This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3061719, IEEE Transactions on Circuits and Systems for Video Technology

PhyDAA: Physiological Dataset Assessing Attention

Victor Delvigne, *Student Member, IEEE,* Hazem Wannous, *Senior Member, IEEE,* Thierry Dutoit, *Member, IEEE,* Laurence Ris and Jean-Philippe Vandeborre, *Senior Member, IEEE*

***Abstract*—Attention Deficit Hyperactivity Disorder (ADHD) is the most prevalent neurodevelopmental disorder among children. It affects patients' lives in many ways: inattention, difficulty with stimuli inhibition or motor function regulation. Different treatments exist today, but these can present side effects or are not effective for all subgroups. Neurofeedback (NF) is an innovative treatment consisting of brain activity display. NF training could consist of a virtual reality (VR) video-game in which the participant's attention affects the game. Attention being assessed through physiological signals, one of the main steps is to design an estimator for the attention state. We present a novel framework able to record physiological signals in specific attention states and able to estimate the corresponding attention state. We propose a database composed of electroencephalography signals (EEG), and an eye-tracker labelled with a score representing the attention span for 32 healthy participants. Different features are extracted from the signals and machine learning (ML) algorithms are proposed. Our approach exhibits high accuracy for attention estimation, which corroborates a correlation between attention state and physiological signals (i.e. EEG, eye-tracking signals). The dataset has been made publicly available to promote research in the domain and we encourage other scientists to use their own approach for attention estimation.**

***Index Terms*—Brain-Computer Interface, Virtual Reality, Deep Learning, Attention Estimation.**

I. Introduction

ATTENTION plays an important role in human interaction with the surrounding environment. The attention state is a behaviour corresponding to the cerebral state where the participant is allocating a certain amount of resources to carry out a task. The attention state can be assessed by different methods consisting of questionnaires, neurophysiological tasks to study the primary and automatic answer to stimuli appearance and/or inhibition (these stimuli may be visual, auditory and even olfactory), and physiological recordings expressing the brain activity related to specific attention. In the context of attention deficit and hyperactivity disorder (ADHD), several research projects have shown that it is possible to assess symptoms with specific questionnaires and tasks (e.g. Go/noGo task, as described in the works of Blume et al. [10]), but also to classify ADHD and the related symptoms vs. a control group with biomedical recording as shown in the works of Liechti et al. [31] and Alchalabi et al. [1]. In these works the authors use an electroencephalogram (EEG) from a recreative [1] or biomedical [31] recorder to investigate the possible correlation between ADHD diagnosis and EEG records. An EEG is a signal acquisition method measuring the brain's electrical activity with an electrode cap placed on the participant's scalp.

One of the use cases of attention estimation from physiological recording is for neurofeedback (NF) training. NF consists of a real-time display of brain activity in an understandable form to teach participants how to control specific behaviour. For instance, in the case of people with ADHD symptoms, possible NF maybe a plant growing if the participant's attention is increasing or fading if the participant is not focused on the task. The existing works aiming to assess attention tend to show that it is possible to estimate the attention state with spectral information from EEG, as shown in the works of Bioulac et al. [9] and Cueli et al. [14]. However, Arns et al. [4] show that this relationship is not validated for every participant' group, and especially in the specific case of an ADHD subgroup. In this context, novel methods for estimating attention from EEG may be interesting, these improvements could be based on novel feature extraction and/or classification methods, as presented by Lotte et al. [33].

In another domain, projects aiming to promote interaction between humans and machines have seen increased interest in recent years. A specific subgroup of these works is Brain and Computer Interfaces (BCI) that aim to help humans interact with a computer through their brain. This interaction can be invasive or not depending on the considered signal acquisition method, e.g. Electrocorticography/intracranial EEG (ECoG) vs. traditional EEG, and more or less easy to use, e.g. MRI (magnetic resonance imaging) vs. EEG. Today, the existing BCI have different purposes: medical [38], entertainment [36], (neuro)marketing [47], etc. As explained above, BCI application can also be used to help people with ADHD to self-regulate their symptoms as, shown in the work of Bioulac et al. [9].

A possible improvement to BCIs is their use in virtual reality (VR) environments, which is motivated by the fact that they provide higher safety and freedom for creation compared to reality, and are more ecologically valid compared to 2D screens, as shown by Pollak et al. [39]. In their work, Bashiri et al. [5] present a review of all the works using VR in the context of ADHD research. The research project presented in their paper can be classified into different categories: assessment, diagnosis and neurofeedback for ADHD in an immersive environment, side effect studies of VR use and other works aiming to help people with a neurodevelopmental disorder, e.g. task training for people with autistic spectrum disorder,

V. Delvigne is with Information, Signal and Artificial Intelligence (ISIA) departement of Engineering Faculty of University of Mons, Mons, Belgium and IMT Lille Douai and CRIStAL UMR CNRS 9189, Lille, France. (E-mail: victor.delvigne@umons.ac.be)

T. Dutoit is with Information, Signal and Artificial Intelligence (ISIA) departement of Engineering Faculty of University of Mons, Mons, Belgium.

H. Wannous and J-P. Vandeborre are with IMT Lille Douai, CRIStAL UMR CNRS 9189, Lille, France

L. Ris is with Neuroscience Department Lab of Faculty of Medicine and Pharmacy of University of Mons, Mons, Belgium.

Manuscript received xx, 20 xx; revised xx , 20xx.

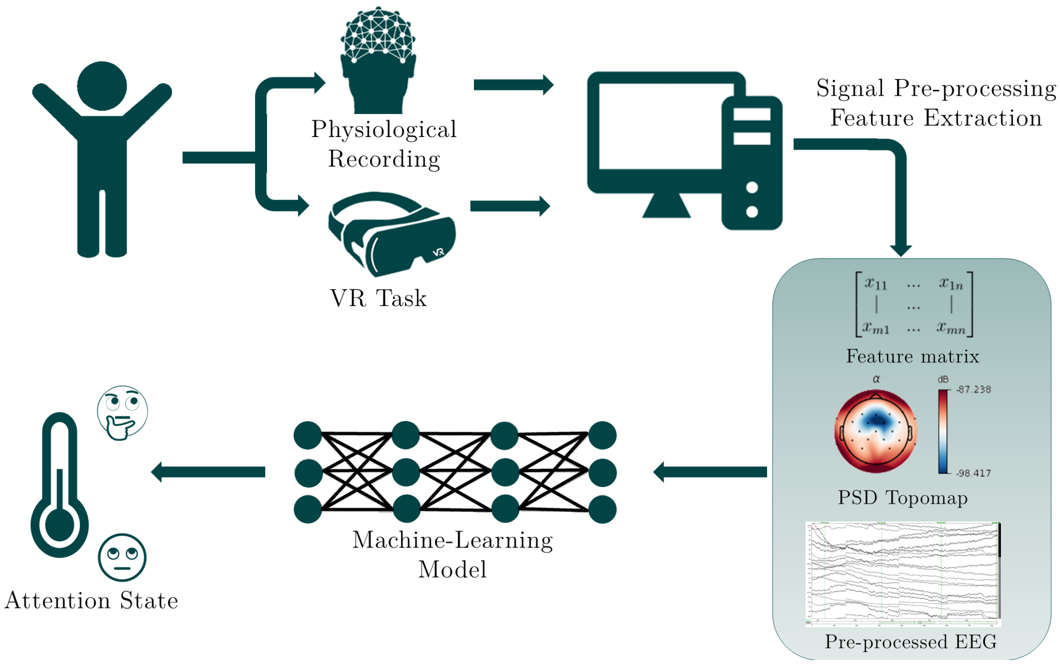

Fig. 1. Proposed framework recording and estimating attention. EEG and eye-tracking are simultaneously recorded during an attentional task. The results of the tasks are processed to estimate an attention score and a feature vector is estimated from the EEG segments. The different features that have been computed are used to train/test machine learning models to retrieve the attention score computed with the task results.

as presented by Ip et al. [25].

During the last decade, artificial intelligence (AI) and machine learning (ML) have seen an increase in their use. Various reasons explain this trend: robustness, higher accuracy, increase in the amount of data collection being able to be used in this context, etc. Although ML algorithms are known to be used in image processing and natural language processing, an increase of their use for medical application has been noted, as reported in Ravi et al.. These applications concern different topics, including drug synthesis [22], DNA understanding [2] and disease detection from medical imaging [3].

As explained above, a large quantity of signals is necessary to train deep learning algorithms. However, a lack of large datasets in the context of attention tasks has been noted. The main issue about dataset assessing attention from EEG (also for other EEG datasets) is that they have not been created to be used for machine learning applications. For instance, they do not promote reproducible works (the datasets being not always publicly available) and generic models (the labels being sometimes wrongly and/or not clearly defined for ML application). In the context of our research project, no datasets not presenting these issues were found.

In this context, for the purpose of this paper, it was decided to present a dataset consisting of EEG and eye-tracking signals during attentional tasks in virtual reality. The goal of this dataset is to characterise physiological signals (i.e. EEG, eye-tracking) corresponding to a degree of attention computed from attention assessment based on eye-tracking signals. The latter may be considered as a score ($\in [0 - 100\%]$) or a class, e.g. focused vs. distracted. Moreover, a framework to estimate the attentional state is also proposed. Its representation is shown in Fig 1. The goal of the dataset and framework is to have a better understanding of the attention mechanism through the brain activity of healthy participants. This knowledge is transferable for comparison with ADHD participants and to design NF to reduce their symptoms. The main contributions proposed by this paper are:

- We present a framework promoting attention aiming to record physiological signals and label them with selective and sustained attention assessment. A set of signals from 32 participants is also publicly available[1]. Moreover, we propose different approaches to retrieve the attentional state (computed from physiological assessments based on eye-tracking signals) from EEG.
- We propose a dataset helping to study the paradigm between EEG and eye-tracking. The experiment being carried out in VR, a large variability is proposed for the eye-tracking recording compared to 2D screen recording.
- The proposed dataset and its multimodal aspect may be of use for a great number of further work possibilities: visual attention map estimation, participant identification, attention estimation, etc.

This paper is organised as follows: a review of most attention-related works, teating ADHD and biomedical signal processing, is presented in Section II. The recording part of the framework is reported in Section IV, and Section III describes the estimation part of the framework in more details. Finally, the results of the attention estimation are presented in Section V, and the conclusion and further works are summarised in Section VI.

[1] https://vdelv.github.io/dataset.html

II. Related Work

This research project covers a wide range of scientific research and, in this section we will focus on the existing work which aims to estimate attention with questionnaires and physiological signals, and the general signal processing pipeline for EEG analysis.

A. Attention Assessment

There are different assessments to estimate a participant's attention state in several neurophysiological research projects. In most cases, these tasks consist of short exercises where the participant is asked to answer as quick as possible and/or to inhibit their reaction when presented with specific stimuli. The stimuli can take different forms and be: visual, auditory or combined.

To assess the attention state, there are different neurological tests:

- Go/noGo test, where participants are asked to react as quickly as possible in response to the appearance of a simple stimulus (e.g. an X on a screen). The mean response time is then computed and analysed to classify the participant's attention in the right group, as shown in Blume et al. [10].
- Conner's Continuous Performance Test (CPT), as shown in Eom et al. [19]. During this task the participant has to answer as fast as possible if a specific stimulus or a sequence of stimuli appeared. They are asked to not answer if a wrong stimulus or sequence of stimuli appeared. The number of stimuli that have been correctly (and wrongly) detected is then measured to assess the participant's attention.
- The Wisconsin Card Sorting Test where four cards are presented to the participant and they are asked to find the correct relationship between the cards. This assessment has already been used in research work aiming to investigate ADHD symptoms as in Mullane and Corkum [34].

These tests can be used to confirm a diagnosis of ADHD or to classify the participant in the correct subgroup of ADHD (i.e. hyperactive, inattentive or combined) [9], [10], [45]. As these tests have been validated by scientific studies, they can also be used to assess the attention of healthy participants. Assessing attention can also be an interesting approach to study the evolution of a disease affecting attention and/or the effect of a treatment as shown in the works of Muhlberger et al. [35] were they compare the performances of different ADHD subgroups (i.e. medicated, unmedicated ADHD children and healthy children) on a CPT. Moreover, it could also be interesting to consider defined attention assessments to compare them with novel methods/tests.

B. Physiological based Attention Assessment

In addition to neurophysiological assessments, it is also possible to find insights from physiological recordings. Signal acquisition has to be done in a specific environment/during a specific task. In the context of attention estimation, different signals can be considered:

- EEG signals representing the electrical activity of the brain over a period of time for specific scalp location depending on the location of the electrodes. Several works show that EEG can provide good insights for attention assessment. Most of these works present the relationship between the frequential behaviour of EEG and the attentional state as in the works of Coelli et al. [12] and Klimesch et al. [27]. However, Arns et al. [4] show that EEG spectral features may not be relevant to estimating attention in every situation (e.g. specific ADHD subgroups). In another domain, Zheng and Lu [53] and Cao et al. [11] both propose a dataset composed of EEG records assessing attention during a driving task. These works aim to help the detection of attention loss in a car driving environment. Moreover, other publications based on these dataset have proven that there is a correlation between physiological signals and vigilance or attention state [20], [50].
- Functional near-infrared spectroscopy (fNIRS) represents the blood oxygenation in several regions of the brain which has a direct relationship with cerebral activity. Blume et al. [10] present a novel treatment based on physiological records from fNIRS from children with ADHD.
- Eye-tracking signals and other information from the eye can also be relevant to assessing attention state. Varela Casal et al. [46] show that it is possible to assess attention from eye vergence information (i.e. feature representing the eyes' focal properties). Moreover, Garcia Boas et al. [21] present a neurofeedback training, based on eye-tracking, to reduce the symptoms of ADHD.
- Other biological signals: Electromyogram (EMG), electrocardiogram (ECG) or GSR (galvanic skin response) may also be interesting biomarkers for attention estimation. These biological signals have already been presented in a previous work aiming to estimate emotion with MPED: *"A Multi-Modal Physiological Emotion Database for Discrete Emotion Recognition"* by Song et al. [44].

C. Virtual Reality and Attention

Virtual Reality (VR) is taking more and more place in different research works related to attention estimation and/or ADHD. Bashiri et al. [5] present in a review all the works combining virtual reality and ADHD and the pros and cons of its use. In this context, VR can be used for several applications, such as treatment, attention assessment or NF training. As Pollak et al. [39] show, VR presents several benefits, including greater safety compared to reality, freedom for environment definition and control, and higher ecological validity compared to 2D screens.

Tan et al. [45] and Eom et al. [19] both present a VR environment through which they assess the attention state of participants with and without ADHD symptoms or diagnosis. The goal of these VR environments is to help classify participants into ADHD/non-ADHD groups by considering the results of a specific task. For instance, Eom et al. [19] show a high correlation between DSM-IV results, the standardised guide to establish a diagnosis of ADHD by the American Psychiatric Association, and the VR-CPT task defined in their work. Another approach can be to study the score evolution during neurophysiological assessments or related tasks to study attention over time or during trials.

Another use of VR in the context of attention is the development of NF training aimed at reducing symptoms and/or increasing the inhibition of perturbators. As a reminder, neurofeedback training is a technique where brain activity is displayed in a more understandable form (e.g. video game evolving in function of specific brain activation). Some recent works have presented NF in an immersive environment, as in Blume et al. [10] where the inattention pattern is detected with combined EEG and fNIRS in a virtual classroom. Other works have shown different methods to detect ADHD patterns based on gaze direction and head position, as described in Tan et al. [45].

D. Biomedical Signal Analysis

With applications aiming to create an interaction between the human brain and machines being more and more popular in today's world, there is an increasing amount of research in this field. To maintain clarity, we will only consider the works using EEG in brain-computer interfaces (BCI). In most cases, EEG signal analysis follows a predefined pathway separated into different steps, as shown in Lotte et al. [32]. These steps are: signal acquisition, step during which the EEG is recorded in a specific context; signal pre-processing, where the noise and artefacts are removed from the signals (this step is sometimes merged with the previous one); feature extraction, where the most relevant signal information is extracted; signal classification/regression, during which the feature vectors are used to classify the signal into categories or assign them to a specific score. In the case of our problematic, attention can be considered as a class (focused vs. not focused) or a score computed from regression (degree of attention between 0-100%).

Signals can be acquired by entertainment EEG recorders [26] or Biomedical recorders [51]. Both of these approaches have their advantages and disadvantages, with the first providing lower signal precision and robustness against noise, but with easier use with more flexibility, and the second being more precise but needing longer preparation by a trained operator. The pre-processing step is not mandatory, some applications based on deep learning skip this step, as shown in Lawhern et al. [29]. However, when pre-processing is applied it often consists of bandpass filtering to remove the continuous gains and artefacts which occur at a specific frequency band (e.g. electrical noise at 50 Hz). In other works, mean filtering is also applied to remove residual noise components. Another approach used to remove ocular artefacts is the use of independent component analysis (ICA) to remove the signal contributions corresponding to eye movements as shown in ZuCo from Hollenstein et al. [24].

Several feature extraction methods have been designed for BCI applications and these can be classified into two main categories:

- Spectral features that can be extracted from the frequential signal behaviour. The most commonly considered is the power spectral density (PSD) which represents the signal power distribution in different frequency bands. In the context of neurofeedback, the spectral information is mainly used to characterise the power of different bandwidths (e.g. the θ - wave [$\sim 5-7$ Hz] and β - wave [$\sim 15-30$ Hz] powers for ADHD neurofeedback, as presented in Ko et al. [28]). There are different ways to extract spectral features, but the most commonly used method is Fast Fourier Transform (FFT). Different neurofeedback for ADHD symptom reduction using spectral features from EEG signals have already been developed [9], [10] [8]. Bazanova et al. [8] present motivating results assuming that the neurofeedback has been well designed (e.g. avoiding issues caused by wrong frequency band definition, as shown by Bazanova and Aftanas [7]).
- Temporal features are mainly used for the evoked related potential (ERP) based application. EEG temporal variation can be quantified and later used as a feature to classify signals in different classes. Temporal features generally consist of downsampling signals samples after low pass filtering, the goal is to have the temporal representation of the signal, as described in Lotte et al. [32]. Other features can be deduced from the signal's temporal evolution with, for example, the differential entropy (DE), used in the work of Zheng and Lu [52], which is directly correlated with the signal amplitude and variance, or the Hjorth parameters representing the signal's activity, mobility and complexity as used in the works of Song et al. [44].

Other feature extraction algorithms can also be applied to EEG signals. These may express spatial information (e.g. electrodes position or activated brain regions) or features representing the signal's disorder with chaos theory-based feature extraction methods (e.g. Higuchi fractal dimension use for EEG analysis, as explained by Harne [23]).

Many ML algorithms are used to classify or regress the feature vectors from biomedical signals. From the traditional machine learning algorithms, the most commonly used are Support Vector Machine (SVM), Random Forest (RF) and Gaussian Naevian Binary Classifier (GNBC). These can be used to classify records during motor activity [42], estimate the emotional state [51] or analyse the sleeping state [18]. During the last ten years, an increasing amount of work using a specific field of Machine Learning called Deep Learning has emerged as reported by Lawhern et al. [29] and Lotte et al. [33]. The main deep learning algorithms used for BCI applications are:

- Deep Neural Networks (DNN) are the simplest deep learning model. They are constituted of a sequence of layers composed by a linear combination and an activation function(i.e. a perceptron). For this reason, they are also known as multi-layer perceptrons (MLP). Zheng and Lu [52] used this kind of network to investigate the best feature for emotion classification.
- Convolutional Neural Networks (CNN) are models orig-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3061719, IEEE Transactions on Circuits and Systems for Video Technology

inally used for image processing. In these networks, different kernels (with learnable weights) are convolved on images to extract the most significant information. In their work, Lawern et al. [29] use a CNN to classify evoked potential from EEG.

- Recurrent Neural Networks (RNN) are networks taking a sequence as input and computing a hidden output vector taking into account the relationship between the elements of the input sequence. The relationship can be spatial, e.g. two successive elements are spatial neighbours, or temporal, e.g. the elements of the input sequence are temporally correlated. Bashivan et al. [6] present a model based on the combination of CNN and RNN considering a sequence of EEG for motor imagery classification.
- Auto-Encoding Networks (AE) are specific unsupervised deep-learning architectures aiming to express an input in a lower representation subspace. These models have the same input and output, and the compressed representation of information (or latent representation) is at the centre of the network. During the training of an autoencoder, a metric aims to minimise the difference between the input and its reconstruction, i.e. output. Wen and Zhang present an AE model showing motivating results for epileptic seizure detection.
- Graph-Based Network is another approach considering an EEG signal as a graph evolving over time. The electrodes are considered as nodes, with their edges being proportional to their distance. Zhong et al. present a graph-based model beating the state-of-the-art approaches for emotion estimation from EEG.

It is important to note that this list is not exhaustive and that other approaches have also been considered, whether for feature extraction algorithms or classification/regression models.

In this context, the lack of physiological datasets assessing sustained and selective attention has been a great motivation. Moreover, these could be beneficial for future research projects within the scientific community.

III. Experimental setup

A. VR Environments

In the context of this research project, five different environments representing everyday life were developed: a bedroom, a gym-hall, an amusement park, a living room and a forest. This variety was used with the aim of reducing the parasitic effect that can be caused by stress, anxiety or negative feelings. Moreover, the choice of avoiding a classroom, as has been used in previous works [10], is justified by the fact that it could induce anxiety for certain participant subgroups. At the beginning of the recording, participants were asked to choose the environment where they felt the most comfortable. After choosing an environment, the participants were asked to complete three tasks. All the stimuli appearing during the tasks were related to the chosen environments:

- Basketball and football in the gym-hall
- Butterflies of different colours in the forest
- Cats and dogs in the bedroom
- Balloons of different colours in the living room
- Animals in the amusement park

The five environments were developed with the Unity game engine for VR headsets. This allowed us to simultaneously record the eye-tracking and EEG signals during the tasks. The framework is freely available on github[2] and is described in more detail in Delvigne et al. [16].

B. Task descriptions

To assess the participant's attention, three tasks were designed in collaboration with specialists in the field of cognitive psychology and neuropsychology. Each of these exercises had a duration of five minutes. The choice of this duration was motivated by two factors: the duration had to suit children and had to be sufficient for novel signal processing application, as reported in Roy et al. [41].

As already mentioned, the recording framework was divided into three tasks:

- A relaxation task during which we asked the participants to be as quiet as possible. The goal was to consider the mean baseline activity while resting. The aim of this task was to begin the attention tasks with a specific attentional state and to have the best participant characterisation for the resting state. At random times during the relaxation task three different perturbators related to the environment appear (one visual, one auditory and one combined).
- A selective attention task. During this task, the participant was asked to look at a specific stimulus (considered as the right one) and to avoid looking at the other (considered as the wrong). The stimuli corresponding to the environments are listed in the previous subsection. The difference between the wrong and the right stimulus can be based on its nature, i.e. different objects, its texture or its colour. The interstimulus interval (ISI)[3] corresponding to the appearance time between a couple of stimuli (right and wrong) followed a normal distribution of 3000 ms and a standard deviation of 250 ms, i.e. ISI $\in \mathcal{N}(3000, 250)$. A representation of stimuli appearance is shown in Fig 2 with the basketball and football corresponding to right and wrong stimuli respectively.
  During the second task, a score representing the selective attention state for each trial was computed, i.e. appearance of stimuli couple. This score was computed with the following equation:

$$\begin{aligned} t_{task2}(k) &= t_{target}(k) - t_{wrong}(k) \\ score_{task2}(k) &= a * t_{task2}(k) + b \end{aligned} \tag{1}$$

  with $t_{target}(k)$ (resp. $t_{wrong}(k)$) being the amount of time during which the target (resp. the wrong stimulus) is looked for during the trial k. a and b are parameters that were experimentally computed to have the mean difference time corresponding to 50% and the lowest (resp. highest) time corresponding to 20% (resp. 80%).

[2]https://github.com/VDelv/VERA

[3]For both attention tasks, the ISI was chosen from the attentional assessment literature review.

Fig. 2. Overview of the selective attention task representation in the VR environment. In the figure the right stimuli (i.e. basketball), the wrong stimuli (i.e. football) and the sight (not visible for the participant) on the fixation point at the centre of the screen are represented.

- A sustained attention task. During this task, only the right stimulus appears. The goal was to direct the sight in the direction of the stimulus. The ISI for the third task also followed a normal distribution with a mean of 3000 ms and standard deviation of 500 ms, i.e. ISI $\in \mathcal{N}(3000, 500)$.
  From the recordings made during the third task a score representing the sustained attention for each trial was computed using the following formula:

$$\begin{aligned} t_{task3}(k) &= t_{elapsed}(k) \\ score_{task3}(k) &= a * t_{task3}(k)^2 + b * t_{task3}(k) + c \end{aligned} \quad (2)$$

  with $t_{elapsed}(k)$ the time elapsed by the participant to move their gaze towards the appearing stimulus at trial k. a, b and c are experimentally chosen parameters that match the mean elapsed time with a score equal to 50%, the lowest (resp. highest) time to a score of 80% (resp. 20%). During this task, a high attention state corresponds to a short duration, contrary to the previous task where the goal was to have the longest time as possible.

With this methodology, the attention level in percent has been defined according to the mean reaction time for both tasks. For instance, a score of 50% corresponds to the mean reaction time for the selective and/or sustained attention time for all the participants. In this study, we have considered a participant as focused (resp. distracted) if his results to the attention tasks were better (resp. worse) than mean attention results.

At the end of the three tasks, the score for each trial of the second and third tasks was computed and registered in a corresponding file. It was then possible to study the score evolution to identify any patterns corresponding to attention loss or gain.

C. EEG Recordings

EEG was recorded at a sampling rate of 500 Hz using 32 electrode actiCHamp recorder from Brain Vision. The electrode placements followed the 10/20 electrodes disposition, as presented in Oostenveld and Praamstra [37]. However, due to the space required by the VR headset at the back of the head, three electrode positions were modified: P3 $\rightarrow$ AF3; Pz $\rightarrow$ FCz; P4 $\rightarrow$ AF4. Along with the EEG records, physiological measurements were also taken with the VR headset: head position, eye position, pupil diameter and blinking time.

Thirty-two healthy participants (44% of female), aged between 19 and 30 took part in this study. Before the start of the experiment, a short explanation of the research project and experiment was given and the participants were asked to sign a consent form. After the information, the electrodes were placed, an impedance check was made to verify that it was lower than 10 kOhm. After the signal check, in the VR headset eye-tracker was calibrated by the experimenter. Then the physiological recording start and the experimenter left the participant's field of view.

As stated earlier the experiment was separated into three different steps:

- A 5-minute relaxation task to record the baseline.
- A 5-minute selective attention task.
- A 5-minute sustained attention task.

The appearance of stimuli and perturbators was directly annotated on the EEG recordings with annotation tools provided by the Brain Recorder software. At the end of the experiment, the participant was debriefed to identify any possible issues and/or improvements. Three files were recorded during the acquisition: the EEG sampled at 500 Hz with annotations, the physiological records at 5 Hz and the evolution of the score for the second and third exercises.

IV. Methodology

After recording the physiological signals, one of the framework goals was to analyse the signal to check if it would be possible to estimate the attentional state from the physiological recordings. The pathway followed to estimate attention is similar to that proposed by Lotte et al. [32] as described in the previous section: signal pre-processing, feature Extraction and attention estimation with a classification algorithm.

A. Signal Preprocessing

During the pre-processing step, a bandpass filtering between 0.1 and 50 Hz was applied to the signal to remove the continuous components (< 0.1 Hz) and artifacts that may be caused by electrical noise or muscles (> 50 Hz). The mean filtering was also applied to the EEG by subtracting the mean of the signal for each electrode. Finally, the signals were cut into segments of 4 seconds, 1 second before the stimuli appearance and 3 seconds after. The motivation to consider only bandpass filtering between 0 and 50 Hz to remove noise contribution are dual: 1) As shown by Roy et al. [41], the most considered denoising step in ML-related EEG research is based on frequential filtering of EEG without considering a more complicated method for artificacts handling in most of the cases. Moreover, this denoising method does not affect frequential based features (which are the most commonly considered), 2) It was decided to let the choice for further works to consider other denoising methods for instance based on ICA for ocular artifacts. In parallel, the physiological recordings including pupil diameter and head movements were also segmented around the stimuli appearance. From the head linear and rotational position, the mean acceleration was computed in each direction to consider only the parameter variation. The variation in the pupil diameter was also computed.

Moreover, an array of corresponding scores and participant IDs for each trial were created. At the end of the pre-processing step the following files were stored:

- Preprocessed EEG of dimension $[n_{trials} \times duration * f_{sampling} \times n_{channels}]$
- Physiological Signals $[n_{trials} \times n_{physiofeatures}]$
- Participant $[n_{trials} \times 1]$
- Score $[n_{trials} \times 1]$

With n_{trials} being the sum of all the trials completed during the whole experimental study, $n_{channels}$ being the number of channels of the EEG recorder, $duration$ being the duration of a considered trial in the dataset (i.e. 4 seconds), $f_{sampling}$ being the sampling frequency $= 500$Hz and $n_{physiofeatures}$ being the length of the feature vector for physiological signals ($n_{physiofeatures} = 7$).

B. Feature Extraction Methods

a) Frequential feature: One of the most used features for BCI application is spectral information. In their work, Song et al. [44] considered the use of frequential features computed with different algorithms: Fast Fourier transform (FFT) and Hilbert Huang Spectrum (HHS). By computing the power spectral density (PSD), it is possible to split EEG power into specific frequency bands each of these corresponding to a specific activity: $\delta, \theta, \alpha, \beta$ and γ band. For this project, we considered the θ, α and β bands. This choice is justified by the fact that the delta band is mainly enhanced in the sleeping state as reported in the works of Collura and Siever [13] and that gamma bands are characterised by stress state (which we tried to avoid in our experiments) and are more affected by noise due to their high frequency.

In the context of attention estimation, we considered frequential features for the signal processing, however, Arns et al. [4] show that these features are not representative in all cases. It is for this reason that other signal information may also be considered.

b) Temporal feature: Temporal features are extracted from time series. They represent the time evolution of the signal (e.g. maximas, minimas, signal's energy, variation, etc). Vidaurre et al. [48] and Song et al. [44] proposed a method aiming to compute the Hjorth parameters representing activity, complexity and mobility of a signal as expressed in the following equations:

$$\begin{aligned} Activity &= var(x(t)) \\ Mobility &= \sqrt{\frac{var(x'(t)}{var(x(t))}} \\ Complexity &= \frac{Mobility(x'(t)}{Mobility(x(t))} \end{aligned} \tag{3}$$

The motivation for the use of these parameters was to study a possible correlation between the evoked potential waveform and the attentional state.

Modality	Extracted Features	Dimensions
EEG	Power Spectral Density (FFT)	$[\mathrm{n}_{trials} \times 3 \times n_{channels}]$
	Hjorth Parameters	$[\mathrm{n}_{trials} \times 3 \times n_{channels}]$
	Latent-space representation (AE)	$[\mathrm{n}_{trials} \times 16]$
	Image representation	$[\mathrm{n}_{trials} \times 3 \times 32 \times 32]$
Eye-tracking	Variation pupils diameter [mm]	$[\mathrm{n}_{trials} \times 1]$
Physiological	Head linear and rotationnal acceleration	$[\mathrm{n}_{trials} \times 6]$

TABLE I
DATASET FILE ORGANISATION AND CORRESPONDING DIMENSIONS FOR EACH FEATURE EXTRACTION METHODS.

c) Image-based representation: From the arrays constituted of temporal and spectral features, it is possible to consider a more *"visual"* representation of the signal as presented in the work of Bashivan et al. [6]. In their approach, they first considered the position of the electrodes in 3D. After the representation of information in 3D was a project in a corresponding 2D frame with geodesic projection. A value for each couple of points (x, y) was represented in the projected 2D frame, however, the information being discrete not all the values were covered between the x and y maximas. To construct the final image, a bicubic interpolation was given for each channel of each trial. The entire method is also described in [6]. This transformation has been applied for the temporal, frequential and preprocessed EEG. For the third application, we considered the EEG down-sampled after low-pass filtering. The ratio between the initial sampling frequency and the down-sampled frequency was equal to five. The image generation step is also represented in the left of Fig 5.

d) Final file list: After the feature extraction step, different arrays were created and stored in the dataset:

- Temporal Array $[n_{trials} \times n_{temp-feat} \times n_{channels}]$: array with temporal features for each trial. The considered temporal information were the Hjorth parameters: activity, complexity and mobility.
- Frequential Array $[n_{trials} \times n_{freq-feat} \times n_{channels}]$: array with spectral features for each trials. The considered features were the power spectral density for three frequency bands: alpha, beta and theta.
- Temporal Images $[n_{trials} \times n_{temp-feat} \times 32 \times 32]$: array with interpolated information from the temporal array to image representation.
- Frequential Images $[n_{trials} \times n_{freq-feat} \times 32 \times 32]$:s array with interpolated information from the frequential array to image representation.
- EEG Images $[n_{trials} \times n_{down} \times 32 \times 32]$, array with interpolated information from downsampled EEG array to images representation.

With $n_{freq-feat}$ being the dimension of the spectral feature vector, $n_{temp-feat}$ the dimension of the temporal feature vector and n_{down} =duration $*\frac{f_{sampling}}{ratio}$, the length of the downsampled signal ($ratio = 5$). A summary of all the signals, the corresponding extracted features and their dimension is listed in Table I.

C. Machine Learning Models for attention estimation

ML algorithms have seen an increase in their use over the last 10 years [40]. In the context of this project, we considered models aiming to classify the attentional state from neurophysiological signals.

We first considered traditional ML approaches with support vector machine (SVM) and random forest (RF) to constitute a benchmark baseline. The goal of the designed models is to estimate if the participant was or was not focused during a certain trial. For this purpose, features from EEG and/or physiological information were considered as input and the attention state as output depending on the task result. For the deep learning (DL) approaches, four models were considered:

a) Hierarchical Recurrent Neural Network: In this approach, we used a hierarchical recurrent neural network (RNN) to exploit the spatial relationship between electrodes. This approach was inspired by the works of Yong Du et al. [15], where they considered different layers of RNN to predict action from skeleton information. In this context, we used a similar approach where different layers of RNN were used to consider each brain region. Similar models have already been applied to biomedical signals, for instance, in the research project of Li et al. [30] where two RNN were used to exploit the vertical and horizontal relationship between electrodes to estimate emotion. For this purpose, different layers of RNN extracting information at different levels were considered.

From an EEG sample $X \in \mathbb{R}^{l \times c}$ with l the length of the signals (corresponding to the number of samples) and c the number of electrodes considered, we extracted a feature array $X^{feat} \in \mathbb{R}^{n \times c}$ with n the number of features depending on the feature extraction method. It is important to note that with some feature extraction methods the array may have different dimensions.

For each electrode region, we separated the two hemispheres. As such, we finally had $x_i^{l,feat}$ and $x_i^{r,feat}$ for each brain region, representing the feature for one of the four considered regions for the left and right hemispheres. This feature array was then passed through three stage of RNN. The hidden state from the different RNN stages can be formulated as

$$h_i = H(W_x \times x_i + W_h \times h_{i-1} + b_i) \tag{4}$$

with H being a non-linear function , h_i the hidden sequence at instant i, W the weight matrices and b a bias vector.

After computing the hidden state for each selected electrode, these were paired with a pairwise operation (corresponding to subtraction) and passed through the second stage of RNN with X_i^{hidden} as input, as shown in Fig 3.

Then the information was passed trough a linear function to merge the contribution in each region to form X_i^{last}, which was passed through the last RNN stage. Finally, from the last hidden representation, an attentional state was estimated with a fully-connected layer.

b) Graph-based approach: Graph neural networks are a type of DL model considering the input as graph evolving in function of a modality. In this context, EEG is considered as a graph with electrodes corresponding to nodes and the edges being proportional to their distances. Each node can take a value equal to the considered feature (e.g. temporal or frequential). This approach has already been considered in BCI applications, as in the work of Zhong et al. [54] where EEG was used to estimate emotional state.

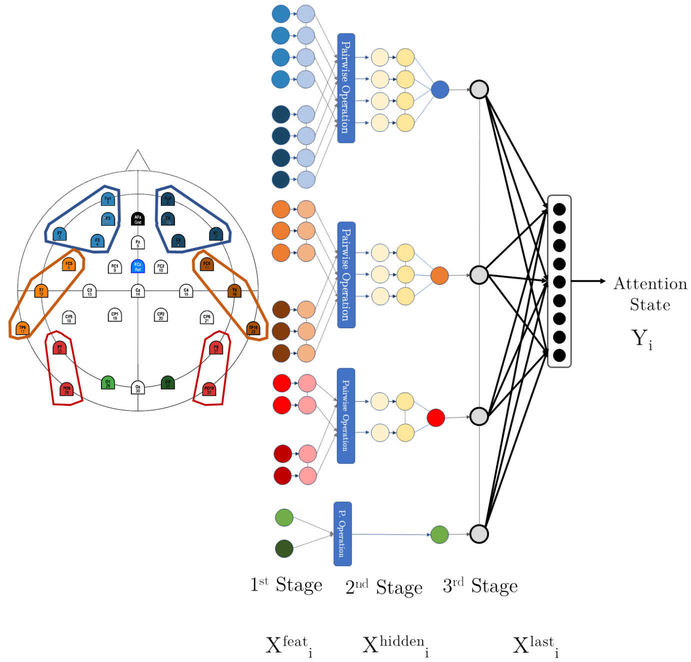

Fig. 3. Hierarchical RNN consisting of three stages of recurrent neural network, the first stage capturing information in the brain region, the second taking into account the relationship between the two hemispheres and the third considering the relationship between the regions. After the RNN stages, a fully-connected network was used to estimate the attention state.

In our approach, represented in Fig 4, we consider a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ composed of vertices $\mathcal{V}$, corresponding to the electrodes, and edges $\mathcal{E}$, corresponding to the electrodes distances. The edge information can be expressed as an adjacency matrix $A \in \mathbb{R}^{c \times c}$ with c corresponding to the number of EEG channels. In our case, the element of the adjacency matrix was proportional to the electrode distance, i.e. $A_{ij} \propto d(\text{electrode}_i, \text{electrode}_j)$. In this context, we considered a feature array $X \in \mathbb{R}^{n_{feat} \times c}$ with n_{feat} the number of features, this features array being convolved and pooled in a basic graph neural network (GNN). The convolution in the layer l being expressed

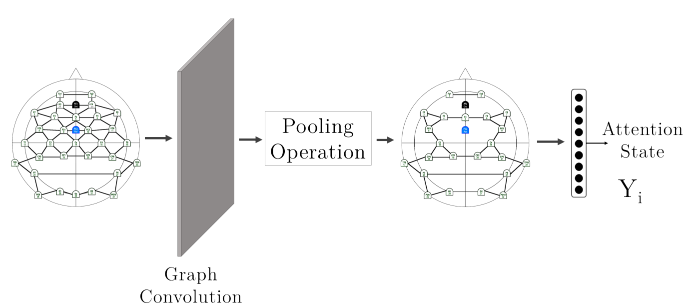

Fig. 4. Graph-based model for attention estimation from EEG. The figure represents the model architecture with the initial graph, the graph convolution, the pooling operation and the fully-connected layer to estimate attention.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3061719, IEEE Transactions on Circuits and Systems for Video Technology

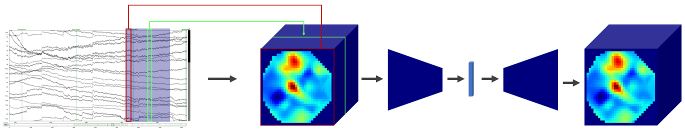

Fig. 5. Autoencoder-based framework. On the left, the image-based EEG representation, and on the right, the autoencoding network aiming to represent EEG in the latent space.

as

$$X^{l+1} = \sigma(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}X^lW^l) \quad (5)$$

with σ being a non-linear function, W^l a weight matrix for the layer l and D the diagonal degree matrix of A, as reported in Zhang et al. [49]. After completing the convolution, it is possible to perform a pooling operation by keeping only the most relevant nodes, i.e. those not being able to be represented by their neighbours. For this purpose, we computed the distance between the actual nodes and their reconstruction from from neighbours, also as described in [49] with

$$p = ||(I - D^{-1}A)X||_1 \quad (6)$$

with $||...||_1$ the ℓ_1 norm. Moreover, a node is considered as relevant if the distance between its reconstruction from the neighbours and its actual coordinates is above a certain threshold. We considered a GNN composed of two convolutions and one graph-pooling [49]. After this step, the computed array is flattened and passed through a fully-connected network to estimate the attentional state.

c) Convolutional Neural Network: Convolutional neural networks (CNN) are networks using updatable kernels aiming to extract features through convolutions. These features are used to make estimations through regression or classification layers. As explained earlier, CNNs have already been used in the context of BCI applications. Lawhern et al. [29] considered a CNN convolving on EEG signals to detect specific evoked potential (P300). Furthermore, Bashivan et al. [6] used a CNN considering image representation of EEG as input to predict motor activation. In our work, we consider a similar approach with images based on temporal and frequential features from EEG passing through a VGG-based network, as presented by Simonyan and Zisserman [43].

More specifically, we consider a model composed of three convolution layers, each of them with a kernel size = 3, a ReLU function as the activation function and a max pool layer between each layer. Each layer is composed of $4, 2$ and 1 convolution stacks respectively, and the number of channels at the end of each layer was: $16, 64$ and128 respectively. The figures were chosen after experimentation with different CNN architectures and after considering a preliminary study for attention regression from image representation of EEG, as presented in Delvigne et al. [17].

d) Autoencoder: This method aims to project the signal in a subspace with less dimensions than the input. An AE reproduces the input as output, and during the training the kernel weights are updated to minimise the mean error rate for reconstruction, i.e. mean squared error, between the input and output. After the training, it is possible to represent the signal in the smaller subspace, the latent-space. The latent representation extracts the high-level information from the EEG signal. The latent space representation is then used to train a ML classifier, e.g. MLP, SVM and RF. In Fig 5, we represent the autoencoder used in our framework.

ML Models / Feature	**Accuracy (μ/σ)[%]**
RF/Frequential	51.12 ± 3.83
RF/Temporal	60.87 ± 3.51
SVM/Frequential	61.16 ± 6.97
SVM/Temporal	61.34 ± 6.48
RNN/Frequential	66.12 ± 7.34
RNN/Temporal	65.42 ± 6.42
Graph/Frequential	70.57 ± 6.51
Graph/Temporal	$\mathbf{72.41 \pm 5.51}$
CNN/Frequential	65.24 ± 5.34
CNN/Temporal	64.93 ± 5.41
RF/AE Feature	63.8 ± 4.43
SVM/AE Feature	63.14 ± 5.81

TABLE II
MODEL PERFORMANCE FOR PARTICIPANT INDEPENDENT CROSS-VALIDATION FOR DIFFERENT FEATURE EXTRACTION METHODS.

V. RESULTS

In this section, we present the framework's ability to estimate attention from EEG signals with ML algorithms. Different comparisons will also be made in function of the considered input and/or the considered models. EEG has been used to estimate a binary attention state, with the latter being computed in function of the attentional score for the two attentional tasks, as explained earlier. In the first time, the two tasks having the same aim: assessing attention, it was decided to consider them merged. In a second time, it was decided to consider them separately to inspect the task dependencies on attention estimation. If the participant's score is above the subject independent mean score we consider that they are focused. In any other case, we considered that the participant was not focused.

A. Subject-Independent Classification

The considered protocol to evaluate the accuracy of the models is the leave-one-subject-out (LOSO) cross-validation. With this protocol, we consider all the N participants except one to train the ML model, i.e. length$(X_{train}) = N - 1$, and we use the remaining subject for the validation. This step is repeated for each participant and the mean and standard deviations are computed. For clarity, the LOSO cross-validation is also called patient independent cross-validation in this paper.

To assess the improvement brought by deep-learning approaches, the estimation was first done with traditional ML algorithms: Support Vector Machine (SVM) and Random Forest (RF). The estimation accuracy of these models was considered as a baseline for attention state estimation.

As shown in Table II, all the models based on deep-learning approaches present similar performances, which are higher for all of them compared to the basic approach with SVM and RF on the feature array. Moreover, better performances are noted for the models based on graph theory, the two approaches

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3061719, IEEE Transactions on Circuits and Systems for Video Technology

ML Models / Feature	**Accuracy** (μ/σ)[%]	
	Task 2	Task 3
RF/Frequential	66.26 ± 7.44	74.19 ± 5.34
RF/Temporal	69.45 ± 6.1	77.85 ± 4.81
SVM/Frequential	67.98 ± 6.9	78.88 ± 8.18
SVM/Temporal	68.26 ± 7.58	78.73 ± 8.59
RNN/Frequential	73.66 ± 19.08	82.82 ± 15.88
RNN/Temporal	77.55 ± 13.86	85.8 ± 11.98
Graph/Frequential	71.53 ± 20.59	83.03 ± 15.51
Graph/Temporal	72.57 ± 19.81	83.66 ± 15.45
CNN/Frequential	76.66 ± 15.35	84.14 ± 15.02
CNN/Temporal	76.79 ± 15.86	85.18 ± 13.34
Combined model/Frequential	73.49 ± 18.65	83.08 ± 15.68
Combined model/Temporal	75.38 ± 16.13	84.82 ± 15.88
RF/AE Feature	71.56 ± 8.06	83.51 ± 5.11
SVM/AE Feature	72.67 ± 9.45	84.09 ± 8.67

TABLE III
MODELS PERFORMANCES FOR PARTICIPANT INDEPENDENT AND TASK DEPENDANT CROSS-VALIDATION FOR DIFFERENT FEATURE EXTRACTION METHODS.

based on the graphs with the two best results compared to other models.

For the comparison of the considered features, each seems to present similar results for all the considered ML models sometimes with better results for the temporal features, as noted in the Table II for RF and graph models.

B. Performances of task dependant Models

To investigate the possible causes of misclassification, we considered task-dependent training for the attention estimation. In this scenario, the models were trained and validated by considering only the attention score from one of the two attentional tasks (i.e. sustained and selective attention). The LOSO cross-validation accuracies for attention estimation for the ML models are listed in Table III. As the table shows, the best model for estimating selective attention (resp. task 2) and sustained attention (i.e. task 3) is the RNN using temporal features. The score for task-dependant training is slightly higher compared with the task-independent cross-validation. However, the standard deviation was also increased, which can be explained by the fact that the amount of data used for the training was halved. Moreover, it was decided to consider a novel approach merging the different feature representation methods (hierarchical-RNN, images in CNN and graph) to notice the possible improvements. The fusion corresponding to a concatenation of the last layer of each network before the final attention state estimation. As shown in Table III, the implementation present similar results than the previously presented approaches with higher accuracy compared to traditional machine-learning approaches. The representation combination increases the accuracy compared to the graph approach.

As noted, the estimation score is higher when the attentional tasks are considered separately. This can be explained by the fact that the brain-mechanism behind sustained and selective attention is different, and therefore attention is not expressed in the same way in both cases.

C. Hardware and Software

The framework recording part was developed in C# on Unity and acquisition was carried out with an HTC-Vive Pro Eye (with Tobii eye-tracker integration) VR headset coupled with the BrainVision actiCHamp Plus biomedical EEG recorder. For the attention estimation part, the parameters chosen for the different network were: hidden dimensions of $n_{hidden} = n_{last} = 64$ for the hierarchical RNN, $n_{hidden} = 16$ for the hidden dimension of the Graph-based model and $n_{latent} = 16$ for the latent-space representation. The four different models were implemented in python 3.8 with Pytorch 1.5 on an Nvidia RTX Titan GPU card.

VI. CONCLUSION

In this paper, we present a framework for recording and assessing attention with physiological records: EEG, eye-tracker and head movement. This framework, and the related dataset, could help to have a better understanding of the paradigms between attention and brain mechanisms. Moreover, we also propose different feature extraction methods and ML models presenting promoising results for attention estimation, which constitutes an interesting benchmark for further works.

In the future works, acquisition with participants from other groups will be made (e.g. ADHD children). It may also be interesting to consider other feature extraction methods and ML algorithms to investigate possible improvements. Another investigation may be to consider novel modality fusion. Moreover, considering a study based on the effect of each electrode on our estimation may be interesting for future projects with a lighter EEG recording headset.

ACKNOWLEDGMENTS

The authors would like to thank all the participants to this study and the Cognitive Psychology and Neuropsychology research group of the Faculty of Psychology and Education at the University of Mons for their help in the definition of the attentional tasks. This research was carried out in collaboration with the CiPsE research centre[4].

REFERENCES

[1] A. E. Alchalabi, S. Shirmohammadi, A. N. Eddin, and M. Elsharnouby. FOCUS: Detecting ADHD Patients by an EEG-Based Serious Game. *IEEE Transactions on Instrumentation and Measurement*, 67(7), July 2018.

[2] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nature Biotechnology*, 33(8), August 2015.

[3] V. Andrearczyk, V. Oreiller, M. Vallières, J. Castelli, H. Elhalawani, M. Jreige, S. Boughdad, J. O. Prior, and A. Depeursinge. 3D Head and Neck Tumor Segmentation in PET/CT. Zenodo, March 2020.

[4] M. Arns, C. K. Conners, and H. C. Kraemer. A decade of EEG Theta/Beta Ratio Research in ADHD: a meta-analysis. *Journal of Attention Disorders*, 17(5), July 2013.

[5] A. Bashiri, M. Ghazisaeedi, and L. Shahmoradi. The opportunities of virtual reality in the rehabilitation of children with attention deficit hyperactivity disorder: a literature review. *Korean Journal of Pediatrics*, 60(11), November 2017.

[6] P. Bashivan, I. Rish, M. Yeasin, and N. Codella. Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks. *arXiv:1511.06448 [cs]*, November 2015. arXiv: 1511.06448.

[4] https://web.umons.ac.be/sante/fr/centres-de-recherche-associes/cipse/

[7] O. M. Bazanova and L. I. Aftanas. Individual EEG Alpha Activity Analysis for Enhancement Neurofeedback Efficiency: Two Case Studies. *Journal of Neurotherapy*, 14(3), August 2010.

[8] O. M. Bazanova, T. Auer, and E. A. Sapina. On the Efficiency of Individualized Theta/Beta Ratio Neurofeedback Combined with Forehead EMG Training in ADHD Children. *Frontiers in Human Neuroscience*, 12, 2018.

[9] S. Bioulac, D. Purper-Ouakil, T. Ros, H. Blasco-Fontecilla, M. Prats, L. Mayaud, and D. Brandeis. Personalized at-home neurofeedback compared with long-acting methylphenidate in an european non-inferiority randomized trial in children with ADHD. *BMC Psychiatry*, 19(1), August 2019.

[10] F. Blume, J. Hudak, T. Dresler, A.-C. Ehlis, J. Kühnhausen, T. J. Renner, and C. Gawrilow. NIRS-based neurofeedback training in a virtual reality classroom for children with attention-deficit/hyperactivity disorder: study protocol for a randomized controlled trial. *Trials*, 18(1), January 2017.

[11] Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin. Multi-channel EEG recordings during a sustained-attention driving task. *Scientific Data*, 6(1), April 2019.

[12] S. Coelli, R. Sclocco, R. Barbieri, G. Reni, C. Zucca, and A. M. Bianchi. EEG-based index for engagement level monitoring during sustained attention. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, August 2015.

[13] T. F. Collura and D. Siever. Chapter 8 - Audio-visual entrainment in relation to mental health and EEG. In Thomas H. Budzynski, Helen Kogan Budzynski, James R. Evans, and Andrew Abarbanel, editors, *Introduction to Quantitative EEG and Neurofeedback (Second Edition)*. Academic Press, San Diego, January 2009.

[14] M. Cueli, C. Rodríguez, P. Cabaleiro, T. García, and P. González-Castro. Differential Efficacy of Neurofeedback in Children with ADHD Presentations. *Journal of Clinical Medicine*, 8(2), February 2019.

[15] Yong D., W. Wang, and L. Wang. Hierarchical recurrent neural network for skeleton based action recognition. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[16] V. Delvigne, L. Ris, T. Dutoit, H. Wannous, and J-P. Vandeborre. VERA: Virtual Environments Recording Attention. In *2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH)*, August 2020.

[17] V. Delvigne, H. Wannous, J-P. Vandeborre, L. Ris, and T. Dutoit. Attention estimation in virtual reality with eegbased image regression. In *2020 IEEE 3rd International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 2020.

[18] S. Devuyst, T. Dutoit, P. Stenuit, and M. Kerkhofs. Automatic sleep spindles detection — Overview and development of a standard proposal assessment method. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, August 2011.

[19] H. Eom, K. Kim, S. Lee, Y.-J. Hong, J. Heo, J.-J. Kim, and E. Kim. Development of Virtual Reality Continuous Performance Test Utilizing Social Cues for Children and Adolescents with Attention-Deficit/Hyperactivity Disorder. *Cyberpsychology, Behavior, and Social Networking*, 22(3), January 2019.

[20] Z. Gao, X. Wang, Y. Yang, C. Mu, Q. Cai, W. Dang, and S. Zuo. EEG-Based Spatio–Temporal Convolutional Neural Network for Driver Fatigue Evaluation. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9), September 2019.

[21] A. García-Baos, T. D'Amelio, I. Oliveira, P. Collins, C. Echevarria, L. P. Zapata, E. Liddle, and H. Supèr. Novel Interactive Eye-Tracking Game for Training Attention in Children With Attention-Deficit/Hyperactivity Disorder. *The primary care companion for CNS disorders*, 21(4), July 2019.

[22] E. Gawehn, J. A. Hiss, and G. Schneider. Deep Learning in Drug Discovery. *Molecular Informatics*, 35(1), 2016.

[23] B. P. Harne. Higuchi Fractal Dimension Analysis of EEG Signal before and after OM Chanting to Observe Overall Effect on Brain. *undefined*, 2014.

[24] N. Hollenstein, J. Rotsztejn, M. Troendle, A. Pedroni, C. Zhang, and N. Langer. ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading. *Scientific Data*, 5(1), December 2018.

[25] H. H. S. Ip, S. W. L. Wong, D. F. Y. Chan, J. Byrne, C. Li, V. S. N. Yuan, K. S. Y. Lau, and J. Y. W. Wong. Virtual Reality Enabled Training for Social Adaptation in Inclusive Education Settings for School-Aged Children with Autism Spectrum Disorder (ASD). In Simon K.S. Cheung, Lam-for Kwok, Junjie Shang, Aihua Wang, and Reggie Kwan, editors, *Blended Learning: Aligning Theory with Practices*, Lecture Notes in Computer Science, Cham, 2016. Springer International Publishing.

[26] S. Katsigiannis and N. Ramzan. DREAMER: A Database for Emotion Recognition Through EEG and ECG Signals From Wireless Low-cost Off-the-Shelf Devices. *IEEE Journal of Biomedical and Health Informatics*, 22(1), January 2018.

[27] W. Klimesch, M. Doppelmayr, H. Russegger, T. Pachinger, and J. Schwaiger. Induced alpha band power changes in the human EEG and attention. *Neuroscience Letters*, 244(2), March 1998.

[28] L.-W. Ko, O. Komarov, W. D. Hairston, T.-P. Jung, and C.-T. Lin. Sustained Attention in Real Classroom Settings: An EEG Study. *Frontiers in Human Neuroscience*, 11, 2017.

[29] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, Ch. P. Hung, and B. J. Lance. EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces. *Journal of Neural Engineering*, 15(5), October 2018. arXiv: 1611.08024.

[30] Y. Li, L. Wang, W. Zheng, Y. Zong, L. Qi, Z. Cui, T. Zhang, and T. Song. A Novel Bi-hemispheric Discrepancy Model for EEG Emotion Recognition. *IEEE Transactions on Cognitive and Developmental Systems*, 2020.

[31] M. D. Liechti, L. Valko, U. C. Müller, M. Döhnert, R. Drechsler, H.-C. Steinhausen, and D. Brandeis. Diagnostic Value of Resting Electroencephalogram in Attention-Deficit/Hyperactivity Disorder Across the Lifespan. *Brain Topography*, 26(1), January 2013.

[32] F. Lotte. A Tutorial on EEG Signal-processing Techniques for Mental-state Recognition in Brain–Computer Interfaces. In Eduardo Reck Miranda and Julien Castet, editors, *Guide to Brain-Computer Music Interfacing*. Springer London, London, 2014.

[33] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger. A review of classification algorithms for EEG-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, 15(3), April 2018.

[34] J. C. Mullane and P. V. Corkum. The Relationship Between Working Memory, Inhibition, and Performance on the Wisconsin Card Sorting Test in Children With and Without ADHD. *Journal of Psychoeducational Assessment*, 25(3), September 2007.

[35] A. Mühlberger, K. Jekel, T. Probst, M. Schecklmann, A. Conzelmann, M. Andreatta, A. A. Rizzo, P. Pauli, and M. Romanos. The Influence of Methylphenidate on Hyperactivity and Attention Deficits in Children With ADHD: A Virtual Classroom Test. *Journal of Attention Disorders*, 0(0).

[36] Anton Nijholt. BCI for Games: A 'State of the Art' Survey. In S. M. Stevens and S. J. Saldamarco, editors, *Entertainment Computing - ICEC 2008*, Lecture Notes in Computer Science, Berlin, Heidelberg, 2009. Springer.

[37] R. Oostenveld and P. Praamstra. The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, 112(4), April 2001.

[38] R. Ortner, B. Z. Allison, G. Korisek, H. Gaggl, and G. Pfurtscheller. An SSVEP BCI to Control a Hand Orthosis for Persons With Tetraplegia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 19, February 2011.

[39] Y. Pollak, H. B. Shomaly, P. L. Weiss, A. A. Rizzo, and V. Gross-Tsur. Methylphenidate effect in children with ADHD can be measured by an ecologically valid continuous performance test embedded in virtual reality. *CNS spectrums*, 15(2), February 2010.

[40] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang. Deep Learning for Health Informatics. *IEEE journal of biomedical and health informatics*, 21(1), 2017.

[41] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering*, 16(5), August 2019.

[42] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller. Characterization of four-class motor imagery EEG data for the BCI-competition 2005. *Journal of Neural Engineering*, 2(4), August 2005.

[43] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs]*, April 2015. arXiv: 1409.1556.

[44] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui. MPED: A Multi-Modal Physiological Emotion Database for Discrete Emotion Recognition. *IEEE Access*, 7, 2019.

[45] Y. Tan, D. Zhu, H. Gao, T. Lin, H. Wu, S. Yeh, and T. Hsu. Virtual Classroom: An ADHD Assessment and Diagnosis System Based on Virtual Reality. In *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, May 2019.

[46] P. Varela Casal, F. Lorena Esposito, I. Morata Martínez, A. Capdevila, M. Solé Puig, N. de la Osa, L. Ezpeleta, A. Perera i Lluna, S. V. Faraone, J. A. Ramos-Quiroga, H. Supèr, and J. Cañete. Clinical Validation of Eye

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2021.3061719, IEEE Transactions on Circuits and Systems for Video Technology

Vergence as an Objective Marker for Diagnosis of ADHD in Children. *Journal of Attention Disorders*, 23(6), April 2019.

[47] G. Vecchiato, L. Astolfi, F. De Vico Fallani, J. Toppi, F. Aloise, F. Bez, D. Wei, W. Kong, J. Dai, F. Cincotti, D. Mattia, and F. Babiloni. On the use of EEG or MEG brain imaging tools in neuromarketing research. *Computational Intelligence and Neuroscience*, 2011, 2011.

[48] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl. Time Domain Parameters as a feature for EEG-based Brain–Computer Interfaces. *Neural Networks*, 22(9), November 2009.

[49] Z. Zhang, J. Bu, M. Ester, J. Zhang, C. Yao, Z. Yu, and C. Wang. Hierarchical Graph Pooling with Structure Learning. *arXiv:1911.05954 [cs, stat]*, December 2019. arXiv: 1911.05954.

[50] W.-L. Zheng, K. Gao, G. Li, W. Liu, C. Liu, J.-Q. Liu, G. Wang, and B.-L. Lu. Vigilance Estimation Using a Wearable EOG Device in Real Driving Environment. *IEEE Transactions on Intelligent Transportation Systems*, 21(1), January 2020.

[51] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki. EmotionMeter: A Multimodal Framework for Recognizing Human Emotions. *IEEE Transactions on Cybernetics*, 49(3), March 2019.

[52] W.-L. Zheng and B.-L. Lu. Investigating Critical Frequency Bands and Channels for EEG-Based Emotion Recognition with Deep Neural Networks. *IEEE Transactions on Autonomous Mental Development*, 7(3), September 2015.

[53] W.-L. Zheng and B.-L. Lu. A multimodal approach to estimating vigilance using EEG and forehead EOG. *Journal of Neural Engineering*, 14(2), February 2017.

[54] P. Zhong, D. Wang, and C. Miao. EEG-Based Emotion Recognition Using Regularized Graph Neural Networks. *IEEE Transactions on Affective Computing*, 2020.

Victor Delvigne received the BSc and MSc degree in electral engineering from Engineering Faculty of Mons (UMons), Belgium in 2017 and 2019. Currently, he is a joint-PhD candidate at Numediart Institute, University of Mons, Mons, Belgium and CRIStAL, Villeneuve d'Ascq, France. His research interest include the areas of brain-computer interfaces, signal processing and computer vision. He is student member of the IEEE.

Hazem Wannous is a Professor of Computer Science with the IMT (Lille, France). He is also a member of the Centre de Recherche en Informatique, Signal et Automatique de Lille (CRIStAL - UMR CNRS 9189). From 2010 to June 2020, he was an Associate Professor of Computer Science at the University of Lille, France. He received his Ph.D. degree in Computer Sciences in 2008 from University of Orléans, France in 2008. His current research interests are mainly focused on computer vision, machine learning, virtual and augmented reality and human-computer-interaction. He serves as a regular reviewer for a number of top journals in the field (IEEE Trans. on PAMI, IEEE Trans. on Image Processing, IEEE Trans. on Medical Imaging, IEEE Trans. on Multimedia). He co-organized and was the general Chair of the International Workshop on Understanding Human Activities through 3D Sensors: UHA3DS 2015 in connection with the IEEE FG 2015 and UHA3DS 2016 hosted by ICPR 2016. He is a PC member for a number of international conferences and workshops and he served also as Area Chair for the IAPR ICPR 2014 conference.

Thierry Dutoit is graduated as an electrical engineer and Ph.D. in 1988 and 1993 from the Faculté Polytechnique de Mons (now UMONS), Belgium, where he teaches Circuit Theory, Signal Processing, and Speech Processing. In 1995, he initiated the MBROLA project for free multilingual speech synthesis, which provided speech synthesis for 29 OS and 75 voices, in 38 languages. Between 1996 and 1998, he spent 16 months at ATT-Bell Labs, in Murray Hill (NJ) and Florham Park (NJ), working on the NextGen Speech synthesizer. He is the author of several books on Speech Synthesis and Applied Signal Processing, of more than 15 chapter in books, 7 patents, and he wrote or co-wrote more than 40 journal papers, and more than 120 papers on speech processing, biomedical signal processing, and digital art technology. He has been an Associate Editor of the IEEE Transactions on Speech and Audio Processing (2003-2006) and the president of ISCA's SynSIG interest group on speech synthesis, from 2007 to 2010. He is an associate editor of the Journal on Multimodal User Interfaces (2007-), and a member of the Editorial Board of Speech Communication (2005-). T. Dutoit is a member of the IEEE Signal Processing society, and has been part of the Speech and Language Technical Committee of the IEEE (2009-2011). He was an ISCA Distinguished Lecturer in 2010-2011. He is involved in collaborations between UMONS and ACAPELA-GROUP, a European company specialized in TTS products. Recently he founded the UMONS/NUMEDIART Institute for Creative Technology, and the CLICK Living Lab, of which he is the director.

Laurence Ris After a Master degree in biology at the University of Mons, L. Ris has started her PhD thesis as a FNRS fellow under the direction of Prof.Godaux in the department of Neurosciences at the University of Mons. The subject of her thesis was the intrinsic plasticity of the vestibular neurons after an unilateral labyrinthectomy. The technique used was the extracellular in vivo recording in awake animals. Just after her thesis, she went to Paris in the laboratory of Prof. Vidal to learn the technique of intracellular recordings in acute brainstem slices. This technique was used in Mons to further study the excitability of vestibular neurons during her "aggregation" thesis. In 2003, she went to laboratory of K.P. Giese at UCLondon to do behavioural measurements and biochemistry on transgenic mice and became Research Associate of the FNRS in the laboratory of Neurosciences (UMONS). During these last 10 years, she received grants from Foundations, Belgian national and regional funds for Scientific Research and European funds (Era-Net). In 2011, she took the direction of the laboratory of Neuroscience and became a full-time lecturer. She is now responsible of the biomedical section and vice-dean of the Faculty of Medicine and Pharmacy at the University of Mons and has created the Centre for Psychophysiology and Electrophysiology of cognition of UMONS (CiPSE).

Jean-Philippe Vandeborre is a Full Professor of Computer Science at IMT Lille Douai (Institut Mines-Télécom Lille Douai), France. He received the M.S. degree in 1997, and the Ph.D. degree in Computer Science in 2002, both from the University of Lille, France. He also holds the French degree Habilitation à Diriger des Recherches (HDR) from the University of Lille, France, since 2012. He is also a researcher of CRIStAL (UMR CNRS 9189), the Computer Science Department of the University of Lille linked to the CNRS (Centre National de la Recherche Scientifique). His current research interests are mainly focused on 3D-model analysis, and include 3D-mesh segmentation, multimedia indexing and retrieval from content and their applications. He is also investigating the topic of human gesture analysis with the help of depth-cameras. He is the co-author of several publications in high-impact journals and conferences, and serves as program committee member or reviewer for international conferences and workshops in the 3D-model processing and computer vision fields.