World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:14, No:6, 2020

Modeling Engagement with Multimodal Multisensor Data: The Continuous Performance Test as an Objective Tool to Track Flow

Mohammad H. Taheri, David J. Brown, Nasser Sherkat

Abstract**—**Engagement is one of the most important factors in determining successful outcomes and deep learning in students. Existing approaches to detect student engagement involve periodic human observations that are subject to inter-rater reliability. Our solution uses real-time multimodal multisensor data labeled by objective performance outcomes to infer the engagement of students. The study involves four students with a combined diagnosis of cerebral palsy and a learning disability who took part in a 3-month trial over 59 sessions. Multimodal multisensor data were collected while they participated in a continuous performance test. Eye gaze, electroencephalogram, body pose, and interaction data were used to create a model of student engagement through objective labeling from the continuous performance test outcomes. In order to achieve this, a type of continuous performance test is introduced, the Seek-X type. Nine features were extracted including high-level handpicked compound features. Using leave-one-out cross-validation, a series of different machine learning approaches were evaluated. Overall, the random forest classification approach achieved the best classification results. Using random forest, 93.3% classification for engagement and 42.9% accuracy for disengagement were achieved. We compared these results to outcomes from different models: AdaBoost, decision tree, k-Nearest Neighbor, naïve Bayes, neural network, and support vector machine. We showed that using a multisensor approach achieved higher accuracy than using features from any reduced set of sensors. We found that using high-level handpicked features can improve the classification accuracy in every sensor mode. Our approach is robust to both sensor fallout and occlusions. The single most important sensor feature to the classification of engagement and distraction was shown to be eye gaze. It has been shown that we can accurately predict the level of engagement of students with learning disabilities in a real-time approach that is not subject to inter-rater reliability, human observation or reliant on a single mode of sensor input. This will help teachers design interventions for a heterogeneous group of students, where teachers cannot possibly attend to each of their individual needs. Our approach can be used to identify those with the greatest learning challenges so that all students are supported to reach their full potential.

Keywords**—**Affective computing in education, affect detection, continuous performance test, engagement, flow, HCI, interaction, learning disabilities, machine learning, multimodal, multisensor, physiological sensors, Signal Detection Theory, student engagement.

M. H. Taheri is a Research Fellow at the Interactive Systems Research Group at Nottingham Trent University, Nottingham, NG14FQ, UK (e-mail: mohammad.taheri@ntu.ac.uk).

D. J. Brown is the Research Group Leader of the Interactive Systems Research Group at Nottingham Trent University, Nottingham, NG14FQ, UK (e-mail: david.brown@ntu.ac.uk).

N. Sherkat is a Professor and Head of the Engineering and Mathematics Department at Sheffield Hallam University, Sheffield, S11WB, UK (e-mail: n.sherkat@shu.ac.uk).

I. Introduction

It is often a challenge to keep children engaged in learning activities, especially if the activity requires them to retain focus and active participation for a continuous period of time. Researchers reported that students with learning disabilities do not display any significant attention deficiency compared to non-disabled students – these students can complete the same activities if given more processing time [1]. Despite this outcome, student engagement can vary greatly depending on the activity, and understanding when the student is engaged, and when they are not, is not a straightforward task.

While research has focused significantly on the ability of children with learning difficulties to recognize [2], perceive [3] and interpret [4] emotional cues, there is little to no research on the recognition of the emotional state of these students. The importance of carers being able to interpret the emotional cues and states of such students has been documented in [5]. It is found that carers made significantly more critical and 'fundamental attribution' [6] errors in the emotional expression of their clients with learning disabilities in comparison to their clients without learning disabilities. This affects the quality and quantity of their client's treatment [5] and has a negative effect on the provisional treatment [7], [8]. Currently, carers rely on their expert understanding and personal experience of the students to interpret their voices, expressions, and gestures. Dependent on the personal experience with a particular client, a carers internal modeling of the emotional expression of that client can vary widely and demonstrate inter-rater reliability issues.

One of the main ways to measure engagement in students with special educational needs is to use the Special Schools and Academies Trust (SSAT) Engagement Scale [9]. The Engagement Profile and Scale is a classroom tool developed through SSAT's research into effective teaching and learning for children with complex learning difficulties and disabilities. It allows educators to focus on the child's engagement as a learner and create personalized learning pathways [10]. The authors describe seven components of engagement namely, awareness, curiosity, investigation, discovery, anticipation, persistence, and initiation. Teachers assign a score out of four for each component giving a total score out of 28. One potential issue with the use of this scale is that teachers assign a subjective rating to each component, which will be subject to inter-rater variability.

The scale has been used to assess the impact of new technologies in special education – especially in studies investigating the suitability of humanoid robots to support learning in students with Profound and Multiple Learning Disabilities (PMLD). The approach of using an engagement scale to create personalized learning pathways has been examined by others [11]-[13].

One way to overcome the variation in observer inter-rater reliability in tracking emotional expression is to introduce a reliable indicator of that emotion. In this research, a robust methodology for tracking engagement levels of children with PMLD or Cerebral Palsy (CP) is proposed using Signal Detection Theory (SDT) [14]. The application of this theory gives quantifiable information on the improvement of deterioration or attention in response to a Continuous Performance Test (CPT) specifically adapted to the abilities of such students [1], [15]. Performance in this test will provide objective labels to train machine learning algorithms using sensor data (e.g., on eye gaze and body pose) collected whilst the students are interacting with a PC. After obtaining a labeled dataset, machine learning models can be applied to the data so that in the future new unlabeled data can be presented to the model and engagement can be inferred.

Many traditional interactive systems use devices such as a keyboard and mouse and are constructed to emphasize the transmission of explicit messages while ignoring implicit information about user interaction. The emerging science of affective computing can only be accelerated with the abundance of sensor data [16], [17] and wearables [18]. These multimodal human cues [19]-[21] provide the multimodal multisensor data points necessary for enhanced emotional modeling. Multimodal multisensor data have been instrumental in determining user affective states [19], [22]-[28] including engagement [29]-[31]. There are a number of challenges to develop such a model including understanding the relationship between the terms used in educational contexts (e.g., ‘flow’ and ‘engagement’), developing appropriate CPTs suitable for the abilities of students with the most profound learning disabilities, selection of appropriate sensors and features derived from these data streams from which emotional states can be inferred, finding a suitable population of end-users to collect data with to train the machine learning algorithms, and finally comparing the performance of a range of machine learning methods to infer flow and engagement. This paper addresses each of these challenges.

II. Engagement, Flow and Learning

In education, the use of the term ‘engagement’ is more familiar to teachers than flow. D’Mello and Graesser [32] see considerable overlap between the two terms: “we conceptualize engagement/flow as a state of engagement with a task such that concentration is intense, attention is focused, and involvement is complete” (p.146). Contrary to engagement, the concept of flow is well defined in Csíkszentmihályi’s works [33], [34]. One is in flow when one is engaged [31], and steady performance has been maintained at the comfortable limits of one’s skill limitations [35], [36] for the duration of time - making flow the optimal psychological state of engagement. This results in immersion, concentrated focus and deep learning [37], [38]. The relationship between flow and engagement has been illustrated in Bianchi-Berthouze’s [31] engagement model, a simplified version that has been shown in (1):

$$Attention \rightarrow Flow \rightarrow Engagement \tag{1}$$

Performance trend tracking can be used as an indicator of flow [36]. This approach has been used in [39]-[41] as a model for relating affect (flow/engagement) to user performance in a pre-defined activity/task challenge.

Engagement’s crucial role in learning was recognized by Carpenter [42], stating that “Sustainable learning can occur only when there is meaningful engagement”. Learner engagement in the classroom is the single, most reliable indicator of deep learning [40], [43], [44] and learner satisfaction [33], [37], [38]. Its role is central to classroom performance and the achievement of learning outcomes [45]-[48]. For these reasons, flow, a sub-state of Engagement [31], [33], [49], is a more suitable measure to follow or track the quality of experience; firstly it can be objectively monitored, and secondly, through its monitoring, engagement is also established. Flow is the optimal state of engagement, where engagement meets productivity [37], [50]. Maintaining flow in learning is especially significant because it is the most reliable indicator for determining successful learning [36], [45]-[48]. In the absence of learner engagement, deep conceptual learning is also not present [48], [51], which is an essential attribute to long-term learning and new skill achievement [51].

III. Challenges in Understanding Engagement in Students with Learning Disabilities

Abrams stated [52], "The vast majority of children with learning disabilities have some emotional problem associated with the learning difficulty.” Generally, however, teachers have prioritized the diagnosis and remediation of learning disabilities [53].

Studies have considered self-reported affect states as the ground truth for inter-rater agreement studies [54], [55]. These studies have looked at the level agreement and correlation between self-rated affect states and peers, clinicians, and long-term partners. The level of correlation even though significant between the 40^{th} and 70^{th} percentiles [54], [55], still leaves room for improvement. In addition, self-rated affect states may carry bias or not be representative of the true affect state. Therefore, an automated method that would base its ground truth on self-rated affect states would thus be impacted by such bias and unknown reliability factors. The validity of a machine learning method based on clinician, or peer-rated affect states would inherit even greater bias, reliability, and interrater reliability uncertainty, as it is one more level separated. Importantly, a machine learning method with 100% classification accuracy trained with clinician-rated affect data would at best achieve around 70% correctness of the self-rated affect states. Furthermore, the self-rated affect states may themselves have a bias or be unrepresentative. This creates a problem for both the clients and care workers as it has been shown that observation is not a reliable method of determining a person's mood and affect state [6]. This can only be more intensified with PMLD and CP users, as their behaviors, body language and voice may not have the same cues as mainstream people. Moreover, the levels of skill and experience between care workers and teachers vary widely, as does their capacity and accuracy of interpretation of others' behaviors. This uncertainty of interpretation and inaccuracy in the observation of the affect state of a person experiencing PMLD or CP (mood and emotional well-being) can be detrimental to their quality of life [5], [7], [8]. Hence, the well-being of a student with PMLD or CP can be improved if their levels of interest and engagement could be determined and tracked by more independent and repeatable means, such as using technology, and in our case sensors. This added interpretation of a student's state of affect is not meant to replace teachers' or carers' interpretation, but more to augment this judgment.

Monitoring a person's level of interest and engagement in activity allows carers, teachers, and parents to be responsive to those levels. In this study, we investigate the ability of sensor-based technology to detect and track sustained attention in a repetitive demanding activity, with a multimodal multisensor platform. This allows us to make inferences on the attention level of the student throughout the length of this activity through their responses to the challenges presented in the repetitive activity.

An objective approach to the reporting of engagement is the use of a standardized test to monitor for indicators of flow. We demonstrate the possibility of tracking and then modeling body movements, eye gaze, electroencephalogram (EEG) and interaction data from students with PMLD and CP to estimate their level of engagement, as a good indicator of what interests them and positively influences the quality of that experience.

IV. A Platform to Measure Engagement Using Multimodal Multisensor Data for PMLD

A gamified *platform* is proposed that monitors the qualities of flow, namely engagement through performance tracking using SDT [14] measures and outcomes. For the remainder of the paper, we will refer to this engagement tracking platform as *'the platform'*.

The participant is required to pay continuous *attention* to a computer screen where an *interactive* game provides them with a pre-defined signal detection *challenge*. The participant is in *control* of the response they give, and *feedback* is given to them regarding the correctness of their response to the *challenge*. This is the basis for Swanson's CPT [15]. The CPT is an integral component of *the platform,* and we have therefore created a version, the 'Seek-X' type. This test has been created to be used specifically as *an objective tool* for engagement tracking using the CPT test outcomes to label multisensor data.

We have named this CPT 'Seek-X type' because the participant is asked to *seek* the target image between other non-target images acting as a matrix of noise. 'Type-Seek-X' exercises engage eye gaze as a crucial element of answering the SDT challenge. The Seek-X type CPT is of the non-rare target type, see (2):

$$\text{CPT test types} = \begin{cases} By\ challenge \begin{cases} Type - X \longrightarrow Seek - X \\ Type - AX \longrightarrow Seek - AX \end{cases} \\ By\ target\ frequency \begin{cases} Rare \\ Non - rare \end{cases} \end{cases} \quad (2)$$

In summary, the period of sustained engagement is marked by participants' *attention* and *interest* being maintained in an interactive interaction. Maintaining sustained attention indicates the key foundation for recognizing lasting engagement. For this reason, this work explores classical methods for attention tracking using a neuropsychological test that measures a person's sustained and selective attention (the CPT) [15]. The CPT is reported to be the most popular measure of sustained attention or vigilance—the ability to sustain attentional focus and remain alert to stimuli over time [56], [57]. The first attempt to objectively evaluate the relationship between maintaining attention in students with learning disabilities using CPTs was introduced by Swanson in [15], [58] and later expanded by Eliason and Richman [1]. Using SDT [14], [15], [58]-[62], quantifiable objective data on the improvement or deterioration of attention are collected and analyzed using SDT detailed in [60], [61].

A. Data Collection

Four students were recruited for data collection (see *Participants*). They took part in an 11-week long study with up to four sessions weekly, depending on participant availability.

Each session included 48 challenges. Each test lasted between 6-32 minutes depending on participant readiness or other setting-up challenges. Every session recorded nearly 4 minutes of data. A total of 59 sessions of the CPT test were carried out (average of 15 sessions per participant). A series of 48 slides with pauses in between were displayed for each participant.

This CPT test design was based on Rosvald and Mirsky's original paper [63]. Recommended time alterations to the experiment length were made to match the shorter length activities that students with PMLD are accustomed to at school [15]. The CPT test was therefore shortened to about 4 minutes for our participants, and the whole process takes around 15 minutes. This is compared to other research, which suggests a 30-minute test for neurotypical participants [64].

The difficulty of the CPT was also adapted for each participant by making the maximum response time (slide display time + blank slide display time) shorter or longer or by adjusting the image matrix grid size. These times are initially 1.8 s and 1 ± 0.1 s, respectively, and are increased or decreased depending on participant capacity. These times (seen in Table I) were established in a series of pilot tests where the aim was to reach close to the 85% rule for learning, where the participant makes around 15% mistakes and 85% correct responses [35] when in flow. The Seek-X type CPT slide timeline is demonstrated in Fig. 1.

TABLE I
CPT SETTINGS ADJUSTED PER PARTICIPANT CAPACITY

Participant alias	P scales mean	Slide display time/ Stimulus duration (s)	Blank slide display time/ Interstimulus interval (s)
Will	6.93	1.8	1.1 ± 0.1
Jen	19.45*	1	1.1 ± 0.1
Mark	3.7	8	2.1 ± 0.1
Rick	6.76	1.8	1.1 ± 0.1

*Jen is enrolled in the National Curriculum.

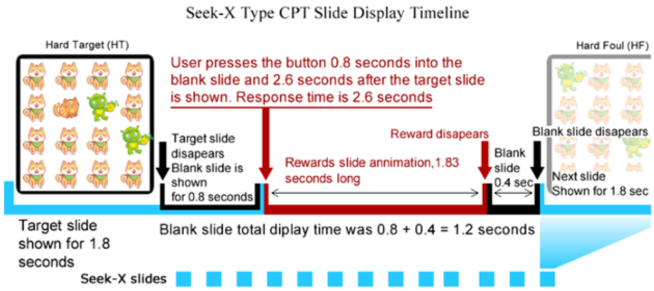

Fig. 1 Seek-X type CPT slide timeline

It is important in SDT that the participant can demonstrate that they understand the difference between the target and noise, given enough time. To establish this, the game objective was re-introduced to the participants at the start of every session using a paper-based mockup to test the participants' understanding of the challenge and validate their response.

B. Experimental Platform and the CPT

The platform tracks student performance in a repetitive game, which rewards them with exciting visual and audio feedback when they answer correctly, but ultimately fatigues the student by being exhausting over a long period. The student is required to pay attention to the game dynamic, which challenges them to pay selective and sustained attention to the elements on the screen and respond appropriately. This induces different states of affect, with lower levels of valence, as the game carries on and the students' attention capacity naturally decreases. During this game, real-time multimodal multisensor data are collected within the experimental platform, which are used later to create a machine learning model of flow. The experimental platform was developed in MATLAB to collect data from various consumer-grade sensor hardware. The experimental platform and the relative student position are visualized in Fig. 2.

The type of CPT, 'Seek-X', was designed for this study. Each slide has a mixture of three images, comprising of the target image, the target imitation and the contrast image, as seen in Fig. 3. The target imitation bears a close resemblance to the target image (similar colors, general shape); however, the contrast image can easily be identified.

The ratio of the mixture of the main image to the filler image in all slide types is always 9 to 1 or as close as possible to this ratio, depending on the grid size and limited spaces available. We found that for our test user group, a grid of 4 x 4 introduced enough difficulty to allow for participant responses, without being so easy that the participant would not make any mistakes when fatigued.

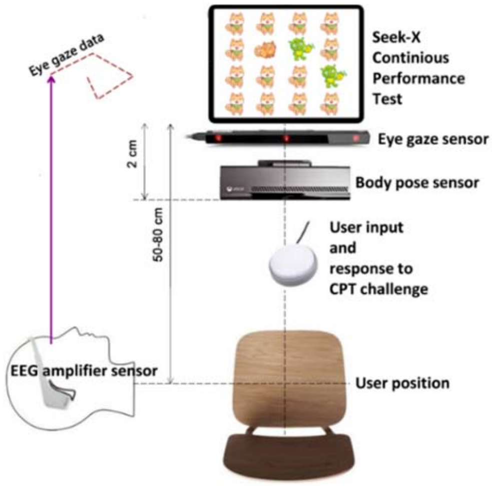

Fig. 2 The multimodal multisensor experimental platform with the eye gaze, body pose, EEG sensor and the CPT

Target image Target imitation image Contrast image

Fig. 3 CPT image types

The distribution of the Hard Target (HT) pattern among the other random patterns has an occurrence probability of 50%. The other CPT occurrences are standardized [63] as Hard Foul (HF), Easy Target (ET) and Easy Foul (EF). These patterns and their corresponding labels are seen in Table II.

TABLE II
THE DISTRIBUTION OF PATTERNS IN THE SEEK-X TEST

Pattern	HT	HF	ET	EF
Distribution	50%	25%	12.5%	12.5%
CPT Label	HT: Target image mixed in with imitations targets with a few contrast images.	HF: Imitation target images with some contrast images.	ET: Target Image mixed in contrast images with some imitation targets.	EF: Contrast images with some imitation targets.

The participants were seated in a chair in front of a 20" computer monitor, at a controlled distance of 50 cm to 80 cm from the screen. Each participant was asked to press the keyboard spacebar, or a big button if wheelchair-bound, whenever they saw the target image on the screen, and not to press the button when they did not see the target image on the screen. During this activity, participant eye gaze, body pose, EEG measurements and button interaction data were continuously recorded.

The participant was then presented with 48 instances of images displayed in a controlled random sequence on the screen. Each image was displayed for a stimulus duration (slide display time) followed by a blank slide displayed for an interstimulus interval.

Real-time eye gaze position using Tobii EyeX [65], body pose data using Kinect v2, EEG data from the Muse headband

[66] and interaction data from the USB button are recorded in MATLAB [67]. The Muse EEG headband streams 16-bit voltage data in microvolt (μV) units at 500 Hz, which are equal or comparable to medical-grade EEG specifications [68]. The Tobii EyeX eye gaze tracking controller [69] uses near-infrared light to track the eye movements and gaze point of a student [70]. It works in variable light conditions and allows for student head movement while maintaining accuracy, which is crucial for our target user group. It has a frequency of 70 Hz and uses backlight assisted near-infrared (NIR 850 nm + red light (650 nm)) to achieve a 95% tracking population [71]. The Kinect 2 sensor [72] is a motion-sensing peripheral for body tracking. Using structured light and machine learning it can infer body position [72]. Kinect 2 is reported with an average depth accuracy of under 2 mm in the central viewing angle and increases to 2-4 mm in the range of up to 3.5 mm [73]. The furthest distance captured by Kinect 2 is 4.5 mm, where the average error typically increases beyond 4 mm. The experimental platform was designed to replicate the majority of the CPT test variations reported in relevant studies [1], [15], [63], [74]-[78]. The features extracted from these sensor data streams are described under feature extraction.

C. Participants

Four participants with PMLD were recruited to collect labeled sensor data whilst using the gamified platform. These four participants have a wide range of abilities, from extreme mobility restrictions to moderate learning disabilities. Our four participants are given pseudonyms, referred to in this paper as Will, Jen, Mark, and Rick.

The four participants are made up of three boys, and one girl, aged 16 to 19 years. Information leaflets were sent to the special educational needs school from which they were recruited to inform staff and parents about the project.

Students were selected based on their performance in scales, which represent a set of descriptions used to record and assess the progress of children who have special educational needs (P-scales) [79], [80] (see Table I). Permission for the study was given by Nottingham Trent University's ethics committee. The user characteristics of each participant are now described in detail.

Will is 18 years old, has a diagnosis of global development delay (GDD) and learning disability. These impact on his speech, language, and social interaction with others. This means his ability to concentrate on a single activity for an extended period is limited, which in turn, limits his sustained attention. His body mobility is not restricted, however slightly imprecise. His speech sounds imprecise and is limited in the selection of words. His capability in conducting particular tasks in quick succession is good; however, he struggles to maintain sustained attention.

Jen is 19 and has a rare form of epilepsy. She is one of the more capable students at the school; she is very cooperative and shows an interest in being involved in the study. She also talks about music and theater and has interests in fashion and celebrities.

Rick is 19 and has a global delay, a rare form of epilepsy and a severe learning difficulty. Rick has problems processing information and communication. His attention is usually committed to a single concept (an activity, a memory, a sound). He is incredibly reliant on routine, and he will try to avoid any disruptions to it. He enjoys loud motor sounds, power tools, and garden work. He often reflects on activities he has done in the past or will do in the future with single words or short phrases. His mobility is not constrained but is delayed and processing time needs to be allowed for any response. Physical objects and sounds help him associate with new concepts.

Mark is 16 years old and has myotonic dystrophy; this makes his muscles very weak. Myotonic dystrophy is a progressive and life-limiting condition. Mark uses a wheelchair and is at risk of chest infections and sudden heart failure. He uses a specialized CP wheelchair for body support and transportation. The wheelchair supports his body frame and keeps him upright and secure with a safety belt. His head is rested against his right ear on a padded headrest. His mobility disability is extreme; however, he has some imprecise movement in his neck and arms. At the school, he uses both eye gaze technology and switches to interact with computer interfaces. Mark uses his voice to communicate; he likes sharing his sense of humor, he laughs when things go wrong, and makes the sound 'uh-oh' to signal mistakes. He enjoys making choices and can become frustrated when he is not offered choices. Mark likes interacting with computers, however, shows sensitivity to anything resting on his forehead like the EEG headband. Because of his CP, he required a member of staff to be present during the study. Mark shows a definite progression with communication and is now very accepting of and participating in a wider variety of activities, events, and opportunities in school.

D. Feature Extraction

Brain-Computer Interfaces (BCIs) represent a novel mode of communication that has been used in emotional classification [81], and cognitive aware applications [82]. BCIs are also considered unique in augmentative and alternative communication (AAC) as they do not require physical movement from a user. This makes BCIs a suitable AAC method for people with Severe Speech and Physical Impairments (SSPI) [83], or CP [84]-[87] who do not have access to conventional means of communication including speech and typing [86].

The quality of a BCI — to offer a direct mode of information from the brain — makes it especially ideal as an element in potential real-time affective user state detection [88], computer interaction for rehabilitation [89] and in brain multimedia interaction [90]. A BCI can also be a complementary source of information towards multimodal interaction systems as well, used in conjunction with other modalities such as gesture, facial expressions, gaze and body posture [91]-[93].

EEG frequency has been used as a feature to determine the active brain state [94]-[96]. In this study, five channels of

EEG data are recorded, TP9, AF7, FPz, AF8 and TP10 [97] at a frequency of 500 Hz. EEG Kalman filtering has been shown to be useful in removing EMG induced artifacts [98]-[104]. A robust Adaptive Autoregressive (AAR) model with an order of six detailed in [104] was used. The AAR model estimate of the EEG Kalman filter was utilized to reduce the impact of Electromyography (EMG) spikes from body movement, eye blinks and other facial muscle movements. These EMG spikes are isolated in a few samples, which make the data ideal for AAR Kalman filtering. In Fig. 4, we see that it has removed the EMG artifact that can be seen between samples points A and B, enhanced the EEG spikes, and revealed an EEG peak between C and D.

By using an AAR Kalman filter on the data, we estimate the EEG wave during the EMG incident artifacts using surrounding neighboring EEG samples and correct those affected samples. This is done by evaluating a moving set of samples and checking for EMG contamination. The contamination is then removed by estimating a normal rate of progression for the signal to reach from point A to point B using a sliding window for the length of the recording.

Studies [105]-[108] show that the EEG beta rhythm (14–30 Hz) is activated when the brain is in a state of arousal. In other EEG studies, mental fatigue related features are associated with decreased alpha band (8-13 Hz) power at one or more parietal locations (e.g., P7 and P8). Liu et al. [109] connected these two factors in their study and showed that alertness can be measured by the signal power of α divided by the signal power of β. McMahan et al. [110] also demonstrated that the ratio is related to arousal.

Using the signal power of α divided by the signal power of β as the EEG feature, the EEG recordings are labeled with the CPT outcomes. A Butterworth bandpass filter was employed to extract the frequency response of the α and β bands from the EEG signal as demonstrated in [111]. Discrete Fourier Transform (DFT) was used to calculate the Power Spectral Density (PSD) of the α and β time series.

DFT periodogram methods for estimating the spectrum power density are prone to variation [112]. Periodogram estimate variation is correlated to the square of the value of the spectrum itself. Welch's method reduces this variance by averaging independent periodogram estimates. Each Welch window covers 50% of the next, which results in the smoothed-out average of independent periodogram spectrum estimations. We use a Hamming window as it produces the least amount of overshoot $\delta_{\text{Hamming}} < \delta_{\text{Hann}} < \delta_{\text{Bartlett}}$ [112] with the most accurate results for EEG data [111], [113].

A Hamming window of M = 100 samples was chosen with a 50% overlap, and since the EEG frequency is 500 Hz, this Hamming window is equivalent to 200 ms of data. To help illustrate, an average data interval length is 2.3 seconds long and would have $2300 \div 200 \times 2 = 23$ overlapping Hamming windows. Let $\{xd(n)\}$ be the sequence, $d = 1, 2, 3 \cdots L$ signal intervals and M the interval length. Welch's method to estimate the power spectrum discrete time sequence is shown in (3) where U is the normalization factor (4) and the Hamming window calculation is shown in (5). Using the Welch method, the ratio of the alpha band power f_α to the beta band power f_β can be simplified as (6).

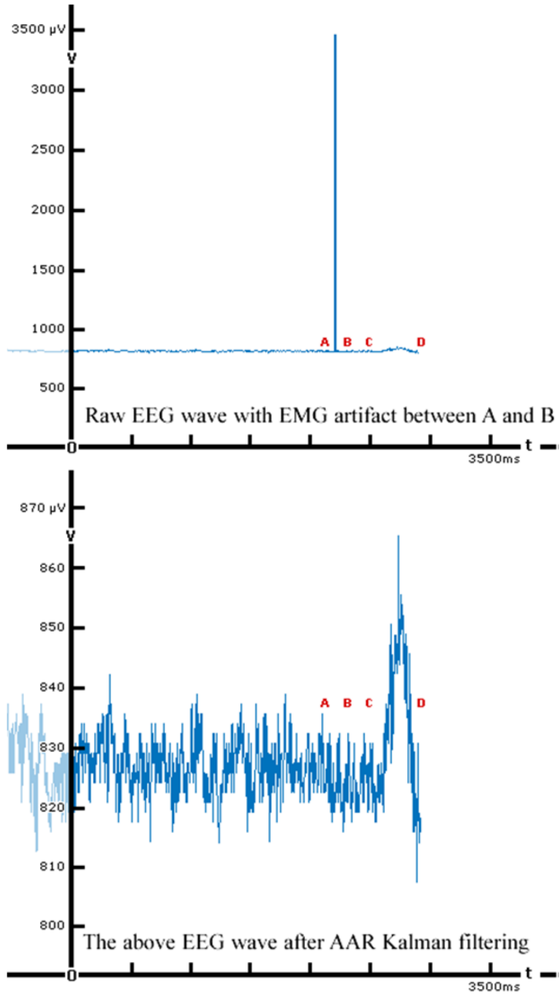

Fig. 4 AAR Kalman filtering reduces EMG noise and enhances EEG spikes

Welch Method:

$$\hat{p}d(f) = \frac{1}{MU}\left|\sum_{n=0}^{M-1} xd(n)w(n)e^{-j2\pi f}\right|^2 \qquad (3)$$

U is the normalization factor for Welch Method:

$$U = \frac{1}{M}\sum_{n=0}^{M-1}|w(n)| \qquad (4)$$

Hamming window:

$$w[n] = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{M}\right), & 0 \le n \le M, \\ 0 & otherwise. \end{cases} \qquad (5)$$

The EEG Alertness feature:

$$\text{Alertness} = \frac{\hat{p}d(f_\alpha)}{\hat{p}d(f_\beta)} \qquad (6)$$

Body pose can be one of the strongest communication channels [114]. Body pose is acquired through the Kinect v2.0 SDK [72], which will provide joint tracking data at 30 Hz. Tracking of the head, neck, mid-spine, right and left shoulders and left and right hands are recorded. Lower joints are not included as occlusion from the table as part of the platform prevents such recordings. Studies have shown that body posture and gesture can communicate affective modalities and also specific emotional categories [115]. They have also been indicators of a firm or weak correlation of engagement during Human-Computer Interaction in gameplay [31]. In this study, the student is positioned in front of a computer system and is challenged to press a button when they identify the target. This type of interaction setup restricts the range of body movements and gestures a student can engage in. Numerous studies [116]-[121] have investigated the importance of body fidgeting in detecting attention for students with PMLD. Fidgeting is an indicator of the onset of attention loss, boredom and engagement deterioration [116], [122]-[125]. We calculate rapid body movement from body pose to assess fidgeting levels. The equation to extract this feature is seen in (7) where Δd_j is the displacement vector of joint j out of N joints and Δt is the time passing between the displacement samples.

$$\text{Body fidgeting} = \frac{1}{N}\sum_{j=1}^{N}\frac{|\Delta d_j|}{\Delta t} \quad (7)$$

Eye gaze data are recorded at 70 Hz. The data include Cartesian information regarding the eye gaze location relative to the bottom left corner of the screen. We track gaze, which is both on and off-screen. The combination of off-screen gaze tracking and eye detection provides information on when the user turns their head away from the screen. Three features were extracted from the eye gaze data: 'eye scanning', 'eye dwelling' and 'eyes off-screen'. These features are commonly used in eye gaze technologies to understand attention, interest and engagement [126], [127].

Scanning represents the eye gaze behavior of when the gaze tracks across more than one image element. The scanning feature is calculated in (8) and represents the sum of the inverse distance from the center of each element where r_{in} is that distance; from the eye gaze location to the center of image i out of $I = 16$ total image elements, for sample n, out of N total discrete sensor samples. This is demonstrated in Fig 5.

$$\text{Scanning} = \sum_{n=1}^{N}\sum_{i=1}^{I}\frac{1}{r_{in}} \quad (8)$$

Dwelling represents the eye gaze behavior of when the gaze stays relatively in the same position for a duration of time. This behavior is independently calculated from the location of image elements on the screen. The dwelling feature is calculated in (9), which is the sum of the inverse distance from each eye gaze position to the next where n is the sample number out of N total discrete sensor samples, and Δd is the distance the eyes have moved since the previous sample, as demonstrated in Fig. 6.

$$\text{Dwelling} = \sum_{n=1}^{N}\frac{1}{\Delta d_n} \quad (9)$$

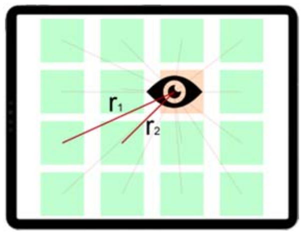

Fig. 5 Scanning calculation with respect to the active elements on the screen

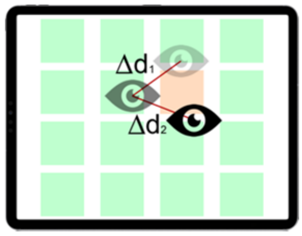

Fig. 6 Dwelling calculation independent of active elements on the screen

The third feature extracted from the eye gaze data is 'eyes off screen'. This continuous but binary feature determines if the participant is looking within the screen area, regardless of whether there was a slide or blank slide on the display. This feature is calculated as in (10):

$$\text{Eyes off screen} = \begin{cases} 1 & eyes\ off\ screen \\ 0 & eyes\ on\ screen \end{cases} \quad (10)$$

Interaction data features were extracted from the participants' behavior activating a button press. The type of pressing, including quick presses or repetitive presses, was recorded as were other sensor data with a view to *behavior*, not just input, but as an independent sensor mode. This makes our approach unique as the input device is considered not only as an objective indicator of attention but also as a separate mode of interaction. We remain impartial to which slide is displayed and only consider the interaction behavior. How the button is pressed, specifically how fast the button is pressed, and how many times it is pressed is of interest. From button presses, we extract two features: single fast button presses and repetitive button presses. Single fast button presses are calculated using (11), with the caveat that they are only calculated if the participant presses the button once and only once during the response time duration. In other instances, the value for this feature is zero. Maximum press count is the second feature extracted from the button press data shown in (12). This value is calculated for only the allowed response time interval and is zero when the button is not pressed.

$$\text{Single fast press} = \frac{1}{response\ time} \quad (11)$$

$$\text{Max press count} = \text{total press attempts} \quad (12)$$

High-Level Compound Features (HLCF) were created to create a higher dimensionality in the feature space as described in the Mudra multimodal framework [128]. The first feature is a compound feature, which is simply a normalized mean of the features that traditionally serve indicators of attention. The High-level Attention feature is calculated as the mean of the normalized features of single fast presses, eye dwelling, eye scanning and EEG alertness, which are seen in (13):

$$\text{HLA} = \frac{1}{4}(norm.Single\ fast\ press + norm.Dwelling + norm.Scanning + norm.Alertness) \quad (13)$$

High-level Distraction feature is calculated as the mean of normalized features of body fidgeting, eyes off-screen and press count as seen in (14):

$$\text{HLD} = \frac{1}{3}(norm.Body\ fidgeting + norm.Eyes\ off\ screen + norm.Max\ press\ count) \quad (14)$$

V. Experimental Results

A. Labelling and Data Fusion

The CPT provides an objective means of labeling the multimodal sensor data. The CPT outcome measures (correct commissions/Hits, False Alarms (FA), correct omissions and misses) are objective outcomes of the participant's attention and engagement with the game. Without these labels, there would be no objective measure or automated way of performing a supervised learning method on the data. An overview of how the data streams are collected and labeled against CPT outcome measures is shown in Fig. 7. Each slide from the moment it is displayed until the moment of the first button press, or until the moment of a new slide being shown (in case of no press), represents a sample of data. Overall, there were 2615 samples collected from the 59 sessions of data collection. The data from all four participants were collated together.

B. Machine Learning Results

A robust cross-validation method ensures that the results are not subject to overfitting. Leave-one-out [129] classification is a state of the art cross-validation methodology and is widely accepted not to be susceptible to overfitting. We show (regardless of the classification method), that there is a relationship between affective state and the multimodal multisensor data features. In this study, 2615 frames, over the length of 59 sessions, were collected and classified into two categories (engaged and disengaged) using nine features (7 low-level and 2 HLCF). The aim of classification is to determine the affective state by predicting the CPT outcome. With two classes, the random classifier classification accuracy to beat is 50%. The overall approach used to evaluate the fit of the different architectures was leave-one-out cross-validation. Impartial scoring metrics were used to competitively compare the performance of the machine learning architectures as these methods normalize across categories (and are suitable for imbalanced datasets). The evaluation parameters used for determining the comparative performance of the machine learning architectures were Area Under the ROC Curve (AUC), Negative Log-likelihood and Kappa. The software used to create this architecture is Python 3.7 and two high-performance computers, which ran in parallel over several weeks. The two PCs were both equipped with Intel i7-7700HQ 2.80 GHz CPUs, and 16 GB of DDR4 RAM. The CPU was benchmarked at 82 Gigaflops, with 15 GB/s memory transfer rate and 1 GB/s SSD disk transfer rate.

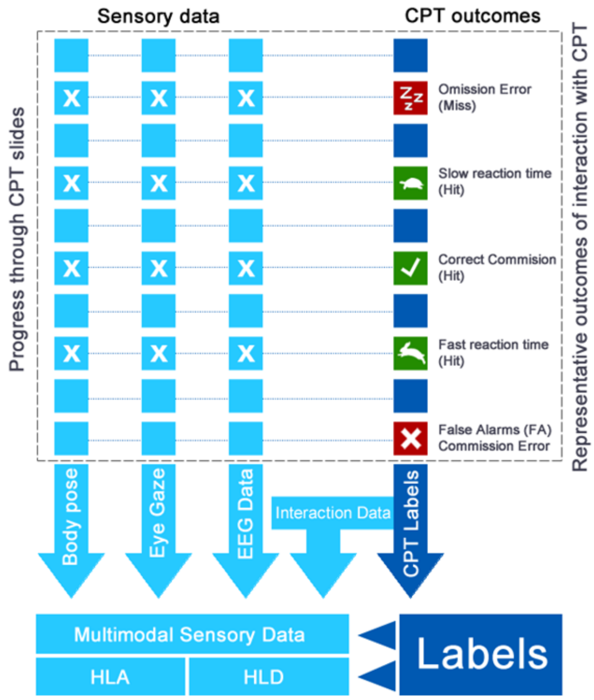

Fig. 7 Multimodal fusion diagram shows the temporal connectivity between the samples and multi-level feature fusion

The summary of results is shown in Table III. Overall, the random forest classification approach achieved the best classification results in all modes of data. This was either when including high-level features, or when only using a sub-set of the data modes. This finding is supported by other studies [130], which suggest that random forest provides consistent pairwise similarity, crucial for multimodal data. Pairwise-similarity facilitates the combination of features, adding higher dimensionality to the feature space whilst being less sensitive to data sample size [131]. The best method, random forest, used both high- and low-level features and achieved 93.3% classification for flow and a 42.9% accuracy for non-flow. The random forest method incorporated 100 trees and all nine features were included at each of the 255 nodes, with 128 leaves in total. AdaBoost (another ensemble method) outperformed random forest for the single modality feature classification. However, in every example, using any machine learning method, multimodal data features delivered significantly better classification results than any single modality.

When compared to the second-best classification method, random forest outperforms neural network on the classification of non-flow classes with a margin of 16.5% and has an 11.7% better coverage in AUC (see Table III). Besides neural networks, other machine learning methods were also assessed; AdaBoost, decision tree, k-Nearest Neighbor, naïve Bayes, and support vector machine, however, all had inferior performance when compared to random forest.

Including the two high-level [128] handpicked features (HLA and HLD) in the feature space, improved the classification in every sensor combination, and every machine learning methodology. In the random forest model including HLCF increased the AUC by 1.5% more coverage, and the classification of True Positives (TP) by 0.9%, and True Negatives by 2.8%. On average, if only two modes of sensor input were available, including interaction data improves the outcome of AUC coverage by 16.8%, compared to any other two modes of data, making interaction data the single most important secondary feature. The single most important mode of data on its own however is eye gaze, with 3.2% better AUC coverage compared to interaction data.

TABLE III
BEST CLASSIFICATION RESULTS ACHIEVED WITH RANDOM FOREST USING MULTI-LEVEL FEATURE FUSION

Features	Best Classification Method Found	Negative log-likelihood for Flow (Less is better)	Negative log-likelihood for Non-Flow (Less is better)	Kappa	AUC	TP	TN	F1	Precision	Recall
All features	Random Forest	0.1377	1.0149	0.418	0.803	93.8%	42.9%	0.819	0.817	0.833
Low-Level		0.1440	0.9753	0.374	0.788	92.9%	40.1%	0.806	0.802	0.820
High-Level		0.1250	0.9547	0.237	0.686	93.1%	26.4%	0.768	0.762	0.794
All features	Neural Network	0.1237	0.8191	0.300	0.773	93.6%	31.5%	0.786	0.783	0.808
Low-Level		0.1203	0.7910	0.273	0.767	95.3%	26.4%	0.781	0.783	0.811
All features	AdaBoost	0.2775	2.1803	0.388	0.794	93.3%	40.7%	0.810	0.807	0.824
Low-Level		0.2756	2.0422	0.335	0.765	90.7%	39.7%	0.791	0.785	0.802
All features	Naïve Bays	0.1341	0.7574	0.233	0.712	91.3%	28.7%	0.764	0.755	0.784
Low-Level		0.1097	0.7463	0.095	0.728	98.0%	8.50%	0.732	0.748	0.796
All features	k-NN	0.1105	1.4139	0.207	0.746	96.4%	19.2%	0.763	0.771	0.804
Low-Level		0.1115	1.5077	0.169	0.730	96.5%	15.9%	0.752	0.760	0.799
All features	Tree	0.2545	1.6061	0.309	0.706	89.9%	38.4%	0.782	0.776	0.793
Low-Level		0.1157	1.6414	0.258	0.686	89.4%	34.0%	0.767	0.760	0.780
All features	SVM	0.1107	0.6620	0.086	0.454	76.2%	33.2%	0.686	0.701	0.673
Low-Level		0.1026	0.6750	0.059	0.467	72.7%	34.0%	0.667	0.693	0.647
Eye + EEG + Inter.	Random Forest	0.1429	1.0202	0.349	0.765	92.4%	38.4%	0.793	0.793	0.812
Eye + Body + Inter.	Random Forest	0.1433	1.0784	0.371	0.781	91.6%	41.9%	0.803	0.798	0.814
EEG + Body + Inter.	Random Forest	0.1619	1.5544	0.318	0.730	91.9%	36.0%	0.788	0.783	0.804
Eye + EEG	Random Forest	0.1335	0.7164	0.277	0.679	95.0%	27.1%	0.781	0.783	0.810
Eye + Body	Random Forest	0.1619	1.5544	0.318	0.708	93.7%	27.9%	0.776	0.772	0.801
EEG + Body	AdaBoost	0.4902	2.2224	0.122	0.610	84.2%	27.4%	0.719	0.713	0.725
Eye + Inter.	Random Forest	0.1380	1.1820	0.308	0.765	93.6%	32.2%	0.788	0.785	0.810
Body + Inter.	AdaBoost	0.2579	2.0817	0.327	0.692	83.5%	52.1%	0.776	0.783	0.770
EEG + Inter.	AdaBoost	0.3002	2.0323	0.246	0.708	85.7%	38.3%	0.756	0.753	0.759
EEG	AdaBoost	0.2821	0.8682	0.100	0.559	84.6%	24.8%	0.714	0.706	0.723
Eye gaze	AdaBoost	0.2646	1.3051	0.255	0.637	89.2%	34.0 %	0.766	0.758	0.778
Body	AdaBoost	0.3091	1.0059	0.003	0.488	93.2%	7.10%	0.702	0.674	0.754
Interaction	AdaBoost	0.2491	0.6092	0.035	0.605	95.8%	6.70%	0.713	0.694	0.774
All Features Low-Level	Constant Classifier	0.1004	0.6854	0.000	0.000	100%	0.00%	0.702	0.630	0.794

The system developed using these machine learning models would not be affected by both sensor fallout and occlusions. At best (all high- and low-level features using random forest) 80.3% AUC coverage is achieved. Using a sub-set of three sensor modes 78.1%-73% AUC coverage is achieved, whilst with a subset of two sensor modes (including interaction) 76.5%-69.2% AUC coverage is achieved. Using a subset of two sensor modes (not including interaction) 70.8%-61.0% AUC coverage is achieved, and with only a single mode of sensor data between 63.7%-48.8% AUC coverage is achieved.

VI. CONCLUSIONS

An approach to labeling multimodal sensor data to train machine-learning algorithms to infer the engagement and flow of students with profound and multiple disabilities has been presented. We posit that this approach can overcome the variation in observer inter-rater reliability when using standardized scales in tracking the emotional expression of students with such profound disabilities. The accuracy of our approach increases with multiple modes of sensor input, and our method is robust to sensor occlusion and fall-out. Multiple sources of sensor input are provided, to accommodate a wide variety of users and their needs. Our model can reliably track the flow of students with profound disabilities, regardless of the sensors available. A system incorporating this model can help teachers design personalized interventions for a very heterogeneous group of students, where teachers cannot possibly attend to each of their individual needs. This approach could be used to identify those with the greatest learning challenges, to guarantee that all students are supported to reach their full potential.

This research was conducted as part of a Ph.D. program of research at Nottingham Trent University (NTU) and has been adopted as part of the Erasmus+ KA201 Pathway+ project to determine the affective state of students with mild and moderate learning disabilities (2017-1-UK01-KA201-036761) [132].

References

[1] M. J. Eliason and L. C. Richman, "The Continuous Performance Test in learning disabled and nondisabled children," *J. Learn. Disabil.*, vol. 20, no. 10, pp. 614–619, 1987.

[2] E. Bloom and N. Heath, "Recognition, expression, and understanding facial expressions of emotion in adolescents with nonverbal and general learning disabilities," *J. Learn. Disabil.*, vol. 43, no. 2, pp. 180–192, Mar. 2010.

[3] V. L. Petti, S. L. Voelker, D. L. Shore, and S. E. Hayman-Abello, "Perception of Nonverbal Emotion Cues by Children with Nonverbal Learning Disabilities," *J. Dev. Phys. Disabil.*, vol. 15, no. 1, pp. 23–36, 2003.

[4] H. B. Holder and S. W. Kirkpatrick, "Interpretation of emotion from facial expressions in children with and without learning disabilities," *J. Learn. Disabil.*, vol. 24, no. 3, pp. 170–177, Mar. 1991.

[5] L. Weigel, P. E. Langdon, S. Collins, and Y. O'Brien, "Challenging behaviour and learning disabilities: The relationship between expressed emotion and staff attributions," *Br. J. Clin. Psychol.*, vol. 45, no. 2, pp. 205–216, 2006.

[6] F. Heider, *The psychology of interpersonal relations.* Hoboken: John Wiley & Sons Inc, 1958.

[7] D. Dagnan, P. Trower, and R. Smith, "Care staff responses to people with learning disabilities and challenging behaviour: A cognitive-emotional analysis," *Br. J. Clin. Psychol.*, vol. 37, no. 1, pp. 59–68, Feb. 1998.

[8] R. Sharrock, A. Day, F. Qazi, and C. R. Brewin, "Explanations by professional care staff, optimism and helping behaviour: an application of attribution theory," *Psychol. Med.*, vol. 20, no. 4, pp. 849–55, Nov. 1990.

[9] "SSATrust - Engagement profile & scale." (Online). Available: http://complexld.ssatrust.org.uk/project-resources/engagement-profile-scale.html.

[10] B. Carpenter, "Engaging young people with complex learning difficulties and disabilities in learning," no. Cldd.

[11] J. P. Roscoe, "Can the use of a humanoid robot enhance the learning of children with learning disabilities?" University of Nottingham, 2014.

[12] P. Standen *et al.*, "Engaging Students with Profound and Multiple Disabilities Using Humanoid Robots," Springer, Cham, 2014, pp. 419–430.

[13] T. Hughes-Roberts *et al.*, "Examining engagement and achievement in learners with individual needs through robotic-based teaching sessions," *Br. J. Educ. Technol.*, vol. 50, no. 5, pp. 2736–2750, Sep. 2019.

[14] J. I. Marcum, "A statistical theory of target detection by pulsed radar," Dec. 1947.

[15] L. Swanson, "Vigilance deficit in learning disabled children: a signal detection analysis," *J. Child Psychol. Psychiatry.*, vol. 22, no. 4, pp. 393–399, Oct. 1981.

[16] A. Lanatà, G. Valenza, and E. P. Scilingo, "Eye gaze patterns in emotional pictures," *J. Ambient Intell. Humaniz. Comput.*, vol. 4, no. 6, pp. 705–715, Dec. 2013.

[17] R. N. Khushaba, C. Wise, S. Kodagoda, J. Louviere, B. E. Kahn, and C. Townsend, "Consumer neuroscience: Assessing the brain response to marketing stimuli using electroencephalogram (EEG) and eye tracking," *Expert Syst. Appl.*, vol. 40, no. 9, pp. 3803–3812, Jul. 2013.

[18] R. W. Picard, "Affective Computing," *MIT Press*, no. 321, pp. 1–16, 1995.

[19] M. Pantic, A. Nijholt, A. Pentland, and T. S. Huang, "Human-Centred Intelligent Human-Computer Interaction (HCI^2): how far are we from attaining it?," *Int. J. Auton. Adapt. Commun. Syst.*, vol. 1, no. 2, pp. 168–187, 2008.

[20] Z. Zeng, M. Pantic, G. I. Roisman, et al, and T. S. Huang, "A survey of affect recognition methods: audio, visual, and spontaneous expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 1, pp. 39–58, Jan. 2009.

[21] M. Pantic, A. Pentland, A. Nijholt, and T. Huang, "Human Computing and Machine Understanding of Human Behavior: A Survey," 2006.

[22] Beckman Institute, "Multimodal Human Computer Interaction towards a Proactive Computer," 2001.

[23] R. Ishii, Y. Shinohara, I. Nakano, and T. Nishida, "Combining Multiple Types of Eye-gaze Information to Predict User's Conversational Engagement," *Hum. Factors*, pp. 1–8, 2011.

[24] A. Jaimes and N. Sebe, "Multimodal human–computer interaction: A survey," *Comput. Vis. Image Underst.*, vol. 108, no. 1–2, pp. 116–134, Oct. 2007.

[25] M. Pantic, N. Sebe, J. F. Cohn, and T. Huang, "Affective Multimodal Human-Computer Interaction," *Proc. 13th Annu. ACM Int. Conf. Multimed.*, pp. 669–676, 2005.

[26] R. Sharma, V. I. Pavlovic, and T. S. Huang, "Toward multimodal human-computer interface," *Proc. IEEE*, vol. 86, no. 5, pp. 853–869, May 1998.

[27] A. Kleinsmith and N. Bianchi-Berthouze, "Recognizing Affective Dimensions from Body Posture," in *Affective Computing and Intelligent Interaction*, 2007, pp. 48–58.

[28] N. Savva, A. Scarinzi, and N. Bianchi-Berthouze, "Continuous recognition of player's affective body expression as dynamic quality of aesthetic experience," *IEEE Trans. Comput. Intell. AI Games*, vol. 4, no. 3, pp. 199–212, 2012.

[29] S. Asteriadis, K. Karpouzis, and S. Kollias, "The Importance of Eye Gaze and Head Pose to Estimating Levels of Attention," *2011 Third Int. Conf. Games Virtual Worlds Serious Appl.*, no. September 2015, pp. 186–191, 2011.

[30] Z. Zeng, J. Tu, B. M. Pianfetti, and T. S. Huang, "Audio-visual affective expression recognition through multistream fused HMM," *IEEE Trans. Multimed.*, vol. 10, no. 4, pp. 570–577, 2008.

[31] N. Bianchi-Berthouze, W. W. Kim, and D. Patel, "Does Body Movement Engage You More in Digital Game Play? and Why?," in *Affective Computing and Intelligent Interaction*, 2007, pp. 102–113.

[32] S. D'Mello and A. Graesser, "Dynamics of affective states during complex learning," *Learn. Instr.*, vol. 22, no. 2, pp. 145–157, 2012.

[33] M. Csikszentmihalyi and M. Csikzentmihaly, "Flow: The psychology of optimal experience," vol. 41, no. May, 1991.

[34] M. Csikszentmihalyi, "The Flow of Creativity," *Creat. Flow Psychol. Discov. Invent.*, pp. 107–126, 1996.

[35] R. C. Wilson, A. Shenhav, M. Straccia, and J. D. Cohen, "The Eighty Five Percent Rule for optimal learning," *Nat. Commun.*, vol. 10, no. 1, pp. 1–9, 2019.

[36] M. Taheri, D. Brown, P. Standen, N. Sherkat, G. Cosma, and C. Langensiepen, "State Diagram for Affective Learning in an Educational Platform," pp. 4–6, 2018.

[37] M. Csikszentmihalyi, "Creativity: Invention, Flow and the psychology of discovery and Invention," in *Personnel Psychology*, vol. 51, no. 3, 1997, p. 16.

[38] P. H. Mirvis and M. Csikszentmihalyi, "Flow: The Psychology of Optimal Experience," *Acad. Manag. Rev.*, vol. 16, no. 3, p. 636, 1991.

[39] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Boredom, engagement and anxiety as indicators for adaptation to difficulty in games," no. January 2008, p. 13, 2008.

[40] M. Taheri, D. Brown, P. Standen, N. Sherkat, G. Cosma, and H. Street, "State Diagram for Affective Learning in an Educational Platform," pp. 4–6, 2018.

[41] A. R. Basawapatna, A. Repenning, K. H. Koh, and H. Nickerson, "The zones of proximal flow," *Proc. ninth Annu. Int. ACM Conf. Int. Comput. Educ. Res. - ICER '13*, p. 67, 2013.

[42] B. Carpenter, S. Rose, H. Rawson, and J. Egerton, "The rules of engagement," *SEN Mag.*, no. 54, pp. 34–54, 2011.

[43] B. Carpenter *et al.*, *Engaging Learners with Complex Learning Difficulties and Disabilities: A resource book for teachers and teaching assistants*. Taylor & Francis, 2015.

[44] R. Iovannone, G. Dunlap, H. Huber, and D. Kincaid, "Effective Educational Practices for Students with Autism Spectrum Disorders," *Focus Autism Other Dev. Disabil.*, vol. 18, pp. 150–165, 2003.

[45] M. J. Lee and B. Ferwerda, "Personalizing online educational tools," *Humaniz. 2017 - Proc. 2017 ACM Work. Theory-Informed User Model.*

Tailoring Pers. Interfaces, co-located with IUI 2017, pp. 27–30, 2017.

[46] C. R. Henrie, L. R. Halverson, and C. R. Graham, "Measuring student engagement in technology-mediated learning: A review," *Comput. Educ.*, vol. 90, pp. 36–53, 2015.

[47] A. K. Wiggins. J. B. Grafsgaard. J. F. Boyer. K. E. E. N. Lester. J. C. Vail, "The Affective Impact of Tutor Questions: Predicting Frustration and Engagement," *Int. Educ. Data Min. Soc.*, 2016.

[48] S. Aslan *et al.*, "Investigating the Impact of a Real-time, Multimodal Student Engagement Analytics Technology in Authentic Classrooms," in *Conference on Human Factors in Computing Systems - Proceedings*, 2019.

[49] J. Hamari, D. J. Shernoff, E. Rowe, B. Coller, J. Asbell-Clarke, and T. Edwards, "Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning," *Comput. Human Behav.*, vol. 54, pp. 170–179, 2016.

[50] P. H. Mirvis, "Flow: The Psychology of Optimal Experience," *Acad. Manag. Rev.*, vol. 16, no. 3, pp. 636–640, 1991.

[51] "Engaging children with complex learning difficulties and disabilities in the Primary Classroom. Barry Carpenter," no. 2002, 2011.

[52] J. C. Abrams, "On learning disabilities: Affective considerations," *J. Reading, Writing, Learn. Disabil. Int.*, vol. 2, no. 3, p. 190, Jan. 1986.

[53] B. Hiebert, B. Y. Wong, and M. Hunter, "Affective influences on learning disabled adolescents," *Learn. Disabil. Q.*, vol. 5, no. 4, pp. 334–343, 1982.

[54] R. R. McCrae, "Consensual validation of personality traits: Evidence from self-reports and ratings," *J. Pers. Soc. Psychol.*, vol. 43, no. 2, pp. 293–303, 1982.

[55] D. Watson and L. A. Clark, "Self- Versus Peer Ratings of Specific Emotional Traits: Evidence of Convergent and Discriminant Validity," *J. Pers. Soc. Psychol.*, vol. 60, no. 6, pp. 927–940, 1991.

[56] J. S. Warm, R. Parasuraman, and G. Matthews, "Vigilance requires hard mental work and is stressful," *Hum. Factors*, vol. 50, no. 3, pp. 433–41, Jun. 2008.

[57] G. J. DuPaul, A. D. Anastopoulos, T. L. Shelton, D. C. Guevremont, and L. Metevia, "Multimethod Assessment of Attention-Deficit Hyperactivity Disorder: The Diagnostic Utility of Clinic-Based Tests," *J. Clin. Child Psychol.*, vol. 21, no. 4, pp. 394–402, Dec. 1992.

[58] H. L. Swanson, "A developmental study of vigilance in learning-disabled and nondisabled children," *J. Abnorm. Child Psychol.*, vol. 11, no. 3, pp. 415–29, Sep. 1983.

[59] W. Peterson, T. Birdsall, and W. Fox, "The theory of signal detectability," *Trans. IRE Prof. Gr. Inf. Theory*, vol. 4, no. 4, pp. 171–212, Sep. 1954.

[60] N. A. Macmillan and C. D. Creelman, *Detection Theory: A User's Guide*. Psychology Press, 1990.

[61] J. A. Swets and R. M. Pickett, *Evaluation of diagnostic systems : methods from signal detection theory*. Academic Press, 1982.

[62] W. P. Tanner and J. A. Swets, "A decision-making theory of visual detection," *Psychol. Rev.*, vol. 61, no. 6, pp. 401–409, 1954.

[63] H. E. Rosvold, A. F. Mirsky, I. Sarason, E. D. Bransome, and L. H. Beck, "A continuous performance test of brain damage.," *J. Consult. Psychol.*, vol. 20, no. 5, pp. 343–350, 1956.

[64] D. H. Sykes, V. I. Douglas, G. Weiss, and K. K. Minde, "Attention in hyperactive children and the effects methylphenidaten (Ritalin)," *J. child Psychol. psychiatry*, vol. 12, no. 2, pp. 129–139, 1971.

[65] "Be first to get a Tobii EyeX Dev Kit Want to develop for Android or Linux ? Get a," vol. 595, p. 795.

[66] I. Software, "Muse : The Brain Sensing Headband Tech Spec Sheet."

[67] "MATLAB - The Language of Technical Computing - MathWorks United Kingdom." (Online). Available: http://uk.mathworks.com/products/matlab/. (Accessed: 07-Mar-2016).

[68] M. Teplan, "Fundamentals of EEG measurement," *Meas. Sci. Rev.*, vol. 2, pp. 1–11, 2002.

[69] "Tobii EyeX Controller – get your own eye tracker." 30-Jul-2015.

[70] "An introduction to Tobii EyeX | Tobii Developer Zone." (Online). Available: http://developer.tobii.com/an-introduction-to-the-tobii-eyex-sdk/. (Accessed: 03-Mar-2016)].

[71] "Specifications for EyeX – Tobii Eye Tracking Support." (Online). Available: https://help.tobii.com/hc/en-us/articles/212818309-Specifications-for-EyeX. (Accessed: 09-Apr-2020).

[72] "Kinect hardware." (Online). Available: https://dev.windows.com/en-us/kinect/hardware. (Accessed: 03-Mar-2016).

[73] Q. Wang, Q. Wang, and G. Kurillo, "Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect Evaluation of Pose Tracking Accuracy in the First and Second Generations of Microsoft Kinect," no. November, 2015.

[74] M. Rosenberg, S. Noonan, J. Degutis, and M. Esterman, "Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task," *Atten. Percept. Psychophys.*, vol. 75, no. 3, pp. 426–39, 2013.

[75] C. Riccio, "The continuous performance test: a window on the neural substrates for attention?," *Arch. Clin. Neuropsychol.*, vol. 17, no. 3, pp. 235–272, Apr. 2002.

[76] S. H. Klee and B. D. Garfinkel, "The computerized continuous performance task: a new measure of inattention.," *J. Abnorm. Child Psychol.*, vol. 11, no. 4, pp. 487–495, 1983.

[77] C. K. Conners, "The computerized continuous performance test," *Psychopharmacol. Bull.*, vol. 21, no. 4, pp. 891–2, Jan. 1985.

[78] C. A. Riccio, J. J. M. Waldrop, C. R. Reynolds, and P. Lowe, "Effects of stimulants on the continuous performance test (CPT): implications for CPT use and interpretation," *J. Neuropsychiatry Clin. Neurosci.*, vol. 13, no. 3, pp. 326–335, Aug. 2001.

[79] "Teacher assessment: using P scales to report on pupils with SEN - Detailed guidance - GOV.UK." (Online). Available: https://www.gov.uk/guidance/teacher-assessment-using-p-scales. (Accessed: 07-Mar-2016).

[80] "P scales: attainment targets for pupils with SEN - Publications - GOV.UK," *Department for Education*. (Online). Available: https://www.gov.uk/government/publications/p-scales-attainment-targets-for-pupils-with-sen. (Accessed: 07-Mar-2016).

[81] T. Affective and O. F. E. Emotion, "Brain – Computer Interface: Fundamentals and Analysis of EEG-Based Emotion Classification."

[82] A. Bulling and T. O. Zander, "Cognition-Aware Computing," *IEEE Pervasive Comput.*, vol. 13, no. 3, pp. 80–83, 2014.

[83] M. Akcakaya *et al.*, "Noninvasive brain-computer interfaces for augmentative and alternative communication," *IEEE Rev. Biomed. Eng.*, vol. 7, no. c, pp. 31–49, 2014.

[84] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–91, Jun. 2002.

[85] M. Ahsan, M. Ibrahimy, and O. Khalifa, "EMG signal classification for human computer interaction: a review," *Eur. J. Sci. …*, vol. 33, no. 3, pp. 480–501, 2009.

[86] I. Daly *et al.*, "On the control of brain-computer interfaces by users with cerebral palsy," *Clin. Neurophysiol.*, vol. 124, no. 9, pp. 1787–1797, 2013.

[87] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans.," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. 51, pp. 17849–54, 2004.

[88] T. O. Zander and C. Kothe, "Towards passive brain-computer interfaces: applying brain-computer interface technology to human-machine systems in general.," *J. Neural Eng.*, vol. 8, no. 2, p. 025005, Apr. 2011.

[89] E. Haselsteiner and G. Pfurtscheller, "Using time-dependent neural networks for EEG classification," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 457–63, Dec. 2000.

[90] Y. P. Lin *et al.*, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1798–1806, 2010.

[91] D. P.-O. O. Bos *et al.*, "Human-computer interaction for BCI games: Usability and user experience," *Proc. - 2010 Int. Conf. Cyberworlds, CW 2010*, pp. 277–281, Oct. 2010.

[92] C. Mühl *et al.*, "Bacteria Hunt," *J. Multimodal User Interfaces*, vol. 4, no. 1, pp. 11–25, Aug. 2010.

[93] C. Mühl *et al.*, "Bacteria Hunt: A multimodal, multiparadigm BCI game," *enterface2009*, no. 1, pp. 1–22, 2009.

[94] [94] T. Egner and M. B. Sterman, "Neurofeedback treatment of epilepsy: from basic rationale to practical application.," *Expert Rev. Neurother.*, vol. 6, no. 2, pp. 247–57, Feb. 2006.

[95] O. M. Bazanova and D. Vernon, "Interpreting EEG alpha activity," *Neurosci. Biobehav. Rev.*, vol. 44, pp. 94–110, 2013.

[96] M. Adjounadi, M. Cabrerizo, I. Yaylali, and P. Jayakar, "Interpreting EEG functional brain activity," *IEEE Potentials*, vol. 23, no. 1, pp. 8–13, 2004.

[97] Trans Cranial Technologies Ltd., "10 / 20 System Positioning Manual," p. 20, 2012.

[98] A. Kubler, E. Holz, T. Kaufmann, and C. Zickler, "A User Centred Approach for Bringing BCI Controlled Applications to End-Users," in *Brain-Computer Interface Systems - Recent Progress and Future Prospects*, 2013.

[99] V. P. Oikonomou, a T. Tzallas, and D. I. Fotiadis, "A Kalman filter

based methodology for EEG spike enhancement," *Comput. Methods Programs Biomed.*, vol. 85, no. 2, pp. 101–8, Feb. 2007.

[100]F. Morbidi, S. Member, A. Garulli, S. Member, C. Rizzo, and S. Rossi, "Application of Kalman filter to remove TMS-induced artifacts from EEG recordings," pp. 1–7.

[101]M. Arnold, W. H. Miltner, H. Witte, R. Bauer, and C. Braun, "Adaptive AR modeling of nonstationary time series by means of Kalman filtering.," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 5, pp. 553–62, May 1998.

[102]A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005.," *J. Neural Eng.*, vol. 2, no. 4, pp. L14-22, Dec. 2005.

[103]V. P. Oikonomou, A. T. Tzallas, S. Konitsiotis, D. G. Tsalikakis, and D. I. Fotiadis, "The Use of Kalman Filter in Biomedical Signal Processing," no. April, 2009.

[104]G. Pfurtscheller, C. Neuper, a Schlögl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters.," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 316–25, Sep. 1998.

[105]L. Trejo *et al.*, "EEG-based estimation of mental fatigue: Convergent evidence for a three-state model," *Found. Augment. Cogn.*, pp. 201–211, 2007.

[106]A. Gevins and M. E. Smith, "Detecting transient cognitive impairment with EEG pattern recognition methods," *Aviat. Sp. Environ. Med.*, vol. 70, no. 10, pp. 1018–1024, Oct. 1999.

[107]T. C. Hankins and G. F. Wilson, "A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight," *Aviat. Space. Environ. Med.*, vol. 69, no. 4, pp. 360–7, Apr. 1998.

[108]D. Ming *et al.*, "Electroencephalograph (EEG) signal processing method of motor imaginary potential for attention level classification," *Proc. 31st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. Eng. Futur. Biomed. EMBC 2009*, vol. 2009, pp. 4347–4351, Jan. 2009.

[109]N. H. Liu, C. Y. Chiang, and H. C. Chu, "Recognizing the degree of human attention using EEG signals from mobile sensors," *Sensors (Switzerland)*, 2013.

[110]T. McMahan, I. Parberry, and T. D. Parsons, "Evaluating Player Task Engagement and Arousal Using Electroencephalography," *Procedia Manuf.*, vol. 3, no. Ahfe, pp. 2303–2310, 2015.

[111]A. Ameera, A. Saidatul, and Z. Ibrahim, "Analysis of EEG Spectrum Bands Using Power Spectral Density for Pleasure and Displeasure State," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 557, no. 1, 2019.

[112]A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, Second. Pearson, 1999.

[113]J. Musson and J. Li, "A Comparative Survey of PSD Estimation Methods for EEG Signal Analysis," *Student Capstone Conf. Proc.*, pp. 1–6, 2010.

[114]N. Bianchi-Berthouze, "The Affective Body Argument in Technology Design," in *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '16*, 2016, pp. 3–6.

[115]A. Kleinsmith and N. Bianchi-Berthouze, "Recognizing Affective Dimensions from Body Posture," in *Affective Computing and Intelligent Interaction*, 2007, pp. 48–58.

[116]J. Farley, E. F. Risko, and A. Kingstone, "Everyday attention and lecture retention: The effects of time, fidgeting, and mind wandering," *Front. Psychol.*, vol. 4, no. SEP, p. 619, Sep. 2013.

[117]M. D. Smith and M. S. Barrett, "The Effect of Parent Training on Hyperactivity and Inattention in Three School-Aged Girls with Attention Deficit Hyperactivity Disorder," *Child Fam. Behav. Ther.*, vol. 24, no. 3, pp. 21–35, Oct. 2002.

[118]C. K. Conners, L. Eisenberg, and A. Barcai, "Effect of Dextroamphetamine on Children," *Arch. Gen. Psychiatry*, vol. 17, no. 4, p. 478, Oct. 1967.

[119]P. L. Holborow and P. S. Berry, "Hyperactivity and Learning Difficulties," *J. Learn. Disabil.*, vol. 19, no. 7, pp. 426–431, Aug. 1986.

[120]E. Benedetto-Nasho and R. Tannock, "Math computation, error patterns and stimulant effects in children with Attention Deficit Hyperactivity Disorder," *J. Atten. Disord.*, vol. 3, no. 3, pp. 121–134, Oct. 1999.

[121]S. L. Smith, *No easy answer : the learning disabled child at home and at school*. Bantam Books, 2012.

[122]B. Fenesi, K. Lucibello, J. A. Kim, and J. J. Heisz, "Sweat So You Don't Forget: Exercise Breaks During a University Lecture Increase On-Task Attention and Learning," *J. Appl. Res. Mem. Cogn.*, vol. 7, no. 2, pp. 261–269, 2018.

[123]S. Lis, N. Baer, C. Stein-En-Nosse, B. Gallhofer, G. Sammer, and P. Kirsch, "Objective measurement of motor activity during cognitive performance in adults with attention-deficit/hyperactivity disorder," *Acta Psychiatr. Scand.*, vol. 122, no. 4, pp. 285–294, 2010.

[124]A. M. F. Zinno, G. Douglas, S. Houghton, V. Lawrence, J. West, and K. Whiting, "Hyperactivity Disorder (ADHD) during computer video game play," *Br. J. Educ. Technol.*, vol. 32, no. 5, pp. 607–618, 2001.

[125]N. Bianchi-Berthouze and Nadia Bianchi-Berthouze, "Understanding the role of body movement in player engagement," *Human-Computer Interact.*, vol. 28, no. 1, pp. 40–75, 2013.

[126]J. A. Vigo, "Eye Gaze Tracking for Tracking Reading Progress," p. 100, 2013.

[127]P. Fattahi, "Eyetracking Definitions," 2011.

[128]L. Hoste, B. Dumas, and B. Signer, "Mudra: A Unified Multimodal Interaction Framework," 2011, pp. 97–104.

[129]H. Zhang, Y. Zhu, J. Maniyeri, and C. Guan, "Detection of variations in cognitive workload using multi-modality physiological sensors and a large margin unbiased regression machine.," *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2014, pp. 2985–8, Jan. 2014.

[130]K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *Neuroimage*, vol. 65, pp. 167–175, 2013.

[131]S. Hor and M. Moradi, "Scandent Tree: A Random Forest Learning Method for Incomplete Multimodal Datasets," Springer, Cham, 2015, pp. 694–701.

[132] "Mobile Pedagogical Assistant to develop meaningful pathways to personalised learning." (Online). Available: https://pedagogics-pathway.eu/. (Accessed: 09-Apr-2020).