Estimating cognitive workload using a commercial in-ear EEG headset

Journal of Neural Engineering

PAPER

Estimating cognitive workload using a commercial in-ear EEG headset

Christoph Tremmel[1,*], Dean J Krusienski[2] and mc schraefel[1]

[1] Wellthlab, School of Electronics and Computer Science, University of Southampton, Southampton, United Kingdom
[2] Department of Biomedical Engineering, Virginia Commonwealth University, Richmond, VA, United States of America
* Author to whom any correspondence should be addressed.

E-mail: christoph.tremmel@southampton.ac.uk

Keywords: in-ear EEG, cognitive workload, EEG, BCI

OPEN ACCESS

RECEIVED
1 May 2024

REVISED
11 October 2024

ACCEPTED FOR PUBLICATION
5 November 2024

PUBLISHED
2 December 2024

Abstract

Objective. This study investigated the potential of estimating various mental workload levels during two different tasks using a commercial in-ear electroencephalography (EEG) system, the IDUN 'Guardian'. *Approach.* Participants performed versions of two classical workload tasks: an n-back task and a mental arithmetic task. Both in-ear and conventional EEG data were simultaneously collected during these tasks. In an effort to facilitate a more comprehensive comparison, the complexity of the tasks was intentionally increased beyond typical levels. Special emphasis was also placed on understanding the significance of γ band activity in workload estimations. Therefore, each signal was analyzed across low frequency (1–35 Hz) and high frequency (1–100 Hz) ranges. Additionally, surrogate in-ear EEG measures, derived from the conventional EEG recordings, were extracted and examined. *Main results.* Workload estimation using in-ear EEG yielded statistically significant performance levels, surpassing chance levels with 44.1% for four classes and 68.4% for two classes in the n-back task and was better than a naive predictor for the mental arithmetic task. Conventional EEG exhibited significantly higher performance compared to in-ear EEG, achieving 80.3% and 92.9% accuracy for the respective tasks, along with lower error rates than the naive predictor. The developed surrogate measures achieved improved results, reaching accuracies of 57.5% and 85.5%, thus providing insights for enhancing future in-ear systems. Notably, most high frequency range signals outperformed their low frequency counterparts in terms of accuracy validating that high frequency γ band features can improve workload estimation. *Significance.* The application of EEG-based Brain–Computer Interfaces beyond laboratory settings is often hindered by practical limitations. In-ear EEG systems offer a promising solution to this problem, potentially enabling everyday use. This study evaluates the performance of a commercial in-ear headset and provides guidelines for increased effectiveness.

1. Introduction

Electroencephalography-based (EEG) brain–computer interfaces (BCIs) are a widely investigated method for acquiring and interfacing brain signals [1], with applications largely confined to laboratory and medical settings. The limited use outside these environments is attributed to the predominant focus on direct BCI control paradigms [2] which often feature non-intuitive human–computer-interaction, coupled with the lack of practical commercial EEG devices that do not require technical expertise for operation, potentially awkward head mounts, or electrode gel, for example [3]. BCIs can be classified into active, passive and reactive interfaces [2]. Active control is achieved by interpreting EEG signals associated with volitional mental imagery such as imagined movements to directly control devices such as computers, wheelchairs, or prosthetics [4]. Reactive BCIs translate EEG signals associated with designated sensory stimuli, such as a grid of flashing letters, to make selections corresponding to the stimuli, as seen in spellers [5–7]. However, both of these methods often require extensive user training [8], rely on sensory stimuli that may feel unnatural or intrusive, and are not always reliable [9]. These issues can

lead to user demotivation and limit the long-term use of BCIs [10]. In contrast, implicit or passive BCI control, which involves passively monitoring the user's cognitive or affective state to influence some auxiliary aspect, can offer a more natural interaction [11–13].

Passive BCIs can be designed to be less sensitive to decoding errors, potentially making them less noticeable and distracting to the user compared to direct BCI control. Multiple affective states can be estimated from EEG such as arousal [14, 15], valence [15, 16] and vigilance [17, 18]. One mental state that has been extensively researched is mental workload [19–26]. Previous studies largely utilized various combinations of traditional EEG power spectral bands θ (5–7 Hz), α (8–12 Hz), β (13–30 Hz), γ (>30 Hz) to characterize mental workload. A meta-analysis of 24 studies concluded that the θ band power, particularly in the frontal region, is the most sensitive to changes in mental workload, although α and β also show significant changes [27]. It is also important to note that the α band power is generally negatively correlated compared to θ and β, but there also have been reports of negative β correlation, which may be due to the motor activity required by the tasks [28, 29]. Due to the limitations of scalp EEG, γ activity has been less frequently reported in cognitive workload studies. However, some studies have shown increased γ band activity with increased workload [30, 31]. It remains unclear whether this increase is solely due to brain activity or also due to increased muscle activity or tension stemming from the frontalis and temporalis muscles associated with changing workload [32]. Furthermore, it should be noted that these tendencies are often only assessed on data averaged across multiple participants. The task-related brain activity can vary substantially across users [24].

As with active and reactive BCIs, the setup of passive BCIs involves the standard EEG cap, which requires a tight fit, wet electrodes filled with electrolyte gel, and technical expertise to set up the system [3]. Some progress has been made in the design of dry electrodes that operate without conductive electrode gel [33]. The downside is that dry electrodes require additional pressure to maintain scalp contact and are more susceptible to noise [34]. An alternative approach for recording EEG signals that has gained popularity in recent years [35] involves the use of electrodes situated in and around the ear. The reduced amount of dense hair in those areas eliminates the need for electrode gel. Since these devices can be worn like headphones or in-ear plugs, they are more user-friendly with increased comfort levels that make long-term measurements more acceptable and more inconspicuous to wear in public. This design, however, also presents specific challenges that must be addressed. The device must achieve a balance between practicality and reliability. It needs to be lightweight and minimally obtrusive to the user, while still collecting sufficient data for accurate analysis. Most study devices incorporate two electrodes in the ear canal [36–39]. However, some studies have found this configuration too limiting and have employed additional electrodes placed either around the ear [26, 40, 41] or by integrating multiple electrodes within the earpieces [42]. These electrodes are typically dry, which introduces common challenges such as susceptibility to noise [34] and the need for pressure to maintain a good connection for positions outside of the ear canal. Additionally, the varying shapes of ear canals can pose problems for electrodes inside the canal, leading some researchers to use custom-made individualized earpiece electrodes [38, 42]. Furthermore, muscle movements around the ear canal, such as those from the auricular muscles, can potentially induce artifacts in the electrode signals.

In the last decade, multiple research teams have reproduced classical EEG paradigms such as event-related potentials [40–42], arousal/valence prediction [39], motor imagery [43], drowsiness prediction [37], steady-state visual evoked potentials [36, 38] and workload [26] using in- or around-ear electrodes. All these studies either modified an existing EEG amplifier or developed an original prototype device to be used in or around the ear.

1.1. Contributions of this study

This study investigates the feasibility of a commercial in-ear EEG system, the IDUN 'Guardian', to estimate different mental workload levels during tasks. In-ear EEG offers a potential solution to the aforementioned challenges, providing a more streamlined procedure that could enhance comfort and enable everyday use. Participants performed versions of two classical workload tasks, an n-back task and a mental arithmetic task. Both in-ear and conventional EEG data were simultaneously collected during these tasks. In an effort to provide a more comprehensive comparison, the complexity of the tasks was intentionally increased beyond what is typically implemented. Specifically, four levels of difficulty were implemented for the n-back task, rather than the typical 2–3 levels used in prior studies [20, 22, 26]. For the mental arithmetic task, a regression-based approach was employed where the difficulty of mathematical operations is based on its q-value [44]. The q-value is a measure of task difficulty that considers the problem size as well as the need for a carry-over operation and has previously been utilized in EEG-based workload prediction [45]. The impact of different workload levels was analyzed for both tasks and the respective performances of the in-ear EEG system and standard EEG systems were compared. Special focus was placed on understanding the significance of γ band activity in workload estimations. Therefore, each signal was analyzed across low frequency (1–35 Hz) and

high frequency (1–100 Hz) ranges. In order to address some of the limitations of in-ear, a surrogate in-ear EEG derived from the conventional EEG recordings was devised and examined. These surrogates consist of multiple channels, allowing them to overcome the limited data generated by just two in-ear electrodes and to apply more advanced denoising algorithms that leverage the additional data. The study concludes with recommendations for improving in-ear EEG systems based on the insights gained from these surrogates, offering guidance for future enhancements and their application in workload estimation.

2. Materials and methods

Participants were recruited from staff and students of the University of Southampton and provided written informed consent in accordance with the University of Southampton's ethical guidelines and the Declaration of Helsinki. Ethical approval was granted by the University's ethics committee in March 2023 (ERGO Application ID: 81043). Each experimental session lasted roughly two hours including preparation time. 17 participants (12 Male, Mean Age 30.8 STD 6.5) enrolled in this study and each received 10 GBP in compensation.

Participants sat comfortably in front of an LCD screen while connected to 26-electrode EEG, 2-electrode electrocardiogram (ECG) and 4-electrode electrooculogram (EOG) using BrainProducts' 'actiChamp plus' system. Additionally, participants wore a 2-electrode wireless in-ear EEG system 'Guardian' by IDUN technologies and were recorded by an ELP webcam model USBFHD01M-SFV. The 'Guardian' is a lightweight EEG headset that is worn around the back of the head and measures electrical brain activity using earbud electrodes. These electrodes are dry and flexible to provide the necessary skin contact in the ear canal. The ground electrode is located at one of the earhooks. It connects to the PC via Bluetooth. Figure 1 shows a schematic image of the headset. The internal measurement unit to acquire head acceleration data was not accessible as the device was received during beta testing. EEG, ECG, and EOG were sampled at 500 Hz, in-ear at 250 Hz, and the video at 30 Hz. All signals were synchronized using LabStreamingLayer [46]. ECG and video were not included in the present analysis since the objective is to compare the performance of the in-ear and conventional EEG signals for estimating mental workload.

As part of the 'ActiChamp' system, 'ActiCap' slim wet electrodes were placed at positions Fp1, F7, F3, Fz, FT9, T7, C3, Cz, TP9, CP1, P7, P3, Pz, O1, FP2, F4, F8, FT10, C4, T8, CP2, TP10, CP2, P4, P8, O2 and Oz according to the International 10–20 system [47]. Two EOG electrodes were placed above and below the right eye to capture the vertical eye movement signal as well as two near the canthus of each eye for the horizontal signal. Before inserting the in-ear earbud electrodes, participants cleaned their ears and coated the inside of their ears with saline solution using a cotton swab to improve conductivity. Good signal quality was ensured for each participant by impedance check and visual inspection for EEG and in-ear EEG. For in-ear in particular the impedance values for all our participants were checked to be below 300 kΩ. One participant had to be excluded because the in-ear EEG impedance was not stable. This decreased the number of participants included in the present analysis to 16.

2.1. Experimental task

The experiment was divided into two experimental blocks, (1) decision confidence and (2) mental workload, with two experimental tasks each. The order of the blocks was alternated across participants. For this study, only the outcome of the workload block is reported. It consisted of two experimental tasks, the n-back task and the mental arithmetic task.

The n-back task was performed in four difficulty levels from 0 to 3-back. Each difficulty level consisted of two blocks of 18 trials each. The first six participants performed the experiment in order of increasing difficulty levels while the rest completed it in pseudo-randomized order of difficulty to examine the effect of trial order on performance. The current n-back level was continuously displayed at the top of the screen to help participants stay on task. Stimuli were randomized sequences of numbers from 0 to 9, presented in the center of the screen for 2.5 s per trial. Each new trial was signaled by a bell sound, providing additional auditory feedback. During this time window, participants had to use the mouse to identify a stimulus as a target or non-target. A left click indicated a target and a right click a non-target. Feedback was provided by displaying the words 'target' or 'nontarget' on the screen based on their response. If participants failed to respond in time, 'Timeout!' was displayed for 50 ms at the end of the trial. For the 0-back condition, the number 0 was the target. Otherwise, a stimulus was a target if it appeared 'n' stimuli earlier in the sequence. Sequences were generated so that there were exactly 5 targets in each block. The total number of trials was 18 (trials per block) $\times$ 2 (blocks per difficulty level) $\times$ 4 (difficulty level) $=$ 144. Before starting the task, participants were instructed to minimize head movements to reduce movement-related artifacts. Additionally, a practice run was conducted where participants had to correctly identify all targets at each difficulty level to ensure task compliance.

The mental arithmetic task consisted of the mental addition of two positive integers ranging from 1 to 876, randomly selected based on the intended difficulty level. Each trial lasted 8.5 s. The

Figure 1. The IDUN 'Guardian' in-ear EEG headset version 2.1.

equation representing the addition to be performed was displayed in the middle of the screen during the first 5 s. Afterward, the participant had 3.5 s to type in the answer using the keyboard and confirm by pressing the 'Enter' key. A confirmed result would change the font of the text to italics to give feedback to the participant. Similarly to the n-back task, 'Timeout!' was displayed for 50 ms at the end of the trial if a participant failed to respond. The experimental setup was inspired by [45]. The difficulty of the equation was altered based on the q-value. This experimental task was performed for eight difficulty levels, starting from a q-value of 0.7 in uniform increments up to 4.55. To sample additional difficulty levels, the q-value of each trial was also randomly perturbed by adjusting the q-value of the uniform increments by ± 0.3. A total of 18 trials were performed for each participant and difficulty level, with the order of difficulty level increasing for the first six participants and pseudo-randomized for the rest. Similar to the n-back task, this yields a total of 144 trials per participant: 18 (trials per difficulty level) $\times$ 8 (difficulty level). Both experiments were designed so that each participant had the exact same number of trials, therefore no trials were excluded from further analysis to ensure balanced datasets.

2.2. Data pre-processing and feature extraction

The pre-processing for both tasks was performed similarly, the only difference was the trial window length. To compare EEG and in-ear recordings, multiple in-ear surrogates were created from different conventional EEG configurations. The 1-channel in-ear surrogate was created by using a bipolar derivation between channels TP9 and TP10. This is the most similar representation by using the conventional electrodes closest to the ear with the same two-electrode derivation as the in-ear. The 2-channel surrogate also consisted of the electrodes TP9 and TP10 but both channels were referenced to the mean of channels T7 and T8. This configuration simulates in-ear EEG with an additional reference electrode. The 4-channel surrogate consisted of the channels FT9, FT10, TP9, and TP10 and was also referenced to the mean of T7 and T9. This configuration simulates an additional pair of channels around the ear. The conventional EEG channels were referenced to the average of the electrodes located at TP9 and TP10, resulting in 24 total channels. EOG channels were derived by subtracting the two vertical EOG electrodes as well as the horizontal EOG electrodes, respectively.

The recorded signals were pre-processed in two configurations: (1) filtered between 0.5 and 35 Hz and down-sampled to 100 Hz and (2) filtered between 0.5 and 100 Hz and down-sampled to 200 Hz. The first is used for direct comparison to literature as most studies have not used the γ band activity, while the second examines the potential of using higher frequencies for the workload estimations. Additionally, a 50 Hz notch filter was applied to all signals to suppress any power-line artifacts.

Since EEG and in-ear signals were recorded from different devices, precise synchrony had to be ensured. Instead of relying on the native timestamps from each device, we used the timestamps provided by LSL. This approach revealed a minor discrepancy in the native timestamps, which would amount to a difference of approximately one second after an hour of recording. To address this, we re-sampled the down-sampled in-ear signal by interpolating it to match the LSL timestamps of the conventional EEG at 100 Hz and 200 Hz, respectively. Afterwards, a cross-correlation between the in-ear signal and the horizontal EOG was performed using a ± 1 s bound. The

in-ear signal was aligned to the other measures using the maximum of this cross-correlation. The horizontal EOG was used because it exhibits a disctinct and robust response for aligning the in-ear and conventional EEG recordings.

Eye movement artifacts were suppressed in the resulting signals by creating a linear regression model to predict vertical and horizontal EOG from each EEG channel [48]. These predictions were then subtracted from the corresponding channels. For in-ear and surrogate channels, this process was repeated using only the horizontal EOG since the positions of the electrodes made it unlikely to capture signals associated with vertical EOG. The removal of EOG signals in EEG and in-ear was applied to isolate and examine brain activity relevant to the task. While EOG has previously been used to estimate workload directly or to support EEG [49], a supplemental analysis showed that the combination of neural and ocular data into a single channel did not significantly change the outcomes of the present study. For the statistical analysis, a Wilcoxon sign-rank test with Benjamini–Hochberg correction was performed to compare the processing pipeline performance with and without the addition of eye movement artifact suppression, while keeping the remaining pipeline identical. This resulted in 20 tests for each pipeline with none indicating significant differences between the two conditions as displayed in table 4 in the appendix. A common average reference filter [50] was applied to conventional EEG and the 4-channel surrogate to suppress spatial noise. This process resulted in 10 different signals:

- At lower frequency range of 0.5–35 Hz: in-ear EEG, 1-channel in-ear surrogate, 2-channel in-ear surrogate, 4-channel in-ear surrogate, and conventional EEG.
- At higher frequency range of 0.5–100 Hz: in-ear EEG, 1-channel in-ear surrogate, 2-channel in-ear surrogate, 4-channel in-ear surrogate, and conventional EEG.

All signals were then divided into epochs based on the task. For the n-back task, 2.5 s non-overlapping epochs were used based on the trial duration. For the mental arithmetic task, 5 s non-overlapping epochs were used, which corresponded to the task presentation time to exclude any movement-related brain activity generated by the keyboard input. The power spectral density (PSD) of these epochs was then calculated by using the median of a moving short-time Fourier transform with 0.5 s overlap and a Hamming window. The median was used instead of the mean because it is less sensitive to potential extreme values that could be acquired by the in-ear EEG since it is more susceptible to motion artifacts due to its location in the ear canal. For the lower frequency range, frequencies from 1 to 35 Hz with 1 Hz resolution were utilized, yielding 35 features for each channel. For the higher frequency range, frequencies from 1 to 100 Hz with 1 Hz resolution were utilized, yielding 100 features for each channel.

2.3. Workload prediction approaches

For the n-back task, a regularized linear discriminant analysis classifier was chosen because of the expected linear relationship between workload and PSD [27] and the relatively low number of trials for EEG experiments. The classification results are examined for predicting all four classes, as well as the binary prediction of the extreme workload levels (0-back and 3-back). A 5-fold cross validation with hyperparameter optimization of the regularization parameter and the linear coefficient threshold was performed. These hyperparameters were especially important for classifying EEG since they counteracted the large number of feature dimensions for multi-channel prediction approaches.

For the mental arithmetic task, ridge regression was chosen to estimate the q-value directly. To give a more intuitive representation the q-value was normalized to a range of 0 and 1, respectively representing the easiest to the most difficult operation possible. A 5-fold cross validation was used with optimization of the regularization parameter, again to counteract the large number of feature dimensions. As an additional measure of quality, the Pearson's correlation between estimated and the true value is reported.

The number of feature dimensions for both tasks varied for each signal and frequency range. For the low-frequency range, the number of feature dimensions was 35 for in-ear and 1-channel in-ear surrogate, 70 for the 2-channel surrogate, 140 for the 4-channel surrogate and 840 for conventional EEG, for the high frequency range it was 100 for in-ear and 1-channel in-ear surrogate, 200 for the 2-channel surrogate, 400 for the 4-channel surrogate and 2400 for conventional EEG for each trial.

3. Results

The average participant task performance (correctly identifying the target) for the n-back task was $97 \pm 2.34\%$ for 0-back, $96 \pm 4.82\%$ for 1-back, $88 \pm 1.91\%$ for 2-back and $82 \pm 3.86\%$ for 3-back, while the average reaction time was 0.74 ± 0.13 s for 0-back, 0.8 ± 0.1 s for 1-back, 1.31 ± 0.11 s for 2-back and 1.36 ± 0.13 s for 3-back. The average participant performance for the mental arithmetic task is reported in figure 2. For low difficulty levels the average performance is at nearly 100% and continually decreases down to 10% at the highest levels. The reaction time behaves opposite and increases from 1–1.5 s for low difficulty to 3 s for high difficulty levels in average.

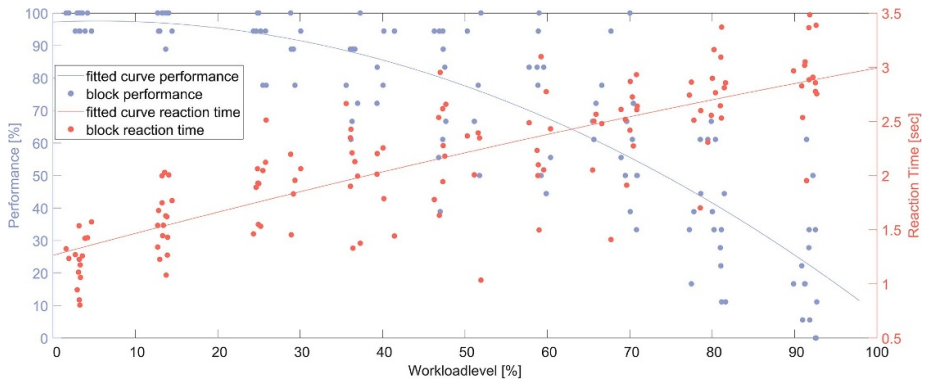

Figure 2. Scatter plots with a fitted regression curves that show the average participant performance during the mental arithmetic task with respect to the task difficulty level. The workload level here represents the q-value normalized to a range of 0%–100%.

To confirm that the trial order did not have any significant impact on the results, a Wilcoxon rank-sum statistical test with Benjamini–Hochberg correction was performed, comparing the average reaction time, accuracy, and all task performance results for the two groups of participants. This resulted in 22 tests for each task with none indicating significant differences between both groups as displayed in table 3 in the appendix.

The performance of the classifier to correctly identify workload levels for the n-back task is illustrated in figure 3. The horizontal lines display the chance level for 4 classes (25%) and 2 classes (50%), respectively. The average, median and most of the individual performances of all analysis conditions were above chance level. Both low frequency in-ear (mean 41.7%, median 36.5%) and 1-channel surrogate (38.2%, 37.4%) demonstrated comparable performance that improved with an expanded frequency range, evidenced by the higher performance of high frequency in-ear (44.4%, 42.4%) and 1-channel surrogate (41%, 42%) compared to their low frequency counterparts. Furthermore, increasing the number of channels enhanced classification accuracy, particularly in the high frequency configuration. The 2-channel low frequency surrogate (43.4%, 43.4%) performed similarly to the high frequency 1-channel surrogate, and the 4-channel low frequency surrogate (48.3%, 49.6%) performed similarly to the 2-channel high frequency surrogate (48.3%, 51.4%). The high frequency 4-channel surrogate (57.7%, 58.7%) yielded the best performance among the surrogates, nearly reaching the performance of conventional low frequency EEG (62.8%, 66.3%). High frequency EEG achieved the highest performance overall (80.3%, 83%).

For 2 classes, the results generally followed the same trends. In-ear (67.7%, 67.7%) and 1-channel surrogate (64.6%, 66.6%) again performed comparably. Increasing the frequency range improved the performance of both in-ear (68.4%, 68.8%) and 1-channel surrogate (69.8%, 73.6%). Similarly, adding more channels improved the results for the surrogates. 2-channel low frequency configuration (72.4%, 70.1%) performed similarly to the high frequency 1-channel surrogate and 4-channel low frequency (75.5%, 75.1%) performed similarly to the 2-channel high frequency surrogate (77.1%, 79.8%). Notably, the only difference presented in the 4-channel high frequency configuration (85.5%, 86.1%) which could outperform the low frequency EEG (84.5%, 85.2%). High frequency range EEG again resulted in the best overall performance (92.9%, 94.5%).

A statistical evaluation of the approaches is presented in table 1. A one-sided Wilcoxon signed-rank test was used to compare the classification results of chance levels, high frequency in-ear, and high frequency 1-channel in-ear surrogate to all approaches. A Benjamini–Hochberg correction was applied to account for the multiple comparison problem. The p-values in the table indicate that the differences in performances of the approaches listed the left side of the table to those listed on the top, based on a distribution with a positive median, i.e. the approaches on the left of the table performed better. It is observed that all approaches performed significantly better than the chance level for four and two classes. For four classes, EEG in both configurations and the 4-channel high frequency surrogate was significantly better than both high frequency in-ear and 1-channel surrogate. However, the 4-channel low frequency surrogate was significantly better than the 1-channel surrogate. For the two-class problem, again EEG in both configurations and the 4-channel high frequency show significantly better results for both approaches.

The performance in the form of mean square error (MSE) and Pearson's correlation (converted to a percentage) is shown in figure 4. The mean estimate line at 8.2 shows the value for always predicting the average performance and serves as a baseline for comparison. Low frequency in-ear performs slightly better than this baseline (mean 7.7, median 7.8) and similarly to the 1-channel low frequency surrogate

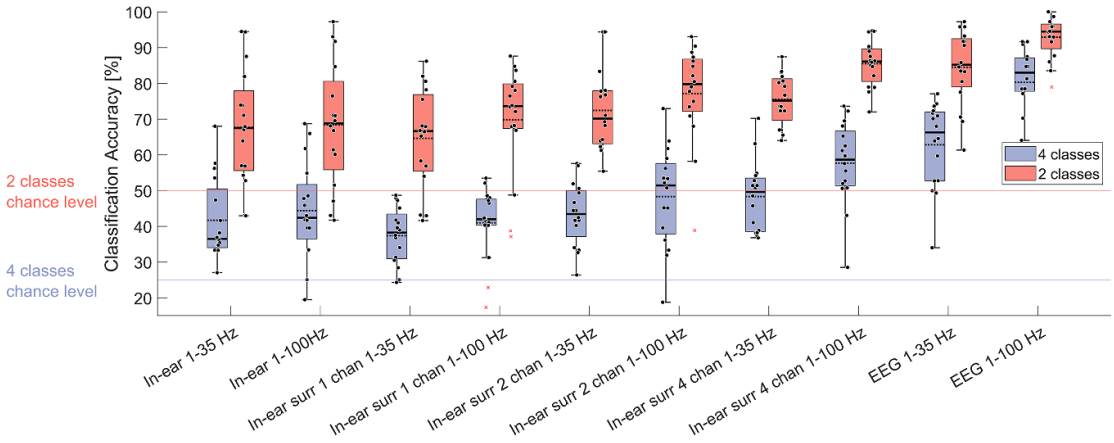

Figure 3. Box plots for the classification accuracy for the n-back task for different pre-processing pipelines and EEG signals. The dotted line denotes the mean and the solid line denotes the median. The blue data shows the accuracy across all four classes, while the red shows the binary accuracy for the extreme workload levels 0-back and 3-back. The scatter points represent the individual performance of each of the 16 participants.

Table 1. This table shows the Benjamini–Hochberg corrected *p*-values for one-sided sign ranked tests between chance/in-ear/1-channel surrogate and all pipelines for the n-back task. A bold *p*-value indicates that the analysis conditions on the left resulted in a significantly higher classification rate.

	4 classes			2 classes		
	Chance (25%)	In-ear 1–100 Hz	In-ear surr 1 chan 1–100 Hz	Chance (50%)	In-ear 1–100 Hz	In-ear surr 1 chan 1–100 Hz
In-ear 1–35 Hz	**0.001**	1	0.425	**0.0036**	0.8477	0.8202
In-ear 1–100 Hz	**0.0011**		0.2541	**0.003**		0.8202
In-ear surr 1 chan 1–35 Hz	**0.0011**	1	1	**0.0046**	1	1
In-ear surr 1 chan 1–100 Hz	**0.0013**	0.9922		**0.0018**	0.4654	
In-ear surr 2 chan 1–35 Hz	**0.001**	0.8202	0.288	**0.001**	0.425	0.4694
In-ear surr 2 chan 1–100 Hz	**0.001**	0.425	0.0522	**0.0011**	0.1519	0.0546
In-ear surr 4 chan 1–35 Hz	**0.001**	0.2048	**0.0087**	**0.001**	0.0773	0.1931
In-ear surr 4 chan 1–100 Hz	**0.001**	**0.0169**	**0.001**	**0.001**	**0.0087**	**0.001**
EEG 1–35 Hz	**0.001**	**0.0034**	**0.0012**	**0.001**	**0.0117**	**0.0037**
EEG 1–100 Hz	**0.001**	**0.001**	**0.001**	**0.001**	**0.0012**	**0.001**

(7.6, 7.8). Adding more channels has a positive impact on the surrogate results for 2 channels (6.8, 6.8) and 4 channels (6.3, 6.0). Increasing the frequency range also positively affects the results of in-ear (7.4, 7), 1-channel (7.1, 7.2), 2-channel (6.3, 6.5), and 4-channel surrogates (5.3, 5.4). Conventional EEG performs best for low (3.9, 4.0) and high frequency bands (3.2, 3.4). For correlation, these results are nearly identical but inverted. Low frequency in-ear gives reasonable correlations (34.7%, 38%), similar to the 1-channel surrogate (34.9%, 38%). These results improve with more channels for the 2-channel surrogate (43%, 46.6%) and 4-channel surrogate (53.9%, 55.2%). Adding higher frequencies has little effect, with each configuration showing only a small increase compared to the lower frequency range. Conventional EEG again performs the best in the low (73.3%, 75.5%) and high frequency configurations (79.1%, 80.3%).

A statistical evaluation of the approaches is presented in table 2. As with the n-back analysis, a one-sided Wilcoxon signed rank test with

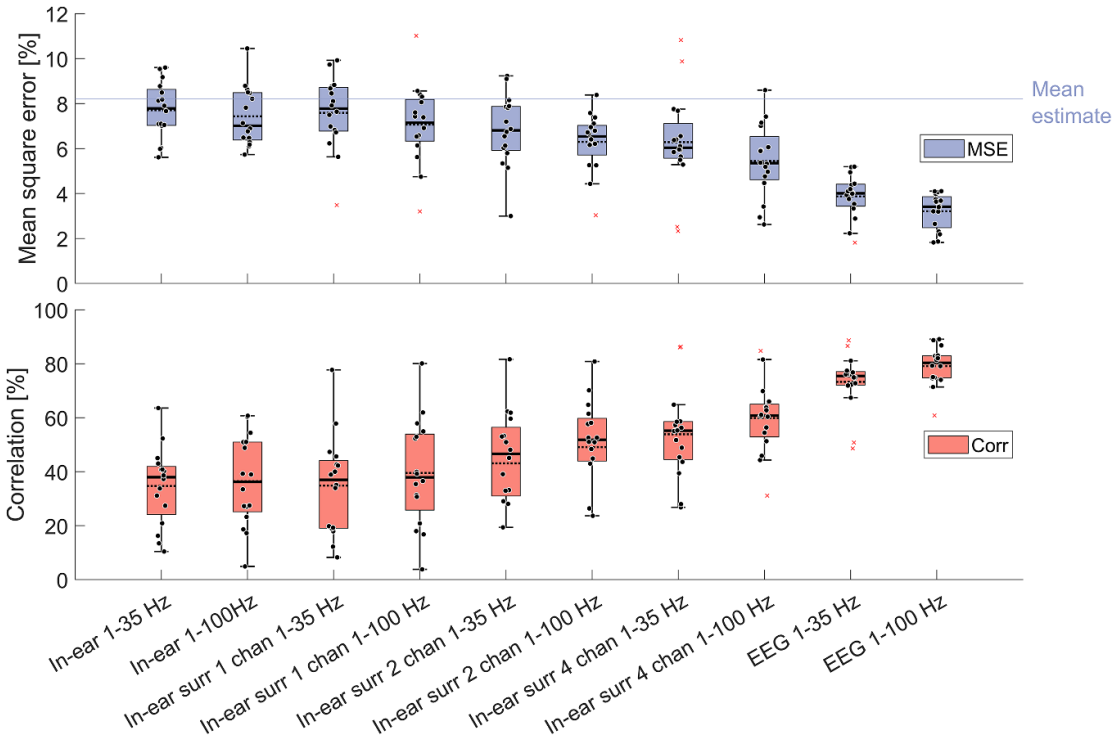

Figure 4. Box plots for the regression mean squared error and correlation between estimate and real difficulty level for the mental arithmetic task for different pre-processing pipelines and EEG signals. The dotted line denotes the mean while the solid line shows the median. The scatter points represent the individual performance of each of the 16 participants.

Table 2. This table shows the Benjamini–Hochberg corrected *p*-values for one-sided sign-ranked tests between the in-ear/1-channel surrogate and all the other conditions for the arithmetic task. A bold *p*-value indicates that the pipelines on the left resulted in a significantly higher classification rate.

	MSE			Corr		
	Mean estimate	In-ear 1–100 Hz	In-ear surr 1 chan 1–100 Hz	Mean estimate	In-ear 1–100 Hz	In-ear surr 1 chan 1–100 Hz
In-ear 1–35 Hz	0.1279	0.9098	1	**0.0008**	0.8892	0.987
In-ear 1–100 Hz	0.0538		0.9971	**0.0008**		0.865
In-ear surr 1 chan 1–35 Hz	0.1501	0.8137	1	**0.0008**	0.6251	0.9936
In-ear surr 1 chan 1–100 Hz	**0.018**	0.1602		**0.0008**	0.4065	
In-ear surr 2 chan 1–35 Hz	**0.0046**	0.1193	0.3532	**0.0008**	0.1858	0.3914
In-ear surr 2 chan 1–100 Hz	**0.0008**	**0.023**	**0.0281**	**0.0008**	**0.0343**	0.0588
In-ear surr 4 chan 1–35 Hz	**0.008**	**0.0343**	**0.0281**	**0.0008**	**0.0121**	**0.004**
In-ear surr 4 chan 1–100 Hz	**0.0008**	**0.0052**	**0.0009**	**0.0008**	**0.0025**	**0.0008**
EEG 1–35 Hz	**0.0008**	**0.0008**	**0.0008**	**0.0008**	**0.0008**	**0.0008**
EEG 1–100 Hz	**0.0008**	**0.0008**	**0.0008**	**0.0008**	**0.0008**	**0.0008**

Benjamini–Hochberg correction was used to compare the performance results of the mean estimate, in-ear, and 1-channel in-ear surrogate to the other approaches. For the mean estimate, a naive predictor that always estimates the average participant performance was used. For correlation, the p-values in the table indicate that the differences in performances of the approaches on the left side of the table, and mean estimate, in-ear, and the 1-channel surrogate originate from a distribution with a positive median, while for MSE the median is negative. For MSE, the low frequency in-ear, 1-channel surrogate, and high frequency in-ear did not perform significantly better than the mean estimate. Even though, it should be noted the p-value for the high frequency in-ear was just outside of the significance level (0.0538). All other approaches performed better. It can also be seen that the 4-channel and EEG signals as well as the 2-channel high frequency surrogate perform significantly better than the high frequency in-ear and 1-channel surrogate. The results for correlation differ in two positions. Since the constant mean estimate always results in zero correlation every estimation approach performs significantly better and only the 2-channel high frequency surrogate performs better than high frequency in-ear. Conventional EEG in both frequency configurations shows the overall lowest p-values, followed by the 4-channel high frequency surrogate.

Figures 5 and 6 show the grand average PSDs for both tasks on the left. Low and high workload conditions are presented with the shaded area indicating the standard deviation. The right side displays the grand average differential PSDs between low and high workload. Additionally, the subject individual differences are shown to highlight the variation between subjects and between approaches. Both sides display three different channels, conventional EEG channel Pz, 1-channel surrogate and in-ear. For visualization, the area above 90 Hz has been removed based on the respective filter range.

For the n-back task, the low workload condition was 0-back and the high workload condition was 3-back. In figure 5, it can be seen that the high workload conditions exhibit a lower α peak for regular EEG and surrogate and also display greater power for higher power starting around approximately 25 Hz for the surrogate. The in-ear plot does not show clear spectral differences but displays a greater standard deviation compared to the other plots, indicating more individualized spectral characteristics. These characteristics are even clearer in the differential plots. The alpha peak is visible in Pz and surrogate, while in-ear displays the greatest variance across participants.

For the mental arithmetic task, the low workload condition is presented as the easiest block, and the high workload condition as the most difficult block. The results, displayed in figure 6, are similar to the n-back task. The increased α peak for low workload is smaller but still visible. For frequencies above 20 Hz, a difference can be seen for the surrogate. The in-ear plot shows again a greater standard deviation compared to the other plots and shows generally a higher power corresponding to higher workloads in this case. The differential plots reinforce these assessments. The variance across participants is again the highest for in-ear EEG.

In both figures it can be seen there is no general behaviour for PSDs for different level of workload. While the averaged PSDs on the left would indicate for example a general increase in the α band for lower levels of workload, the plots on the right show that each participant has a unique behaviour confirming previous results from literature [24]. Finally, it must be noted that in-ear shows a distinct increase in power around 25 Hz and 75 Hz.

The topographical maps in figure 7 show the performance differences at the subject level. They display the percentage of participants where a two-sided Wilcoxon sign rank test with Benjamini–Hochberg correction showed significant performance differences for each channel/frequency band combination. The frequency bands consist of δ, θ, α, β, γ, and high γ, while the values correspond to the average out of all 1 Hz frequency bins in the respective band.

For the n-back task, it can be seen that for the EEG channels, the α band in the central and parietal regions and the γ and high γ bands in the central and temporal regions show an increased number of participants with significant changes. This is likely due to the typical negative correlation of workload with the α band [27] and the increased correlation for higher frequency bands that are potentially caused by muscle activity [32]. The surrogate shows generally an increased number of significant participants in the higher frequency bands starting from the α band. Compared to EEG and surrogate, the in-ear EEG shows a greater number of participants with significant differences in the δ and θ ranges and a comparable number in the other frequency bands to the surrogate.

For the mental arithmetic task, it can be seen that compared to the n-back task, there is no increased significance around the α band for EEG. Instead, the δ band shows more significance. For high frequencies, EEG results are nearly identical to the n-back task. The surrogate behaves nearly exactly the same as with the n-back task, only significance in the α and β bands seems to be slightly reduced. The in-ear channel shows overall significance in all frequency bands, slightly lower than in the n-back task but similar

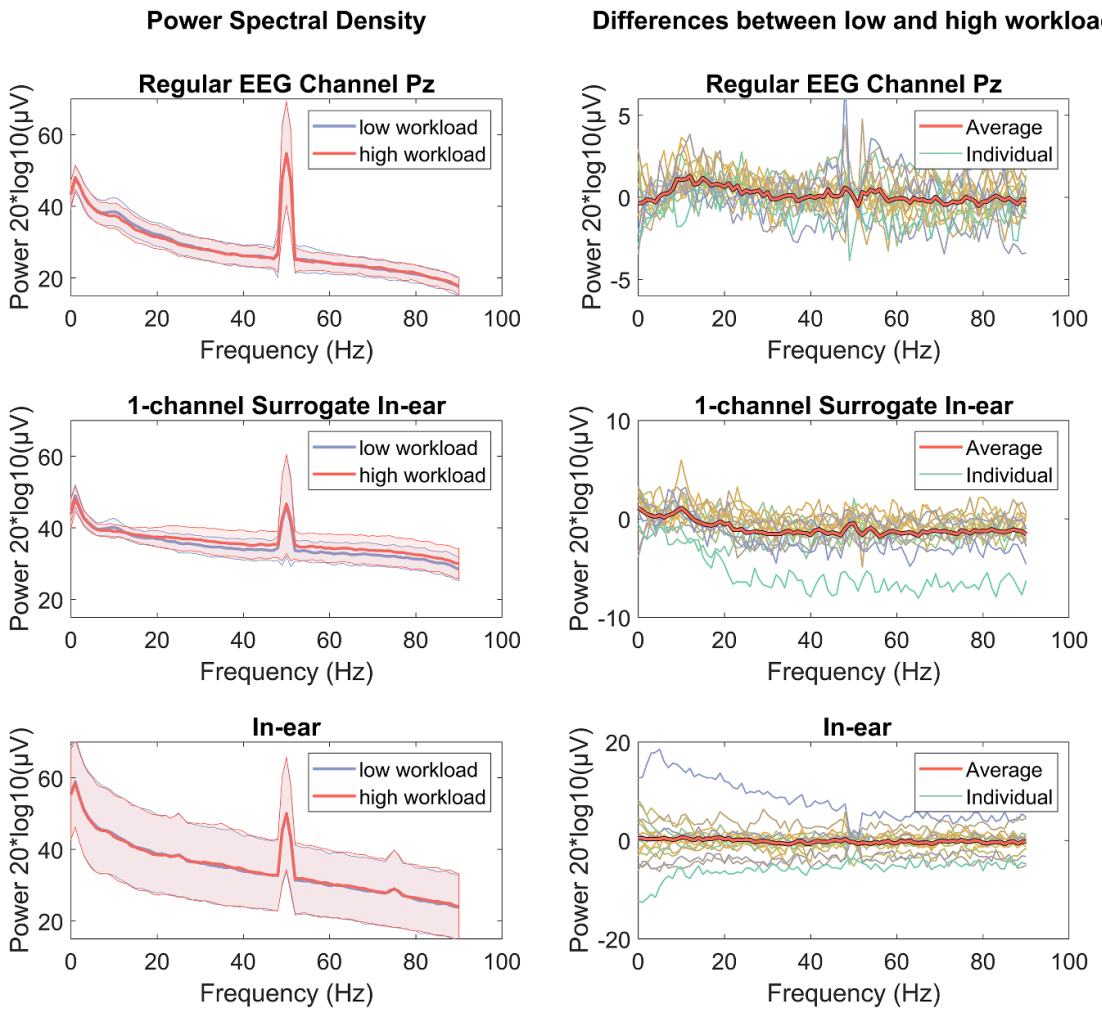

Figure 5. Left: grand average and standard deviation PSD for the n-back task for regular EEG channel Pz, in-ear, and 1-channel surrogate for low and high workload. Right: grand average PSD differences between low and high difficulty in average and individually for each participant for regular EEG channel Pz, in-ear, and 1-channel surrogate.

or higher compared to EEG. The overall number of significant frequency bands compared to the n-back task.

Figure 8 shows topographical plots for both tasks for the conventional EEG frequency bands. The plots show the Pearson's R and R^2 values for correlating the PSDs with workload level for each task. Next to each topographic plot, two circular markers indicate the in-ear and surrogate values.

The R-values for the n-back task follow the typical pattern of a negative correlation of α band PSDs with increased workload. This also overlaps with some channels in the β band. δ and θ only show some increased correlation in the frontal electrodes, potentially stemming from facial movements or tension. β, γ, and high γ show high correlation for electrodes located in the posterior areas. The in-ear channel stays nearly constant across all frequency bands. The surrogate channel shows negative correlation for α bands and positive correlation for β, γ, and high γ. The R^2 values show a slight increase for the α band and a larger increase for γ and high γ, especially in channels located in the temporal regions. In contrast to the R-values, in-ear shows increased R^2 for δ, θ and α with a slightly lower increase in the higher frequencies. The surrogate's R^2 value increases with increasing frequency.

For the mental arithmetic task, the topographical plots show the correlation between the PSDs and the q-values. It exhibits similar patterns for high frequency R and R^2 values for both EEG and surrogate. There are differences in the low frequency bands, notably a weaker negative correlation in the α band and a negative correlation in the δ band, and partly θ band, compared to the n-back task. In-ear shows a stronger positive correlation that increases in higher frequencies. The surrogate performs similarly but shows lower correlation in the δ, θ and α bands. The behavior of in-ear and surrogate is similar for R^2 values.

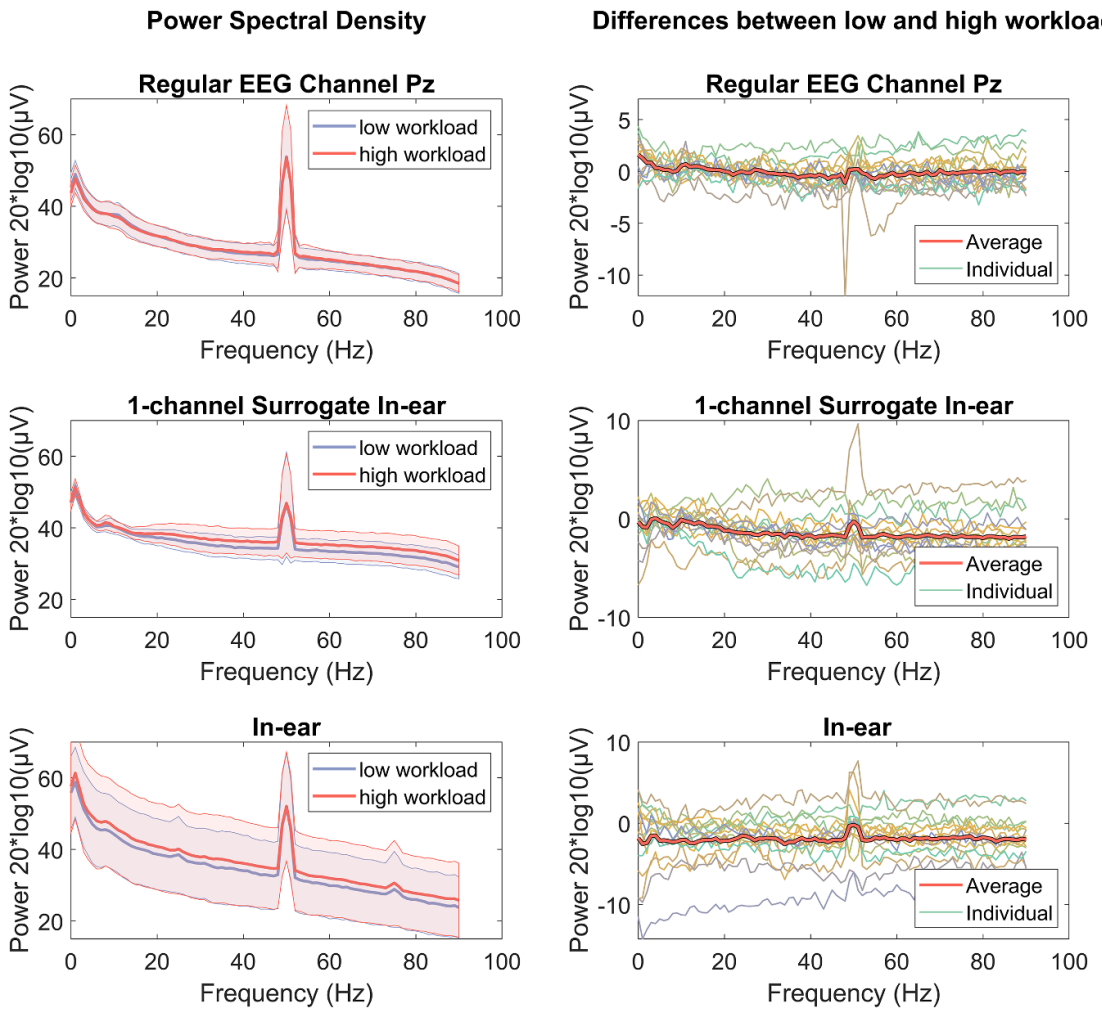

Figure 6. Left: grand average and standard deviation PSD for the mental arithmetic task for regular EEG channel Pz, in-ear and 1-channel surrogate for low and high workload. Right: grand average PSD differences between low and high difficulty in average and individually for each participant for regular EEG channel Pz, in-ear, and 1-channel surrogate.

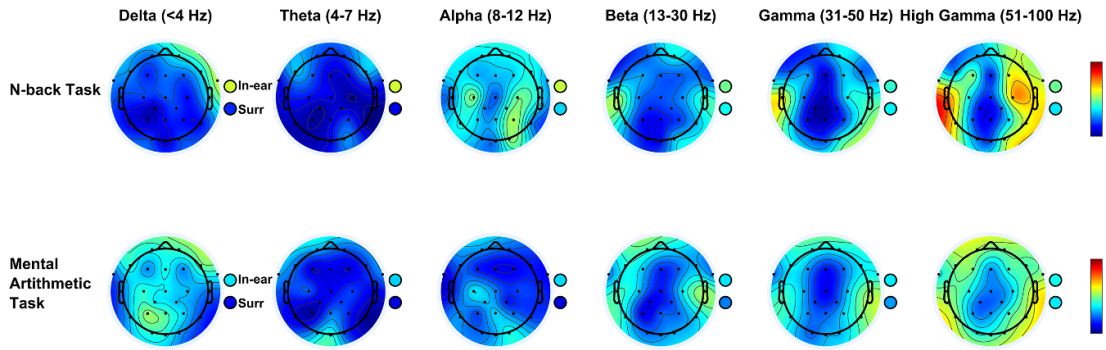

Figure 7. Topographical maps showing the percentage of participants where differences between high and low workload were statistically significant for both tasks.

4. Discussion

4.1. In-ear EEG

The findings of this study demonstrate the feasibility of estimating multiple workload levels with an accuracy exceeding chance level using a commercial in-ear EEG headset, the IDUN 'Guardian'. This is particularly noteworthy given that it extends beyond the simple binary classification problems that are typically examined in this domain [26].

Using the traditional n-back task, we found that in-ear EEG can be used to correctly classify

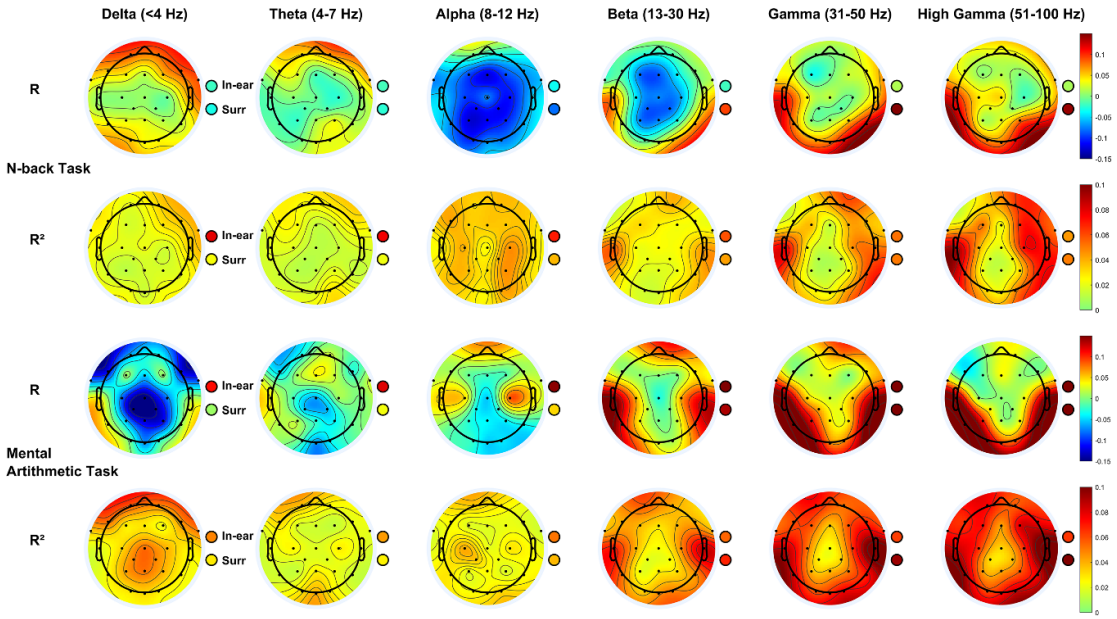

Figure 8. Topographical plots for both tasks. The plots display the correlation and squared correlation between workload/q-value and PSDs in selected EEG frequency bands. In-ear and in-ear surrogate are visualized as small circles next to the right ear.

workload with 44.4% accuracy for four classes and with 68.4% for two classes, significantly better than chance level, as depicted in figure 3. A more complex regression-based workload prediction approach using a mental arithmetic task performed slightly better compared to a naive estimator that only predicts the mean performance as shown in figure 4. This improvement was just outside the range of statistical significance ($p = 0.0538$).

However, minor modifications to the in-ear EEG could potentially enhance its performance. Therefore, we developed surrogate in-ear measures based on EEG electrodes located around the ear. Three different surrogates were created with different numbers of channels. The in-ear headset acquires a single bipolar channel. The 1-channel surrogate was designed to function in the same way as the in-ear using a bipolar derivation of the closest electrodes to the ear and serves as a baseline comparison to validate the surrogates' performance. The 2-channel surrogate was designed to simulate an additional reference electrode and the 4-channel surrogate simulated the addition of two more electrodes around the ear.

We also compared the conventional EEG frequency range from 1–35 Hz to a larger frequency range of 1–100 Hz including the γ band for workload estimation, since there has been evidence [30, 31] that γ band features can improve the classification results. Our findings in figures 3, 4, tables 1 and 2 also support this notion. Nearly all approaches perform better in high frequency configuration compared to low frequency. Comparing in-ear with their surrogates in both frequency ranges shows nearly identical results, a statistical comparison between high frequency in-ear and surrogate also shows that they perform nearly equivalently. Adding more channels further improves the surrogate, so that the 4-channel high frequency in-ear surrogate significantly outperforms in-ear and 1-channel surrogates in every measure and configuration examined. Even the 2-channel high frequency surrogate can outperform all in-ear results in the math task. The in-ear headset is designed with two earbud electrodes and a ground electrode positioned where the headset rests on the ear. Based on the improved results of our surrogate in-ear EEG measures we propose the following design modifications:

- 2-channel surrogate: This could be realized by simply adding a reference electrode at the ear opposite of the ground electrode.
- 4-channel surrogate: This would necessitate the addition of two more electrodes, which could be integrated in a similar manner as inducers in bone conduction headsets. These are situated near the ear, require pressure to properly transmit the vibration signal, and need a location that is not covered by hair.

Incorporating more than a single bipolar channel also offers the advantage of increased spatial information for further enhancing processing. This includes spatial filters [50] like the common average reference used in this study, or blind source separation methods [51] such as independent component analysis. These methods enhance the signal quality and improve the interpretability of the EEG data. While there are no

results for four class classification and regression-based problems, our classification accuracy of 68.4% for the two class problems of the n-back task for in-ear EEG and 77.1%–79.1% for the 2-channel surrogates is comparable to previous research [26]. The key advantage of our used and proposed systems, however, is that all functionality is integrated into a single device. It does not required an expensive biomedical amplifier that needs placement on the head for a physical connection to the electrodes, nor does it demand any tape or adhesives to secure the electrodes in place [26].

Upon examination of the grand average PSDs depicted in figures 5 and 6, it can be seen that the in-ear and surrogate exhibit slightly different behavior. While the surrogate shows approximately similar plots for both tasks, the in-ear plots differentiate. For the n-back task, the curves are nearly identical. For the mental arithmetic task, the high workload condition produces higher PSDs starting from around 5 Hz. One reason for this difference could be the large standard deviation of in-ear that is several times larger than for the surrogate. Since the individual participant responses are varied, they can easily influence these grand averages. This could be due to muscle artifacts, tension, or the differences in the electrode types used. Furthermore, there is a distinct peak in power around 25 Hz and 75 Hz. Due to the in-ear electrodes' higher impedance, they are more susceptible to artifacts than the wet electrodes compared to conventional EEG [52]. The noise in the in-ear signal could be caused by a subharmonic power line artifact or an interaction with the conventional EEG system. The surrogate exhibits a similar increase in PSDs for increased workload, starting at 20 Hz. Notably, the negative correlation with workload around the α peak, which is present in the surrogate spectrum, is entirely absent in the in-ear spectrum.

The topographical maps in figure 7 provide an alternative representation of workload effects, showcasing the relative count of participants that exhibited significant differences in PSD across varying workload levels. In-ear EEG again differentiates between both tasks. For the n-back task, a greater number of participants exhibit significance in the lower frequency bands (1–10 Hz) compared to the higher frequency bands. Conversely, for the mental arithmetic task, the frequency bands (30–100 Hz) display a larger number of significant subjects. The surrogate also exhibits a slight task-based differentiation. The results of the mental arithmetic task mirror the in-ear EEG's behavior, where higher frequencies (40–100 Hz) demonstrate greater significance compared to lower frequencies. However, for the n-back task, the surrogate only occasionally presents a substantial number of participants with significance in the higher frequency range (20–100 Hz).

The correlation coefficients R and R^2 of the PSDs associated with workload, as depicted in figure 8, present a comparable narrative. The surrogate exhibits similar behavior across both tasks and measures, with a low correlation in lower frequencies and a high correlation in higher frequencies. The only substantial difference across both tasks is the negative correlation around the α peak for the n-back task. However, the in-ear EEG again displays considerable differences between the two tasks. The correlation for the n-back task remains constantly neutral, while it is overall positive for the mental arithmetic tasks. Furthermore, while the squared correlation initiates at a high value at low frequencies and diminishes for the n-back task, it remains relatively constant for the mental arithmetic tasks. In summary, the in-ear EEG demonstrates a more individualized response to increased workload, yielding divergent results depending on the task and the user, thereby resulting in a higher variance of responses.

4.2. Conventional EEG

Workload estimation for conventional EEG has been researched extensively and our classification results with 92.9% for the 2-class and 80.3% for the 4-class n-back tasks are slightly higher but comparable to prior work [20, 22–26]. Features in the γ band are often excluded due to the ambiguity of their origin, whether from brain or muscle activity [32]. However, even in the low frequency range from 1–35 Hz, our approach results in 84.5% for 2 classes and 62.8% for 4 classes for the n-back task. For the mental arithmetic task, while our experimental settings differ from those in the literature, it is still possible to compare the correlation coefficient between estimation and real difficulty to previous research. With 73.3% for low frequency EEG and 79.1% for high frequency EEG, it is also in a similar range [45]. The grand average PSDs depicted in figures 5 and 6 exhibit characteristics consistent with existing literature [20, 23, 26].

The topographical maps of statistically significant differences in EEG channel location and frequency bands presented in figure 7 also show a similar spatial distribution as found in prior literature [26]. Finally, the topographical plots for the correlation between workload level and PSD of the n-back task, as illustrated in figure 8 further reinforce these results, demonstrating a negative correlation in the α frequency range and an increased correlation in the γ ranges similar to existing literature [25]. However, we were unable to replicate the topographical plots for the mental arithmetic task. Spüler *et al* [45] reported increased R^2 values predominantly in the α range, with little squared correlation in other frequency bands. This behavior mirrors that of R^2 for the n-back task in our results. For the mental arithmetic task, we observed an increased squared correlation for δ bands and frequencies of β and higher which is a

common behavior in workload tasks [25]. Our results also indicate an atypical negative correlation of low frequencies to increased workload, which could potentially be attributed to participants completing easier calculations quickly and subsequently making subconscious movements.

4.3. Real-world implementation and future work

Our analysis of conventional EEG data aligns closely with established literature, affirming the validity of our experimental setup and supporting the feasibility of assessing various levels of workload using in-ear EEG. Additionally, our experimental design of the mental math task strengthens the confidence of workload estimations. This is achieved through enhanced granularity resulting from its pseudo-randomized eight difficulty levels but also due to the introduction of the added randomized noise within each difficulty block. This approach ensured that features are not dependent on specific trials or blocks, but accurately represent workload levels. A BCI should have a minimum binary classification accuracy between 65%–70% for users to feel in control [53]. With a classification rate of 67.7% for the 2-class problem, in-ear falls into that range. It is also important to highlight that this stated accuracy range was determined for active BCIs where users expect immediate feedback. Since workload estimation is usually used passively, longer windows than the 2.5 s used in the study can be utilized to further increase the performance. However, for the widespread adoption of such devices, further research is crucial. Challenges emerge when these devices are used in less controlled environments, leading to increased artifacts from movements, muscle tension, or electromagnetic radiation, complicating workload identification. Developing improved noise reduction techniques [54, 55], potentially leveraging the integrated accelerometer unit in many headsets, becomes essential. Enhancing the electrode design to reduce sensitivity to noise would also be beneficial. The complexity of workload assessment is also compounded by the introduction of more diverse and intricate tasks, each evoking distinct user responses as shown in our results above. Moreover, the necessity for task and user-specific training data presents an additional obstacle. Most machine learning algorithms require retraining for each user and task, a time-intensive process. While transfer learning techniques [56, 57] across tasks and subjects are becoming more common, they have yet to match the performance of subject-specific methods. Exploring meta-learning strategies, where machine learning algorithms utilize pre-recorded data [57, 58] to streamline the training process, offers a promising avenue until further advancements in transfer learning are realized.The insights gained from our investigations using the n-back task and mental arithmetic task, along with the development of surrogate in-ear measures, underscore the need for continuous research to address these challenges and optimize the performance of in-ear EEG devices.

5. Conclusions

In conclusion, our study successfully demonstrates the feasibility of estimating multiple workload levels in two distinct tasks through in-ear EEG recordings using one of the first commercial available headsets, the IDUN 'Guardian'. Utilizing frequency EEG bands spanning from 1–100 Hz led to classification above chance levels for the n-back task and a lower regression error than a naive mean performance predictor for the mental arithmetic task. The in-ear system did not perform as well as traditional EEG, primarily due to the limited number of channels. However, when comparing the in-ear results to a bipolar derivation of traditional EEG, the performance was similar. Therefore, we introduced in-ear surrogates based on traditional EEG electrode locations around the ear, offering guidance on how to enhance these results to attain performance levels closer to EEG. Our findings emphasize the substantial variability of in-ear EEG recordings among participants, underscoring the importance of individually training machine learning algorithms. In-ear EEG emerges as a promising candidate for expanding EEG-based brain monitoring and interfacing applications into everyday life, providing valuable insights into the user's mental state.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors. https://doi.org/10.5258/SOTON/D3057 [59]. Data will be available from 1 November 2024.

Acknowledgments

The authors would like to acknowledge the University of Southampton's Researcher Development Concordat team for funding consumables in support of this study. The authors would also like to acknowledge IDUN technology for supplying the in-ear headset. The authors declare they have no conflicts of interest. This research is supported by UKRI/EPSRC Grants EP/T007656/1: Health Resilience Interactive Technologies and EP/R029563/1: AutoTrust.

Appendix

Table 3. This table shows the Benjamini–Hochberg corrected *p*-values for two-sided Wilcoxon rank-sum tests comparing multiple measures of the randomized and non-randomized participant groups. None of the *p*-values showed significant differences.

	Reaction Time	Performance	In-ear 1–35 Hz	In-ear 1–100 Hz	In-ear surr 1 chan 1–35 Hz	In-ear surr 1 chan 1–100 Hz	In-ear surr 2 chan 1–35 Hz	In-ear surr 2 chan 1–100 Hz	In-ear surr 4 chan 1–35 Hz	In-ear surr 4 chan 1–100 Hz	EEG 1–35 Hz	EEG 1–100 Hz
N-back task—4 class classification accuracy												
p-value	0.8976	0.8622	0.8976	0.8059	0.8976	0.7945	0.8976	0.8834	0.2253	0.8622	0.8976	0.8622
N-back task—2 class classification accuracy												
p-value			0.8622	0.8622	0.7945	0.8976	0.8976	0.8976	0.7945	0.8622	0.8622	0.7945
Math task—regression error												
p-value	0.7220	0.8749	0.7220	0.4414	0.8222	0.8717	0.4045	0.3516	0.7730	0.4045	0.4045	0.8253
Math task—correlation												
p-value			0.4835	0.6264	0.8749	0.7220	0.4045	0.4045	0.4045	0.4045	0.7220	0.8253

Table 4. This table shows the Benjamini–Hochberg corrected *p*-values for two-sided Wilcoxon sign-rank tests comparing multiple pipelines with and without eye movement suppression. None of the *p*-values showed significant differences.

	In-ear 1–35 Hz	In-ear 1–100 Hz	In-ear surr 1 chan 1–35 Hz	In-ear surr 1 chan 1–100 Hz	In-ear surr 2 chan 1–35 Hz	In-ear surr 2 chan 1–100 Hz	In-ear surr 4 chan 1–35 Hz	In-ear surr 4 chan 1–100 Hz	EEG 1–35 Hz	EEG 1–100 Hz
N-back task—4 class classification accuracy										
p-value	0.1977	0.5382	0.9794	0.9794	0.468	0.3266	0.1977	0.7119	0.4161	0.4696
N-back task—2 class classification accuracy										
p-value	0.5839	0.5836	0.1977	0.8404	0.8404	0.5382	0.4067	0.4067	0.1977	0.1977
Math task—regression error										
p-value	0.3861	0.64	0.7962	0.7119	0.64	0.5603	0.5603	0.7962	0.7119	0.4688
Math task—correlation										
p-value	0.5603	0.6382	0.7119	0.7119	0.5603	0.64	0.6801	0.8361	0.7119	0.3861

ORCID iDs

Christoph Tremmel https://orcid.org/0000-0002-0324-6626
Dean J Krusienski https://orcid.org/0000-0002-4668-5784
mc schraefel https://orcid.org/0000-0002-9061-7957

References

[1] Wolpaw J R, Birbaumer N, McFarland D J, Pfurtscheller G and Vaughan T M 2002 Brain–computer interfaces for communication and control *Clin. Neurophysiol.* **113** 767–91
[2] Zander T O and Kothe C 2011 Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general *J. Neural Eng.* **8** 025005
[3] Ferree T C, Luu P, Russell G S and Tucker D M 2001 Scalp electrode impedance, infection risk and EEG data quality *Clin. Neurophysiol.* **112** 536–44
[4] Pfurtscheller G, Brunner C, Schlögl A and Da Silva F H 2006 Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks *NeuroImage* **31** 153–9
[5] Guan C, Thulasidas M and Wu J 2004 High performance P300 speller for brain-computer interface *IEEE Int. Workshop on Biomedical Circuits and Systems (2004)* (IEEE) pp S3–5
[6] Krusienski D J, Sellers E W, McFarland D J, Vaughan T M and Wolpaw J R 2008 Toward enhanced P300 speller performance *J. Neurosci. Methods* **167** 15–21
[7] Lin Z, Zhang C, Wu W and Gao X 2006 Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs *IEEE Trans. Biomed. Eng.* **53** 2610–4
[8] Schuster C, Hilfiker R, Amft O, Scheidhauer A, Andrews B, Butler J, Kischka U and Ettlin T 2011 Best practice for motor imagery: a systematic literature review on motor imagery training elements in five different disciplines *BMC Med.* **9** 1–35
[9] Shenoy P, Krauledat M, Blankertz B, Rao R P N and Müller K-R 2006 Towards adaptive classification for BCI *J. Neural Eng.* **3** R13
[10] Jeunet C, N'Kaoua B and Lotte F 2016 Advances in user-training for mental-imagery-based BCI control: psychological and cognitive factors and their neural correlates *Brain-Computer Interfaces: Lab Experiments to Real-World Applications* (*Progress in Brain Research* vol 228) (Elsevier) pp 3–35
[11] Roy R N, Bonnet S, Charbonnier S and Campagne A 2013 Mental fatigue and working memory load estimation: interaction and implications for EEG-based passive BCI *2013 35th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) pp 6607–10
[12] Aricò P, Borghini G, Di Flumeri G, Sciaraffa N and Babiloni F 2018 Passive BCI beyond the lab: current trends and future directions *Physiol. Meas.* **39** 08TR02
[13] Myrden A and Chau T 2017 A passive EEG-BCI for single-trial detection of changes in mental state *IEEE Trans. Neural Syst. Rehabil. Eng.* **25** 345–56
[14] Lendner J D, Helfrich R F, Mander B A, Romundstad L, Lin J J, Walker M P, Larsson P G and Knight R T 2020 An electrophysiological marker of arousal level in humans *eLife* **9** e55092
[15] Sourina O and Liu Y 2011 A fractal-based algorithm of emotion recognition from EEG using arousal-valence model *Biosignals* pp 209–14
[16] Galvão F, Alarcão S M and Fonseca M J 2021 Predicting exact valence and arousal values from EEG *Sensors* **21** 3414
[17] Shi L-C and Lu B-L 2013 EEG-based vigilance estimation using extreme learning machines *Neurocomputing* **102** 135–43
[18] Shi L-C, Jiao Y-Y and Lu B-L 2013 Differential entropy feature for EEG-based vigilance estimation *2013 35th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) pp 6627–30
[19] Berka C, Levendowski D J, Lumicao M N, Yau A, Davis G, Zivkovic V T, Olmstead R E, Tremoulet P D and Craven P L 2007 EEG correlates of task engagement and mental workload in vigilance, learning and memory tasks *Aviat. Space Environ. Med.* **78** B231–44
[20] Brouwer A-M, Hogervorst M A, Van Erp J B F, Heffelaar T, Zimmerman P H and Oostenveld R 2012 Estimating workload using EEG spectral power and ERPs in the n-back task *J. Neural Eng.* **9** 045008
[21] Gerjets P, Walter C, Rosenstiel W, Bogdan M and Zander T O 2014 Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach *Front. Neurosci.* **8** 385
[22] Hogervorst M A, Brouwer A-M and Van Erp J B F 2014 Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload *Front. Neurosci.* **8** 322
[23] Wang S, Gwizdka J and Chaovalitwongse W A 2015 Using wireless EEG signals to assess memory workload in the *n*-back task *IEEE Trans. Hum.-Mach. Syst.* **46** 424–35
[24] Tremmel C, Herff C, Sato T, Rechowicz K, Yamani Y and Krusienski D J 2019 Estimating cognitive workload in an interactive virtual reality environment using EEG *Front. Hum. Neurosci.* **13** 401
[25] Chakladar D D, Dey S, Roy P P and Dogra D P 2020 EEG-based mental workload estimation using deep BLSTM-LSTM network and evolutionary algorithm *Biomed. Signal Process. Control* **60** 101989
[26] Crétot-Richert G, De Vos M, Debener S, Bleichner M G and Voix J 2023 Assessing focus through ear-EEG: a comparative study between conventional cap EEG and mobile in-and around-the-ear EEG systems *Front. Neurosci.* **17** 895094
[27] Chikhi S, Matton N and Blanchet S 2022 EEG power spectral measures of cognitive workload: a meta-analysis *Psychophysiology* **59** e14009
[28] Pesonen M, Hämäläinen H and Krause C M 2007 Brain oscillatory 4–30 hz responses during a visual n-back memory task with varying memory load *Brain Res.* **1138** 171–7
[29] Scharinger C, Soutschek A, Schubert T and Gerjets P 2017 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.* **11** 6
[30] Fitzgibbon S, Pope K, Mackenzie L, Clark C and Willoughby J 2004 Cognitive tasks augment gamma EEG power *Clin. Neurophysiol.* **115** 1802–9
[31] Tallon-Baudry C, Bertrand O, Peronnet F and Pernier J 1998 Induced γ-band activity during the delay of a visual short-term memory task in humans *J. Neurosci.* **18** 4244–54
[32] Goncharova I I, McFarland D J, Vaughan T M and Wolpaw J R 2003 EMG contamination of EEG: spectral and topographical characteristics *Clin. Neurophysiol.* **114** 1580–93
[33] Lopez-Gordo M A, Sanchez-Morillo D and Valle F P 2014 Dry EEG electrodes *Sensors* **14** 12847–70
[34] Mathewson K E, Harrison T J L and Kizuk S A D 2017 High and dry? Comparing active dry EEG electrodes to active and passive wet electrodes *Psychophysiology* **54** 74–82
[35] Juez J Y, Moumane H, Nassar M, Molina-Salcedo I, Segura-Quijano F E, Valderrama M and Le Van Quyen M 2024 Ear-EEG devices for the assessment of brain activity: a review *IEEE Sens. J.* **24** 31606–23
[36] Goverdovsky V, Looney D, Kidmose P and Mandic D P 2015 In-ear EEG from viscoelastic generic earpieces: robust and unobtrusive 24/7 monitoring *IEEE Sens. J.* **16** 271–7

[37] Hwang T, Kim M, Hong S and Park K S 2016 Driver drowsiness detection using the in-ear EEG *2016 38th Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) pp 4646–9
[38] Kappel S L, Rank M L, Toft H O, Andersen M and Kidmose P 2018 Dry-contact electrode ear-EEG *IEEE Trans. Biomed. Eng.* **66** 150–8
[39] Athavipach C, Pan-Ngum S and Israsena P 2019 A wearable in-ear EEG device for emotion monitoring *Sensors* **19** 4014
[40] Debener S, Emkes R, De Vos M and Bleichner M 2015 Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear *Sci. Rep.* **5** 16743
[41] Pacharra M, Debener S and Wascher E 2017 Concealed around-the-ear EEG captures cognitive processing in a visual Simon task *Front. Hum. Neurosci.* **11** 290
[42] Bleichner M G, Lundbeck M, Selisky M, Minow F, Jäger M, Emkes R, Debener S and De Vos M 2015 Exploring miniaturized EEG electrodes for brain-computer interfaces. An EEG you do not see? *Physiol. Rep.* **3** e12362
[43] Kim Y-J, Kwak N-S and Lee S-W 2018 Classification of motor imagery for ear-EEG based brain-computer interface *2018 6th Int. Conf. on Brain-Computer Interface (BCI)* (IEEE) pp 1–2
[44] Thomas H 1963 Communication theory and the constellation hypothesis of calculation *Q. J. Exp. Psychol.* **15** 173–91
[45] Spüler M, Walter C, Rosenstiel W, Gerjets P, Moeller K and Klein E 2016 EEG-based prediction of cognitive workload induced by arithmetic: a step towards online adaptation in numerical learning *ZDM* **48** 267–78
[46] Boulay C 2024 *Labstreaminglayer GithubRepository* (available at: https://github.com/sccn/labstreaminglayer) (Accessed 12 February 2024)
[47] Sharbrough F 1991 American electroencephalographic society guidelines for standard electrode position nomenclature *Clin. Neurophysiol.* **8** 200–2
[48] Croft R J and Barry R J 2000 Removal of ocular artifact from the EEG: a review *Neurophysiol. Clini./Clin. Neurophysiol.* **30** 5–19
[49] Belkhiria C and Peysakhovich V 2021 EOG metrics for cognitive workload detection *Proc. Comput. Sci.* **192** 1875–84
[50] McFarland D J, McCane L M, David S V and Wolpaw J R 1997 Spatial filter selection for EEG-based communication *Electroencephalogr. Clin. Neurophysiol.* **103** 386–94
[51] Kachenoura A, Albera L, Senhadji L and Comon P 2007 ICA: a potential tool for BCI systems *IEEE Signal Process. Mag.* **25** 57–68
[52] Shad E H T, Molinas M and Ytterdal T 2020 Impedance and noise of passive and active dry EEG electrodes: a review *IEEE Sens. J.* **20** 14565–77
[53] Fard P R and Grosse-Wentrup M 2014 The influence of decoding accuracy on perceived control: a simulated BCI study (arXiv:1410.6752)
[54] Gwin J T, Gramann K, Makeig S and Ferris D P 2010 Removal of movement artifact from high-density EEG recorded during walking and running *J. Neurophysiol.* **103** 3526–34
[55] Tremmel C, Herff C and Krusienski D J 2019 EEG movement artifact suppression in interactive virtual reality *2019 41st Annual Int. Conf. IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE) pp 4576–9
[56] Wan Z, Yang R, Huang M, Zeng N and Liu X 2021 A review on transfer learning in EEG signal analysis *Neurocomputing* **421** 1–14
[57] Ko W, Jeon E, Jeong S, Phyo J and Suk H-I 2021 A survey on deep learning-based short/zero-calibration approaches for EEG-based brain–computer interfaces *Front. Hum. Neurosci.* **15** 643386
[58] Tremmel C, Fernandez-Vargas J, Stamos D, Cinel C, Pontil M, Citi L and Poli R 2022 A meta-learning BCI for estimating decision confidence *J. Neural Eng.* **19** 046009
[59] Tremmel C 2024 Dataset in support of the publication 'Estimating cognitive workload using a commercial in-ear EEG headset' *Soton e-Prints* 495292 (https://doi.org/10.5258/SOTON/D3057)