Real-time estimation of EEG-based engagement in different tasks

(Article begins on next page)

17 May 2024

Real-time estimation of EEG-based engagement in different tasks

Angela Natalizio[1,2,3,*], Sebastian Sieghartsleitner[1,4,*], Leonhard Schreiner[1,5], Martin Walchshofer[1], Antonio Esposito[2,6], Josef Scharinger[4], Harald Pretl[5], Pasquale Arpaia[2,7,8], Marco Parvis[3], Jordi Solé-Casals[9,10], Marc Sebastián-Romagosa[11], Rupert Ortner[11], Christoph Guger[1,11]

[1] g.tec medical engineering GmbH, Schiedlberg, Austria
[2] Augmented Reality for Health Monitoring Laboratory (ARHeMLab), Università degli Studi di Napoli Federico II, Naples, Italy
[3] Department of Electronics and Telecommunications (DET), Polytechnic of Turin, Turin, Italy
[4] Institute of Computational Perception, Johannes Kepler University, Linz, Austria
[5] Institute for Integrated Circuits, Johannes Kepler University, Linz, Austria
[6] Department of Engineering for Innovation University of Salento, Lecce, Italy
[7] Department of Electrical Engineering and Information Technology (DIETI), Università degli Studi di Napoli Federico II, Naples, Italy
[8] Centro Interdipartimentale di Ricerca in Management Sanitario e Innovazione in Sanità (CIRMIS), Università degli Studi di Napoli Federico II, Naples, Italy
[9] Data and Signal Processing Research Group, University of Vic-Central, University of Catalonia, Vic, Catalonia, Spain
[10] Department of Psychiatry, University of Cambridge, Cambridge, UK
[11] g.tec medical engineering Spain SL, Barcelona, Spain
[*] Shared co-first authorship

E-mail: `angela.natalizio@polito.it`, `sieghartsleitner@gtec.at`

September 2023

Abstract.
Objective. Recent trends in brain-computer interface (BCI) research concern the passive monitoring of brain activity, which aim to monitor a wide variety of cognitive states. Engagement is such a cognitive state, which is of interest in contexts such as learning, entertainment or rehabilitation. This study proposes a novel approach for real-time estimation of engagement during different tasks using electroencephalography (EEG). *Approach.* Twenty-three healthy subjects participated in the BCI experiment. A modified version of the d2 test was used to elicit engagement.Within-subject classification models which discriminate between engaging and resting states were trained based on EEG recorded during a d2 test based paradigm. The EEG was recorded using eight electrodes and the classification model was based on Filter-Bank Common Spatial Patterns and a Linear Discriminant Analysis. The classification models were evaluated in cross-task applications, namely when playing Tetris at different speeds (i.e., slow, medium, fast) and when watching two videos (i.e., advertisement and landscape video). Additionally, subjects' perceived engagement was

quantified using a questionnaire. *Main results.* The models achieved a classification accuracy of 90 % on average when tested on an independent d2 test paradigm recording. Subjects' perceived and estimated engagement were found to be greater during the advertisement compared to the landscape video (p = 0.025 and p < 0.001, respectively); greater during medium and fast compared to slow Tetris speed (p < 0.001, respectively); not different between medium and fast Tetris speeds. Additionally, a common linear relationship was observed for perceived and estimated engagement ($r_{rm} = 0.44$, p < 0.001). Finally, theta and alpha band powers were investigated, which respectively increased and decreased during more engaging states. *Significance.* This study proposes a task-specific EEG engagement estimation model with cross-task capabilities, offering a framework for real-world applications.

1. Introduction

A brain-computer interface (BCI) provides a human-computer communication channel based solely on brain activity, without the need to physically move body parts [1]. For decades, researchers widely investigated BCI systems to help patients with severe motor disabilities to communicate with the outside world and indirectly restore motor function. However, direct access to brain activity can also be used to improve and support human-machine interaction in general, providing an additional implicit communication channel.

So-called passive BCIs are based on brain activity that is not voluntarily modulated by the user to control an application [2]. This activity is then interpreted so that the computer can adapt to the user's mental state. In general, users can perform their daily activities and a system records changes in their cognitive state. With passive BCIs, a wide variety of cognitive states can be investigated, e.g., fatigue [3], vigilance, working memory load [4], attention, frustration [5], emotions, and engagement.

Engagement is a positive state comprising several dimensions, including workload state, attention, motivation, interest, emotions, and perceived time spent on a given task [6,7]. Measuring the level of engagement is of great interest in many contexts. For example, in a healthcare scenario, the level of patient engagement may decrease during a rehabilitation session and a pause is necessary. Keeping patient engagement high can improve treatment outcomes [8]. In the context of education, engagement detection could be a new input channel of an adaptive automatic teaching platform to improve learning effectiveness [9–11]. In an entertainment context, a video game could adapt to the player's level of engagement to keep it high, for example, by introducing obstacles or increasing the speed of a game [12].

Concerning the measurability of engagement, evaluation grids and self-assessment questionnaires, interviews, and direct observation in the field or in a controlled

environment have traditionally been the most widely used methods [13]. In recent years, engagement is estimated indirectly using biosignals such as galvanic skin response [14], heart rate [15] or electroencephalography (EEG) [7, 16, 17]. Among these modalities, EEG seems to be one of the most promising technologies as a wide range of information on the subject's mental state can be derived from it.

One of the best-known parameters in the literature for measuring EEG-based engagement is the engagement index [7,16–19]. It is defined as the ratio between the beta power and the sum of theta and alpha powers associated with certain EEG channels. It is based on the hypothesis that an increase in beta-band power is correlated with an attentive state of the brain, such as the performance of a mental task, activation of the visual system, and movement planning activity [7]. In contrast, increases in alpha and theta activity correlated with decreased vigilance and mental alertness. In particular, decreases in alpha-band power are related to the processing of important information [20], and increases in this power are associated with a resting state. For these reasons, theta, alpha, and beta bands are among the prominent frequency bands in the estimation of engagement. However, this engagement index seems to not take into account the different dimensions of engagement. In [9], cognitive and emotional engagement in learning was detected using a filter bank common spatial patterns (CSP) approach, which provided superior performance compared to the engagement index. In [21], a hybrid system based on eye tracking and engagement-based BCI is proposed to play a hands-free version of the Tetris game. The level of user engagement is again detected by means of a filter bank CSP and was used to control the speed of the game.

In EEG-based engagement detection, features are extracted from the signal and then classified using machine learning techniques. In a supervised approach, the algorithm has to be trained to distinguish different mental states. Therefore, subjects must be put in these specific states to obtain the training data. Eliciting these mental states can be achieved using psychometric tests. These tests are based on the assumption that what they measure is also required to perform the test. However, there is no specific test that measures engagement. Instead several different tests are commonly used for this aim. Examples of common tests are the continuous performance task, a test used in neuropsychology for the assessment of sustained and selective attention [9, 22]; the Stroop test that is used in psychology to determine the flexibility of cognitive thinking [23]; the d2 test that is a neuropsychological measure of selective and sustained attention and visual scanning speed [24, 25]; the n-back test that measures the working memory capacity [26]; other memory tests [21].

This study focuses on the research and development of a passive BCI that continuously estimates participants' engagement using neural oscillations recorded by EEG. A computerized and modified version of the d2 test is proposed to increase the level of user engagement and elicit distinctive features for within-subject classifier training. Subsequently, the generalizability of the trained classifier to estimate the user's level of engagement was evaluated online during different tasks.

2. Materials and Methods

2.1. Subjects

Twenty-three healthy subjects [S01-S23, seven females and sixteen males, age: 34 (7) years in mean (SD)] participated in the experiment. Subjects did not receive any prior training regarding this experiment and had normal or corrected to normal vision.

The study was conducted in accordance with the principles embodied in the Declaration of Helsinki and approved by the Comitè d'Ética de la Recerca de la UVic-UCC (code 156/2021), Catalonia, Spain. All subjects provided written informed consent before taking part in the experiment.

2.2. Experimental protocol

Figure 1 shows the experimental procedure. The experiment included two recordings of the d2 test paradigm followed by two tasks: playing Tetris and watching videos. The d2 test paradigm consists of an alternation between engaging and resting state. Specifically, the d2 test and a fixation cross were used for this purpose.
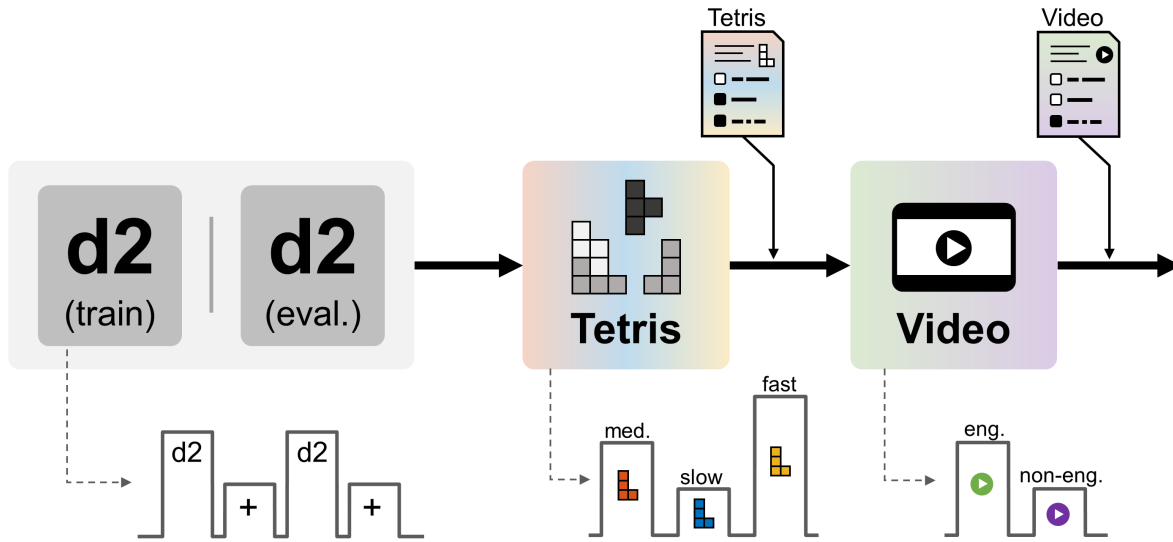

Figure 1: Experimental procedure. Subjects performed two d2 test paradigms, resulting in two recordings. The first recording was used to train a classification model to discriminate between the engaging (d2 test) and resting (fixation cross) states. The second recording was used as an independent data set to evaluate the predictive power of the classification model. Then, the classification model was used to estimate subjects' engagement during two different tasks in real-time. Specifically, subjects played Tetris at three different speeds and watched an engaging and non-engaging videos. After both the Tetris and video paradigms subjects filled out a subjective questionnaire quantifying their perceived engagement.

The d2 test is a neuropsychological measure of selective and sustained attention and visual scanning speed [25]. The task for the user is to select any letter "d" with two marks above it, or with two marks below it, or even with one mark above and one mark

below it. The surrounding distractors are similar to the target stimulus, for example the letter "d" with one, three or four marks, and also the letter "p" with one to four marks above and below it. The test comprises 14 lines with a maximum time of 20 seconds/line.

In the present study, some modifications were made to the classic d2 pencil and paper test. Specifically, a new computerized version was developed to try to elicit engagement in the subjects. First, compared to the traditional d2 test, a likely more difficult version of the d2 test was performed, including also the letters "b" and "q" with one to four marks above and below them as distractors. Moreover, the test proposed in this paper comprises 12 lines and 57 items per line. A total of one minute was given for the test. Second, the computerized version provides the user with feedback related to the correct d2, making the task more engaging. Specifically, correctly selected d2s are shaded in blue, as incorrect selections are shaded in red. In addition, a blue bar indicates to subjects how much time they have left to complete the task. This bar was added with the aim of increasing motivation and engagement during the d2 test but it may also induce stress. An example of the proposed d2 test is shown in Figure 2.
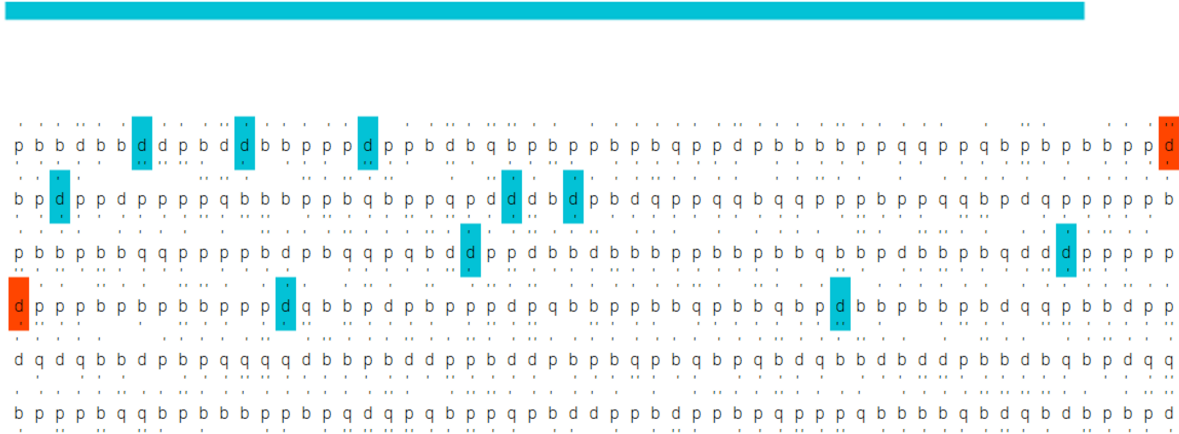

Figure 2: Example of the proposed d2 test showing the first 6 out of 12 lines with 57 items per line. Subjects were instructed to click on all d2s as quickly as possible. Correctly selected d2s are shaded in blue, whereas incorrect selections are shaded in red. The blue bar at the top indicates the time (fills from left to right).

Before the experiments, subjects were instructed on how to perform the d2 test with an example. They also played Tetris to get familiar with the keys on the keyboard to use.

The first part of the experiment consisted of EEG data collection to train the classification model to discriminate between the engaging and resting states of the subject. Subjects were instructed to perform the d2 test line by line as fast as possible, from left to right, to keep their level of engagement high. Conversely, they were instructed to relax as much as possible during the fixation cross while keeping their eyes open and fixated on the cross. In total, two d2 test paradigms were recorded

(Figure 1). The first recording was used to train the classification model, whereas the second was used as an independent dataset to evaluate the predictive power of the classification model.

Each recording comprised two 1-min blocks of the d2 test and two 1-min blocks of the fixation cross. The conditions were alternated; however, the starting condition was randomized for each of the two recordings. Between each block a 5-second pause was inserted for subjects to briefly relax.

At the end of the first recording, the developed application estimated how accurate the model was in detecting the two mental states. If the classification accuracy was above 75 %, the second d2 test paradigm was recorded (see Section 2.4.1). Otherwise, the first recording was repeated and the model was re-trained based on solely the newly acquired data. Subjects who did not not reach the specified threshold, even after re-training, nevertheless proceeded with the rest of the experiment.

After the two recordings of the d2 test paradigm, the same classification model was used to estimate the subjects' level of engagement during different tasks, namely playing Tetris and watching videos.

During the Tetris paradigm, subjects played Tetris at three speeds: slow, medium, and fast. Each speed level lasted one minute with a 30-second pause between the games. The order of the Tetris speeds was randomized. Subjects were instructed to move the blocks left and right or rotate them using the arrow keys on the keyboard, but they could not push the blocks down.

Regarding videos, two types were provided. One was expected to be engaging, as the scenes changed rapidly and it was a short advertisement ‡, the other as non-engaging, as it was a video of a landscape §. Each video lasted one minute and the order of the two videos was randomized. Both videos were provided without sound. No specific instructions were given to the subjects for this task.

Also the Tetris and video paradigms themselves were randomized.

After both the Tetris and video paradigms subjects filled out a subjective questionnaire in order to quantify their perceived engagement. Based on its definition [6,7], the questions investigated five dimensions of perceived engagement using a 7-point Likert scale [27]:

- The [X] was challenging.
- I felt that time passed too quickly during the [X].
- What happens in the [X] moved me emotionally.
- It was easy to focus on the [X].
- The [X] was visually pleasing.

where [X] is replaced for each of the three speeds of Tetris (i.e., slow, medium, and fast) and the two videos (i.e., short advertisement and landscape).

‡ https://www.youtube.com/watch?v=vgM4jfii4B8
§ https://www.youtube.com/watch?v=8vyVw0qcpe4

2.3. BCI system description

2.3.1. EEG device The EEG was recorded with the Unicorn Hybrid Black system (g.tec medical engineering GmbH, Austria). It comprises 8 hybrid EEG electrodes, i.e., it is possible to acquire EEG both with and without gel. Instead of the standard device assembly with the electrodes positioned at Fz, C3, Cz, C4, Pz, PO7, Oz, PO8, the parieto-occipital region of the brain was covered. Hence, electrodes were placed at P3, P4, PO3, POz, PO4, O1, Oz and O2. This choice was made to reduce the possibility of adversely biasing the classification model and the reason was twofold. Firstly, electrodes over motor areas may be influenced by motor actions due to mouse movements during the d2 test. Secondly, the chosen posterior electrode locations are less affected by eye movements and blinking compared to more anterior electrodes. Moreover, previous studies have shown involvement of parieto-occipital areas in cognitive processes such as working memory, processing of visual and sensory information, particularly in relation to arousal [18, 28–31].

The ground and reference were placed on the subject's left and right mastoids, respectively, using disposable adhesive surface electrodes. The EEG signal was acquired at a sampling rate of 250 Hz and a resolution of 24 bits.

The device was connected to a personal computer via the integrated Bluetooth interface. In this study, the EEG signal was acquired with wet electrodes. Therefore, the hybrid electrodes were filled with conductive gel.

2.3.2. Application A dedicated application for the stimuli presentation, EEG acquisition and online processing of the acquired data was developed. Specifically, the application processes the EEG data, estimates the level of users' engagement, and is able to provide them with feedback (i.e., current engagement estimate) in real-time.

Prior to the measurements, the signal quality can be assessed in the application. In addition, information on the number of noisy channels is available throughout the use of the application. The application allows for the selection of the paradigm to perform. As described in Section 2.2, the d2 test paradigm is exploited to train the classification model. The paradigm consists of the d2 test and the fixation cross.

Reliable performance scores [25, 32, 33] are calculated by the application after the d2 test paradigm:

(i) total processed items: sum of the number of items processed;

(ii) error of omission: sum of the number of target items not canceled;

(iii) total correctly processed: total items processed minus total errors made;

(iv) accuracy: total correctly processed items divided by all processed items;

(v) concentration performance: total number of correctly canceled minus total number incorrectly canceled items.

In the classic paper and pencil version, the scores are separately calculated for each of the 14 lines of the test [33]. In the proposed computerized version, the performance

scores were calculated considering all the processed items (i.e., from the first item until the last selected one) in one minute.

Regarding the fixation cross, the application shows a cross in the center of the screen. As with the d2 test, the time passed is indicated by the blue bar.

After the first d2 test paradigm, the application assesses the goodness of the classification model in discriminating between the engaging (d2 test) and resting (fixation cross) states. Additionally, the application offers the possibility to re-train the model (see Section 2.4.1 for details). This calibration phase, including gelling of the electrodes, d2 test paradigm and training of the classification model takes approximately 5 min.

Once the model is trained, it is possible to perform any task, even outside the application, and to check the user's level of engagement during these tasks in real-time.

2.4. Proposed classification model

The computational cost of the EEG processing pipeline was kept low as it is intended to be used in online (i.e., real-time) BCI experiments. Therefore, a filter bank CSP-based approach was used for estimating engagement. Models were trained in a within-subject manner, meaning every subject had their own model which was exclusively trained on their EEG data.

The raw EEG recordings were notch-filtered at 50 Hz by means of a 2nd order Butterworth filter. Then, data were filtered using a filter bank with 4 Hz to 8 Hz, 6 Hz to 10 Hz, and 8 Hz to 12 Hz band-pass filters. Higher frequency components (e.g., beta) were not included as they may be contaminated by muscle artifacts. Specifically, as subjects were looking at a computer screen while performing a visually demanding task (i.e., d2 test), the tension in the neck and back muscles may contaminate higher frequency components in the EEG.

The EEG data was segmented into non-overlapping 1-second windows in order to also achieve a 1-second resolution for the online processing. This resulted in 120 segments for each the d2 test and resting classes. Features were extracted for these EEG segments using the CSP algorithm [34]. Specifically, CSPs maximize the variance between the two classes and reduces the number of features. A similar approach was also proposed in [9] and was found to be successful in detecting engagement. Finally, the features obtained were used to train the Linear Discriminant Analysis (LDA) [35]. The LDA returns two outputs, namely the LDA label and the LDA score. The former is a binary value reflecting the predicted class and can thus be used to assess the accuracy, whereas the latter is a continuous numerical value and was used to estimate subjects' engagement.

2.4.1. Goodness of the classification model After the calibration phase (i.e., first d2 test paradigm), the goodness of the proposed classification model was assessed. Specifically, the non-overlapping 1-second EEG segments of this first paradigm were used to obtain

a 10-fold cross-validation accuracy. In the current study, the model was re-trained once if the cross-validation accuracy was below 75 %. Note that temporal correlations may still persist between consecutive non-overlapping 1-second EEG segments, attributable to filtering processes. However, here these 1-second segments can be assumed to be statistically independent as they only contain frequency components above 4 Hz (see Section 2.4).

2.4.2. Classification accuracy for independent recording To evaluate the predictive power of the classification model, the classification model trained on the first d2 test paradigm was tested on an independent recording (see Figure 1). For this purpose, the LDA labels were used to assess the accuracy for the second d2 test paradigm.

For comparative purposes, the results obtained with the proposed approach were compared with the ones obtained by means of a classical approach, namely the engagement index [7,18]. Specifically, the engagement index was computed as the ratio between the beta (13 Hz to 30 Hz) power and the sum of theta (4 Hz to 8 Hz) and alpha (8 Hz to 12 Hz) powers according to equation 1. The band powers were computed for the same non-overlapping 1-second windows and channels. The resulting engagement indices were then used as features for the LDA.

$$EI = \frac{\beta}{\alpha + \theta} \tag{1}$$

Again, the LDA labels were used to assess the accuracy for the second d2 test paradigm. The accuracy obtained was then compared with the one obtained using the proposed approach.

2.4.3. Online engagement estimation After the two recordings of the d2 test paradigm, the classification model was used online to estimate the subjects' engagement during the Tetris and Video recordings (see Figure 1). Again, the features were computed based on non-overlapping 1-second windows as described in Section 2.4. The LDA scores were used to quantify the estimated engagement.

2.5. Statistical analysis

The statistical analyses were performed using MATLAB R2021b (MathWorks Inc., United States). The normality of data was tested using the Shapiro–Wilk test [36]. The statistical test was chosen according to the normality of the sample, so either the paired t-test or the Wilcoxon signed rank test was used.

Descriptive statistics are reported as mean and the standard deviation (SD), or median and the interquartile range (IQR) (i.e., 25th and 75th percentile). The Bonferroni correction was used in the case of multiple comparisons to control for type I errors.

The difference in subjects' d2 test performance between the two d2 test recordings was investigated. Therefore, the obtained performance scores were compared between the two recordings and the p-values were Bonferroni corrected. The perceived

engagement was analyzed based on the sum of the engagement dimensions of the Tetris game questionnaire. Additionally, the scores of the classification model during the four conditions (rest period, low speed, medium speed, and fast speed) were compared in pairs recordings and the p-values were Bonferroni corrected. Perceived engagement during the two videos was also analyzed based on summing up engagement dimensions of the questionnaire for the two videos, as well as the scores of the classification model during the two conditions (engaging and non-engaging video).

Finally, a correlation analysis was performed to investigate the relationship between perceived engagement and estimated engagement. As subject-specific models were generated, the scores provided by these models are also subject specific. In other words, scores should not be compared between subjects but must be compared within subjects or in a paired manner as previously described. In this analysis, each subject's estimated engagement during the five conditions (low speed, medium speed, fast speed, engaging video and non-engaging video) were standardized to achieve a common scale across subjects. Then, the median z-score (i.e., median standardized estimated engagement) during each condition was calculated for each subject. Finally, a repeated-measures correlation was performed to investigate the common linear relationship between perceived and estimated engagement [37–39]. The repeated-measures correlation was utilized as each subject was present five times in the dataset due to the five conditions. Therefore, the observations in the dataset are no longer be independent, violating the assumption of independence of the classic regression/classification [40].

2.6. Topography plots

A group level analysis was carried out to investigate changes in the frequency bands, which were used for creating the classification models. Specifically, the theta and alpha frequency bands were further investigated.

In this offline analysis, each subjects' EEG data during the training, Tetris and video paradigms were processed using the FastICA algorithm [41]. This was done to extract and remove components that reflect artifacts such as muscle, movement, and cardiac artifacts. Before applying the FastICA, EEG data were notch filtered at 50 Hz and high-pass filtered at 1 Hz using a 2nd order Butterworth filter. After removing components, which showed obvious artifacts, the data were back-projected to electrode (i.e., channel) level and the short-time Fourier transform was used to estimate the power. A 4-second Hamming window with 50 % overlap was employed, resulting in a frequency resolution of 0.25 Hz. The frequency bins for each of the two frequency bands (i.e., theta and alpha bands) were then averaged and log-transformed, resulting in the two log-transformed band power signals for each channel. Finally, the median log-transformed band power was computed during each condition (e.g., d2 test and resting condition in the d2 paradigm).

The following comparisons were performed for each frequency band and channel: during the training recording of the d2 paradigm, the difference in log-transformed band
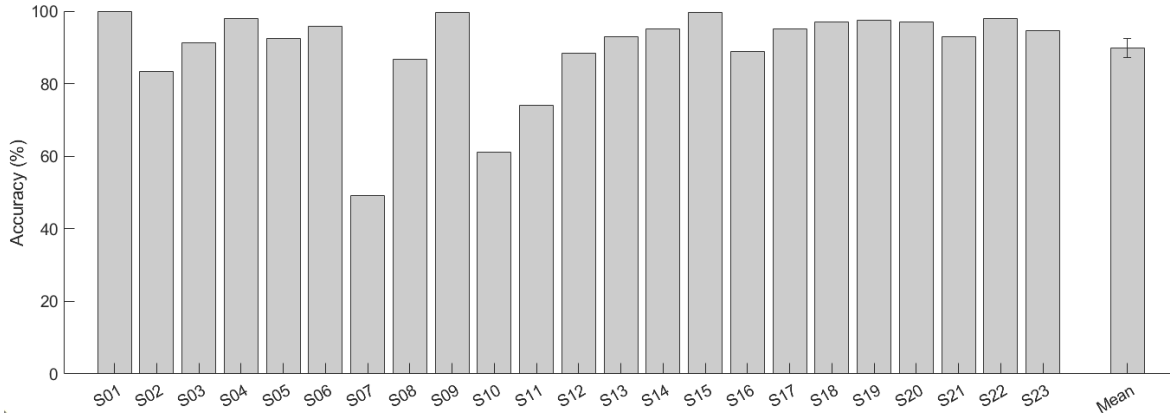
Figure 3: Classification accuracy for each subject (S01 to S23) discriminating between engaging (d2 test) and resting (fixation cross) states obtained for an independent recording using the proposed model (i.e., filter-bank common spatial patterns). The error bar on the mean reflects the standard error of the mean, and the classification accuracy was 94.6 [88.5, 97.4] % in median [IQR].

power was calculated between d2 test and resting condition. During the Tetris paradigm, the differences between each speed and the resting condition were computed. During the video paradigm, the difference between the engaging and non-engaging video was computed. Differences reflect the mean difference across all subjects. Additionally, Cohen's d_z (one sample, paired statistics) was calculated to quantify the observed effect size [42]. Finally, the computed log-transformed band power differences for each electrode and comparison were visualized using a 3D model of the MNI-152 template brain (Montreal Neurological Institute, Canada). A custom software (g.tec medical engineering GmbH, Austria) was used to visualize the changes in band power according to [43, 44] on the scalp of the 3D model. Additionally, electrodes were color-coded to infer the observed effect sizes.

3. Results

3.1. d2 test paradigm

None of the subjects required re-training of the classification model. In other words, all of them exceeded 75 % estimated classification accuracy based on the 10-fold cross-validation on the first d2 test paradigm.

Figure 3 shows the classification accuracy for discriminating between engaging (d2 test) and resting (fixation cross) states for the independent recording, namely the second d2 test paradigm. Specifically, both the accuracy for individual subjects and the across-subject mean together with the associated type A uncertainty $(90 \pm 3 \%)$ are reported. Twenty out of the twenty-three subjects exceeded 80 % accuracy for their independent recording.

In the offline analysis, the classification accuracy for discriminating between

Measure	Training	Evaluation	Δ	p-value
Total processed items	466.6 (86.7)	456.2 (78.4)	-10.3 (53.4)	1.000
Omissions	6.0 (4.7)	4.1 (3.2)	-1.8 (4.3)	0.281
Total correctly processed	460.3 (85.1)	451.6 (77.1)	-8.7 (52.3)	1.000
Accuracy (%)	97.6 (1.7)	98.2 (1.3)	0.6 (1.7)	0.535
Concentration Performance	18.8 (7.5)	18.4 (7.7)	-0.3 (8.5)	1.000

Table 1: d2 test performance scores computed for the training and evaluation recording. The results are reported across subjects as mean (SD). The difference in performance scores (Δ) is computed as training minus evaluation. Depending on the normality of the differences, the Wilcoxon signed rank test or paired t-test were used for statistical testing. Finally, the p-values were Bonferroni corrected.

engaging and resting states was also calculated using the engagement index. In this case, the across-subject mean accuracy together with the associated type A uncertainty was equal to $81 \pm 3\%$, losing about 10% in mean accuracy compared to the proposed approach.

In order to also take into account how subjects performed the d2 test during the two recordings, the performance scores calculated by the application were evaluated. Table 1 shows the results as mean (SD) for each recording and for the difference of them. No difference in d2 test performance between the two recordings was found.

3.2. Tetris

Figure 4(A) shows the subjective answers for each included engagement dimension. Specifically, the median of the answers across subjects for each speed and dimension is reported. Figure S2 in the Supplementary material presents the same data and results as Figure 4(A), with the inclusion of individual data points for each subject. The sums of the engagement questionnaire dimensions were observed to be 14.5 (3.2), 22.9 (3.8) and 24.9 (4.7) points in mean (SD) for the slow, medium, and fast speed, respectively. Subjects experienced the medium and fast speeds as more engaging than the slow speed ($p < 0.001$, respectively), whereas no difference was observed between medium and fast speeds ($p = 0.081$).

Figure 4(B) shows the results of subjects' estimated level of engagement provided by the classification model as a score for each Tetris condition. The y-axis shows the classification model score. Negative score values reflect states which are closer to the resting state (i.e., fixation cross), whereas positive score values reflect states closer to the engaging state (i.e., d2 test). The boxplots reflect the median scores during each condition and for each subject. Overall, a difference in median score can be observed between the conditions with predominantly negative values during rest between the three speeds. For the other three conditions, the values are predominantly positive. Table 2
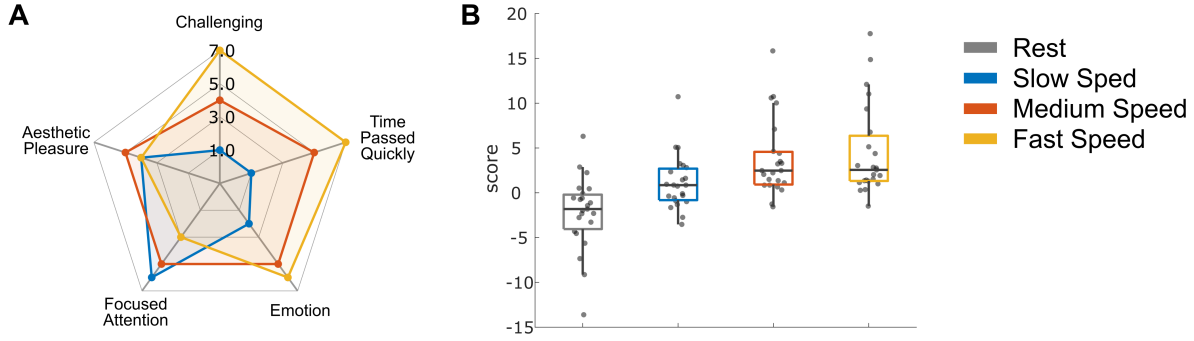
Figure 4: Tetris paradigm: Across subject results. (A) Subjective answers (i.e., perceived engagement) from the questionnaire (7-point Likert) are reported as median value per each question. (B) Objective results from the proposed classification model (i.e., estimated engagement), specifically each boxplot reflects subjects' median LDA scores. In addition, individual data points for each subject are shown, indicating their median LDA score in the respective condition.

Tetris	Δ	p-value
Rest < Slow	2.2 [0.4, 5.5]	< 0.001
Rest < Medium	3.6 [2.1, 6.7]	< 0.001
Rest < Fast	4.2 [2.1, 7.1]	< 0.001
Slow < Medium	1.5 [0.7, 3.8]	< 0.001
Slow < Fast	2.0 [0.6, 3.3]	< 0.001
Medium = Fast	0.1 [-0.5, 0.8]	0.345

Table 2: Tetris paradigm: Across subject statistical analysis using the Wilcoxon signed rank test. Subjects' estimated engagement (median LDA score) obtained by the proposed classification model were used. The difference in estimated engagement (Δ) for the respective comparison was calculated by subtracting the latter condition from the former (e.g., Slow minus Rest). The differenes are reported as median [IQR]. Depending on the normality of the differences, the Wilcoxon signed rank test or paired t-test were used for statistical testing. Finally, the p-values were Bonferroni corrected.

shows the results across subjects, comparing the difference between the four conditions in pairs. As expected, the three Tetris speeds were found to be more engaging than the rest period ($p < 0.001$, respectively). The same applies to the medium and fast speeds compared to the slow speed ($p < 0.001$, respectively). Similarly, as for the subjective questionnaire, no difference was observed between the medium and fast speeds ($p = 0.345$). Taken together, subjects' perceived engagement is in line with the objective evidence put forward by the classification model scores.

3.3. Video

Figure 5(A) shows the subjective answers for each included engagement dimension. Note, that the "Challenging" dimension was set to 1 for comparison purposes, as it was not asked for the videos. The median of the answers across subjects per each video and per each dimension is reported. Figure S3 in the Supplementary material presents the same data and results as Figure 5(A), with the inclusion of individual data points for each subject. The sum of the engagement questionnaire dimensions during the video paradigm was observed to be 16.7 (4.6) and 20.3 (5.1) points in mean (SD) for the landscape and advertisement videos, respectively. Subjects experienced the advertisement video to be overall more engaging than the landscape video (p = 0.025).
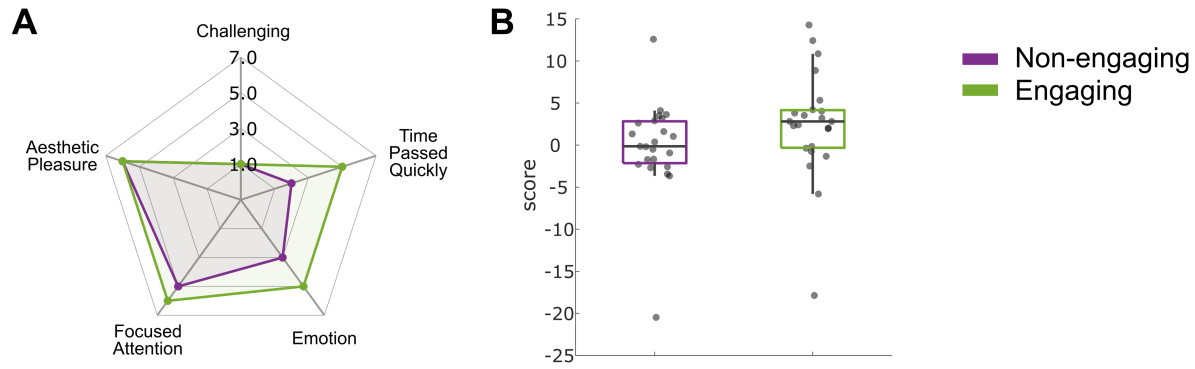

Figure 5: Video paradigm: Across subject results. (A) Subjective answers (i.e., perceived engagement) from the questionnaire (7-point Likert) are reported as median value per each question. As the "Challenging" dimension was not asked for the video paradigm and was set to 1 for comparison purposes. (B) Objective results from the proposed classification model (i.e., estimated engagement), specifically, each boxplot reflects subjects' median LDA scores. In addition, individual data points for each subject are shown, indicating their median LDA score in the respective condition.

Figure 5(B) shows the results of the subjects' estimated level of engagement provided by the classification model as a score for each video condition. The interpretation of this figure is analogous to Figure 4(B). A difference in estimated engagement is evident between the two videos, with predominantly negative or close to 0 values (i.e., associated with resting state) for the landscape video and predominantly positive values (i.e., associated with engaging state) for the advertising video. Finally, the across subject delta resulted in 1.8 [1.2, 3.5] in median [IQR] comparing the difference between the advertising and landscape videos. Consistent with the subjects' perceived engagement, the advertisement was observed to be more engaging by the classification model (p < 0.001).

3.4. Perceived and estimated engagement

Figure 6 shows the relationship between the perceived and standardized estimated engagement. The repeated-measures correlation shows a significant common linear

relationship between the perceived and standardized estimated engagement $r_{rm} = 0.44$ (95 % CI: 0.26 to 0.59), p < 0.001 (Figure 6(A)). Figure 6(A) shows for each subject (color-coded) the respective perceived and estimated engagement, including a linear regression with a common slope across subjects according to [39]. Additionally, Figure 6(B) and Figure 6(C) present the same data again but this time color-coded by condition, showing qualitatively that the perceived as well as estimated engagement are lower for slow Tetris speed and non-engaging video followed by medium and fast Tetris speeds.
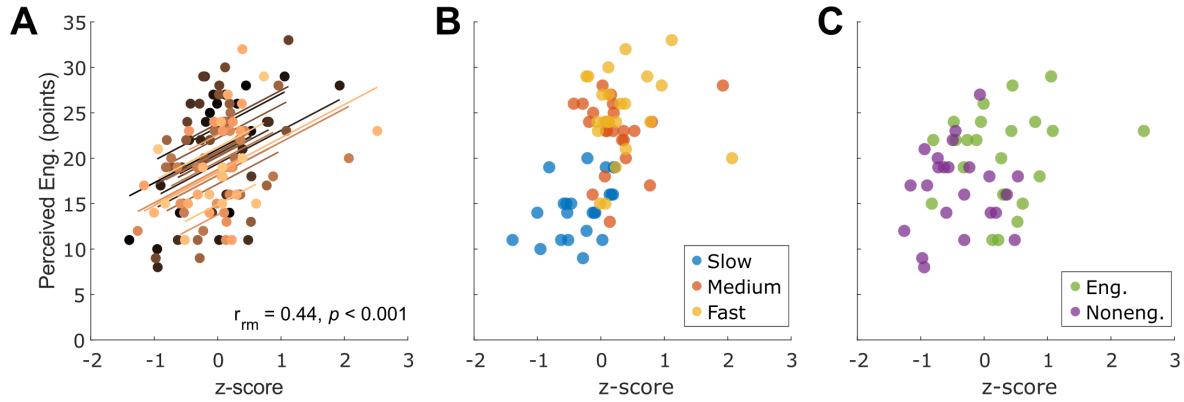

Figure 6: Relationship between perceived and standardized estimated engagement of the proposed classification model for both (A), (B) and (C). (A) Each subject's data are color-coded and separate linear regressions lines with common across subject slope are fit for each subject according to [39]. (B) Data are color-coded for the three Tetris game speed conditions: slow, medium and fast. (C) Data are color-coded for the engaging and non-engaging video.

3.5. Topography plots

Figure 7 shows the mean difference in log-transformed power between the conditions for the theta and alpha frequency bands.

Subjects had greater parieto-occipital theta power and lower parietal alpha power while performing the d2 test in comparison to the resting condition. While not shown here, the same phenomenon was also observed for the subjects' evaluation recording of the d2 paradigm. Furthermore, the changes in these two band powers between resting and d2 test condition were able to predict subjects' d2 test concentration performance (p = 0.041), as shown in a cross-validated regression analysis in the Supplementary material. A positive relationship between theta power and Tetris game speed can be observed over parietal and occipital areas. Specifically, greater Tetris speed was associated with greater theta band power. In contrast, parieto-occipital alpha power was decreased while playing Tetris and was modulated to a lesser extent during the fast Tetris speed.

Finally, parieto-occipital theta power was greater while subjects watched the engaging video, in comparison to the non-engaging video.
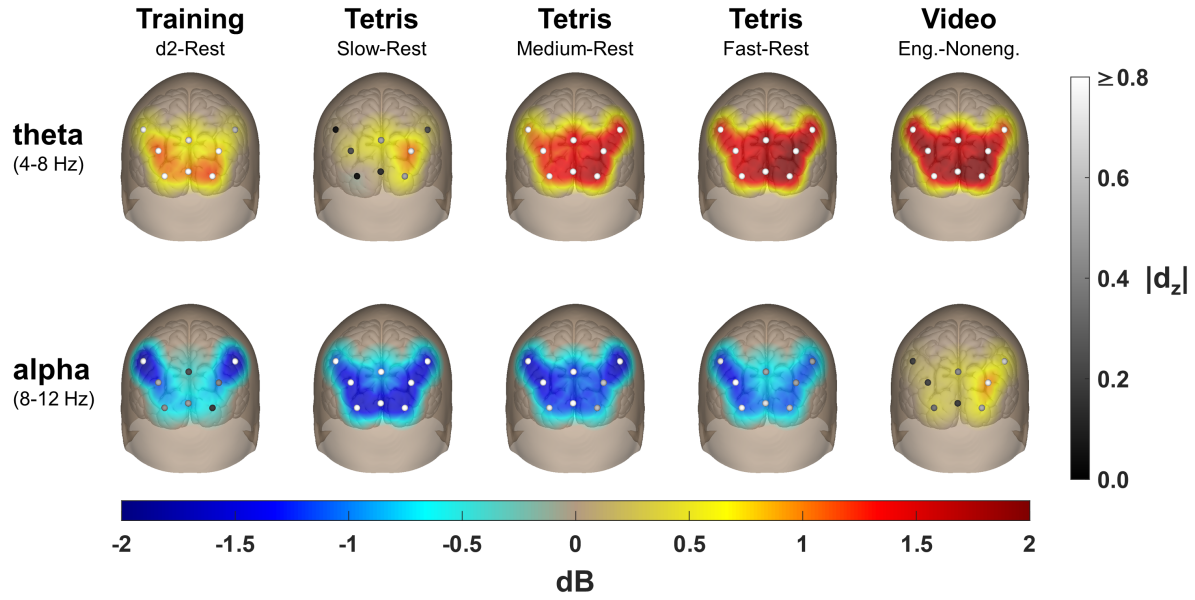
Figure 7: Topographic plots were used to visualize the spatial distribution of power modulation in theta (4 Hz to 8 Hz) and alpha (8 Hz to 12 Hz) frequency bands on the surface of the scalp. Specifically, they show the mean change in log-transformed theta and alpha power between the respective conditions (e.g., d2 test minus Rest) across all subjects with values reflecting dB. The eight points on the scalp represent the positions of the eight electrodes used to acquire the EEG signals. They are color-coded to infer effect size, quantified by the absolute value of Cohen's d_z with brighter colors indicating greater effect size.

4. Discussion

4.1. d2 test paradigm

This work showed that the proposed approach is able to discriminate between states of engagement and rest. Consistent with literature [9], it also confirmed that the engagement index is not as accurate as the filter-bank CSP algorithm. Notably, compared to the proposed approach, the accuracy obtained using the engagement index was about 10 % lower across subjects. This result could be related to (i) modulations in the beta band being smaller over parieto-occipital compared to more frontal regions [45] and/or (ii) theta power being positively related to engagement in the current study.

The scores computed based on the d2 test performance (Table 1) show that the subjects processed approximately 4 lines in one minute with a low error of omission and high accuracy. The concentration performance index was kept constant during the two recordings.

Although the most common performance scores were used [25, 32, 33], a direct comparison with the results in literature can not be made as a modified version of the d2 test was used in the present study. The reported performance scores were calculated considering all processed items during both the d2 tests in each recording (Section 2.3.2), however the test duration was 2 min in the proposed version compared to 4 min in the

original version. Nonetheless, the total processed items, number of omissions and total correctly processed are all comparable to the results in literature, when adjusting for the difference in test duration [25, 32]. On the other hand, the concentration performance score can not be compared due to the increased complexity of the proposed d2 test, which incorporates more distractors compared to the original version. Consequently, the number of correctly canceled items in the current version is substantially lower than the one reported in literature for the original d2 test.

No differences in d2 test performance score was observed between the two recordings, as the d2 tests were performed within minutes of each other.

4.2. Cross-task validity

The algorithm previously trained on the d2 test paradigm was used to estimate the subjects' level of engagement while playing Tetris and watching two videos. Engagement in gaming is of great interest both to improve player experience and because serious games are increasingly used, for example, to promote learning or in rehabilitation [6, 13, 46]. On the other hand, engagement while watching videos or movies could be used to predict population preferences in neuromarketing [47].

In the current study, both perceived and estimated engagement differentiate as well as correlate across different speeds of Tetris and the two video stimuli. For the Tetris game, the slow speed was found to be boring for the subjects as it was very slow. However, it allowed them to find the optimal location for each block and was associated with the greatest focused attention. In contrast, the fast speed was so fast that subjects had difficulty moving and rotating the blocks quickly enough. In comparison to the Tetris game, both videos were found to be engaging. One explanation for this observation could be that subjects have never previously watched these videos. Therefore, they paid attention to the details in the scene. For landscape video, the level of engagement may decrease over time as the scenes are mainly the same. In contrast, during advertisement videos, subjects may feel engaged throughout the duration of the video as the scenes change rapidly.

This was confirmed by the subjective questionnaire answers to the dimensions of challenging and perception of time (Figure 4(A) and Figure 5(A)). The subjects also reported that the medium and fast speeds, as well as the advertisement video, were more emotionally moving compared to the slow speed and the landscape video. In contrast to previous findings, no difference in aesthetic pleasure was found for the two videos [48] and the three Tetris speeds.

Future work could use this approach to adapt applications to users' estimated level of engagement, as proposed in [21]. Notably, while proposing adaptive modifications, [21] did not present any quantitative results.

4.3. Topography plots

The d2 test requires substantial visual processing and attention [24, 25], as subjects are performing visual pattern matching under the pressure of time. It is worth noting that the filter bank exploited in this study included the theta and alpha band as these two frequency bands are associated with cognitive demands, visual processing and spatial attention [49–53]. All of these aspects are expected to be greater during engaging tasks. Additionally, alpha power is inversely related to blood oxygenation level dependent (BOLD) activity, as well as connectivity in the visual system [54]. This is of relevance as visual cortex BOLD activity was recently shown to be able to predict subjective time (i.e., subjective reports of elapsed time) [55], which is a dimension of engagement.

Here, parietal and occipital increases in theta band power were observed during the d2 test, which were greatest over the occipital lobe. This activation of occipital theta is thought to be associated with control of cognitive demands [49] and the visual identification and recognition of patterns [50].

The same association was found with increasing Tetris game speed and for the engaging compared to the non-engaging video. This is to be expected, as the required cognitive demands, such as attention, increased with Tetris speed. Similarly, cognitive demands were greater during the engaging video, as it has changes in scenes, as well as short storylines including a person for the engaging video.

Additionally to increases in theta band power, decreases in parietal alpha band power were observed during the d2 test, which likely reflect spatial attention [51]. Specifically, alpha rhythms over the parieto-occipital area are known to desynchronize during the anticipation and processing of visual stimuli [52, 53] and can even predict visual target performance [56]. In line with [56], a similar result was found in this study, when using both theta and alpha band powers, providing insights regarding a possible causal brain-behaviour relationship. Decreases in alpha band power over parieto-occipital areas were also observed for all Tetris speeds, whereas no difference was observed for the two videos, with only the right occipital area showing a modulation. The results for the videos warrant further investigation, to ensure that it is not a spurious finding. However, the right hemisphere was found to be involved during processing of emotionally arousing stimuli [57, 58].

4.4. Limitations

Although this study provides insights into real-time classification of task-specific training-based engagement and cross-task applications of training models, some limitations should be pointed out.

Engagement involves multiple interdependent processes, including attention, working memory, emotions, decision making, and interest. Consequently, these behaviours are not confined to single brain regions. Rather, they involve a network of brain regions [59]. This network spans from the frontal region to parieto-occipital regions. The former is related to cognitive processes and the interpretation of emotional

engagement, especially in the context of valence, whereas the others contribute to the processing of visual and sensory information, particularly in relation to arousal [18, 29–31]. In this study, the electrode locations were chosen to eliminate or attenuate the influence of ocular artifacts, as well as event-related desynchronization (ERD) over the sensorimotor cortex due to hand movements during the d2 test. However, ERD may still be observed over parietal electrodes (i.e., P3 and P4) due to volume conduction. Future research could explore the integration of the frontal region to further understand the complexities of the engagement network within the brain.

In the current study videos were presented without audio, but future experiments could consider additional sensory modalities, such as the combination of visual and auditory elements. Furthermore, videos including people could also be chosen for the non-engaging video to further refine the modelling process.

The proposed classification model is able to estimate users' level of engagement during two distinct videos and different Tetris game speeds. However, it was trained on the d2 test paradigm, which is primarily a visual searching task. Thus, the model's performance may be influenced by perceptual factors related to the visual stimuli rather than solely reflecting cognitive engagement. One such factor is visual complexity which was found to be positively related to subjects' perceived engagement. Indeed, decoupling engagement from visual complexity would be difficult in the current setup, but future works may investigate this aspect.

5. Conclusion

This work focused on developing and evaluating a novel application for the real-time estimation of EEG-based engagement during different tasks. A classification model is trained to distinguish between users' brain activity during an engaging task (i.e., modified version of d2 test) and resting state. Once the model is trained, users may perform tasks either inside or outside this application, as the application continuously estimates the users' engagement. In this study playing Tetris and watching videos were tested.

The application was validated with an experimental campaign involving 23 subjects. The results show that the proposed classification model estimated levels of users' engagement consistent with users' perceived engagement. Furthermore, perceived and estimated engagement were found to be linearly related to each other. Finally, theta band power over parietal-occipital areas increased with increasing task engagement. The opposite association was found for alpha band power. These results are in line with studies investigating visual processing and attention.

The ability of the proposed method to estimate engagement should be further investigated by applying it to additional tasks. Furthermore, instead of using the application to estimate engagement, the application itself may even modulate different aspects of a game (e.g., difficulty) according to its estimate.

Declarations

Competing interests

SS, LS, MW, MS-R and RO were employed by g.tec medical engineering. AN was a visiting PhD student at g.tec medical engineering. CG was the CEO of g.tec medical engineering.

The remaining authors declare no conflict of interests concerning the publication of this manuscript.

Acknowledgment

References

[1] McFarland D J and Wolpaw J R 2011 Brain-computer interfaces for communication and control *Communications of the ACM* **54** 60–66

[2] Alimardani M and Hiraki K 2020 Passive brain-computer interfaces for enhanced human-robot interaction *Frontiers in Robotics and AI* **7** 125

[3] Cao T, Wan F, Wong C M, da Cruz J N and Hu Y 2014 Objective evaluation of fatigue by eeg spectral analysis in steady-state visual evoked potential-based brain-computer interfaces *Biomedical engineering online* **13** 1–13

[4] Mora-Sánchez A, Pulini A A, Gaume A, Dreyfus G and Vialatte F B 2020 A brain–computer interface for the continuous, real-time monitoring of working memory load in real-world environments *Cognitive Neurodynamics* **14** 301–321

[5] Reuderink B, Nijholt A and Poel M 2009 Affective pacman: A frustrating game for brain-computer interface experiments *Intelligent Technologies for Interactive Entertainment: Third International Conference, INTETAIN 2009, Amsterdam, The Netherlands, June 22-24, 2009. Proceedings 3* (Springer) pp 221–227

[6] Martey R M, Kenski K, Folkestad J, Feldman L, Gordis E, Shaw A, Stromer-Galley J, Clegg B, Zhang H, Kaufman N *et al* 2014 Measuring game engagement: multiple methods and construct complexity *Simulation & Gaming* **45** 528–547

[7] Ismail L E and Karwowski W 2020 Applications of eeg indices for the quantification of human cognitive performance: A systematic review and bibliometric analysis *Plos one* **15** e0242857

[8] Brett C E, Sykes C and Pires-Yfantouda R 2017 Interventions to increase engagement with rehabilitation in adults with acquired brain injury: A systematic review *Neuropsychological rehabilitation* **27** 959–982

[9] Apicella A, Arpaia P, Frosolone M, Improta G, Moccaldi N and Pollastro A 2022 Eeg-based measurement system for monitoring student engagement in learning 4.0 *Scientific Reports* **12** 5857

[10] Dewan M, Murshed M and Lin F 2019 Engagement detection in online learning: a review *Smart Learning Environments* **6** 1–20

[11] Khedher A B, Jraidi I and Frasson C 2019 Tracking students' mental engagement using eeg signals during an interaction with a virtual learning environment *Journal of Intelligent Learning Systems and Applications* **11** 1–14

[12] Biercewicz K, Borawski M and Duda J 2020 Method for selecting an engagement index for a specific type of game using cognitive neuroscience *International Journal of Computer Games Technology* **2020**

[13] Hookham G and Nesbitt K 2019 A systematic review of the definition and measurement of engagement in serious games *Proceedings of the australasian computer science week multiconference* pp 1–10

[14] Di Lascio E, Gashi S and Santini S 2018 Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* **2** 1–21

[15] Darnell D K and Krieg P A 2019 Student engagement, assessed using heart rate, shows no reset following active learning sessions in lectures *PloS one* **14** e0225709

[16] Nuamah J and Seong Y 2018 Support vector machine (svm) classification of cognitive tasks based on electroencephalography (eeg) engagement index *Brain-Computer Interfaces* **5** 1–12

[17] Kosmyna N and Maes P 2019 Attentivu: an eeg-based closed-loop biofeedback system for real-time monitoring and improvement of engagement for personalized learning *Sensors* **19** 5200

[18] Pope A T, Bogart E H and Bartolome D S 1995 Biocybernetic system evaluates indices of operator engagement in automated task *Biological psychology* **40** 187–195

[19] Dehais F, Dupres A, Di Flumeri G, Verdiere K, Borghini G, Babiloni F and Roy R 2018 Monitoring pilot's cognitive fatigue with engagement features in simulated and actual flight conditions using an hybrid fnirs-eeg passive bci *2018 IEEE international conference on systems, man, and cybernetics (SMC)* (IEEE) pp 544–549

[20] Sokoliuk R, Mayhew S D, Aquino K M, Wilson R, Brookes M J, Francis S T, Hanslmayr S and Mullinger K J 2019 Two spatially distinct posterior alpha sources fulfill different functional roles in attention *Journal of Neuroscience* **39** 7183–7194

[21] Krol L R, Freytag S C and Zander T O 2017 Meyendtris: A hands-free, multimodal tetris clone using eye tracking and passive bci for intuitive neuroadaptive gaming *Proceedings of the 19th ACM International Conference on Multimodal Interaction* pp 433–437

[22] Gaume A, Dreyfus G and Vialatte F B 2019 A cognitive brain–computer interface monitoring sustained attentional variations during a continuous task *Cognitive neurodynamics* **13** 257–269

[23] MacLeod C M 1992 The stroop task: The" gold standard" of attentional measures. *Journal of Experimental Psychology: General* **121** 12

[24] Schmidt M, Benzing V and Kamer M 2016 Classroom-based physical activity breaks and children's attention: Cognitive engagement works! *Frontiers in psychology* **7** 1474

[25] Bates M E and Lemay E P 2004 The d2 test of attention: construct validity and extensions in scoring techniques *Journal of the International Neuropsychological Society* **10** 392–400

[26] Kulikowski K and Orzechowski J 2019 Working memory and fluid intelligence as predictors of work engagement—testing preliminary models *Applied Cognitive Psychology* **33** 596–616

[27] Joshi A, Kale S, Chandel S and Pal D K 2015 Likert scale: Explored and explained *British journal of applied science & technology* **7** 396

[28] Chang W S, Liang W K, Li D H, Muggleton N G, Balachandran P, Huang N E and Juan C H 2023 The association between working memory precision and the nonlinear dynamics of frontal and parieto-occipital eeg activity *Scientific reports* **13** 14252

[29] Apicella A, Arpaia P, Isgrò F, Mastrati G and Moccaldi N 2022 A survey on eeg-based solutions for emotion recognition with a low number of channels *IEEE Access* **10** 117411–117428

[30] Li F, Zhang G, Wang W, Xu R, Schnell T, Wen J, McKenzie F and Li J 2016 Deep models for engagement assessment with scarce label information *IEEE Transactions on Human-Machine Systems* **47** 598–605

[31] Sauseng P, Klimesch W, Schabus M and Doppelmayr M 2005 Fronto-parietal eeg coherence in theta and upper alpha reflect central executive functions of working memory *International journal of Psychophysiology* **57** 97–103

[32] Steinborn M B, Langner R, Flehmig H C and Huestegge L 2018 Methodology of performance

scoring in the d2 sustained-attention test: Cumulative-reliability functions and practical guidelines. *Psychological Assessment* **30** 339

[33] Baghaei P, Ravand H and Nadri M 2019 Is the d2 test of attention rasch scalable? analysis with the rasch poisson counts model *Perceptual and Motor Skills* **126** 70–86

[34] Ang K K, Chin Z Y, Wang C, Guan C and Zhang H 2012 Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b *Frontiers in neuroscience* **6** 39

[35] Xanthopoulos P, Pardalos P M, Trafalis T B, Xanthopoulos P, Pardalos P M and Trafalis T B 2013 Linear discriminant analysis *Robust data mining* 27–33

[36] Yap B W and Sim C H 2011 Comparisons of various types of normality tests *Journal of Statistical Computation and Simulation* **81** 2141–2155

[37] Bland J M and Altman D G 1995 Statistics notes: Calculating correlation coefficients with repeated observations: Part 1—correlation within subjects *Bmj* **310** 446

[38] Bland J M and Altman D G 1995 Calculating correlation coefficients with repeated observations: Part 2—correlation between subjects *Bmj* **310** 633

[39] Bakdash J Z and Marusich L R 2017 Repeated measures correlation *Frontiers in psychology* **8** 456

[40] Cohen J, Cohen P, West S G and Aiken L S 2013 *Applied multiple regression/correlation analysis for the behavioral sciences* (Routledge)

[41] Hyvarinen A 1999 Fast and robust fixed-point algorithms for independent component analysis *IEEE transactions on Neural Networks* **10** 626–634

[42] Lakens D 2013 Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas *Frontiers in psychology* **4** 863

[43] Kubanek J and Schalk G 2015 Neuralact: a tool to visualize electrocortical (ecog) activity on a three-dimensional model of the cortex *Neuroinformatics* **13** 167–174

[44] Lee H S, Schreiner L, Jo S H, Sieghartsleitner S, Jordan M, Pretl H, Guger C and Park H S 2022 Individual finger movement decoding using a novel ultra-high-density electroencephalography-based brain-computer interface system *Frontiers in Neuroscience* **16** 1009878

[45] Capilla A, Arana L, García-Huéscar M, Melcón M, Gross J and Campo P 2022 The natural frequencies of the resting human brain: an meg-based atlas *NeuroImage* **258** 119373

[46] Abdul Jabbar A I and Felicia P 2015 Gameplay engagement and learning in game-based learning: A systematic review *Review of educational research* **85** 740–779

[47] Hoshi A, Hirayama Y, Saito F, Ishiguro T, Suetani H and Kitajo K 2023 Spatiotemporal consistency of neural responses to repeatedly presented video stimuli accounts for population preferences *Scientific Reports* **13** 5532

[48] Welke D and Vessel E A 2022 Naturalistic viewing conditions can increase task engagement and aesthetic preference but have only minimal impact on eeg quality *NeuroImage* **256** 119218

[49] Raghavachari S, Lisman J E, Tully M, Madsen J R, Bromfield E and Kahana M J 2006 Theta oscillations in human cortex during a working-memory task: evidence for local generators *Journal of neurophysiology* **95** 1630–1638

[50] Cartier C, Bittencourt J, Peressutti C, Machado S, Paes F, Sack A T, Basile L F, Teixeira S, Salles J I, Nardi A E *et al* 2012 Premotor and occipital theta asymmetries as discriminators of memory-and stimulus-guided tasks *Brain Research Bulletin* **87** 103–108

[51] Capotosto P, Babiloni C, Romani G L and Corbetta M 2009 Frontoparietal cortex controls spatial attention through modulation of anticipatory alpha rhythms *Journal of Neuroscience* **29** 5863–5872

[52] Steriade M and Llinás R R 1988 The functional states of the thalamus and the associated neuronal interplay. *Physiological reviews* **68** 649–742

[53] Klimesch W, Doppelmayr M, Russegger H, Pachinger T and Schwaiger J 1998 Induced alpha band power changes in the human eeg and attention *Neuroscience letters* **244** 73–76

[54] Scheeringa R, Petersson K M, Kleinschmidt A, Jensen O and Bastiaansen M C 2012 Eeg alpha power modulation of fmri resting-state connectivity *Brain connectivity* **2** 254–264

[55] Sherman M T, Fountas Z, Seth A K and Roseboom W 2022 Trial-by-trial predictions of subjective

time from human brain activity *PLOS Computational Biology* **18** e1010223

[56] Thut G, Nietzel A, Brandt S A and Pascual-Leone A 2006 α-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection *Journal of Neuroscience* **26** 9494–9502

[57] Canli T, Desmond J E, Zhao Z, Glover G and Gabrieli J D 1998 Hemispheric asymmetry for emotional stimuli detected with fmri *Neuroreport* **9** 3233–3239

[58] Lang P J, Bradley M M, Fitzsimmons J R, Cuthbert B N, Scott J D, Moulder B and Nangia V 1998 Emotional arousal and activation of the visual cortex: an fmri analysis *Psychophysiology* **35** 199–210

[59] Provenza N R, Paulk A C, Peled N, Restrepo M I, Cash S S, Dougherty D D, Eskandar E N, Borton D A and Widge A S 2019 Decoding task engagement from distributed network electrophysiology in humans *Journal of neural engineering* **16** 056015