fig. 7). Finally, a control analysis confirmed that neither learning nor fatigue effects influenced the main outcomes. No significant *Device* × *Order* interactions were found in any neurometric or condition, and repeating the analyses considering only the first recording session per participant yielded comparable results.

4 Discussion

This study presents a systematic comparison between two widely used consumer-grade wearable EEG systems—Emotiv EPOC X and Muse S—and a research-grade system, the Mindtooth Touch, in the context of passive BCI applications under ecologically valid conditions. The aim was to assess whether popular consumer devices, originally designed for accessibility and ease of use, can meet the performance standards required for more demanding cognitive monitoring scenarios. The evaluation covered signal quality, neurometric reliability, responsiveness to changes in mental states, and alignment with subjective reports of effort and usability. The results highlight distinct performance profiles that reflect the devices' differing design purposes and technical capabilities, offering practical guidance for selecting EEG systems based on application-specific requirements and constraints.

4.1 Comparison with prior work

Our findings complement and extend earlier comparative studies that typically evaluated headset performance

Table 2 Summary of the results emerged from the analysis for identifying the minimum temporal resolution for obtaining a non-significant p-value when comparing the considered gold-standard (i.e., the Mindtooth Touch) and each of the Emotiv EPOC X and Muse S

Neurometric	Experimental condition	Minimum resolution Mindtooth Touch-Emotiv EPOC X	Minimum resolution Mindtooth Touch-Muse S
Vigilance	Auditory	10 s	10 s
	Visual	5 s	10 s
	Phone	10 s	5 s
	Easy Math	10 s	10 s
	Hard Math	5 s	5 s
	Easy Multitasking	10 s	10 s
	Hard Multitasking	10 s	10 s
	Easy Driven	10 s	15 s
	Hard Driven	10 s	15 s
Mental Workload	Auditory	10 s	10 s
	Visual	5 s	10 s
	Phone	10 s	5 s
	Easy Math	10 s	10 s
	Math hard	5 s	5 s
	Easy Multitasking	10 s	10 s
	Multitasking hard	10 s	10 s
	Easy Driven	10 s	15 s
	Hard Driven	10 s	15 s
Attention	Neutral Spot	5 s	10 s
	Emotional Spot	5 s	5 s
	Auditory	10 s	10 s
	Visual	5 s	10 s
	Phone	10 s	5 s
	Easy Math	10 s	10 s
	Hard Math	5 s	5 s
	Easy Multitasking	10 s	10 s
	Hard Multitasking	10 s	10 s
	Easy Driven	10 s	15 s
	Hard Driven	10 s	15 s
Approach Withdrawal	Neutral Spot	5 s	5 s
	Emotional Spot	10 s	5 s

Table 3 Summary of the results emerged from the Pearson's correlations analysis between the Mindtooth Touch-Emotiv EPOC X and Mindtooth Touch-Muse S pairs in terms of neurometrics

Comparison	Neurometric	Without Moving Average (R)	5 s-Moving Average (R)	10 s-Moving Average (R)	15 s-Moving Average (R)	30 s-Moving Average (R)
Mindtooth Touch-Emotiv EPOC X	Vigilance	0.043	0.066	0.089	0.099	0.076
	MWL	0.041	0.065	0.095	−0.007	0.042
	Attention	0.007	0.007	−0.002	0.110	0.123
	AW	0.015	0.022	0.027	0.020	0.034
Mindtooth Touch-Muse S	Vigilance	0.011	0.046	0.085	0.011	0.088
	MWL	0.008	0.027	0.064	0.041	0.060
	Attention	0.009	0.018	0.037	0.081	0.094
	AW	0.011	0.018	0.021	0.016	0.027

on resting-state or simple ERP tasks [20, 23]. Consistent with those reports, the research-grade system displayed higher spectral stability and lower artefact incidence. Yet, our ecological design revealed additional divergences: during multitasking and driving conditions, consumer-grade devices exhibited larger drops in PSD consistency and weaker neurometric discrimination between cognitive-load levels. These effects appear primarily linked

Table 4 Summary of the results emerged from the analysis performed for identifying the ability of each considered device in discriminating among the proposed experimental conditions

Neurometric	Task	Device	Results	
			Without Moving Average	15-s Moving Average
MWL	Math easy vs. Math hard	Mindtooth Touch	–	$p=0.06$
		Emotiv EPOC X	–	–
		Muse S	–	–
	Multitasking easy vs. hard	Mindtooth Touch	$p=0.02$, W = 0.26	–
		Emotiv EPOC X	–	–
		Muse S	–	–
	Driving easy vs. hard	Mindtooth Touch	$p=0.04$, W = 0.19	$p=0.02$, W = 0.29
		Emotiv EPOC X	–	–
		Muse S	–	–
Attention	Math easy vs. Math hard	Mindtooth Touch	–	$p=0.006$, W = 0.39
		Emotiv EPOC X	–	–
		Muse S	–	–
	Multitasking easy vs. hard	Mindtooth Touch	–	$p=0.02$, W = 0.27
		Emotiv EPOC X	–	–
		Muse S	–	–
	Driving easy vs. hard	Mindtooth Touch	$p=0.04$, W = 0.20	$p=0.002$, W = 0.44
		Emotiv EPOC X	–	$p=0.1$, W = 0.11
		Muse S	–	–
Vigilance	Multitasking	Mindtooth Touch	Visual	Visual
		Emotiv EPOC X	–	Visual
		Muse S	–	–
	Multitasking easy vs. hard	Mindtooth Touch	–	$p=0.04$, W = 0.19
		Emotiv EPOC X	–	–
		Muse S	–	–
	Driving easy vs. hard	Mindtooth Touch	$p=0.07$, W = 0.14	$p=0.01$, W = 0.27
		Emotiv EPOC X	–	$p=0.09$, W = 0.12
		Muse S	–	–
AW	Video spot (Emotional vs. Neutral)	Mindtooth Touch	$p=0.02$, W = 0.28	$p<.001$, W = 0.48
		Emotiv EPOC X	–	
		Muse S	$p=0.1$, W = 0.10	$p=0.002$, W = 0.45

to motion sensitivity and reduced electrode coverage, aspects that were rarely tested in previous studies. By jointly analysing signal-level, neurometric, and subjective data, this study bridges the gap between traditional laboratory validations and applied real-world EEG use cases. In fact, EPOC X and Muse S are among the most prevalent consumer wearables in the literature [14, 55], and Mindtooth Touch is validated for research-grade operation [7, 24]. This ensured that our conclusions would be generalizable to devices that practitioners actually use in ecological monitoring and that observed performance gaps reflect category-level trade-offs rather than idiosyncratic choices.

4.2 Signal quality and temporal stability

One of the main challenges in out-of-the-lab EEG applications is maintaining stable and clean signal acquisition in the presence of movement and environmental variability, so limiting the number of artifacts coupled to the signal. Across all experimental conditions, Mindtooth Touch demonstrated superior signal quality, with significantly lower artefact incidence compared to both Emotiv EPOC X and Muse S. This difference became particularly evident in more dynamic tasks, such as the driving simulation, where the consumer devices exhibited an increased number of artefactual segments, likely due to less stable electrode contact and higher sensitivity to motion artefacts.

Temporal stability was assessed through intra-device and inter-device correlations of the EEG power spectral density (PSD) computed on separate temporal segments. The Mindtooth Touch consistently showed higher internal consistency, with strong correlations between the first and second halves of the same recording, a proxy for signal stationarity. In contrast, the Muse S and Emotiv EPOC X devices yielded significantly lower correlations both within and across devices, suggesting that their recordings are more affected by transient artefacts or less stable signal acquisition. These results underscore the necessity of high signal fidelity for reliable EEG-based cognitive monitoring in uncontrolled environments.

4.3 Reliability of the neurometrics

The ability to extract meaningful neurophysiological indices from EEG data is a core requirement for passive BCI applications. This study examined four key neurometrics: mental workload, attention, vigilance, and approach-withdrawal. Mindtooth Touch provided robust neurometric time series with low proportions of missing data due to artefact rejection. In contrast, Emotiv EPOC X and Muse S exhibited significantly higher loss of valid data, especially during physically or cognitively demanding tasks.

The application of moving average filters was explored as a strategy to mitigate the impact of artefacts. A temporal smoothing window of at least 10 s was necessary for Emotiv EPOC X and Muse S to approach the artefact resistance of Mindtooth Touch. However, this approach reduces temporal granularity, making these devices less suitable for applications requiring real-time

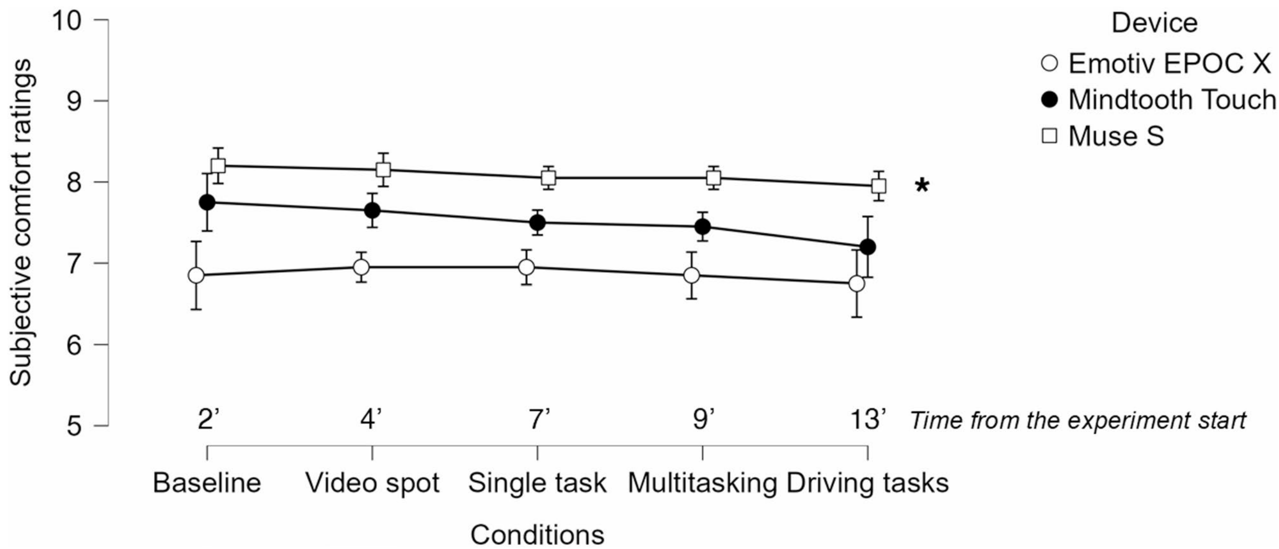


Fig. 5 Subjective comfort ratings provided by the participants along the experimental session when wearing each of the considered EEG systems

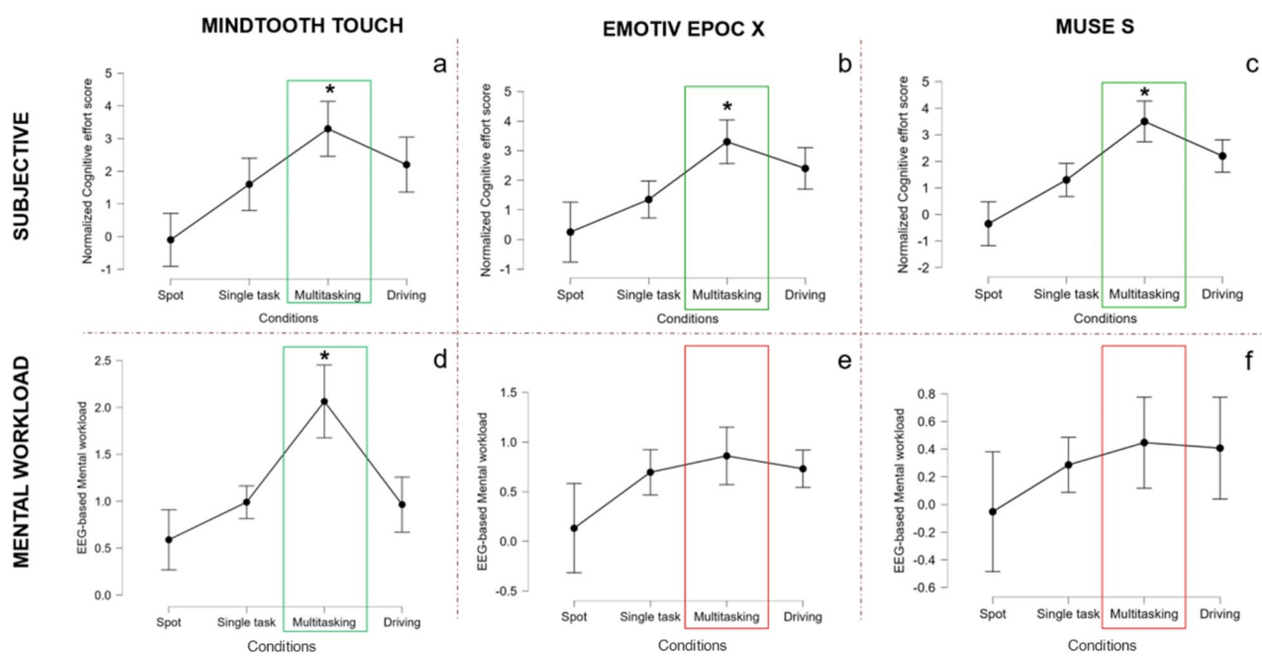


Fig. 6 Representation of the concordance between the neurophysiological and subjective Mental Workload-related measurements for each of the three considered EEG systems

or fine-grained tracking of cognitive states. Furthermore, even after filtering, neurometrics derived from Emotiv EPOC X and Muse S showed low correlation with those extracted from the reference device. This lack of convergence suggests that, beyond noise, differences in hardware design (e.g., electrode type, impedance control, channel placement) affect the fidelity of the extracted cognitive markers. Moreover, to ensure a fair comparison across devices with different electrode layouts, all neurometrics were computed using only the electrodes shared between the reference and each consumer-grade system (AF3–AF4 for EPOC X, AF7–AF8 for Muse S). This approach allowed maintaining methodological consistency while isolating the contribution of hardware differences to overall performance. In addition, the systematic evaluation of smoothing-window duration (Supplementary Figure F1) revealed that performance stabilization occurred within a 10–15 s range for both consumer-grade systems, while the Mindtooth Touch remained largely unaffected by smoothing. These results corroborate the interpretation that consumer-grade devices, though more sensitive to transient noise, can achieve

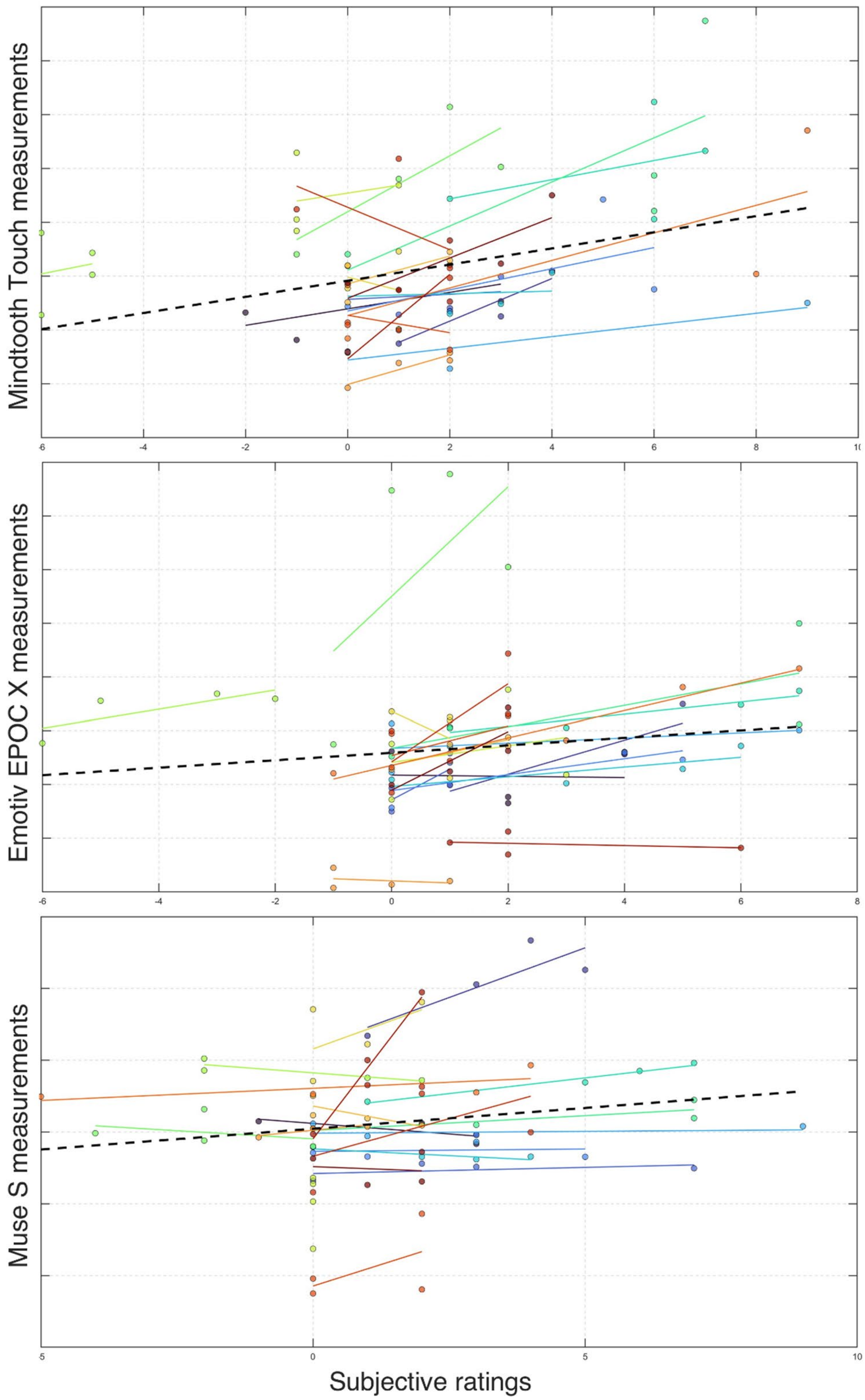


Fig. 7 The repeated measure correlation analysis performed between the Mental Workload measurements performed by each of the three considered EEG systems and the subjective ratings highlighted how the ones evaluated by Mindtooth Touch were more significantly coherent with the participants' perceptions

marginal acceptable neurometric reliability through moderate temporal averaging. This empirical evidence provides practical guidance for optimizing signal processing in future real-time applications of wearable EEGs outside laboratory settings.

4.4 Passive BCI responsiveness

A key objective of this study was to evaluate the devices' ability to detect changes in cognitive state in response to varying task demands, a fundamental requirement for passive BCIs. Statistical comparisons between experimental conditions (e.g., math easy vs. hard, multitasking easy vs. hard, driving complexity) revealed that only the Mindtooth Touch was able to consistently detect significant differences in neurometrics across tasks, aligning with expected cognitive load variations.

Emotiv EPOC X showed marginal improvement in some conditions following the application of a 15-s moving average but still failed to reliably discriminate between all task contrasts. Muse S demonstrated the weakest performance, with neurometrics that were largely unresponsive to task difficulty. This pattern is consistent with the observed performance gradient across devices, which appears to align with their respective price points.

4.5 Why do devices differ?

This section also extends the discussion by comparing how each device's hardware choices specifically influenced its performance in our results. The performance differences observed across headsets can be largely explained by their hardware architecture and mechanical design. Research-grade systems, such as the Mindtooth Touch, employ semi-dry active electrodes with built-in impedance monitoring and high-resolution front-ends (24-bit ADCs), which ensure a stable contact and reduce motion-related artefacts [32, 55–57]. In contrast, most consumer-grade systems rely on saline or fabric-based passive electrodes, which are more prone to drying and micro-movements, leading to increased low-frequency drifts and sporadic signal dropouts [14, 15]. The channel layout also plays a major role: devices with limited frontal coverage or without parietal electrodes cannot compute canonical fronto-parietal contrasts typically required for workload and engagement indices [58–61], thereby reducing their neurometric sensitivity. Additionally, sampling resolution and front-end noise characteristics differ considerably across systems—research-grade amplifiers usually offer a higher dynamic range and better common-mode rejection [62]—which affects both the stability and fidelity of spectral features. Mechanical aspects are equally critical: lightweight headsets with flexible frames improve comfort but are mechanically less stable, producing micro-slippage and electrode impedance fluctuations during movement [13, 16]. Finally, data transmission protocols can influence reliability: Bluetooth Low-Energy (BLE) headsets are more vulnerable to packet loss and jitter, particularly under electromagnetic interference, which may explain the slightly reduced temporal coherence observed for consumer systems [21].

Altogether, these hardware- and design-related factors explain why performance gaps become especially evident in dynamic, movement-rich tasks, such as multitasking or driving, where signal stability and temporal integrity are crucial.

4.6 Subjective measures and neurophysiological concordance

The integration of subjective measures with physiological data provides an important cross-validation of neuroergonomic indices. Participants reported their perceived mental effort after each task, and these ratings were compared with EEG-derived workload indices. For Mindtooth Touch, workload neurometrics closely correlate with subjective reports, particularly during high-demand phases such as multitasking and driving under pressure. This alignment supports the validity of the device's neurometric output and its potential for user-centred adaptive systems.

In contrast, the workload indices computed by using the Emotiv EPOC X and Muse S showed weaker alignment with participants' self-reported effort levels. While some degree of correlation was observed for the Emotiv system, particularly under high-demand conditions, this association was less consistent and notably lower than that observed with the Mindtooth Touch. The Muse S device exhibited even lower correspondence, with workload estimates often failing to reflect the expected increases during cognitively demanding tasks such as multitasking. These discrepancies may be attributed to limitations in signal stability.

From a usability standpoint, Muse S was perceived as more comfortable than Emotiv EPOC X, likely due to its lightweight structure and soft-band design. However, this comfort advantage did not appear to be accompanied by improvements in signal quality or cognitive state tracking. These observations highlight the ongoing challenge of balancing comfort and data quality in wearable EEG design, particularly for prolonged use in applied settings.

4.7 Limitations

Although the results obtained offered interesting indications on the reliability and applicability of wearable EEG devices in real-world settings, it is necessary to consider some methodological limitations that may have influenced the interpretation and generalizability of the data. First, the size of the experimental sample was limited, mainly due to the voluntary and unpaid nature of participation. A second limitation was related to the need to adopt methodological compromises in the calculation of EEG metrics, due to the absence of complete and shared electrode coverage among all devices. Since each system used had a different electrode configuration, the analysis was limited to the exclusive use of common channels between the pairs of devices (AF3 and AF4 between Mindtooth Touch and Emotiv EPOC X, AF7 and AF8 between Mindtooth Touch and Muse S). This inevitably reduced the richness and variety of the analysable signals, partly limiting the representativeness of the neurometrics considered and forcing conservative choices in terms of comparison between the systems. An additional limitation of this study concerns the temporal invariance of the recordings. Due to the physical design and wearing constraints of the selected devices, it was not feasible to collect EEG data from all three systems simultaneously. As a result, recordings were performed in separate sessions, potentially introducing variability related to time-dependent factors such as fatigue, learning effects, or fluctuations in cognitive state. However, these potential sequence-related effects were statistically controlled. As reported in the Performed analyses section, the order of device use was randomized, and additional mixed-model analyses confirmed the absence of significant order or learning effects. This supports the robustness of the between-device comparisons despite the sequential testing design. Concerning the sample size, a post-hoc sensitivity analysis (G*Power 3.1) indicated that with $N = 24$ and $\alpha = 0.05$, the study could detect medium effects (Kendall's $W \approx 0.20$, $r \approx 0.45$) with 80% power. While the sample size is typical for EEG validation studies, results should be generalized cautiously to broader populations and task types. Finally, it is worth noting that only two consumer-grade devices were considered in this study—albeit among the most widely used—while several other commercial EEG systems are available on the market. Further comparisons including a broader range of devices would be needed to generalize the present findings more comprehensively. While EPOC X and Muse S capture dominant consumer segments and are highly representative in the literature [14, 55], we acknowledge that other consumer systems exist; future work should broaden coverage to confirm whether the category-level patterns reported here generalize across additional headsets.

5 Conclusion

This study systematically compared a higher-cost, research-grade EEG system (Mindtooth Touch) with two widely used consumer-grade alternatives (Emotiv EPOC X and Muse S), aiming to assess their suitability for passive BCI applications in ecologically valid, real-world scenarios. The evaluation focused on signal quality, neurometric reliability, responsiveness to cognitive demands, and alignment with subjective measures. Across these dimensions, the Mindtooth Touch consistently demonstrated more stable and reliable performance, confirming a benchmark for research-grade expectations.

In particular, the Mindtooth Touch was able to provide EEG recordings with lower artefact incidence, neurometrics that aligned more closely with subjective effort reports, and greater sensitivity in detecting cognitive state variations. While both consumer-grade devices demonstrated some capability, signal quality and measurement reliability were generally more variable. Post-processing techniques such as temporal smoothing helped to mitigate some issues but may reduce temporal resolution—an important factor for real-time applications. Furthermore, neurophysiological alignment with self-reported perceptions was less pronounced in the consumer-grade systems, particularly in the Muse S, suggesting that additional refinement may be needed for reliable state tracking.

Importantly, these findings do not imply that consumer-grade devices lack value. On the contrary, their accessibility, affordability, and ease of use make them attractive for exploratory studies, educational contexts, or low-stakes applications. However, it is crucial to be aware of their current limitations in signal fidelity and real-time responsiveness when considering their deployment in adaptive or operational BCI scenarios.

Future work should aim to reduce the trade-offs between usability and signal quality. This could involve developing next-generation wearable EEG devices that integrate user-friendly designs with improved sensor performance, enhanced artefact rejection techniques, and optimized electrode placements. At the same time, the adoption of standardized benchmarking protocols and validation frameworks will be essential to support rigorous comparisons and broader technology transfer.

Overall, while high-fidelity EEG systems continue to represent the most reliable option for demanding applications, the evolving landscape of wearable neurotechnology holds promising potential. With continued development, consumer-grade solutions may increasingly contribute to the expansion of passive BCIs beyond controlled laboratory environments.

6 Fundings

This work was co-funded by HORIZON 2.5 project "CODA: COntroller adaptive Digital Assistant" (GA n. 101,114,765), SESAR 3 Joint Undertaking project "TRUSTY: TRUStworthy inTellingent sYstem for remote digital tower" (GA n. 101,114,838). We acknowledge financial support under the grants "TCI" (Bando Ateneo Grande 2024) provided by Sapienza University of Rome to Pietro Aricò (SAP_RICERCA_2024_TCI_ARICÒ_P_01), "2bSafe—2bSafe—Neurophysiological characterization and modelling of "readiness to safely work", for an enhanced workers' training in virtual reality", for an enhanced workers' training in virtual reality", provided by Sapienza University of Rome to Vincenzo Ronca (AR2241907596EF17), and "Social Empa-Team" provided by Sapienza University of Rome to Rossella Capotorto (AR12419077EEFFA5) are also acknowledged.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40708-025-00290-x.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

This work was funded by the "NeuroUX 5.0" project, within the program BRIC, funded by the National Institute for Insurance against Accidents at Work, CUP B83C24006240005.

Author contributions

V.R. and M.C. wrote the main manuscript text, prepared figures, performed the neurophysiological and subjective data analysis and supported the data acquisition. R.C., A.G., D.G., G.D.F., G.B. supported the participants recruitment, and the data collection and analysis. V.R. and P.A. worked on the protocol design and supported the data acquisition. G.D.F. and P.A. supported and supervised the software implementation for the neurophysiological data analysis. F.B. and P.A. supervised the research project and validated the main manuscript text. All authors reviewed the manuscript.

Data availability

The data that support the findings of this study might be available from the corresponding authors upon reasonable request.

Declarations

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Computer, Control, and Management Engineering "Antonio Ruberti", Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy
[2]Department of Anatomical, Histological, Forensic & Orthopedic Sciences, Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy
[3]Department of Molecular Medicine, Sapienza University of Rome, Viale Regina Elena, 291, 00161 Rome, Italy
[4]Department of Physiology and Pharmacology, "Vittorio Erspamer", Sapienza University of Rome, Piazzale Aldo Moro 5, 00185 Rome, Italy
[5]Department of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China

Received: 1 August 2025 / Accepted: 11 December 2025
Published online: 28 December 2025

References

1. Arico P, Borghini G, Di Flumeri G, Sciaraffa N, Babiloni F (2018) Passive BCI beyond the lab: current trends and future directions. Physiol Meas. https://doi.org/10.1088/1361-6579/aad57e
2. Douibi K et al (2021) Toward EEG-based BCI applications for Industry 4.0: challenges and possible applications. Front Hum Neurosci 15:705064
3. Udovičić G, Topić A, Russo M (2016) Wearable technologies for smart environments: A review with emphasis on BCI. In: 2016 24th International conference on software, telecommunications and computer networks, SoftCOM 2016 https://doi.org/10.1109/SOFTCOM.2016.7772186
4. Mahmoudi A, Khosrotabar M, Gramann K, Rinderknecht S, Sharbafi MA (2025) Using passive BCI for personalization of assistive wearable devices: a proof-of-concept study. IEEE Trans Neural Syst Rehabil Eng 33:476–487
5. Zhang J et al (2023) Recent progress in wearable brain–computer interface (BCI) devices based on electroencephalogram (EEG) for medical applications: a review. Health Data Sci. https://doi.org/10.34133/hds.0096
6. Aricò P et al (2019) How neurophysiological measures can be used to enhance the evaluation of remote tower solutions. Front Hum Neurosci 13:469137
7. Sciaraffa N et al (2022) Validation of a Light EEG-based measure for real-time stress monitoring during realistic driving. Brain Sci 12:304
8. Di Flumeri G et al (2022) EEG-based index for timely detecting user's drowsiness occurrence in automotive applications. Front Hum Neurosci. https://doi.org/10.3389/fnhum.2022.866118
9. Ronca V et al (2022) Validation of an EEG-based Neurometric for online monitoring and detection of mental drowsiness while driving. Annu Int Conf IEEE Eng Med Biol Soc 2022:3714–3717
10. Frey J, Mühl C, Lotte F, Hachet M (2013) Review of the use of electroencephalography as an evaluation method for human-computer interaction. In: PhyCS 2014 - Proceedings of the international conference on physiological computing systems 214–223 https://doi.org/10.5220/0004708102140223.
11. Vézard L, Legrand P, Chavent M, Faïta-Aïnseba F, Trujillo L (2015) EEG classification for the detection of mental states. Appl Soft Comput 32:113–131
12. Monteiro TG, Skourup C, Zhang H (2019) Using EEG for mental fatigue assessment: a comprehensive look into the current state of the art. IEEE Trans Human-Mach Syst 49:599–610
13. Gao Z, Cui X, Wan W, Qin Z, Gu Z (2022) Signal quality investigation of a new wearable frontal lobe EEG device. Sensors 22:1898
14. Ratti E, Waninger S, Berka C, Ruffini G, Verma A (2017) Comparison of medical and consumer wireless EEG systems for use in clinical trials. Front Hum Neurosci. https://doi.org/10.3389/fnhum.2017.00398
15. Mikhaylov D et al (2024) Comparison of EEG signal spectral characteristics obtained with consumer- and research-grade devices. Sensors 24:8108
16. Arpaia P et al (2025) A systematic review of techniques for artifact detection and artifact category identification in electroencephalography from wearable devices. Sensors 25:5770
17. Mahdid Y, Lee U, Blain-Moraes S (2020) Assessing the quality of wearable EEG systems using functional connectivity. IEEE Access 8:193214–193225
18. Ekandem JI, Davis TA, Alvarez I, James MT, Gilbert JE (2012) Evaluating the ergonomics of BCI devices for research and experimentation. Ergonomics 55:592–598
19. Muvvala VK et al (2025) Comparative analysis of signal quality and usability for a novel wireless, wearable EEG sensor. Clin Neurophysiol Pract 10:292–300
20. Sabio J, Williams NS, McArthur GM, Badcock NA (2024) A scoping review on the use of consumer-grade EEG devices for research. PLoS ONE 19:e0291186
21. Niso G, Romero E, Moreau JT, Araujo A, Krol LR (2023) Wireless EEG: a survey of systems and studies. Neuroimage. https://doi.org/10.1016/j.neuroimage.2022.119774
22. Duvinage M et al (2013) Performance of the Emotiv Epoc headset for P300-based applications. BioMed Eng OnLine 12:1–15
23. Riedl R, Minas RK, Dennis AR, Müller-Putz GR (2020) Consumer-grade EEG instruments: insights on the measurement quality based on a literature review and implications for neurois research. Lecture Notes Inf Syst Organ 43:350–361
24. Di Flumeri G et al (2023) A neuroergonomic approach fostered by wearable EEG for the multimodal assessment of drivers trainees. Sensors 23:8389
25. Di Flumeri G et al (2019) The dry revolution: evaluation of three different EEG dry electrode types in terms of signal spectral features, mental states classification and usability. Sensors 19:1365
26. Sciaraffa N et al (2022) Evaluation of a new lightweight EEG technology for translational applications of passive brain-computer interfaces. Front Hum Neurosci 16:458
27. Gerner N, Pickerle D, Höller Y, Hartl A (2025) Neurophysiological markers of design-induced cognitive changes: a feasibility study with consumer-grade mobile EEG. Brain Sci 15:432
28. Krigolson OE, Williams CC, Norton A, Hassall CD, Colino FL (2017) Choosing MUSE: validation of a low-cost, portable EEG system for ERP research. Front Neurosci 11:243179

29. Ghosh K et al (2023) Mindfulness using a wearable brain sensing device for health care professionals during a pandemic: a pilot program. J Prim Care Community Health 14:21501319231162308
30. Klimesch W (1999) EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res Rev 29:169–195
31. Lopez-Gordo MA, Sanchez Morillo D, Pelayo Valle F (2014) Dry EEG electrodes. Sensors 14:12847–12870
32. Ouzir M, Chakir Lamrani H, Bradley RL, El Moudden I (2024) Neuromarketing and decision-making: classification of consumer preferences based on changes analysis in the EEG signal of brain regions. Biomed Signal Process Control 87:105469
33. Khondakar MFK et al (2024) A systematic review on EEG-based neuromarketing: recent trends and analyzing techniques. Brain Inf 11:1–25
34. Vozzi A, Ronca V, Cherubino P, Trettel A, Babiloni F (2021) Measuring the emotional and cognitive consumers' responses during interaction with marketing stimuli. Contrib Manag Sci. https://doi.org/10.1007/978-3-030-67020-7_8/COVER
35. Koelstra S et al (2012) DEAP: a database for emotion analysis; using physiological signals. IEEE Trans Affect Comput 3:18–31
36. Di Flumeri G, Arico P, Borghini G, Colosimo A, Babiloni F (2016) A new regression-based method for the eye blinks artifacts correction in the EEG signal, without using any EOG channel. In: Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, EMBS vols 2016-Octob 3187–3190 (institute of electrical and electronics engineers Inc.).
37. Ronca V et al (2024) o-CLEAN: a novel multi-stage algorithm for the ocular artifacts' correction from EEG data in out-of-the-lab applications. J Neural Eng 21(5):056023
38. Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134:9–21
39. Skrandies W (1990) Global field power and topographic similarity. Brain Topogr 3:137–141t
40. Vecchiato G et al (2014) How to measure cerebral correlates of emotions in marketing relevant tasks. Cognit Comput 6:856–871
41. Bates D, Mächler M, Bolker BM, Walker SC (2014) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48
42. Molina E, Correa A, Sanabria D, Jung TP (2013) Tonic EEG dynamics during psychomotor vigilance task. In: International IEEE/EMBS conference on neural engineering, NER 1382–1385 https://doi.org/10.1109/NER.2013.6696200.
43. Sebastiani M et al (2020) Neurophysiological vigilance characterisation and assessment: laboratory and realistic validations involving professional air traffic controllers. Brain Sci 10:48
44. Wickens CD (2008) Multiple resources and mental workload. Hum Factors 50:449–455
45. Longo L (2018) Experienced mental workload, perception of usability, their interaction and impact on task performance. PLoS ONE 13:e0199661
46. Gaillard AWK, Boucsein W, Stern J (1996) Psychophysiology of workload. Biol Psychol 42:245–247
47. Babiloni F (2019) Mental workload monitoring: new perspectives from neuroscience. In communications in computer and information science vol. 1107 3–19 Springer
48. Aricò P et al (2014) Towards a multimodal bioelectrical framework for the online mental workload evaluation. Conf Proc IEEE Eng Med Biol Soc 2014:3001–3004
49. Davidson RJ, Ekman P, Saron CD, Senulis JA, Friesen WV (1990) Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology I. J Pers Soc Psychol 58:330–341
50. Briesemeister BB et al (2013) Approach the good, withdraw from the bad: a review on frontal alpha asymmetry measures in applied psychological research. Psychology 4:261–267
51. Harmon-Jones E, Gable PA, Peterson CK (2010) The role of asymmetric frontal cortical activity in emotion-related phenomena: a review and update. Biol Psychol 84:451–462
52. Di Flumeri G et al. (2017) EEG-based Approach-Withdrawal index for the pleasantness evaluation during taste experience in realistic settings. In: Proceedings of the annual international conference of the IEEE **engineering in medicine and biology societ**y, EMBS 3228–3231 https://doi.org/10.1109/EMBC.2017.8037544.
53. Inguscio BMS et al (2021) Smoke signals: a study of the neurophysiological reaction of smokers and non-smokers to smoking cues inserted into antismoking public service announcements. Int J Psychophysiol 167:22–29
54. Bakdash JZ, Marusich LR (2017) Repeated measures correlation. Front Psychol 8:456
55. Rogala J, Kublik E, Krauz R, Wróbel A (2020) Resting-state EEG activity predicts frontoparietal network reconfiguration and improved attentional performance. Sci Rep. https://doi.org/10.1038/s41598-020-61866-7
56. Gargiulo G et al (2010) A new EEG recording system for passive dry electrodes. Clin Neurophysiol 121:686–693
57. Mihajlovic V, Grundlehner B, Vullers R, Penders J (2015) Wearable, wireless EEG solutions in daily life applications: what are we missing? IEEE J Biomed Health Inform 19:6–21
58. Aricò P et al (2016) Adaptive automation triggered by EEG-based mental workload index: a passive brain-computer interface application in realistic air traffic control environment. Front Hum Neurosci 10:217761
59. Borghini G, Astolfi L, Vecchiato G, Mattia D, Babiloni F (2014) Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci Biobehav Rev 44:58–75
60. Ronca V et al (2023) Neurophysiological Assessment of an innovative maritime safety system in terms of ship operators' mental workload, stress, and attention in the full mission bridge simulator. Brain Sci 13:1319
61. Borghini G et al (2022) Real-time pilot crew's mental workload and arousal assessment during simulated flights for training evaluation: a case study. Annu Int Conf IEEE Eng Med Biol Soc 2022:3568–3571
62. Casson AJ, Yates DC, Smith SJM, Duncan JS, Rodriguez-Villegas E (2010) Wearable electroencephalography. What is it, why is it needed, and what does it entail? IEEE Eng Med Biol Mag 29:44–56

Publisher's Note