applied sciences

MDPI

Article

Promise for Personalized Diagnosis? Assessing the Precision of Wireless Consumer-Grade Electroencephalography across Mental States

Amedeo D'Angiulli [1,2,*], Guillaume Lockman-Dufour [1] and Derrick Matthew Buchanan [1]

[1] NICER Lab, Carleton University, Ottawa, ON K1S5B6, Canada; guillaumelockmandufo@cmail.carleton.ca (G.L.-D.); matthewbuchanan@cmail.carleton.ca (D.M.B.)
[2] Department of Neuroscience, Carleton University, Ottawa, ON K1S5B6, Canada
* Correspondence: amedeo.dangiulli@carleton.ca

Abstract: In the last decade there has been significant growth in the interest and application of using EEG (electroencephalography) outside of laboratory as well as in medical and clinical settings, for more ecological and mobile applications. However, for now such applications have mainly included military, educational, cognitive enhancement, and consumer-based games. Given the monetary and ecological advantages, consumer-grade EEG devices such as the Emotiv EPOC have emerged, however consumer-grade devices make certain compromises of data quality in order to become affordable and easy to use. The goal of this study was to investigate the reliability and accuracy of EPOC as compared to a research-grade device, Brainvision. To this end, we collected data from participants using both devices during three distinct cognitive tasks designed to elicit changes in arousal, valence, and cognitive load: namely, Affective Norms for English Words, International Affective Picture System, and the *n*-Back task. Our design and analytical strategies followed an ideographic person-level approach (electrode-wise analysis of vincentized repeated measures). We aimed to assess how well the Emotiv could differentiate between mental states using an Event-Related Band Power approach and EEG features such as amplitude and power, as compared to Brainvision. The Emotiv device was able to differentiate mental states during these tasks to some degree, however it was generally poorer than Brainvision, with smaller effect sizes. The Emotiv may be used with reasonable reliability and accuracy in ecological settings and in some clinical contexts (for example, for training professionals), however Brainvision or other, equivalent research-grade devices are still recommended for laboratory or medical based applications.

Keywords: Emotiv EPOC; brainvision; brain-computer interface BCI; reliability; accuracy; arousal; valence; mental load; consumer-grade EEG; research-grade EEG

check for updates

Citation: D'Angiulli, A.; Lockman-Dufour, G.; Buchanan, D.M. Promise for Personalized Diagnosis? Assessing the Precision of Wireless Consumer-Grade Electroencephalography across Mental States. *Appl. Sci.* **2022**, *12*, 6430. https://doi.org/10.3390/app12136430

Academic Editors: Fabio La Foresta and Serena Dattola

Received: 30 March 2022
Accepted: 17 June 2022
Published: 24 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Emotiv EPOC Wireless EEG Device in Context: Current Research and Applications

Since its inception, the field of electroencephalography (EEG) and its potential uses have developed considerably. In laboratory research settings, EEG has been used extensively for cognitive neuroscience research, and in medical settings, it is considered the most robust method of diagnosing seizures and epilepsy [1,2]. In the last decade, however, there has been significant growth in the interest and application of using EEG outside of the laboratory and medical settings [3]. EEG has become the most widely used neuroimaging technique for brain-computer interfaces (BCI). Some of these extended uses of EEG include military operations such as controlling weapons or drones [4–8], educational classroom applications such as monitoring student's attention/other mental states or helping them engage with material [9–13], cognitive enhancement such as increasing cognitive load or focus [12,14,15], and consumer based games such as computer games or physical toys controlled via brain waves [2,15–19].

One challenge in developing outside lab EEG applications, however, is that traditional research- and medical-grade EEG devices are costly (~US$50,000+), typically stationary, fully wired, and time consuming to properly set up, in that electrode application may typically take 30–60 min depending on the type and number of wet (gel-based) sensors [3]. For example, one research-grade EEG device is the Brainvision EEG headset. Its most typical configuration includes 32 channels and must be connected by wires to an amplifier, a computer, and a monitor, for data input and analysis. Its electrodes require the use of electro-conductive gel prior to the experiment, which allows it to minimize impedance and be more sensitive to the electrical activity produced by the brain. While these factors are a strength of Brainvision in regard to the accuracy and reliability of data collection, they can also be interpreted as limitations in other capacities. For instance, all of these factors make using traditional EEG challenging in more ecological settings (i.e., educational and training health settings). In response to these challenges, over the last decade or so, wearable, portable, cost-efficient EEG headsets have been developed [3]. To date, the most recent and exhaustive review of consumer grade EEG devices impressively covering applications in cognition, brain computer interfaces, education research, and game development has been published by Sawangjai et al. [20]. Their review dealt primarily with the validity associated with the wireless EEG devices, that is, with the issues related to whether what was measured by wireless consumer-grade devices reflected the same constructs (for example the same EEG or/and ERP components or signatures) or the same entities/features intended to be identified and measured as the ones traditionally obtained with research-grade EEG devices. However, Sawangjai et al. did not explicitly differentiate issues of validity from those of reliability, therefore, their assessment did not adequately address related issues of personalization, these are the central objective and contribution of our work.

A commercial consumer-gradedevice that emerged in response to the limiting factors associated with stationary research-grade systems was the Emotiv EPOC. The EPOC headset is much more affordable; as of fall 2021, the newer 14-channel version of the Emotiv EPOC (now called the Emotiv EPOC X) can be purchased for less than US $1000. These headsets boast much quicker and easier installation and removal times of 3–5 min, in addition to being wireless. This device also has low weight and a battery life of up to 12 hours. Such devices have the advantage of being more user and consumer-friendly making them well-suited for situations where ergonomics and patient or participant movement is involved; as well as situations where scalability is of interest (e.g., having 30 students in a classroom all wearing an Emotiv, which would be highly impractical using a research device like Brainvision). Crucially, what set EPOC apart from the other available commercial devices is that at the time of its release it had one of the largest set of wireless electrodes, making it comparable to the electrode set sizes used in a significant part of literature representing research with traditional laboratory wired EEG equipment.

Although there are some clear advantages to using a device like the Emotiv headsets in consumer/ecological and in some clinical settings, it does not come without its own challenges. Consumer grade EEG devices typically compromise data resolution for shorter setup time, affordability, and portability [21–26]. Since portable headsets have less electrodes and channels than standard research-grade EEGs, scalp coverage and spatial resolution decrease [27]. The use of dry electrodes can also pose a problem as impedance can hinder the detection of electrical activity and lead to lower data values. This can be mitigated by using saline solution, but its effect is minimal compared to electro-conductive gel [24]. Overall, concerns around portable EEGs are centred on the accuracy and reliability of their data as compared to standard, research-grade EEGs.

There is an important growing body of scientific literature investigating the Emotiv under different conditions utilizing various EEG signatures such as the P300 [24] which has demonstrated its poor reliability/accuracy compared to research-grade devices. However, other studies have demonstrated that EPOC can accurately detect P300 and N200 [26,27]. That said, the study from Barham et al. (2017) utilized a modified Emotiv with electrodes that had been upgraded to research-grade quality [26] which may have played a significant

role in the success of the Emotiv application. Another study from Badcock et al. (2015) [28] further demonstrated that the morphology of EEG signals such as the P1, N1, P2, N2, and P3 were highly correlated between the Emotiv and research-grade Neuroscan system; however there were some differences in the amplitude of the signal. Needless to say, there remains a disparity in the literature which requires more research to be done to elucidate the truth about the Emotiv. There are so many factors that may influence the success of an Emotiv application: Notably, the study by Duvinage et al. (2012) [24] which reported poor quality of the Emotiv for detecting P300 utilized a visual odd-ball paradigm, whereas the studies that were more successful [26–28] utilized auditory odd-ball paradigms. Therefore, the modality of the task used to elicit the EEG signal may play an important role. However, a more recent study from Fouad (2021) [29] was able to successfully capture and utilize subject's P300 signal with reasonable accuracy for a brain-computer interface speller (a visual paradigm) when combined with machine learning such as support vector machines.

A critical emerging issue related to wireless portable EEG devices such as EPOC is the application to personalized healthcare [30]. Personalization implies that the quality of repeated measurements of desired EEG features have an acceptable to optimal level of **precision** or **reliability** (i.e., *measurement or instrumental precision*) at the level of an individual (i.e., the so-called "N-of-1") (see [31]). Although a dominant mainstream view is that some aspects of the classification and processing of the signal are deemed to be handled by the integration with machine learning on the basis of massive iterative training applied to large within subject repeated measures, the latter types of approaches are no substitutes for the quality of EEG signal acquisition from the individual subject, that is, they do not directly address the possible problems associated with *weak measurement* at the individual level. Therefore, the issues partly require approaches and techniques that address small-N or N-of-1 experimental designs, and neither can be *only* solved with the current mainstream data mining science based on large samples nor with parametrization referenced to a population distribution (see [31]).

The issues in EEG personalization largely overlap with the current debate in the multidisciplinary field of measurement which contrasts the **idiographic** person-oriented approach [32,33] versus the **nomothetic** population distribution referenced approach (for extensive overview and review and discussion within the context of psychological and brain sciences see [34]). Personalized predictions can only be made based on prior data from the individual for whom a prediction is to be made (idiographic data) and not with aggregated data from other individuals (nomothetic data) [35]. This is an implication from the classic ergodicity theorems [36], from which it follows that intra-individual (within subjects) variation (IAV) is equivalent to inter-individual (between subjects) variation (IEV), that is, they are ergodic, only if IAV is homogeneous in time, that is, if there are no fluctuations, trends or other types of time-dependent changes in the IAV time series). If the structure of IAV is heterogeneous, then it can no longer be examined by switching to the IEV perspective—as in most of the approaches relying on population distributions—because the two types of variations are incommensurable. In terms of relevance for neurocognitive processes pertinent to EEG, the non-ergodic implications also apply to functional brain connectivity [37,38] and the underlying dynamics of neural networks (e.g., see [39]).

Thus, employing small-N designs that focus on the individual participant as the *replication unit* and the *control of him/herself* is a sound alternative or/and supplementary tool to assess the precision of personalization EEG devices for the measurement of individual mental states. Indeed, there are alternative approaches that derive and conform to the tradition of behavioral [40] and psychophysical [31,41] experimental designs. One is the numerical method known as Vincentization or Vincentized average [40], in which data from small samples (N = 3–5) are binned so that the derived distribution reflects approximately the same shape as each individual, and a second approach is the item-wise (or by-item) analysis approach [42] in which the dependent variable is aggregated across subjects for a particular item (i.e., stimulus or channel) and consequently the item becomes the unit or case which the statistics is applied to, usually requiring a completely nested

within-subjects design. These two methods are complementary and when used together (i.e., item-wise analysis of vincentized data) they offer the opportunity to integrate intra-individual and inter-individual/group variability by empirically constraining or restricting the range of heterogeneity of the combined variance, so that variability can be assumed to be reasonably ergodic.

In summary, there seems to be some disparity in the literature regarding the reliability, reproducibility, and accuracy of the Emotiv EPOC headset. Moreover, the literature is specifically lacking in studies investigating other characteristics of the EEG signal such as power and amplitude. The literature is also lacking in studies investigating the precision of EEG features related to a wider variety of mental states such as arousal, valence, and mental load. Most important, so far there have been only a handful of studies (reviewed above) which compared measurements collected from the *same* subjects, serving as controls of themselves, and done on *both* consumer-grade EPOC Emotiv and research-grade devices.

To start filling some of the gaps in the literature we here report a preliminary empirical study assessing the precision of the Emotiv EPOC consumer device, with a particular eye to implications for EEG personalization. That is, the empirical question we address here is one of **reliability** or **precision**: *can EEG signal measured with EPOC from an individual A, while he/she is in a defined psychological state X, consistently replicate the EEG signal from the same individual A in the same psychological state X when measured with a stationary traditional device such as Brainvision assumed as the "gold standard"*. This is independent from the validity question, that is, whether the EEG signal measured with both types of EEG devices from A can be validly classified as a correlate of the same state X *in the general population*.

1.2. The Present Study

The present study aimed at investigating the accuracy and reliability of consumer-grade portable and wireless EEG as compared to standard, wired research-grade EEG, using the Emotiv EPOC 14-channel headset and the Brainvision 32-channel headset. While, as already mentioned, there are many empirical studies and exhaustive reviews which have focused on addressing the validity of wireless consumer-grade EEG devices including Emotiv EPOC, few have directly addressed the issue of reliability of measurement, independent of the content-validity of the measurement; among the latter small literature, even fewer have compared EEG features in data collected from the same subjects using both types of devices, that is, using a repeated measures (i.e., within-subject sample-matching) design.

Accordingly, to advance current research in the field, the present study contributed a novel approach for the direct assessment of the reliability and accuracy of wireless consumer-grade EEG devices, using as paradigmatic example case Emotiv EPOC. Specifically, we investigated the features of EEG power and amplitude measurements for specific frequency bands of matched electrodes (the 14 channels these headsets share are AF3 (FP1), F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 and AF4 (FP2)) between the Emotiv device and the Brainvision device during different mental states in the same participants. In the next sections, we describe the EEG features and the level of psychological states considered.

1.2.1. EEG Features: Event-Related Band Power and Amplitude

Typically, a clear separation is made between the analysis of the EEG in the frequency domain (Spectral analysis; measurements: power, coherence, phase, etc.) and the analysis of the EEG in the time domain, whose most common form is the Event-Related Potentials (ERPs), obtained by averaging activity associated with events of the same nature (reported measurements: amplitude, latency, topography).

However, most recently, hybrid approaches have also been devised in which time and frequency domain EEG features are examined as covariates (for example see [43–46]), particularly, to investigate spectral changes associated with stimulus or task time-course, as opposed to resting brain activity. In line with these approaches, in the present study we focused on two EEG event-locked or task-related features: Event-related band frequency amplitude and power. As described by Klimesch [47], the event-related **amplitude** is the

'magnitude' of an oscillation, reflecting the distance between the maximal positive- and negative-going points (phase) of an oscillatory cycle. The related feature of **power** is the under-the-curve integral of the event-related amplitude response (to a stimulus and/or task) within the entire frequency band during a test period.

1.2.2. Definition of Low vs. High Mental States

Three distinct psychological tasks were used which involve emotional and cognitive processes and have been designed and validated to elicit varying levels of arousal, valence, and mental load: Affective Norms for English Words (ANEW), International Affective Picture System (IAPS), and the *n*-Back task. These tasks were selected because they demonstrated reasonable validity in distinguishing a relative difference in low as opposed to high psychological states as defined by the subjective self-ratings of the participants and as reflected by the state-dependent changes in EEG correlated with those low vs high levels.

ANEW is a set of 1034 English-language words associated with emotional values [48]. This dataset was made to complement the IAPS database described below [49,50] to assess attentional and emotional processing in the auditory and verbal/linguistic domains. The words are assessed by self-ratings based on arousal, which goes from calm to excited, and valence, which ranges from unpleasant to pleasant. For example, "war" would evoke high arousal and low valence since it is stressful and displeasing, while "rollercoaster" would evoke high arousal and high valence due to its mostly pleasant and exciting nature [49]. Low level as opposed to high level of word arousal and valence has been previously demonstrated to alter cortical activity in ways measurable by EEG [51].

The IAPS is a repository of 956 photos chosen to be standardized visual stimuli in experiments of emotion and attention. Just like the ANEW system, the photos are chosen based on their ability to cause emotions and are primarily classified using valence and arousal [52]. Photos can also be rated based on dominance, but the latter variable was not considered in this study. Similarly to the ANEW, this task can create noticeable differences in EEG in relation to changes from low as compared to high levels of arousal and valence, as previous experiments have been able to detect such differences in brain activity during image processing, depending on pleasantness and unpleasantness [53].

The *n*-Back task was initially designed to include up to three items [54], and then upgraded to six items [55]. In this task, the participant is exposed to a stream of stimuli and must indicate whether the current stimuli matches the one *n* trials before [56]. For example, in a 1-back task, participants must indicate whether or not each letter matches the one right before; in a 3-back task, participants must indicate whether or not each letter matches the one 3 letters before; etc. The change in attention from low as opposed to high cognitive or mental load required for this task can be detected by EEG [57].

1.2.3. Hypothesis and Predictions

Our default working hypothesis was that the *wireless consumer-grade device we considered (Emotiv EPOC) should perform as accurately as the selected gold standard stationary device (Brainvision)*. Consequently, we predicted that the low as compared to the high levels in mental psychological states—as defined by psychological self-report in the ANEW (arousal), IAPS (valence) and n-Back (cognitive load)—should show similar within-person and homogeneous-group (i.e., vincentized) patterns of *relative* differences/changes between the corresponding EEG features (amplitude and power) associated with each respective level (i.e., low vs. high mental state). These patterns of changes should occur across all homologous pairs of compared electrodes in both types of devices. Thus, to the extent in which this default prediction would be supported for all, or some of, those psychological states, it would provide an assessment of the degree of reliability of the wireless consumer-grade device.

2. Materials and Methods

2.1. Sample and Data Re-Analysis Approach

The data utilized in this study was derived from a previous study by the same authors [20] which investigated global averages of EEG power across the scalp comparing EEG signal pooled across all electrodes (14 for EPOC vs. 32 for Brainvision). The database used repeated long recording session samples from "professional observers" in the span of 48 hours. Participants were three healthy university graduate students (mean age 24.66 years old). They all gave verbal consent to participate and waived the requirement of signed consent. The study was part of a grant funded pilot project which was retrospectively assessed and approved by Carleton University Institutional Research Ethics Board (section Human Subjects) under the regulation set by the Canadian Tri-Council [58].

All participants were right-handed males with corrected-to-normal or normal vision, and none reported any history of neurological impairment or were currently using psychoactive medications. The average number of hours of sleep on the previous night was 7.5 h. Testing was conducted from 10 a.m. to 4 p.m. with a one-hour break in between on two separate days. The pre-experimental short adult version of the *mood and feelings questionnaire* [5] was administered by an independent research assistant unaware of the hypotheses and goals of the study to screen for mood differences or emotional changes before the experimental sessions between days. No remarkable differences were reported as all participants scored similarly (overall score range: 3–5) in the two days and consistently well below the recommended clinical cut-off (possible maximum = 26; clinical cut-off $\geq$ 11).

Prior to the experiment, all participants gained familiarity with the actual stimuli and conditions of the tasks during several study sessions, since they were involved in selecting the stimuli and the conditions, they designed the computer programs for the tasks, and they test piloted the delivery and the performance of the task computer programs and the instrumentation. Since all these activities involved extensive exposure and practice over time, it can be assumed the materials of the tasks were overlearned and practice effects, particularly due to the order in which the devices were used, were washed away. This compensated for lack of counterbalancing due to small sample.

The present study utilized a different analytical strategy going beyond global EEG patterns, to focus specifically on the accuracy and reliability of different EEG signals at electrode sites matched between the two devices; these present analyses and results are novel unpublished findings. Thus, the unit of analysis, or "subjects" were the 14 EEG electrodes. This type of within-subject approach is usually traditionally known as by-item or item-wise analysis [42,59] or, in keeping with EEG terminology, an *electrode-wise* analysis.

The item-wise approach corresponds to a random effects model on the items, and a fixed-effects model on subjects. Therefore, the effects from statistical tests can be generalized to new items and task from the same subjects but cannot be used to make a reliable prediction for new subjects that would be generalized to the population [60]. The mean-wise group approach is ordinarily used to obtain the highest possible effect. In contrast, the item-wise approach centers around the pursuit of replicable effects based on weak but stable interindividual correlation coefficients, that is, on relatively homogeneous intraindividual variance. The weak correlation for an item is generally due to excessive interindividual noise. To reduce such noise heterogeneity, in our study the participants' data were first vincentized, that is, they were partitioned according to time-series bins with same data density per interval of time by averaging the participants' quantile functions, as previously done by D'Angiulli et al. (2020) [61]. Successively, bin-by-bin means were estimated for each electrode (i.e., across participants). This procedure insured the definition of group quantiles from which a reliable distribution function could be constructed for the data from each electrode even with N = 3 (see [62]). In essence and more simply, the outcome of the vincentization procedure is obtaining an average distribution that reflects closely the distribution of each individual subject, and as mentioned this optimizes ergodicity in the data. It is of critical importance to point out here that the statistical tests of accuracy were applied to the repeated measures linked with the set of electrodes, hence, the size of the

study coincides with the size of the repeated measurements (N = 14) not the actual sample size (number of participants).

In sum, we used an approach through which, by combining electrode-wise and vincentization techniques, we obtained a completely nested within-subject design which tends to homogenize (i.e., limit heterogeneity of) the combined inter- and intraindividual variance. The latter set up therefore permits to draw inferences more confidently at the idiographic or individual person prediction level, which ultimately is the purpose of the main applications of consumer-grade devices such as EPOC.

2.2. Electrophysiological Measures

The portable consumer grade device used in this study was the 14-electrode Emotiv EPOC. The research grade device used in this study was the 32-electrode Brainvision system. EPOC was utilized first followed by Brainvision.

For the Brainvision actiCHamp,EEG activity was first amplified and then sampled at 1 kHz. An online bandpass filter from 1 to 100 Hz was applied. During offline processing, the Brainvision data was downsampled using EEGLAB [63] to 128 Hz to make it compatible with the EPOC sampling rate.

For the EPOC, EEG signals were initially preprocessed as specified by the manufacturer were high-pass filtered with a 0.16 Hz cut-off, pre-amplified and low-pass filtered at an 83 Hz cut-off. The data were then digitized at 2048 Hz. The digitized signal was further filtered using a 5th-order sinc notch filter, and down-sampled to 128 Hz following standard practices used within the community of users (for example, see [64]).

For both systems, the average offline EEG signal was subtracted from each electrode for each time point (offline re-referencing to a common mean reference). The offline averaging was done in a way that allowed separate averages for the **Event-Related Band Potentials (ERBPs)** in each experimental condition (such as the different tasks described below in Sections 2.3 and 2.4) and stimulus type for electrodes with epochs ranging from −200 ms pre-stimulus to 1000ms post-stimulus. The trials where excessive peak-to-peak deflection occurred at non-ocular electrode placements were considered contaminated and excluded from the averages. Following artefact correction and removal, less than 15% of trials were rejected.

To extract the EEG features, we applied the Fast Fourier Transform (FFT) algorithm to transform time recordings into frequency recordings in order to compute the continuous input EEG data for each condition. We tabulated and computed the total spectral content of the electrical activity of each electrode, using a −1024 to 1024 ms time window. We computed FFTs for each participant on a grand average for each condition (high arousal, low arousal, high valence, low valence) on both the ANEW and IAPS tasks, and low and high load for the two versions of the n-Back task. FFT computation was performed through the BESA spectral analysis pipeline (http://wiki.besa.de/index.php?title=BESA_Research_Spectral_Analysis, accessed on 20 May 2022). The FFT was applied to the marked regions of the blocked data windows to examine the frequency content of the signal. FFTs were normalized and the frequency power and amplitude bands from 1–50 Hz were exported for further analysis (FFT was viewed as amplitude and power in frequency bands, and the tabulated results were saved in ASCII data files). The power Band frequencies were defined as follows: Delta (1.0 Hz–4.0 Hz), Theta (4.0 Hz–8.0 Hz), Alpha (8.0 Hz–14.0 Hz), Beta (14.0 Hz–30.0 Hz), Gamma (30.0 Hz–50.0 Hz).

2.3. IAPS and ANEW Procedures

The order in which the three observers carried out the tasks were respectively: (1) ANEW, IAPS, n-Back; (2) n-Back, IAPS, ANEW; (3) IAPS, n-Back, ANEW.

Participants were informed by pre-set computer cues of the expectations for each task, such as whether pictures, words, or letters would be presented and which items required a response. Participants were one meter away from the computer screen where the stimuli were presented. The researcher stood two meters behind the participant and offered verbal

instructions. Participants were primed with instructions to experience emotional connection with every presented image or word. Participants viewed every presented image/word for five seconds during a period of 10 min. After each block, participants were allowed a break for as long as required, which generally lasted five minutes (this was consistent between the testing completed with EPOC and the testing completed with Brainvision). During these breaks we checked electrode contact and stability of measurement. The stability of EEG signal from Brainvision was checked and maintained by monitoring the impedance of active electrodes and ensuring that it was kept below 5kOhms; this was also facilitated by the use of electrogel. The stability of EEG signal from EPOC was maintained by removing the device after each task and reapplying saline solution. Head positions were re-measured to maintain consistency between electrode placements each time the device was replaced on the head.

In total, 50 images/words were randomly presented to participants, with a break after the 25th and 50th image/word. After participants confirmed readiness, the experimenter launched the stimulus presentation program. Stimulus presentation began with a central cross fix for 200 ms to help participants focus their gaze and reduce eye movements. After this, image/word presentation occurred for 4500 ms, followed by a 1300 ms delay before the presentation of the cross fix. Images were based on scores of the standardized IAPS/ANEW scale; they were designated as extremely high or low arousal in a block, and extremely high or low valence in another block. Blocks were sets of 50 image/word presentations. Participants were told to stay still while viewing images/words. Images filled a 48.3 cm (19″) monitor with a 1024 × 768 pixel resolution, while letters had a font size of 48 points and were shown in the center of the monitor.

2.4. n-Back Procedures

Participants performed a memory task that consisted of remembering a letter based on its position in a sequence of continuously presented letters. A cross-fix would also appear at the center of the screen before each letter's presentation. The variations of this task were the 1-back and the 3-back. In the 1-back, participants were told to press a key if the letter displayed matched the one presented right before. The rationale is identical for the 3-back, where participants must indicate whether or not each letter matches the one shown three letters before. Each letter of size 48 was shown for 500 ms in the center of the monitor. EEG was recorded for all answers, valid and invalid alike. Every participant finished six blocks of the two task variations and performed three of each, alternating between variations and starting with the 1-back. Each block had 26 trials for a total of 78 trials for each variation. Subjects performed three blocks of 13 trials for the 1-back condition, and three blocks of 13 trials for the 3-back condition. A total of 117 ERBP averages for the 1-back condition and 117 ERBP averages for the 3-back condition were obtained. No trials were omitted for the *n*-Back task.

2.5. Data Analysis

The estimated marginal means were charted for the absolute value of frequency band power and its amplitude, which were expressed in microvolts (μV), for each headset's electrodes. The estimated marginal means were plotted across participants though the analysis was performed according to the electrode-wise schema (one-to-one paired matching between homologous electrodes). In order to compare headset accuracy and reliability when using the same channels, the Emotiv EPOC's 14 electrodes were matched to Brainvision's 14 correspondingly-located electrodes. Thus, the 14 electrodes used in Emotiv were AF3, AF4, F3, F4, F7, F8, FC5, FC6, T7, T8, P8, P7, O1 and O2, while the 14 ones used in Brainvision were FP1, FP2, F3, F4, F7, F8, FC5, FC6, T7, T8, P8, P7, O1, O2 and Oz (see Figure 1). (Note that the EPOC system doesn't have electrodes placed along the midline of the head, and it uses the CMS and DRL electrodes as dual reference points, instead of a single ground electrode as in the Brainvision system. Pooled reference electrodes are shown in green in Figure 1)

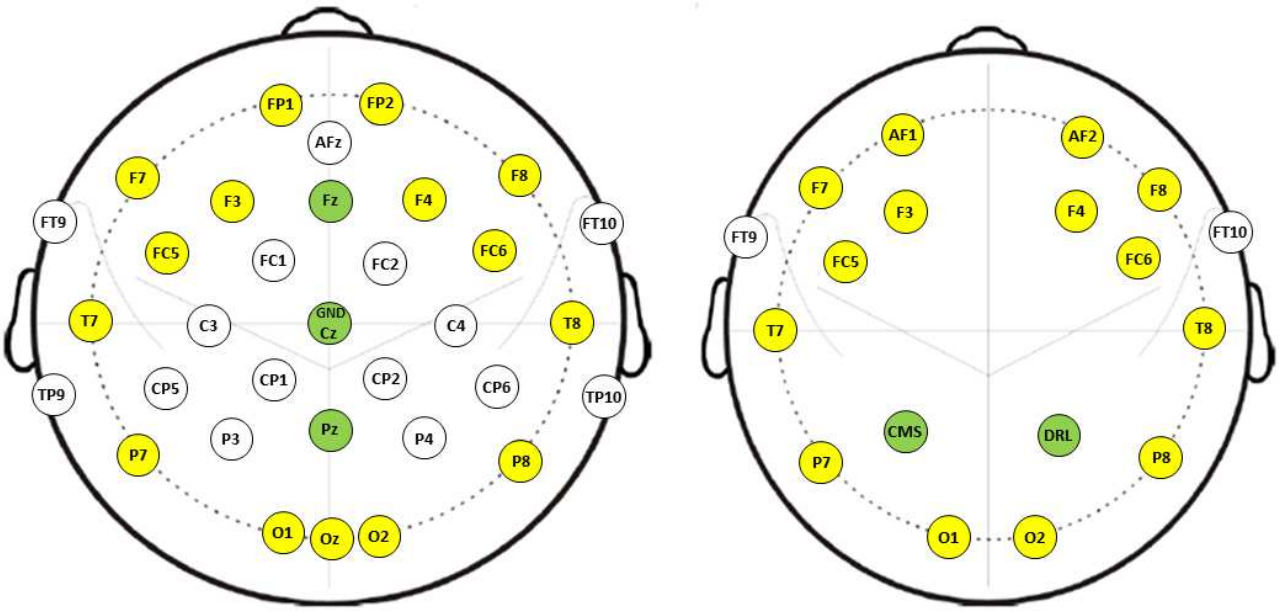


Figure 1. Left Panel: Brainvision (32-electrode) EEG system with ground at Cz. Right Panel: Emotiv EPOC (14-electrode) system with CMS at P3 and DRL at P4.

The estimated marginal means of both amplitude and power were reported in figures to demonstrate visually the detection of arousal, valence and workload, in addition to each EEG's ability to discriminate between high and low psychologically defined states in these three conditions (all these figures are included as Supplementary Materials). For conciseness, we summarize all main findings in the table below (see Table 1A,B).

Table 1. A. Grand averaged event-related power amplitudes and nonparametric contrasts for low vs. high mental states in ANEW, IAPS and n-Back Tasks measured by a consumer-grade (EPOC) and a research grade (Brainvision) EEG system. (Statistics were performed on electrode-wise repeated measurements, N = 14). **B**. Grand averaged cumulated power and nonparametric contrasts for low vs. high mental states in ANEW, IAPS and n-Back Tasks measured by a consumer-grade (EPOC) and a research grade (Brainvision) EEG system. (Statistics were performed on electrode-wise repeated measurements, N = 14).

(A)										
	EPOC					**Brainvision**				
Band	**Mental State**		**z**	***p***	**μ^2**	**Mental State**		**z**	***p***	**μ^2**
	Low	**High**				**Low**	**High**			
					(ANEW, Arousal)					
Delta	2.02	2.16	−1.10	0.271	0.09	10.74	9.34	3.18	0.001	0.72
Theta	1.48	1.57	−1.57	0.116	0.18	8.57	7.044	3.30	<0.001	0.78
Alpha	2.22	2.30	−1.45	0.148	0.15	13.23	11.344	3.30	<0.001	0.78
Beta	2.20	2.11	1.51	0.132	0.16	16.53	18.784	−2.23	0.026	0.36
Gamma	**0.89**	**0.97**	**−2.04**	**0.041**	**0.30**	**14.94**	**18.99**	**−3.23**	**0.001**	**0.75**
					(ANEW, Valence)					
Delta	**1.45**	**1.77**	**−2.07**	**0.039**	**0.31**	**6.65**	**8.19**	**−3.30**	**<0.001**	**0.78**
Theta	2.10	1.72	1.30	0.195	0.12	5.62	5.93	−3.30	<0.001	0.78
Alpha	2.13	1.98	0.88	0.379	0.06	10.28	10.51	−1.38	0.167	0.14
Beta	2.50	2.18	2.10	0.036	0.32	12.50	13.55	−3.30	<0.001	0.78
Gamma	**0.52**	**0.70**	**−3.04**	**0.002**	**0.66**	**11.27**	**12.21**	**−3.17**	**0.002**	**0.72**

Table 1. *Cont.*

(A)

Band	EPOC					Brainvision				
	Mental State		z	*p*	μ^2	Mental State		z	*p*	μ^2
	Low	High				Low	High			
					(IAPS, Arousal)					
Delta	1.96	2.76	−3.30	<0.001	0.78	7.97	7.24	3.17	0.002	0.72
Theta	**1.42**	**1.78**	**−3.11**	**0.002**	**0.69**	**5.80**	**6.00**	**−3.18**	**0.001**	**0.72**
Alpha	2.21	2.48	−2.93	0.003	0.61	8.79	9.03	−1.48	0.140	0.16
Beta	**1.73**	**1.93**	**−3.30**	**<0.001**	**0.78**	**14.57**	**15.39**	**−3.30**	**<0.001**	**0.78**
Gamma	**0.42**	**0.63**	**−3.31**	**<0.001**	**0.78**	**14.53**	**15.22**	**−3.24**	**0.001**	**0.75**
					(IAPS, Valence)					
Delta	1.75	2.80	−3.30	<0.001	0.78	7.66	6.99	3.30	<0.001	0.78
Theta	1.35	2.05	−3.23	0.001	0.75	5.98	5.84	2.00	0.046	0.29
Alpha	**2.04**	**2.78**	**−3.30**	**<0.001**	**0.78**	**10.04**	**10.28**	**−1.92**	**0.055 ***	**0.26**
Beta	1.80	2.30	−3.30	<0.001	0.78	15.07	15.02	0.67	0.506	0.03
Gamma	0.39	0.69	−3.31	<0.001	0.78	14.57	14.15	3.30	<0.001	0.78
					(n-Back 1,3,5)					
Delta	2.70	2.39	2.45	0.014	0.43	2.89	2.72	1.23	0.221	0.11
Theta	1.95	2.05	−1.54	0.124	0.17	3.07	2.95	1.29	0.198	0.12
Alpha	2.75	2.69	0.47	0.638	0.02	3.08	3.07	0.44	0.660	0.01
Beta	2.77	2.75	0.03	0.975	0.00	3.92	4.23	−3.30	<0.001	0.78
Gamma	0.95	0.93	0.59	0.556	0.02	4.25	4.49	−2.86	0.004	0.58
					(n-Back 2,4,6)					
Delta	4.58	2.35	3.30	<0.001	0.78	2.28	2.30	−0.19	0.851	0.00
Theta	2.90	1.98	3.30	<0.001	0.78	2.78	3.11	−2.79	0.005	0.56
Alpha	4.08	2.55	3.30	<0.001	0.78	2.65	2.94	−3.20	0.001	0.73
Beta	3.81	2.72	3.30	<0.001	0.78	4.66	4.64	−2.31	0.021	0.38
Gamma	1.27	0.89	3.30	<0.001	0.78	4.62	4.56	−0.25	0.802	0.00

Note. Estimated marginal means of EEG power amplitude for mental states are reported in μV. *For standard errors please refer to Figures S1–S6 in Supplementary Materials*. Z statistics calculated from Wilcoxon non parametric test for small samples. Rows in bold identify comparisons yielding statistically similar results for both devices. "*" indicates marginal significance.

(B)

Band	EPOC					Brainvision				
	Mental State		z	*p*	μ^2	Mental State		z	*p*	μ^2
	Low	High				Low	High			
					(ANEW, Arousal)					
Delta	2.25	1.88	1.85	0.064	0.24	31.11	27.85	2.10	0.035	0.32
Theta	0.58	0.60	−0.46	0.649	0.02	13.38	9.59	3.30	<0.001	0.78
Alpha	**1.22**	**1.04**	**1.92**	**0.055 ***	**0.06**	**29.82**	**21.29**	**3.30**	**<0.001**	**0.78**
Beta	0.40	0.36	1.69	0.090	0.20	12.63	16.65	−2.17	0.030	0.34
Gamma	0.11	0.11	−0.24	0.812	0.00	9.94	15.09	−3.11	0.002	0.69
					(ANEW, Valence)					
Delta	0.89	1.02	−0.69	0.488	0.03	12.09	20.83	−3.30	<0.001	0.78
Theta	1.17	0.64	1.92	0.055 *	0.26	6.40	6.40	0.21	0.834	0.00
Alpha	0.97	0.60	2.48	0.013	0.44	19.65	18.98	1.41	0.158	0.14
Beta	0.49	032	2.35	0.019	0.39	7.21	8.45	−3.30	<0.001	0.78
Gamma	0.02	0.03	1.00	0.317	0.07	5.25	6.40	−3.30	<0.001	0.78
					(IAPS, Arousal)					
Delta	1.97	2.81	−1.98	0.048	0.28	16.14	13.08	3.30	<0.001	0.78
Theta	**0.54**	**0.70**	**−2.42**	**0.015**	**0.42**	**5.90**	**6.36**	**−2.13**	**0.033**	**0.32**
Alpha	1.27	1.16	0.25	0.807	0.00	11.34	14.77	−2.73	0.006	0.53
Beta	0.21	0.22	−0.92	0.357	0.06	9.57	11.26	−3.30	<0.001	0.78
Gamma	<0.01	<0.01	1.00	0.317	0.07	8.44	9.57	−3.17	0.002	0.72

Table 1. *Cont.*

(B)										
	EPOC					**Brainvision**				
Band	**Mental State**		**z**	***p***	**μ^2**	**Mental State**		**z**	***p***	**μ^2**
	Low	**High**				**Low**	**High**			
					(IAPS, Valence)					
Delta	**1.60**	**2.69**	**−3.02**	**0.003**	**0.65**	**14.10**	**12.14**	**2.79**	**0.005**	**0.56**
Theta	0.47	0.98	−3.05	0.002	0.67	6.29	6.24	0.47	0.638	0.02
Alpha	**0.96**	**1.40**	**−3.11**	**0.002**	**0.69**	**16.98**	**17.95**	**−1.57**	**<0.001**	**0.12**
Beta	0.24	0.33	−3.22	0.001	0.74	10.59	10.68	−0.18	0.861	0.00
Gamma	0.01	0.01	0.00	1.000	0.00	9.08	8.50	3.30	<0.001	0.78
					(n-Back 1,3,5)					
Delta	3.68	2.70	2.35	0.019	0.39	4.63	3.72	1.41	0.158	0.14
Theta	0.99	1.10	−1.82	0.069	0.24	2.50	2.42	0.47	0.638	0.02
Alpha	1.67	1.65	0.03	0.975	0.00	2.12	1.97	1.16	0.245	0.10
Beta	0.67	0.69	−0.25	0.807	0.00	1.17	1.34	3.18	0.001	0.72
Gamma	0.14	0.15	−1.70	0.090	0.21	1.09	1.24	−2.67	0.008	0.51
					(n-Back 2,4,6)					
Delta	10.48	2.64	3.30	<0.001	0.78	2.53	2.39	0.66	0.510	0.03
Theta	2.45	1.05	3.30	<0.001	0.78	2.37	2.79	−2.61	0.009	0.49
Alpha	4.79	1.50	3.30	<0.001	0.78	1.42	1.69	−2.73	0.006	0.53
Beta	1.16	0.61	3.30	<0.001	0.78	1.71	1.72	−0.16	0.875	0.00
Gamma	**0.17**	**0.11**	**3.19**	**0.001**	**0.73**	**1.47**	**1.38**	**1.95**	**0.052**	**0.27**

Note. Estimated marginal means of EEG power for mental states are reported in μV. *For standard errors please refer to Figures S1–S6 in Supplementary Materials*. Z statistics calculated from Wilcoxon non parametric test for small samples. Rows in bold identify comparisons yielding statistically similar results for both devices. "*" indicates marginal significance.

Two-tailed paired Wilcoxon Signed Rank Tests were used to evaluate data to determine if there was a significant difference between the measures of each wave type at high and low states. To balance for Type I error likelihood, given the low statistical power (since the item analysis was limited to n = 14, the number of electrode-associated repeated measurements), we adopted a correction of the p-value with a generalization (family-wise average rate of false discovery rate) of the Hochberg-Benjamini procedure [65] with alpha 0.09, which gives a significance threshold of $p = 0.054$. Eta squared (μ^2) was computed as a measure of the effect size of the difference between high and low states. Finally, the value of EPOC electrode measurements was represented proportionally to the value of electrode measurements obtained by Brainvision, as a measure of EEG signal measurement *efficiency*. For example, if for gamma waves EPOC detected a signal amplitude of 1.5 μV while Brainvision detected a signal amplitude of 15 μV, Emotiv would be represented as sensing 10% of what Brainvision did.

Lastly, two confirmatory additional "control" analyses were performed (their results are not reported here but are available upon request). Firstly, we repeated the same analyses done on the vincentized averages for the data relative to each of the three professional observers. The patterns of results were very similar, supporting the rationale that the vincentized group data reflect valid inferences at N-of-1 personalization level. Secondly, we compared the non-parametric statistics reported here to the parametric version of the same analyses based on paired t-tests comparisons; the results were virtually identical using either parametric or nonparametric procedures.

3. Results

Considering both EEG features of amplitude and power, there were only very few instances in which EPOC and Brainvision converged in yielding a significant difference in the same direction, from low to high mental state, in the three tasks (7 out of 30 comparisons for amplitude, on a binomial test $p = 0.0003$; 5 out of 30 comparisons for power, on a binomial test $p = 0.0052$). The convergent comparisons are shown in bold in Table 1A,B.

Sensitivity to detect individual significant changes between low and high mental state level in any frequency band and any direction were comparable in both systems for most tasks and states except ANEW where EPOC performed sensibly worse, picking up changes only for Gamma as opposed to all electrodes as in Brainvision, and yielding about half of the effect size on average. Additionally, the performance of EPOC and Brainvision showed equally few effects for the 1,2,3 version of the n-Back task.

In terms of absolute measurement of EEG signal, EPOC demonstrated a variable performance as compared to Brainvision depending on the task. EPOC's efficiencies for the tasks were computed as the ratio of EPOC's measurements over Brainvision measurements; they are represented graphically in Figure 2. When compared to Brainvision measurements in ANEW (see Figure 2A), EPOC's efficiency in measuring amplitude changes was modest for arousal, generally hovering around 40-20% and below 10% for Gamma, and low for valence 30–20% and below 5% for Gamma. In the IAPS (see Figure 2A), EPOC's efficiency in measuring power was even lower for both arousal and valence across all frequencies ranging between 20 and 10%, with Gamma being below 2% or undetected as compared to Braivision measurements. However, EPOC's efficiency was much higher in the n-Back task (See Figure 2B) whereby it was variedly distributed across bands going from low/modest (i.e., Gamma) to medium (i.e., Theta, Beta), on par with or even exceeding (i.e., Alpha and Beta) Brainvision. Notably, in the 2,4,6 n-Back high load state EPOC outperformed Brainvison in all frequency bands except Gamma.

Finally, when the patterns of changes are considered across the frequency spectrum, as summarized in Table 2, it can be observed that there is minimal overlap between the results obtained with EPOC and those with Brainvision across tasks.

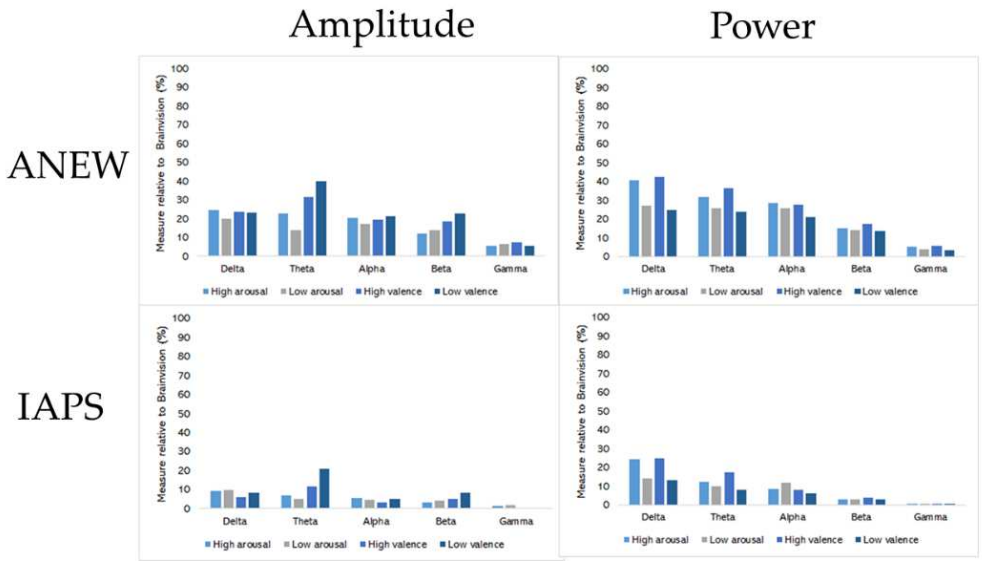
Table 2. Summary of patterns of EEG amplitude and power changes measured with EPOC and Brainvision.

Task	EEG Feature	Mental State	System	Significant Changes from Low to High State
ANEW	Amplitude	Arousal	Epoc	Increase for Gamma
			BV	Decrease for Delta, Theta, and Alpha Increase for Beta and Gamma
		Valence	Epoc	Decrease for Beta Increase for Delta and Gamma
			BV	Increase for all frequency bands
IAPS	Amp	Arousal	Epoc	Increase for all frequency bands
			BV	Decrease for Delta Increase for Theta, Beta and Gamma
		Valence	Epoc	Increase for all frequency bands
			BV	Decrease for Delta, Theta, and Gamma Increase for Alpha *
n-Back	Amp	1-3-5	Epoc	Decrease for Delta
			BV	Increase for Beta and Gamma
		2-4-6	Epoc	Decrease for all frequency bands
			BV	Increase for Theta, Alpha, and Beta
ANEW	Power	Arousal	Epoc	Decrease for Alpha *
			BV	Decrease for Delta, Theta, and Alpha Increase for Beta and Gamma
		Valence	Epoc	Decrease for Theta *, Alpha, and Beta
			BV	Increase for Delta, Beta, and Gamma

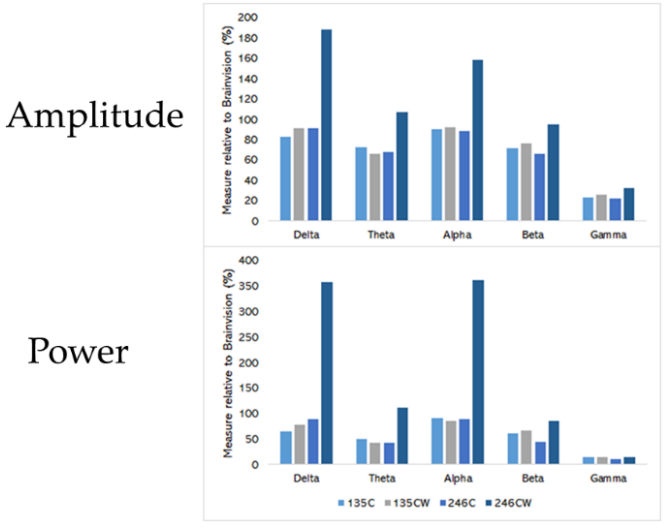
Table 2. *Cont.*

Task	EEG Feature	Mental State	System	Significant Changes from Low to High State
IAPS	Power	Arousal	Epoc	Increase for Delta and Theta
			BV	Decrease for Delta Increase for Theta, Alpha, Beta, and Gamma
		Valence	Epoc	Increase for Delta, Theta, Alpha, and Beta
			BV	Decrease for Delta and Gamma Increase for Alpha
n-Back	Power	1-3-5	Epoc	Decrease for Delta
			BV	Decrease for Beta Increase for Gamma
		2-4-6	Epoc	Decrease for all frequency bands
			BV	Decrease for Gamma Increase for Theta and Alpha

Note. "*" indicates marginally significant effects.



(**A**)



(**B**)

Figure 2. (**A**) Efficiency ratio of EPOC signal strength compared to Brainvision in ANEW and IAPS tasks. (**B**) Efficiency ratio of EPOC signal strength compared to Brainvision in the n-Back tasks.

4. Discussion

The results of our study overall seem to suggest that EPOC is capable of differentiating between high vs low arousal/valence states using the amplitude of different frequency bands, but that the overall signal detected from EPOC was only a fraction of the signal detected by Brainvision. Despite the poorer signal detection, EPOC was still able to differentiate between the mental states via amplitude in most instances, although it had relatively small effect sizes compared to those of Brainvision. For amplitude, it also appears that EPOC most prominently detected differences in low frequency bands such as Delta and Theta. Brainvision on the other hand commonly detected differences of amplitude across all frequency bands.

As above, the same can generally be said about the EPOC for differentiating mental states using EEG power, with a few exceptions. The ability of EPOC to detect Gamma power whatsoever was regularly quite low. Its mean amplitude was 8.7% of the amplitude measured by Brainvision. Even though EPOC was able to identify significant differences between high and low mental states in Gamma waves, the corresponding amplitude values never rose above 1.5 μV, even though Brainvision's measures went up to 13 μV. Since Gamma waves have the lowest amplitude of EEG frequency bands, the difficulty in registering their activity might be caused by the limited sensitivity and limited amount of electrodes on the EPOC headset. Additional factors could also include lower electrode accuracy and higher impedance. EPOC does not use conductive paste but saline solution, which can leave the electrodes more prone to corrosion and dry out more quickly than electrogel or paste.

Unexpectedly, in the instance of the n-Back task 2-4-6 EPOC actually appeared to outperform Brainvision's capability of differentiating high vs low mental load across frequency bands. In this same way, the power detected via EPOC was greater than the power detected by Brainvision. However, Brainvision still performed well in differentiating mental load during the n-Back task with reasonable effect sizes.

Given that the focus of the present study was accuracy and reliability, the tasks we used were selected because they demonstrated validity in distinguishing a relative difference in low and high psychological states as defined by subjective self-ratings of the participants and as reflected by concurrent state-dependent changes in EEG. Our study however has some implications and interesting insights also for research on comparative validity of consumer-grade vs. research-grade EEG devices.

The most EEG recent research focusing on content validity and attempting to relate a particular ERBP feature or signature to a particular mental state has been generally inconclusive, since the findings are contradictory. In particular, the literature on emotional arousal and valence using IAPS yielded as many studies which find a pattern of reduced Alpha power and increased Theta [66] as those which either find null effects, or even show the opposite pattern, i.e., increased Alpha and decreased Delta and/or Theta [67] as well as increased Beta and Gamma [68–70]. One possible plausible conclusion that has been repeatedly proposed is that the specific pattern of results related to EEG power spectral distribution may be therefore dependent on design, stimuli, task, context, and individual differences [71]. The findings also vary according to the electrodes considered, whereby, for example, opposite patterns can be observed in frontal/anterior as compared to parietal/posterior sites. An exhaustive review is outside the scope and space of this paper, but it is reasonable to reach similar conclusions by considering recent meta-analyses on the n-Back task and mental workload in general [69]. The major issue in the context of the present study is overlearning and practice effects in that the professional observers serving as participants in our study underwent a lot of pre-experimental repeated exposure with materials, tasks and both devices. The literature on the effect of the extensive exposure or practice variable on EEG correlates of emotional processing is scant (for example see [72]) but there is evidence for the n-Back task that Alpha and Theta increase with cognitive load [73–75]. Our findings show that both EPOC and Brainvision did not find Alpha desynchronization followed by a concurrent increase of slow frequency oscillations (Delta

or Theta), as reported by some abovementioned literature. In our study, Alpha showed occasional increases or null findings while increases in Delta were rather ubiquitous and inconsistent. The most frequent and consistent changes were increases in Theta, Beta and Gamma going from low to high mental states, which is, as we have cursorily mentioned, in keeping with the literature on practiced mental or emotional processing load. This is both interesting as an extension of research on training of cognitive load and working memory capacity as well as, we believe, a relatively new contribution for the area on emotional processing since there are very few event-related EEG studies on repeated exposure to pleasant or unpleasant stimuli.

In our experiment the sample size was very low, and it was impossible to achieve reliable results that generalize to population distributions. However, because we used an ideographic approach (i.e., electrode-wise analysis of vincentized data) which we can defend to be generalized at the level of person prediction, the actual sample size was relatively unimportant for the explicitly stated scope (tests of reliability) addressed in this paper; this last point holds especially because we took several convergent repeated measurements (N = 14), which were the actual basis for our test of accuracy.

Nonetheless, a possible caveat is that, given our very small sample, it was obviously impossible to properly counterbalance learning effects associated with the device used, namely, EPOC was used first and then Brainvision was used on all subjects, potentially, this order might have induced a learning effect favoring Brainvision. Yet, the rate of significant differences between high and low mental states picked up by EPOC and Brainvision did not show any differential order trend, that is, there were no obvious and consistent differences between sensitivity in detecting significant effects between mental states for the first device used or for the second one, and no advantage of one device over the other can be clearly observed or inferred.

Another related possible caveat might be learning effects of the individual, regardless of the device used. However, as we have pointed out all three participants were graduate student professional observers which were abundantly familiar with stimuli, tasks, and devices. Furthermore, all participants carried out the tasks in a different order. Most likely, these two aspects were sufficient to wash out learning effects due to the task sequence during the experiment.

Our findings would suggest that future applications, such as brain-computer interfaces, requiring a more ecological setup should be able to confidently use EPOC for tasks intending to elicit change in valence/arousal/mental load such as ANEW, IAPS, or n-Back. However, in a laboratory or medical setting the researcher or clinician would still benefit from the quality performance found using Brainvision. Future research should consider ways of optimizing the signal quality of the EPOC so as to obtain larger effect sizes when investigating differences of amplitude/power in response to different mental states. For instance, it would be feasible to leverage machine learning models to enhance the accuracy and classification capabilities of the signals obtained via the EPOC [29]. This would be a promising, although more complex, avenue for increasing the efficacy of the Emotiv device. Modifying the hardware of the EPOC may also be feasible and effective, as Barham et al. (2017) demonstrated that upgrading the EPOC with research-grade electrodes was successful, although the EPOC still had significantly more rejected trials than the research-grade device (Neuroscan) used in their study [26].

5. Conclusions

In summary, the consumer-grade wearable EPOC device appears to be able to differentiate between mental states of valence, arousal, and mental load to some degree, but is overall still not as reliable or accurate as the research-grade Brainvision device. However, EPOC appears to be particularly good at differentiating mental load during an n-Back task via frequency band amplitude/power. Overall, this is in line with previous literature [21], but here we have extended these observations to tasks related to valence, arousal, and mental load.

In principle, EEG measured during high vs low arousal, valence, and mental load, should elicit different EEG signatures (e.g., synchronization/desynchronization and corresponding increases/decreases in amplitude/power), as the mental states require differing levels of emotional or cognitive resources [76–82]. This would be expected from a research-grade EEG device like Brainvision. However, due to the limited amount of evidence for this type of outcome using consumer-grade wearable EEG devices like the EPOC [83,84], further research regarding content validity is still required to confirm this expectation by empirical tests performed using a procedure similar to ours (i.e., by comparing homologous repeated measurements using both devices within the same individuals) and not, as currently done in many review papers (e.g., [20]), by simply assuming that finding a EEG feature or signature with the wireless consumer grade device (e.g., P300) in one sample has the same meaning as reports about the same signature or feature in different samples, even if utilizing comparable experimental paradigms.

Finally, our main objective was to demonstrate an initial methodological benchmark for precision, rather than conducting a comprehensive assessment of all existing consumer-grade EEG devices. However, because the two types of EEG devices compared in the present study are among the most "popular" and widely used for research purposes in labs across the world (especially the Emotiv EPOC, see [85]), the present comparison does contribute a heuristic inductive assessment regarding the precision of, presumably, one of the "best" current EEG consumer instrumentation available as compared to an established (i.e., standard) wired stationary counterpart.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app12136430/s1.

Author Contributions: Conceptualization, A.D.; methodology, A.D., D.M.B. and G.L.-D.; validation, A.D., D.M.B. and G.L.-D.; formal analysis, G.L.-D., D.M.B. and A.D.; investigation, A.D., D.M.B. and G.L.-D.; resources, A.D.; data curation, G.L.-D., A.D. and D.M.B.; writing—original draft preparation, D.M.B., G.L.-D. and A.D.; writing—review and editing, D.M.B., G.L.-D. and A.D.; supervision, A.D.; project administration, A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Defense R&D Canada, by Thales Research and Technology Canada, and by a research partnership grant from the Department of National Defense of Canada and the Natural Sciences and Engineering Research Council of Canada. Open access publication (APC) was funded by a Carleton University Research Impact Endeavour (CURIE) award to A.D.

Institutional Review Board Statement: Ethical review and approval were waived for this study because this was a secondary analysis of completely anonymized archival data.

Informed Consent Statement: Informed verbal consent was obtained from all subjects involved in the original data collection pre-archived studies.

Data Availability Statement: Data may be made available upon request to the authors provided it is for uses in accordance with international ethical, intellectual property and copyright regulations. Public archival of this data is not possible due to privacy restrictions imposed by the funding agencies.

Acknowledgments: The authors would like to acknowledge Jeremy Grant for his role in assisting with the initial data collection and first analysis of this data which was previously published. We thank three anonymous reviewers (especially Reviewer 1) for the numerous insightful comments, criticisms, and suggestions which have helped to greatly improve the paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Smith, S.J.M. EEG in the diagnosis, classification, and management of patients with epilepsy. *J. Neurol. Neurosurg. Psychiatry* **2005**, *76*, ii2–ii7. [CrossRef] [PubMed]

2. Vasiljevic, G.A.M.; De Miranda, L.C. Brain–Computer Interface Games Based on Consumer-Grade EEG Devices: A Systematic Literature Review. *Int. J. Hum. Comput. Interact.* **2019**, *36*, 105–142. [CrossRef]
3. TajDini, M.; Sokolov, V.; Kuzminykh, I.; Shiaeles, S.; Ghita, B. Wireless Sensors for Brain Activity—A Survey. *Electronics* **2020**, *9*, 2092. [CrossRef]
4. Barngrover, C.; Althoff, A.; DeGuzman, P.; Kastner, R. A Brain–Computer Interface (BCI) for the Detection of Mine-Like Objects in Sidescan Sonar Imagery. *IEEE J. Ocean. Eng.* **2015**, *41*, 123–138. [CrossRef]
5. Ganga, R.C.; Vijayakumar, P.; Badrinath, P.; Singh, A.R.; Singh, M. Drone control using EEG signal. *J. Adv. Res. Dyn. Control Syst.* **2019**, *11*, 2107–2113.
6. Munyon, C.N. Neuroethics of Non-primary Brain Computer Interface: Focus on Potential Military Applications. *Front. Neurosci.* **2018**, *12*, 696. [CrossRef] [PubMed]
7. Binnendijk, A.; Marler, T.; Bartels, E.M. *Brain-Computer Interfaces: U.S. Military Applications and Implications, An Initial Assessment*; RAND Corporation: Santa Monica, CA, USA, 2020. [CrossRef]
8. Czech, A. Brain-Computer Interface Use to Control Military Weapons and Tools. In *Control, Computer Engineering and Neuroscience*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 196–204. [CrossRef]
9. Hernandez-Cuevas, B.; Egbert, W.; Denham, A.; Mehul, A.; Crawford, C.S. Changing Minds: Exploring Brain-Computer Interface Experiences with High School Students. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020. [CrossRef]
10. Gnedykh, D. Trends and Prospects of Using Brain-Computer Interfaces in Education. *Sib. Psikhologicheskiy Zhurnal* **2021**, 108–129. [CrossRef] [PubMed]
11. Papanastasiou, G.; Drigas, A.; Skianis, C.; Lytras, M. Brain computer interface based applications for training and rehabilitation of students with neurodevelopmental disorders. A literature review. *Heliyon* **2020**, *6*. [CrossRef]
12. Rohani, D.A.; Puthusserypady, S. BCI inside a virtual reality classroom: A potential training tool for attention. *EPJ Nonlinear Biomed. Phys.* **2015**, *3*, 12. [CrossRef]
13. Liu, N.-H.; Chiang, C.-Y.; Chu, H.-C. Recognizing the Degree of Human Attention Using EEG Signals from Mobile Sensors. *Sensors* **2013**, *13*, 10273–10286. [CrossRef]
14. Thomas, K.P.; Vinod, A.P.; Guan, C. Enhancement of attention and cognitive skills using EEG based neurofeedback game. In Proceedings of the 2013 6th International IEEE/EMBS Conference, San Diego, CA, USA, 6–8 November 2013; pp. 21–24. [CrossRef]
15. Vinod, A.P.; Thomas, K.P. *Neurofeedback Games Using EEG-Based Brain–Computer Interface Technology*; The Institution of Engineering and Technology: London, UK, 2018; pp. 301–329. [CrossRef]
16. Sekhavat, Y.A. Battle of minds: A new interaction approach in BCI games through competitive reinforcement. *Multimedia Tools Appl.* **2019**, *79*, 3449–3464. [CrossRef]
17. Paszkiel, S. Using BCI and VR Technology in Neurogaming. In *Signal Processing and Machine Learning for Brain-Machine Interfaces*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 93–99. [CrossRef]
18. Marshall, D.; Coyle, D.; Wilson, S.; Callaghan, M. Games, Gameplay, and BCI: The State of the Art. *IEEE Trans. Comput. Intell. AI Games* **2013**, *5*, 82–99. [CrossRef]
19. Bos, D.P.; Obbink, M.; Nijholt, A.; Hakvoort, G.; Christian, M. Towards multiplayer BCI games. In Proceedings of the Workshop on Multiuser and Social Biosignal Adaptive Games and Playful Applications, BioS-Play, 2010; Available online: http://www.physiologicalcomputing.net/workshops/biosplay2010/BioSPlay_Gurkok%20et%20al%20(Multiplayer%20BCI).pdf (accessed on 20 May 2022).
20. Sawangjai, P.; Hompoonsup, S.; Leelaarporn, P.; Kongwudhikunakorn, S.; Wilaiprasitporn, T. Consumer Grade EEG Measuring Sensors as Research Tools: A Review. *IEEE Sens. J.* **2019**, *20*, 3996–4024. [CrossRef]
21. Buchanan, D.M.; Grant, J.; D'Angiulli, A. Commercial wireless versus standard stationary EEG systems for personalized emotional brain-computer interfaces: A preliminary reliability check. *Neurosci. Res. Notes* **2019**, *2*, 7–15. [CrossRef]
22. Maskeliunas, R.; Damasevicius, R.; Martisius, I.; Vasiljevas, M. Consumer grade EEG devices: Are they usable for control tasks? *PeerJ* **2016**, *4*, e1746. [CrossRef]
23. Nijboer, F.; Van De Laar, B.; Gerritsen, S.; Nijholt, A.; Poel, M. Usability of Three Electroencephalogram Headsets for Brain–Computer Interfaces: A Within Subject Comparison. *Interact. Comput.* **2015**, *27*, 500–511. [CrossRef]
24. Duvinage, M.; Castermans, T.; Petieau, M.; Hoellinger, T.; Cheron, G.; Dutoit, T. Performance of the Emotiv Epoc headset for P300-based applications. *Biomed. Eng. Online* **2013**, *12*, 56. [CrossRef]
25. Duvinage, M.; Castermans, T.; Dutoit, T.; Petieau, M.; Hoellinger, T.; De Saedeleer, C.; Seetharaman, K.; Cheron, G. A P300-based Quantitative Comparison between the Emotiv Epoc Headset and a Medical EEG Device. In *BioMedical Engineering OnLine*; Springer Nature: Berlin, Germany, 2012. [CrossRef]
26. Barham, M.P.; Clark, G.M.; Hayden, M.J.; Enticott, P.; Conduit, R.; Lum, J. Acquiring research-grade ERPs on a shoestring budget: A comparison of a modified Emotiv and commercial SynAmps EEG system. *Psychophysiology* **2017**, *54*, 1393–1404. [CrossRef]
27. Liu, X.; Chao, F.; Jiang, M.; Zhou, C.; Ren, W.; Shi, M. Towards Low-Cost P300-Based BCI Using Emotiv Epoc Headset. In Proceedings of the UK Workshop on Computational Intelligence, Cardiff, UK, 6–8 September 2017; Volume 650, pp. 239–244. [CrossRef]

28. Badcock, N.A.; Preece, K.A.; de Wit, B.; Glenn, K.; Fieder, N.; Thie, J.; McArthur, G. Validation of the Emotiv EPOC EEG system for research quality auditory event-related potentials in children. *PeerJ* **2015**, *3*, e907. [CrossRef]
29. Fouad, I.A. A robust and reliable online P300-based BCI system using Emotiv EPOC + headset. *J. Med Eng. Technol.* **2021**, *45*, 94–114. [CrossRef]
30. Balanou, E.; van Gils, M.; Vanhala, T. State-of-the-Art of Wearable EEG for Personalized Health Applications. *Stud. Health Technol. Inform.* **2013**, *189*, 119–124. [CrossRef] [PubMed]
31. Smith, P.L.; Little, D.R. Small is beautiful: In defense of the small-N design. *Psychon. Bull. Rev.* **2018**, *25*, 2083–2101. [CrossRef] [PubMed]
32. Molenaar, P.C.M. A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Meas. Interdiscip. Res. Perspect.* **2004**, *2*, 201–218. [CrossRef]
33. Bos, F.M.; Snippe, E.; de Vos, S.; Hartmann, J.A.; Simons, C.J.; van der Krieke, L.; de Jonge, P.B.; Wichers, M. Can we jump from cross-sectional to dynamic interpretationsof networks? Implications for the network perspective in psychiatry. *Psychother. Psychosom.* **2017**, *86*, 175–177. [CrossRef] [PubMed]
34. Molenaar, P.C. On the implications of the classical ergodic theorems: Analysis of developmental processes has to focus on intra-individual variation. *Dev. Psychobiol.* **2007**, *50*, 60–69. [CrossRef]
35. Shah, R.V.; Grennan, G.; Zafar-Khan, M.; Alim, F.; Dey, S.; Ramanathan, D.; Mishra, J. Personalized machine learning of depressed mood using wearables. *Transl. Psychiatry* **2021**, *11*, 1–18. [CrossRef]
36. Birkhoff, G.D. What is the ergodic theorem? *Am. Math. Mon.* **1942**, *49*, 222–226. [CrossRef]
37. Nelson, C.A.; de Haan, M.; Thomas, K.M. *Neuroscience of Cognitivedevelopment: The Role of Experience and the Developing Brain*; Wiley: New York, NY, USA, 2006.
38. Sporns, O. *Networks of the Brain*; MIT Press: Cambridge, MA, USA, 2010.
39. Medaglia, J.D.; Ramanathan, D.M.; Venkatesan, U.M.; Hillary, F.G. The challenge of non-ergodicity in network neuroscience. *Netw. Comput. Neural Syst.* **2011**, *22*, 148–153. [CrossRef]
40. Vincent, S.B. The function of the viborissae in the behavior of the white rat. *Anim. Behav. Monogr.* **1912**, *1*, 84.
41. Atkinson, R.C.; Bower, G.H.; Crothers, E.J. *Introduction to Mathematical Learning Theory*; Wiley: Hoboken, NJ, USA, 1965.
42. Bedny, M.; Aguirre, G.; Thompson-Schill, S.L. Item analysis in functional magnetic resonance imaging. *NeuroImage* **2007**, *35*, 1093–1102. [CrossRef]
43. Makeig, S.; Debener, S.; Onton, J.; Delorme, A. Mining event-related brain dynamics. *Trends Cogn. Sci.* **2004**, *8*, 204–210. [CrossRef] [PubMed]
44. Pfurtscheller, G. and Aranibar, A. Event-related cortical desychronization detected by power measurement of scalp EEG. *Electroencephalograph. Clin. Neurophysiol.* **1977**, *42*, 817–826. [CrossRef]
45. Byczynski, G.; Schibli, K.; Goldfield, G.; Leisman, G.; D'Angiulli, A. EEG Power Band Asymmetries in Children with and without Classical Ensemble Music Training. *Symmetry* **2022**, *14*, 538. [CrossRef]
46. D'Angiulli, A.; Kenney, D.; Pham, D.A.T.; Lefebvre, E.; Bellavance, J.; Buchanan, D.M. Neurofunctional Symmetries and Asymmetries during Voluntary out-of- and within-Body Vivid Imagery Concurrent with Orienting Attention and Visuospatial Detection. *Symmetry* **2021**, *13*, 1549. [CrossRef]
47. Klimesch, W. Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* **2012**, *16*, 606–617. [CrossRef]
48. Stevenson, R.A.; Mikels, J.A.; James, T.W. Characterization of the Affective Norms for English Words by discrete emotional categories. *Behav. Res. Methods* **2007**, *39*, 1020–1024. [CrossRef]
49. Bradley, M.; Lang, P.J. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings (Technical Report C-1)*; Gainesv Cent Res Psychophysiology, University of Florida: Gainesville, FL, USA, 1999.
50. Bradley, M.M.; Lang, P.J. *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings (Technical Report C-2)*; University of Florida: Gainesville, FL, USA, 2010.
51. Imbir, K.K. Affective Norms for 4900 Polish Words Reload (ANPW_R): Assessments for Valence, Arousal, Dominance, Origin, Significance, Concreteness, Imageability and, Age of Acquisition. *Front. Psychol.* **2016**, *7*, 1081. [CrossRef]
52. Lang, P.J.; Bradley, M.M.; Cuthbert, B.N. *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual*; Technical Report A-8; University of Florida: Gainesville, FL, USA, 2008.
53. Hajcak, G.; Dennis, T.A. Brain potentials during affective picture processing in children. *Biol. Psychol.* **2009**, *80*, 333–338. [CrossRef]
54. Kirchner, W.K. Age differences in short-term retention of rapidly changing information. *J. Exp. Psychol.* **1958**, *55*, 352–358. [CrossRef]
55. Mackworth, J.F. Paced memorizing in a continuous task. *J. Exp. Psychol.* **1959**, *58*, 206–211. [CrossRef]
56. Gajewski, P.D.; Hanisch, E.; Falkenstein, M.; Thönes, S.; Wascher, E. What Does the n-Back Task Measure as We Get Older? Relations Between Working-Memory Measures and Other Cognitive Functions Across the Lifespan. *Front. Psychol.* **2018**, *9*, 2208. [CrossRef] [PubMed]
57. Scharinger, C.; Soutschek, A.; Schubert, T.; Gerjets, P. Comparison of the Working Memory Load in N-Back and Working Memory Span Tasks by Means of EEG Frequency Band Power and P300 Amplitude. *Front. Hum. Neurosci.* **2017**, *11*, 6. [CrossRef]

58. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council, Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, December 2018. Available online: https://ethics.gc.ca/eng/documents/tcps2-2018-en-interactive-final.pdf (accessed on 1 June 2022).
59. D'Angiulli, A.; Griffiths, G.; Marmolejo-Ramos, F. Neural correlates of visualizations of concrete and abstract words in preschool children: A developmental embodied approach. *Front. Psychol.* **2015**, *6*, 856. [CrossRef] [PubMed]
60. Zhou, X.; Li, M.; Zhou, H.; Li, L.; Cui, J. Item-Wise Interindividual Brain-Behavior Correlation in Task Neuroimaging Analysis. *Front. Neurosci.* **2018**, *12*, 817. [CrossRef]
61. D'Angiulli, A.; Pham, D.A.T.; Leisman, G.; Goldfield, G. Evaluating Preschool Visual Attentional Selective-Set: Preliminary ERP Modeling and Simulation of Target Enhancement Homology. *Brain Sci.* **2020**, *10*, 124. [CrossRef]
62. Genest, C. Vincentization Revisited. *Ann. Stat.* **1992**, *20*, 1137–1142. [CrossRef]
63. Delorme, A.; Makeig, S. EEGLAB: An Open Source Toolbox for Analysis of Single-Trial EEG Dynamics Including Independent Component Analysis. *J. Neurosci. Methods* **2004**, *134*, 9–21. [CrossRef]
64. Badcock, N.A.; Mousikou, P.; Mahajan, Y.; De Lissa, P.; Thie, J.; McArthur, G. Validation of the Emotiv EPOC® EEG gaming systemfor measuring research quality auditory ERPs. *PeerJ* **2013**, *1*, e38. [CrossRef]
65. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **1995**, *57*, 289–300. [CrossRef]
66. Schubring, D.; Schupp, H.T. Emotion and brain oscillations: High arousal is associated with decreases in alpha-and lower beta-band power. *Cerebral Cortex* **2021**, *31*, 1597–1608. [CrossRef]
67. Güntekin, B.; Femir, B.; Gölbaşı, B.T.; Tülay, E.; Başar, E. Affective pictures processing is reflected by an increased long-distance EEG connectivity. *Cogn. Neurodyn.* **2017**, *11*, 355–367. [CrossRef] [PubMed]
68. Güntekin, B.; Başar, E. Event-related beta oscillations are affected by emotional eliciting stimuli. *Neurosci. Lett.* **2010**, *483*, 173–178. [CrossRef] [PubMed]
69. Strube, A.; Rose, M.; Fazeli, S.; Büchel, C. Alpha-to-beta-and gamma-band activity reflect predictive coding in affective visual processing. *Sci. Rep.* **2021**, *11*, 23492. [CrossRef] [PubMed]
70. Yeo, D.; Choi, J.W.; Kim, K.H. Increased Gamma-band Neural Synchrony by Pleasant and Unpleasant Visual Stimuli. *J. Biomed. Eng. Res.* **2018**, *39*, 94–102.
71. Chikhi, S.; Matton, N.; Blanchet, S. EEG power spectral measures of cognitive workload: A meta-analysis. *Psychophysiology* **2022**, *59*, e14009. [CrossRef]
72. Güntekin, B.; Tülay, E. Event related beta and gamma oscillatory responses during perception of affective pictures. *Brain Res.* **2014**, *1577*, 45–56. [CrossRef]
73. Blacker, K.J.; Negoita, S.; Ewen, J.B.; Courtney, S.M. N-back versus complex span working memory training. *J. Cogn. Enhanc.* **2017**, *1*, 434–454. [CrossRef]
74. Liu, Y.; Ayaz, H.; Onaral, B.; Shewokis, P.A. Neural Adaptation to a Working Memory Task: A Concurrent EEG-fNIRS Study. In Proceedings of the 2015 International Conference on Augmented Cognition, Los Angeles, CA, USA, 7 July–2 August 2015; pp. 268–280. [CrossRef]
75. Gevins, A.; Smith, M.E.; McEvoy, L.; Yu, D. High-resolution EEG mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice. *Cereb. Cortex* **1997**, *7*, 374–385. [CrossRef]
76. Klotzsche, F.; Mariola, A.; Hofmann, S.; Nikulin, V.V.; Villringer, A.; Gaebler, M. Using EEG to Decode Subjective Levels of Emotional Arousal During an Immersive VR Roller Coaster Ride. In Proceedings of the 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Christchurch, New Zealand, 18–22 March 2018; pp. 605–606. [CrossRef]
77. Sarma, P.; Barma, S. Emotion recognition by distinguishing appropriate EEG segments based on random matrix theory. *Biomed. Signal Process. Control* **2021**, *70*, 102991. [CrossRef]
78. Duma, G.M.; Mento, G.; Semenzato, L.; Tressoldi, P. EEG anticipation of random high and low arousal faces and sounds. *F1000Research* **2019**, *8*, 1508. [CrossRef]
79. Gummadavelli, A.; Kundishora, A.J.; Willie, J.T.; Andrews, J.; Gerrard, J.; Spencer, D.D.; Blumenfeld, H. Neurostimulation to improve level of consciousness in patients with epilepsy. *Neurosurg. Focus* **2015**, *38*, E10. [CrossRef] [PubMed]
80. Aftanas, L.; Golocheikine, S. Human anterior and frontal midline theta and lower alpha reflect emotionally positive state and internalized attention: High-resolution EEG investigation of meditation. *Neurosci. Lett.* **2001**, *310*, 57–60. [CrossRef]
81. Müller-Bardorff, M.; Schulz, C.; Peterburs, J.; Bruchmann, M.; Mothes-Lasch, M.; Miltner, W.; Straube, T. Effects of emotional intensity under perceptual load: An event-related potentials (ERPs) study. *Biol. Psychol.* **2016**, *117*, 141–149. [CrossRef]
82. Boring, M.J.; Ridgeway, K.; Shvartsman, M.; Jonker, T.R. Continuous decoding of cognitive load from electroencephalography reveals task-general and task-specific correlates. *J. Neural Eng.* **2020**, *17*, 056016. [CrossRef] [PubMed]
83. Wang, S.; Gwizdka, J.; Chaovalitwongse, W.A. Using Wireless EEG Signals to Assess Memory Workload in the n -Back Task. *IEEE Trans. Human-Machine Syst.* **2015**, *46*, 424–435. [CrossRef]
84. Kutafina, E.; Heiligers, A.; Popovic, R.; Brenner, A.; Hankammer, B.; Jonas, S.M.; Mathiak, K.; Zweerings, J. Tracking of Mental Workload with a Mobile EEG Sensor. *Sensors* **2021**, *21*, 5205. [CrossRef] [PubMed]
85. Williams, N.S.; McArthur, G.M.; Badcock, N.A. 10 years of EPOC: A scoping review of Emotiv's portable EEG device. *BioRxiv* **2020**. [CrossRef]