




Article

An intra-subject approach based on the Application of HMM to Predict Concentration in Educational Contexts from Non-intrusive Physiological Signals in real-world situations

Ana Serrano * ¹ , Miguel Arevalillo-Herráez ²  and Jesus G. Boticario¹ 

¹ UNED, Madrid, Spain; aserrano@dia.uned.es; jgb@dia.uned.es

² Universitat de València, Valencia, Spain; miguel.arevalillo@uv.es

* Correspondence: aserrano@dia.uned.es;

Abstract: Emotion recognition is becoming very relevant in educational scenarios, since previous research has proven the strong influence of emotions on the student's engagement and motivation. There is no standard method for stating student's affect, but physiological signals have been widely used in educational contexts. Physiological signals have been proved to offer high accuracy in detecting emotions because they reflect spontaneous affect-related information, and which is fresh and do not require an additional control or interpretation. However, most proposed works use measuring equipment that limit its applicability in real-world scenarios because of their high cost and their intrusiveness. Expensive material means an economic challenge for schools and reduce the scalability of the experiments. Intrusive equipment can be uncomfortable for the students which can lead to errors in the collected data. In this work, we analyse the feasibility of developing a low-cost non-intrusive device that integrates easy-to-capture signals that guarantee high detection accuracy. The advantage of the approach also lies in using user's centred information sources (intra-subject) in real-world situations, which provide better detection accuracy, by offering models adapted to each subject. To this end, we present an experimental study that aims to explore the potential application of Hidden Markov Models (HMM) to predict the concentration state from 4 commonly used physiological signals, namely heart rate, breath rate, skin conductance and skin temperature. We study the multi-fusion of every possible combination of these four signals and analyse their potential use in an educational context in terms of intrusiveness, cost and accuracy. Results show that a high accuracy can be achieved with three of the signals when using HMM-based intra-subject models. However, inter-subject models, which are meant to obtain subject-independent approaches for affect detection, fail at the same task. This work concludes that the implementation of a low-cost wrist-worn device for recognising relevant emotions from each student is possible and open the way to a wide range of practical applications in the context of adaptive learning systems.

Keywords: Affective Computing ; Physiological sensors ; Non-intrusive ; Learner Modelling; User-centred systems

1. Introduction

Affective computing is being considered as a way to improve learning [1]. Furthermore, there is strong evidence coming from previous research that shows that emotions have an important effect on the student's engagement and motivation, and consequently influence learning outcomes [2–4]. Previous studies place engaged concentration as the most prevalent affect in a classroom context [2]. Engaged concentration is a state of engagement with a task such that concentration is intense, attention is focused, and involvement is complete [5]. Thus, an effective detection of the concentration state over

time is a crucial task for adaptive learning systems that aim to take proper actions in order to improve the student's engagement.

However, detecting emotions in educational context is still a challenge. Many previous works have been published to overcome this problem, but there is still not consensus regarding the methods that best suite for recognising a particular emotion. Signals related to autonomic nervous system (ANS) responses have been widely used due to the significant correlation of their changes and emotional stimuli [6] [7]. Thus, changes in heart rate, blood pressure, temperature, breathing, etc are used as features to build models in affect recognition. This kind of signals have been widely used in many contexts [8], including educational ones [9]. The main advantages of physiological signals over other sources of information such as visual resources are: 1) ANS is mostly automatically and involuntarily activated, so since physiological signals are the result of ANS activity, they are fresh information sources and cannot be easily faked; 2) physiological signals are more effective at extracting individual traits from each subject. However, this type of signals present several drawbacks: 1) their features are less generalizable across individuals than others such as the activation of facial muscles; 2) they are more difficult to collect and its collection is usually more intrusive for the subject, which can condition the experiments and their results; and 3) sensors' cost can be relatively more expensive than when using other capturing methods, such as vision-based systems. These aspects have been a limitation to the use of this type of signals in both real interactive environments and typical educational settings.

Because of the above limitations, it is crucial to provide cost-effective approaches for detecting each person's emotional state from signals that offer more information on relevant states in the given context and can be unobtrusively collected. If intrusiveness is reduced, accuracy is maximised and cost is minimised, the use of physiological signals may become a practical alternative, whether on their own, or combined with other detection methods. This justifies the increasing interest in Commercial-Off-The-Shelf (COTS) wearables for physiological monitoring in different contexts [10,11], including education [12]. The key benefit of those devices are their wearability and thus their low intrusiveness. The use of different low-cost wearables that measure daily subject's activity is becoming very popular in our society, especially wristband meant for exercise tracking. However, commercial devices are not yet completely ready for research purposes [13,14]. This is due to different issues such as low sampling rate, low accuracy, and the fact that most of them only provide few signals that are not enough to predict emotions. In this work, we study the feasibility of using four signals that are considered relevant in affect recognition [15] to detect the concentration state, named heart rate, skin conductance, skin temperature and breath rate. These four signals can be easily captured and they could be all be integrated in low cost wearables in the near future.

Some of the previous research in physiological-based affect detection attempted to build subject-independent (inter-subject) models [16–19]. In such works a common model is built by considering all data as if it were coming from the same subject without taking into account the particularities of each subject. This approach can then be used with unseen subjects, which is the main advantage. However, this generalisation of the subject's emotions representation can damage the detection accuracy, especially when using physiological patterns which can vary from subject to subject and across contexts. Another early research approach is to build individual models adapted to each user, known as subject-dependent (intra-subject) models. The benefits of the subject-dependent models have recently been proved in several cases such as Electroencephalography (EEG) signals in [20–22] or keyboard and mouse signals in [23]. This approach offers better detection accuracy, by offering models adapted to each subject. Unlike the subject-independent models, because of their single subject oriented nature, the resulting model can not be used for unseen subjects but the approach does, as it provides consistent results across different subjects. However, it requires collecting a larger amount of data per subject, which implies following an intra-subject experimentation approach.

Our aim is to advance methodological issues and open the way towards the development of new commercial devices that can be integrated in affect-aware user-centred adaptive systems in real-world educational scenarios. With this in mind, we study the potential use of intra-subjects and inter-subject modelling approaches when they are combined with multimodality fusion methods and fed with a

common set of physiological signals that are relatively easy to capture by using non-intrusive low cost devices. We focus on concentration as one representative state of interest from an educational perspective, and use Hidden Markov Models (HMM) in order to capture the dynamic nature of a cognitive state that builds up along the time axis. HMMs have been widely used in speech synthesis and lately in facial expression synthesis, and have been also recently proposed to be used with physiological signals in affect recognition [24,25] because of their ability to model non stationary signals or events [25]. The results of our analysis suggest that a reasonable accuracy can only be obtained when intra-subject models are used, and outline the need to intensify research efforts on techniques that enable the automatic construction of subject models.

The rest of this paper is organised as follows. Section 2 describes some previous work that relates to the methods and results presented along the paper. Section 3 focuses on the experimental design, also covering the data collection, data labelling and data pre-processing steps. Section 4 reports the experimental results which consist of evaluating the intra-subject and inter-subject models and compare the performance of different multimodal fusion of the four physiological signals. Section 5 provides a discussion of the results. Finally, conclusions and future research activities are drawn in Section 6.

2. Related Work

The multiple reviews on the affect recognition field, e.g. [8,26–29], are a clear indicator of the importance of the topic and the level of research activity that has been reached. Affect recognition has been successfully applied to marketing [30,31], health [32,33] and more recently to learning systems [34,35], including methods to automatically recognise student's engagement [36].

Most of the previous works focus on less accurate sensor-free approach to restrict intrusiveness and reduce costs, as in [37–41]; or are based on visual sources of information, such as facial expressions or eye-tracking, e.g. [18,36,42,43]. Such sources have the advantage that they generalise well between users, and enable the construction of inter-subject models. Many other works rely on a multimodal setup. According to D'mello and Kory [44] affect recognition systems based on multimodal data tend to be almost 10% more accurate than their unimodal counterparts. However sensor-based systems tend to be more intrusive and costly and then their applicability is reduced in a real-world educational scenario.

In this review section we focus on previous works that use physiological data for affect recognition. Feature selection is a crucial step to accurately recognise affective states using physiological signals. Furthermore, the obtrusiveness of the system can also influence detection accuracy. Thus, we have not only focused on those works that study student engagement, but in a set of works that use physiological signals to recognise different emotions through different contexts. There are many between-study differences in the literature review in terms of a number of factors that makes difficult to draw broad conclusions. We have selected the following factors to compare them:

- Target emotion. This factor describes the emotion to be detected by the system proposed in the related work. Both categorical and dimensional models are considered in this factor.
- Physiological signals that were used to model the affective state of the user. Different combinations of physiological signals are used across the related works, some of them are easier to capture and less intrusive than others in a real-world scenario, such as those that can be obtained from the subjects wrist while others require more intrusive devices as a headset or a clamp in the finger.
- Methods/classifiers used. Different classifiers have been used across related works.
- Highest accuracy obtained in the study. When the study compares different approaches, such as different signals, modelling approaches or classifiers, the best accuracy reported is included in this review.
- Dependency of the model on the subject. This indicates whether the study follows an intra-subject or inter-subject approach.

- Number of subjects involved in the study. The number of subjects is especially important for inter-subject approaches in order to manage better generalisation results. In contrast, intra-subject approaches require a big amount of data from the same individual while having just one subject involved would be enough.
- Labelling of emotions. This indicates if the labelling is obtained through a self-report of the user, determined from the stimulus used to elicit emotions or it is otherwise labelled by an expert.

Table 1 shows a summary of the literature survey of recent works in Affect Recognition using Physiological Signals indicating the aforementioned factors for each case.

2.1. Emotions representation

Regarding emotion representation, some works used a dimensional representation whereas others used a categorical one. Within the ones that chose a dimensional representation [45–47], all of them use a common approach that considers the two dimensions proposed by Russell [48], valence and arousal, as the two main dimensions to represent affect. On the other side, many works that use a categorical definition of the affective states are focused on the typically considered as basic emotions [49], such as anger, surprise, happiness, disgust, sadness, and fear [50–53]. However, as stated in [54], these basic emotions are quite infrequent in educational contexts when using learning software tools. Instead, there are more common affective states that the user experiment during the learning tasks, such as boredom, engaged concentration, anxiety, confusion and frustration. Two works [55,56], and the one presented in this paper, are focused on nonbasic states, which share only few features commonly attributed to basic emotions, and that are consequently considered more challenging to be detected. The number of affective states to be recognised is also an important factor as the detection problem becomes more challenging when the number of states increases.

2.2. Physiological signals selected

Since selecting an appropriate set of features is crucial in affect recognition, an important issue is to select which physiological signals are used in order to identify the correct measures that helps to discriminate between affective states. Different physiological measures are used in the reviewed works: Galvanic Skin Response (GSR), also named as Skin Conductance (SC), ElectroDermal Activity (EDA) or ElectroDermal Response (EDR); heart rate variability (HRV), which can be obtained from electrocardiography (ECG); electroencephalographs (EEG); Skin Temperature (ST); Breath Rate (BR); Electromyogram (EMG); and Blood Oxygen saturation (OXY). Among them, there are some signals that are not easy to capture, at least in a non-intrusive manner, such as EEG, EMG, BVP and OXY. Some researchers centred their works on the use of the features obtained from just one signal to minimise intrusiveness, as in [57]. In that work the use of off-the-shelf wearable devices that measures EDA to monitor the emotional engagement of students during lectures is studied. Furthermore authors investigate the physiological synchrony between teachers and students and how it helps to infer students' emotional engagement. Another example is [50] that aims to provide a feature extraction method for a user-independent emotion recognition system from electroencephalograms (EEGs). However, EEG pose strong limitations on the movement of the subjects and, hence, is not really applicable in real-world scenarios. The rest of the works use a set of features obtained from more than one physiological signals, being GSR and ECG the most commonly used in the selected works. Both [55,57] takes the intrusiveness of the physiological sensors in consideration in their works. In [55] they conclude that intrusive physiological sensors used may have influenced their study affecting the way students engaged with the ITS they used. The system proposed in [57] follows a non-intrusive approach and aims at continuously gathering sensor data in classroom settings through a wrist-worn and a lightweight device. The rest of the selected works make use of traditional sensors without taking in account their obtrusiveness. Thus, despite the progress made so far in the previous works it is safe to say that there is still a lot of work to do before affect recognition can be integrated into everyday interfaces and can be more readily deployed into real-world contexts. In this work we aim to analyse

the use of different non intrusive sensors towards the integration of an affect detection system in real-world educational scenarios while not influencing the student behaviour. Thus, the present study uses a subset of the aforementioned signals, namely, heart rate, breath rate, skin conductance and skin temperature and includes a discussion about the accuracy results obtained with different combinations of them.

2.3. Subject dependency

Some of the selected works focused on building user-independent models for physiological-based affect recognition. This is the case of most of them [45,47,50,53,56,57], that aimed to build models that have the potential to generalise to new users as the main advantage. Within these works the best accuracy were obtained when classifying basic emotions, as it happens in [47] where a 90% of accuracy is obtained when discriminating among 5 levels of arousal and valence. In contrast, a lower accuracy is obtained when detecting nonbasic emotions, as it happen in [56] where a maximum of 63% of accuracy is obtained when detecting boredom, engagement and anxiety. This lower accuracy is due to the individual differences in affective experience and expression of nonbasic emotions. Other works as [51] and [52] adopt a user-dependent approach. Two of the selected works, [46,55], compares both approaches. Both of them conclude that a better accuracy is obtained with subject-dependent approaches. In [55] they aim to detect nonbasic emotions such as boredom, confusion, curiosity, delight, flow /engagement, surprise, and neutral, resulting with engagement, boredom and neutral as the best scored in the user-dependent models[2,5,37]. In [46] the classification of four bidimensional emotions (positive/high arousal, negative/high arousal, negative/low arousal, and positive/low arousal) which are induced by music is studied. It is important to point that in this work they used within-participants cross validation for the user-independent approach. That means that there were data from the same user in both the training and testing sets. Classification accuracy is expected to decrease from the one reported, CCR=77%, if a between-subject validation technique is applied, which means that the data from one user appears only in the train or the test set.

2.4. Other aspects

There are other aspects shown in Table 1 that mark some differences between the selected works and that will be summarised in this subsection. There are also many others that could be pointed out but they are out of the scope of this work such as the method used for emotions elicitation, the location of the experiment (natural context vs lab context), and so on.

The number of participants in the study is an important factor to take into account especially for the inter-subject approaches, because the more subjects the more generalizable the obtained models will be. In the case of the intra-subject approaches, it is very important to record data from the same use over time and in different situations, in order to be able to build reliable individual models of each user. The number of subjects involved in the selected work vary from 1 for an intra-subject approach in [51] to 101 for an inter-subject approach in [53].

Regarding classification methods a wide variety of techniques have been used for classifying affective physiological data. These include k-nearest neighbours, decision trees and SVM to name a few. According to related work there are not consensus regarding which one of the algorithms is the most efficient. In this work we propose the use of HMM, which has been previously used for emotion detection with EEG signals [24,25].

Regarding the data annotation, most of the works make use of a self-assessment of the emotion experienced by the subject. Some other works use stimuli to elicit subject's emotions, such as music [46,51] or images [47], and the labelling of emotions are objectively determined by the stimulus type. In [55], which used self-reported judgements, they conclude that the interpretation of the labels may vary among individuals and that results obtained could improve if such self-reported annotations were complemented with judgement from observers or experts. In our study we used this approach combining both self-reports and the judgement of two experienced experts.

Table 1. Summary of the main points of literature Review

Ref	Emotions	Signals ^a	Algorithm ^b	Best. Acc	Dep	N	Labelling
[57]	Engament/Arousal	GSR	SVM	81%	inter	24	self-report
[45]	Arousal / Valence	EEG, HRV (from ECG)	SVM	75%	inter	60	self-report
[55]	boredom, confusion, curiosity, delight, engagement, frustration, surprise, and neutral	ECG, EMG, GSR	SVM, K-nearest, NB, LBNC, LR, C4.5	inter: F1=0.41 intra: F1=.62	intra & inter	27	self-report
[56]	boredom, engagement, anxiety	EEG, GSR, BVP, ST, BR	LDA, QDA, SVM	63%	inter	20	self-report
[50]	happiness, surprise, anger, fear, disgust, and sadness	EEG	QDA, k-nn, MD, SVM	85.17%	inter	16	self-report
[51]	joy, anger, sadness and pleasure.	ECG, EMG, GSR, RSP	C4.5	80%	intra	1	det. by stimulus
[52]	amusement, fear, sadness, joy, anger, and disgust	ECG, EMG	MLP, SVM, BN	98%	intra	100	self-report
[53]	amusement, anger, grief and fear	OXY, GSR, HR	Random Forest	74%	inter	101	self-report
[46]	positive/high arousal, negative/high arousal, negative/low arousal, positive/low arousal	EMG, ECG, GSR, BR	LDA	inter: CCR 77 intra: CCR 95	intra & inter	3	det. by stimulus
[47]	5 leves of aurosal and 5 leves of valence	ECG,GSR, BR	QDC	90%	inter	35	det. by stimulus

^a Signals: (GSR) Galvanic Skin Response; (EEG) Electroencephalogram; (HRV) Hear Rate Variability; (ECG) Electrocardiogram; (EMG) Electromyograph; (BVP) Blood Volume Pulse; (ST) Skin Temperature; (BR) Breath Rate; (OXY) Blood Oxygen Saturation; (HR) Hear Rate.

^b Algorithms: (SVM) Support Vector Machines; (NB) Naive Bayes; (LBNC) Linear Bayes Normal Classifier; (C4.5) C4.5 Decission Trees; (LR) Multinomial Logistic Regression; (LDA) Linear Discriminant Analysis; (QDA) Quadratic Discriminant Analysis (QDA); (MD) Mahalanobis Distance (MD); (MLP) Multi-Layer Perceptron; (BN) Bayesian Network; (QDC) Quadratic Discriminant Classifier.

3. Experimental design

3.1. Physiological signals acquisition

Physiological signals have been widely used to recognise emotions. Some of the signals most commonly used are: blood pressure, skin conductance, galvanic skin response, respiration rate and heart rate variability. Some of them are easier to capture than others in a real-world scenario. The non-intrusiveness of the acquisition system is crucial in an educational context when we want to study the student's behaviour. There are sensors that are too intrusive, and may disturb the correct performing of the learning task. For example, those that have to be wore in the fingers prevents the student to use a keyboard or mouse, because typing or clicking can provoke signal distortions. Furthermore, the behaviour of the student can be conditioned when using uncomfortable wearables.

In order to avoid the aforementioned problems, in this work we have selected four signals that can be gathered from sensors that minimise this intrusiveness and that could be easily integrated in a low-cost device. Those signals were already identified in [15] as valuable information to identify changes in the subject's mental state.

The acquisition system, called PhyAs (PHYsiological Acquisition Shield), corresponds to the third iteration of the AICARP (Ambient Intelligence Context-aware Affective Recommender Platform) system, which takes advantage of the environmental intelligence in order to help the learner in managing their emotional state while performing learning tasks [58]. In this version, many improvements were introduced. For instance, the accuracy of certain sensors as the skin conductance and the response time of the temperature where revised. Some other improvements aimed to reduce the obtrusiveness of the system, as the new pulse-oximeter. The selected signals used in this study and how are they acquired to minimise the cost and intrusiveness of the system is detailed here:

- **Heart rate.** This signal, together with Heart Rate Variability (HRV) and other parameters related to cardiac circle are measured in many works with an electrocardiogram (ECG) [45–47,51,52]. In this work this signal is gathered through a pulse-oximeter. The pulse sensor that in previous versions of the PhyAS used to be a clamp in the finger as it happen in many studies, was changed by an ear clip sensor which allows the fingers to be used in writing tests. This sensor can be also placed in the wrist, so it could be easily integrated in a wrist-worn device, but a better response is get from the lobe of the ear. In this version the acquisition system can provide information regarding instantaneous changes of the heart rate, with a similar behaviour to the HRV (Heart Rate Variability) provided by an electrocardiogram which would be much more intrusive. In this study we compute the number of beats per minute (BPM).
- **Breath rate.** A relaxation state is generally linked with a decrease in the respiration rate, while momentary cessations of the respiration are linked to tense situations [46]. During respiration, the thorax expands and constricts. Hence, commonly a piezoelectric chest belt is used to measure the respiration pattern [56] of the subjects as it is the case in this study. Thus, we compute the BR as the number of breath cycles taken within a minute (BCPM).
- **Galvanic Skin Response.** It is very commonly used in affective computing, especially to detect arousal differences. It provides a measure of the resistance of the skin (electrodermal activity or skin conductance) and it is commonly measured at locations with a high density of sweat glands, for example, palm/finger [55], feet [59] or wrist. In some works this sensor is placed in the fingertips [55], which reduce user mobility to develop writing tasks. In the work proposed in this paper, this sensor is placed in the wrist of the subject to enable user mobility when developing writing tasks.
- **Skin temperature.** Changes in the skin temperature can be measured using an infrared light that does not require to be in contact with the participant's skin. This is the case of the present study that uses a sensor placed closed to the wrist. This sensor also allows to measure the temperature in the environment, which is not used within this study.

More details regarding the acquisition system can be found in [15].

3.2. Data collection

A data collection experience was conducted to evaluate the proposed approach to automatically detect concentration patterns from physiological signals. In order to verify that our methodology provides an added value in a practical setting, data collection took place in a real school, while students interacted with a series of learning tasks. The experience was repeated for a set of 4 different users $U = \{u_i, 1 \leq i \leq 4\}$.

The PhyAS was connected to an Arduino Uno to capture the 4 physiological signals at a rate of 10Hz. At the same time the student interaction was recorded in video, within a framework that was built to support tracking and labelling from a single-subject experiment, including on-site and offline data labelling. A wider explanation of such framework can be found in [60]. The video and physiological signals recordings were synchronised by using the system's clock.

The construction of an individual model per subject usually presents a challenge because it is assumed that requires a large amount of data for each person [20]. However, having the limitations imposed on a real-world learning scenario, where the amount of data from a single user is usually limited [23], we have come up with an affordable setting based on a small number of short-term sessions per user. In particular, to have enough information to build intra-subject models, each student participated in 4 sessions. A different learning activity was carried out in each session. The dataset was hence composed of 4 recordings for each of the 4 subjects involved in the experiment. All experiments of this work were carried out following the rules of the Declaration of Helsinki of 1975 [61], and each participant was asked to sign a consent form. This consent form was given to them in advance so that they could read it without any external pressure, and contained information about the aim of this research work, along with a complete description of the experience. When they signed, they confirmed that they had read the complete form, understood all the information contained therein and authorised us to conduct the experience, including the exploitation of the data gathered during the session.

There were two different types of sessions. The first three ones were focused on detecting the affective state of the user, and the fourth included an additional task where users got feedback when they entered a state of excessive agitation that would hamper their performance. These followed our previous approach [60].

The first two sessions consisted of a series of math exercises with an increasing level of difficulty. The third session included a series of logic exercises that the student had to solve. These first three sessions involved tracking the students with the aforementioned sensors while they were performing the requested tasks using a keyboard and a mouse. The fourth session was an oral test in a second language (English). In this fourth session, feedback to the learner is introduced from a band placed on their wrist that vibrated when the user entered a zone of nervous excitement close to stress.

The first three sessions focused on detecting the affective state and followed a well-defined protocol in order to assure that every session was carried out under the exact same conditions. Such protocol defines all the steps involved in the session. Thus, the session started by entering the work space of a web browser with the corresponding user and configuring the activity for the day. After that, the infrastructure for the experiment was launched. Then, the participant was welcome and informed about the experience. Next, the consent form was signed, and the physiological sensors were placed and tested to verify that they were correctly collecting data. Once everything was ready, the data collection started and lasted 9 minutes, which were distributed as follows: 2 minutes with no activity, 5 minutes solving exercises and 2 more minutes with no activity. During the 2 minutes of recording before and after the exercise session, the participant was asked to just relax. The purpose of these recordings without activity was to obtain a baseline for the participants. Having a baseline of the subject helps to model their physiological state when they are not cognitively involved. When the data collection was finished, the sensors were removed from the participant, and the recording was stopped. Finally, the data was tagged with the support of the participant.

The fourth session, an oral test in a second language, consisted of a voice baseline and two oral activities of 10 minutes each with increasing level of difficulty. The students took them in turn. The session started with a brief presentation explaining the experiment to the participant and confirming their consent for the use of the data. The 10 minutes of each activity were distributed as follows: 2 minutes with the participant reading a text to record the baseline voice, 1 minute for the preparation of the test, 5 minutes of oral presentation and 2 more minutes with no activity. Thereupon each activity, the sensors were removed, and the experts labelled the collected data with the participant. Between the first and second activities, the action rule parameters that control the aforementioned band were obtained while the subject completed the General Self-Efficacy Scale (GSE) questionnaire [62] and the performance was configured in the recorder. Besides, after the second activity, the Personal Data questionnaire, the Big Five Inventory (BFI) [63] and the System Usability Scale (SUS) [64] questionnaires were completed. The BFI questionnaire creates a participant's profile regarding what psychologists define as the five broad dimensions of personality traits (extraversion, agreeableness, conscientiousness,

neuroticism and openness to experience). The SUS questionnaire allows the participant to value the experience environment and was used to evaluate the usability of the infrastructure.

3.3. Data labelling

The video recordings were used to manually label the data. The procedure followed for the labelling was the same as in [65]. Two different experts labelled the data set independently. One expert had a psychology background, and the other had 6 years of experience in the affective computing research field. They identified specific moments in the video where the user seemed to have reached one of a set of specific mental states, and used the application described in [60] to mark the concrete instant in time where they believed the peak was. Both experts labelled the complete dataset independently and then discussed any disagreement regarding a particular label at a given time. After the labelling consensus, a validation meeting was held with each participant, correcting the labelling where appropriate.

3.4. Data preparation

Out of the labels provided by the experts, we have focused our study on the concentration state. This is because of its relation to engagement, and also because this label has shown the highest reporting agreement between the experts and the subjects.

As it happens with most other mental states, concentration does not happen all of a sudden. We hence can reasonably assume concentration on a time frame of 6 seconds around the identified peaks. Based on this hypothesis, we took each time mark reported as a concentration peak by the experts and isolated the physiological measurements of the 3 seconds previous and following to it. With a capture rate of 10Hz, this yielded a matrix of size 4×60 (4 physiological signals \times 60 measurements per signal) for each concentration label reported, containing information about how the signals evolved until the concentration peak was reached. We call each of these matrices a concentrated sample.

In order to generate non-concentrated samples, we discarded the first second of the physiological signals and chose 6 seconds disjoint slots that were at least two minutes apart from any identified concentration peak. We also guaranteed a minimum of a 6 seconds separation space between any two non-concentrated samples. The number of positive and negative samples obtained per user by using this procedure is shown in Table 2.

Table 2. Number of samples per subject

	Concentrated samples	Non-concentrated Samples
subject 1	31	29
subject 2	9	20
subject 3	7	19
subject 4	13	49

3.5. Description of experiments

There were two issues to analyse through the experimentation. First, the validity of inter-subject and intra-subject models to predict concentration. Second, the relevance of the selected physiological features for concentrated state assessment, while studying different combinations of them. In order to analyse both issues, a diversity of experiments were carried out.

3.5.1. User-dependent VS user-independent models

In this first set of experiments we aim to study the performance of both inter-subject and intra-subject approaches at detecting concentration in an educational scenario, by using Hidden Markov Models (HMM) that are trained with physiological data. To this end, we devised 3 different experiments, which were repeated for each of the 4 users in U . In the first two experiments, we used two sets of samples, S^+ and S^- , whose elements differed on whether or not they met a certain criterion

C. We then applied a cross-validation scheme based on the elements in S^+ over 12 500 iterations. For each iteration, we used 75% of the samples in S^+ to train a HMM model. The remaining 25% of the samples were used as part of the test set. The distribution of the samples between training and test was randomly calculated for each iteration.

Each time a HMM model θ was built, we computed the probability $p(X|\theta)$ for each sample in the test set, which was composed of the remaining 25% of the samples in S^+ and all the samples in S^- . The scores produced for each sample in the test set were hence related to the probability that the sample met the specified criterion C.

In order to evaluate the effectiveness of the model to discriminate samples according to whether they met the criterion C, the results from the 12 500 iterations were used together to build a Receiver Operating Characteristic (ROC) curve and compute typical accuracy indicators. The accuracy indicators that were used are the Area Under the ROC Curve (AUC) and the Equal Error Rate (EER). The AUC estimates the probability that the model classifies a random positive sample with a higher rank than a random negative sample. The EER is the point of the ROC curve that corresponds to have an equal probability of miss-classifying a positive or negative sample.

In the first experiment, we aimed to test whether the information contained in the physiological signals of an individual can be used to train a HMM model that is able to detect when the subject is concentrated from an unseen sample of the signals. We repeated the experiment for each subject u_i , always focusing on samples from the same subject. In this case, the criterion C was defined as whether the sample was labelled as concentrated. Thus, S^+ was composed of all concentrated samples for u_i , and S^- contained all non-concentrated samples for the same individual (see definition of concentrated and non-concentrated samples in Section 3.4).

In the second experiment, we studied whether subjects can be easily distinguished from the others by how concentration affects their physiological signals. This time, we only considered concentrated samples and defined the criterion C as whether the sample belongs to a given user u_i . Thus, and again for each u_i , S^+ included all concentrated samples for u_i , and S^- all concentrated samples for the rest of the subjects ($U - \{u_i\}$).

Finally, the third experiment attempted to test an inter-subject HMM model. That is, the viability of accurately predicting concentration for an individual u_i , using data coming from subjects other than u_i . In this case, we followed a slightly different setting. For each user u_i we run a single validation experiment using all concentrated samples from subject u_i as training data. The test set was then built using all concentrated and non-concentrated samples from users in $U - \{u_i\}$. Results were assessed using the same measures as in the previous two experiments.

3.5.2. Performance of physiological signals on engagement assessment

In this second set of experiments we aim to evaluate different multimodal fusion approaches at feature level. Thus, within this set of experiments we wanted to analyse the effectiveness of the use of each physiological signal, or different fusion of them, for the detection of the concentrated state of the subject.

To this end, we run the first two experiments of the previous set across each of the 15 possible combinations of the 4 signals in M , heart rate, skin conductance, breath rate and skin temperature.

$$\sum_{k=1}^M \binom{M}{k} = 2^M - 1$$

Thus, we analysed the performance of each signal separately, as well as the performance of all the possible fusions in an intra-subject approach (one by one $\binom{M}{1}$, in pairs $\binom{M}{2}$ or three by three $\binom{M}{3}$). We repeated each experiment for each subject u_i . We followed the same definitions than in the previous section to run the experiments regarding criterions C to define samples S^+ and S^- , proportion of the samples across training and test sets, number of iterations and the way to compute probability $p(X|\theta)$.

The purpose of this study was to test whether we can keep detecting the concentrated state of the subject without using all of the four selected signals. In this case we wanted to discover which

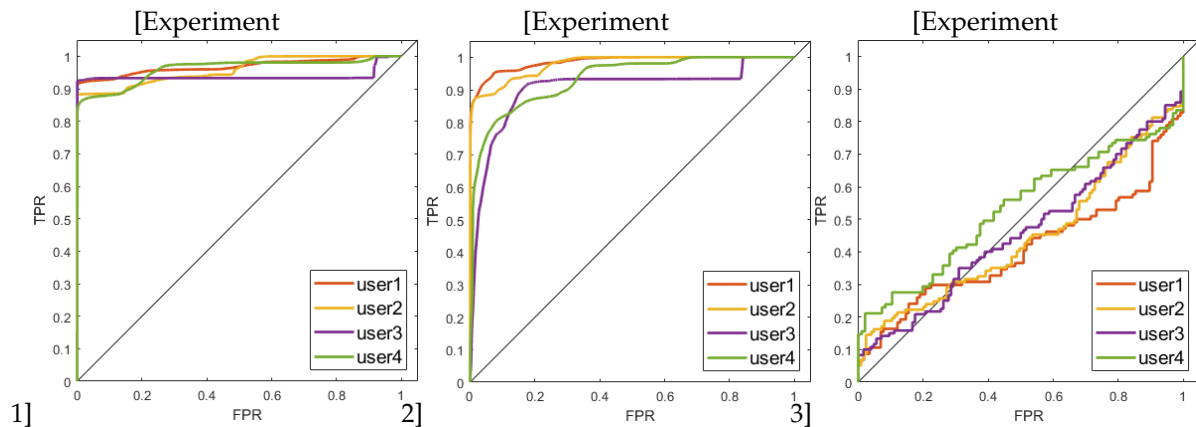


Figure 1. Results for the three experiments carried out represented in ROC curves. Experiment 1: using concentrated samples from u_i as training data and testing on concentrated and non-concentrated samples from the same user. Experiment 2: using concentrated samples from a user u_i and testing on concentrated samples from all subjects in U . Experiment 3: using concentrated samples from all users except u_i for training, and concentrated and non-concentrated samples from user u_i for testing.

multimodal fusion performs better and which of the rest have a significant performance so it could be considered in future works in order to optimise costs and intrusiveness and thus maximise the applicability of the detection system. These empirical results will help to take decisions on the selection of sensors to be integrated for the implementation of Commercial-Off-The-Shelf devices of the near future according to accuracy and cost requirements.

4. Experimental results

4.1. User-dependent VS user-independent models

4.1.1. First Experiment

Fig. 1.a represents the ROC curve for the first of the experiments, in which the model was fit by taking 75% of the concentrated samples from one subject and tested against a set that contained both concentrated and non-concentrated samples of the same subject. For each testing sample X , the probability $p(X|\theta)$ was used as a prediction of whether the user was concentrated. The AUC and EER for each user are shown in Table ?? 3. The lowest AUC is very high, 0.94, and corresponds to the third subject. An average AUC=0.96 suggests that concentration can be predicted in a very accurate way from the physiological signals when the model has been trained with labelled data from the same subject.

The results are consistent across the 4 users under the study and indicate that the HMM model is able to discriminate well between concentrated and non-concentrated samples for the same individual.

Table 3. Accuracy results for the first set of experiments (User-dependent VS user-independent models)

Subject	Experiment 1 (intra)		Experiment 2 (intra)		Experiment 3 (inter)	
	AUC	ERR	AUC	ERR	AUC	ERR
1	0.97	0.08	0.98	0.06	0.41	0.54
2	0.96	0.11	0.98	0.09	0.44	0.53
3	0.94	0.07	0.90	0.13	0.45	0.51
4	0.96	0.11	0.93	0.14	0.53	0.43
Average	0.96	0.09	0.95	0.11	0.46	0.50

4.1.2. Second Experiment

Fig. 1.b shows the results when the model was trained with 75% of the concentrated samples from a user u_i and tested against concentrated samples from all subjects. For each testing sample X , the probability $p(X|\theta)$ was used as a prediction score of whether the sample belonged to the user u_i . The high AUC shown in Table 3, minimum 0.90, and low error rates obtained in all cases clearly indicate that the way concentration reflects on the physiological signals is subject-dependent, to such an extent that from a concentrated sample we can accurately find out which user the sample belongs to.

4.1.3. Third Experiment

The relatively higher accuracy values obtained in the second experiment as compared to those obtained in the first one suggest that the subject's influence in the physiological signals is higher than that caused by the mental state itself. This finding motivated this third experiment to test results when a model is created by using data coming from subjects other than the target. This is, in fact, an inter-subject approach to the detection problem.

Fig. 1.c shows the ROC curve for each user u_i , when the model is trained with concentrated samples from all other users and tested against positive and negative samples from user u_i . AUC and EER values reported in Table 3 for this experiment are close to a random classifier and reinforce the hypothesis that the subject-related component of the physiological signals is stronger than the concentration-related one, and reveal the inadequacy of inter-subject models in this particular context of adaptive systems.

4.2. Performance of physiological signals on engagement assessment

In view of the results obtained in the previous set of experiments and the bad results obtained for inter-subject models (view Section 4.1), the study of the performance for each signal is shown only for the intra-subject models. Thus, in this section we repeat the first two experiments of Section 4.1 to study the performance of each signal in in M as well as every combination of them $\binom{M}{k}$.

4.2.1. First Experiment

Fig. 2 shows the results for the first experiment, which test whether the information contained in the physiological signal or signals gathered from each u_i can be used to detect when the subject is concentrated from an unseen sample of such set of signals. We repeated the experiment for each subject u_i , always focusing on samples from the same subject. In the horizontal axis of the figure are represented the set of signals used in each case, and the accuracy obtained for each user is represented in the vertical axis as well as the average accuracy from the 4 users. Numeric results, including AUC and ERR are detailed in Table 4. The average accuracy results above 0.75, which we consider a relatively high performance that could be enough in some scenarios, have been highlighted with bold text in the table.

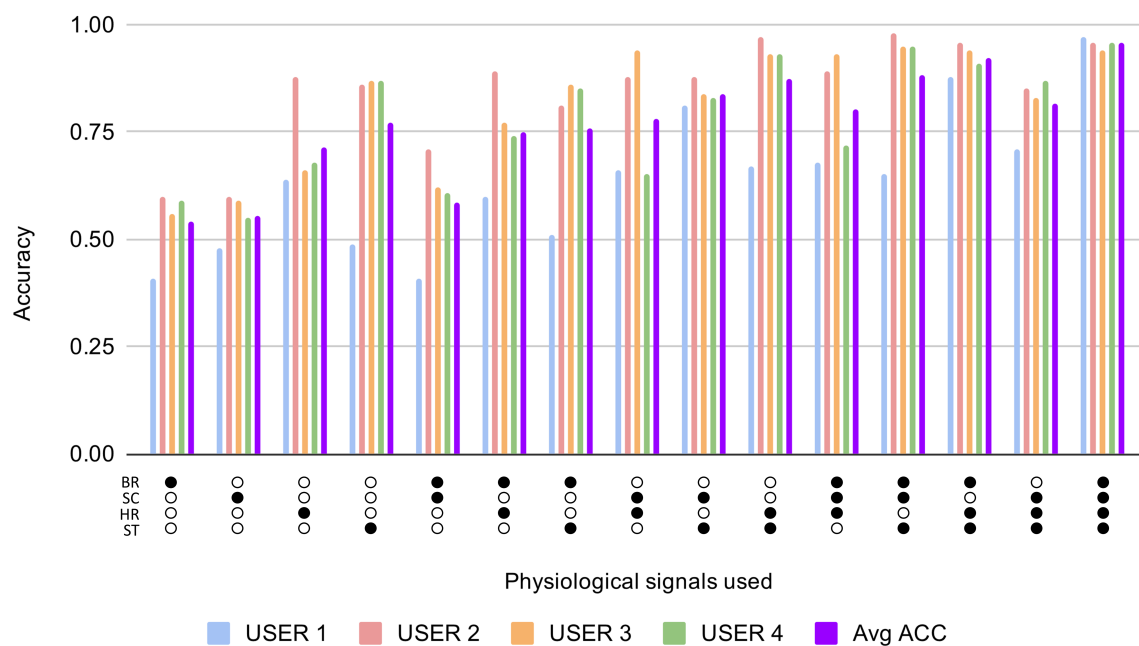


Figure 2. Accuracy results for different combination of the selected signals for Experiment 1: studying the relevance of physiological features to detect when the subject is concentrated from an unseen sample of such set of signals.

(BR) Breath Rate; (SC) Skin Conductance; (HR) Heart Rate; (ST) Skin Temperature

BR	SC	HR	ST	USER 1		USER 2		USER 3		USER 4		Averages	
				ACC	EER	ACC	EER	ACC	EER	ACC	EER	ACC	EER
●				0.64	0.41	0.88	0.17	0.66	0.36	0.68	0.35	0.72	0.32
	●			0.41	0.59	0.6	0.4	0.56	0.43	0.59	0.43	0.54	0.46
		●		0.48	0.52	0.6	0.45	0.59	0.49	0.55	0.48	0.56	0.49
			●	0.49	0.53	0.86	0.2	0.87	0.16	0.87	0.13	0.77	0.26
●	●			0.41	0.56	0.71	0.32	0.62	0.42	0.61	0.38	0.59	0.42
●		●		0.6	0.45	0.89	0.14	0.77	0.21	0.74	0.3	0.75	0.28
●			●	0.51	0.51	0.81	0.21	0.86	0.17	0.85	0.21	0.76	0.28
	●	●		0.66	0.4	0.88	0.18	0.94	0.08	0.65	0.37	0.78	0.26
	●		●	0.81	0.27	0.88	0.15	0.84	0.14	0.83	0.19	0.84	0.19
		●	●	0.67	0.34	0.97	0.09	0.93	0.17	0.93	0.13	0.88	0.18
●	●	●		0.68	0.41	0.89	0.18	0.93	0.1	0.72	0.32	0.81	0.25
●	●		●	0.65	0.4	0.98	0.06	0.95	0.12	0.95	0.11	0.88	0.17
●		●	●	0.88	0.2	0.96	0.11	0.94	0.08	0.91	0.16	0.92	0.14
	●	●	●	0.71	0.38	0.85	0.21	0.83	0.17	0.87	0.19	0.82	0.24
●	●	●	●	0.97	0.08	0.96	0.11	0.94	0.07	0.96	0.11	0.95	0.09

Table 4. Results for the first experiment using different set of signals. (HR) Heart Rate; (BR) Breath Rate; (SC) Skin Conductance; (ST) Skin Temperature

According to results depicted in Fig. 2, when the 4 signals are used separately, skin temperature is the signal that provides more information about the concentrated state of the user for most of them. When we fusion any pair of signals, the average accuracy is more that 75% except in the case when we used breath rate and skin conductance. The pair with the highest performance is the one formed by the signals heart rate and skin temperature. When we combine three signals the behaviour of the model is very similar to the one that we obtained with the total set of the four signals. There is only one fusion of three signals that performs a little bit worse that the rest, and this is the one without

the skin temperature signal. From the obtained results, we can also deduce that the behaviour of the signals across users become more regular when we combine more signals. As expected, the best result obtained in average is the case when using the four signals together. The influence of each signal is also depicted in Fig. 3 which shows the average accuracy obtained across all the combinations when each signal is present. Results for this first experiment are depicted on the left of the figure, where it is shown how skin temperature obtains again the best performance, really close to the one obtained for the heart rate. In summary, from both figures it can be conclude that a high performance, $AUC>0.75$, can be obtained to recognise user concentration in the following cases:

- When the skin temperature signal is used independently or present in any of the multi-fusion cases.
- When any of the following pairs of signals are used: heart rate - breath rate, heart rate - skin conductance, heart rate - skin temperature.
- When any possible combination of three of the four selected signals is used.
- When the four signals are used as it was already confirmed in previous sections.

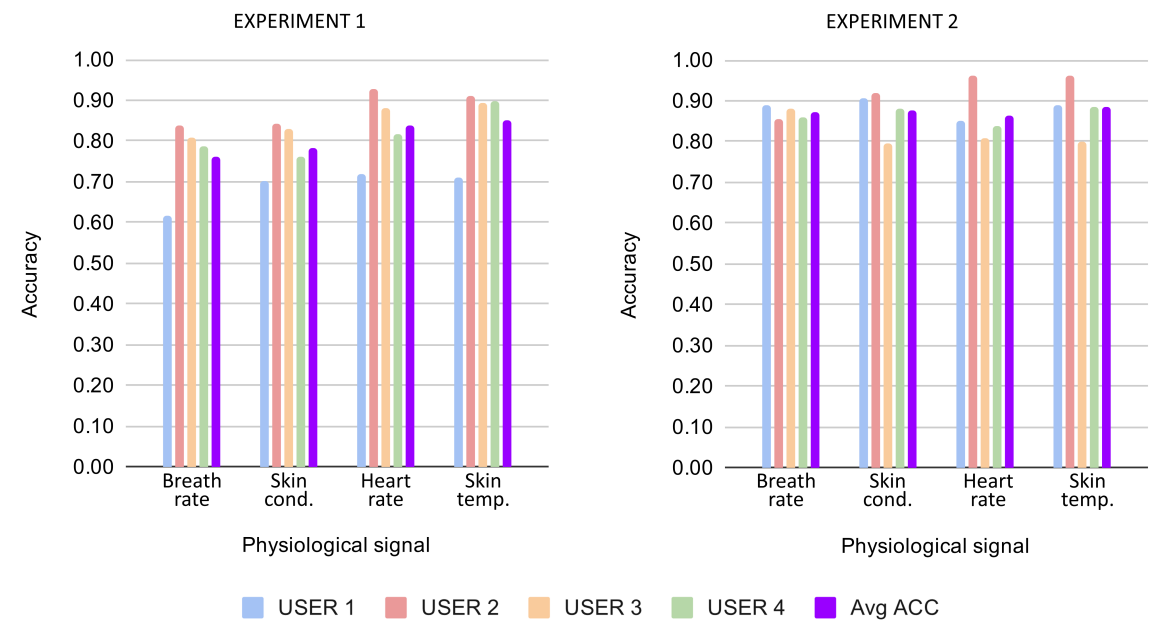


Figure 3. Average accuracy results for both experiments when each physiological signal is present in any of the combinations.

4.2.2. Second Experiment

Fig. 4 shows the results for the second experiment, which study whether subjects can be easily distinguished from the others by how concentration affects different sets of physiological signals. Results are presented in the graph in the same way as the previous experiment. Numeric results of this second experiment are presented in Table 5.

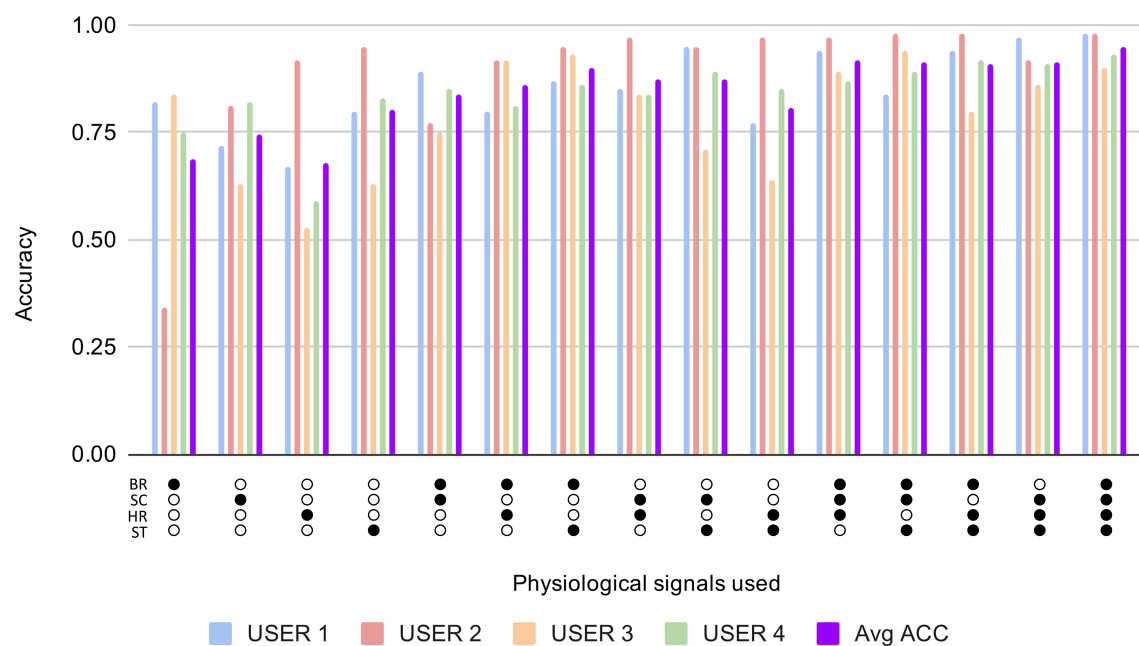


Figure 4. Accuracy results for different combination of the selected signals for Experiment 2: to study the relevance of the set of signals to discriminate one user from the rest by how concentration affects such set of signals.

(BR) Breath Rate; (SC) Skin Conductance; (HR) Heart Rate; (ST) Skin Temperature

BR	SC	HR	ST	USER 1		USER 2		USER 3		USER 4		Averages	
				ACC	EER	ACC	EER	ACC	EER	ACC	EER	ACC	EER
●				0.82	0.24	0.34	0.65	0.84	0.22	0.75	0.29	0.69	0.35
	●			0.72	0.3	0.81	0.22	0.63	0.45	0.82	0.21	0.75	0.30
		●		0.67	0.36	0.92	0.14	0.53	0.51	0.59	0.45	0.68	0.37
			●	0.8	0.29	0.95	0.12	0.63	0.38	0.83	0.2	0.80	0.25
●	●			0.89	0.2	0.77	0.29	0.75	0.29	0.85	0.23	0.84	0.24
●		●		0.8	0.24	0.92	0.12	0.92	0.16	0.81	0.22	0.86	0.19
●			●	0.87	0.22	0.95	0.09	0.93	0.18	0.86	0.2	0.90	0.17
	●	●		0.85	0.2	0.97	0.11	0.84	0.21	0.84	0.23	0.88	0.19
	●		●	0.95	0.13	0.95	0.13	0.71	0.34	0.89	0.23	0.88	0.21
		●	●	0.77	0.29	0.97	0.1	0.64	0.4	0.85	0.2	0.81	0.25
●	●	●		0.94	0.11	0.97	0.11	0.89	0.16	0.87	0.16	0.92	0.14
●	●		●	0.84	0.24	0.98	0.07	0.94	0.14	0.89	0.15	0.91	0.15
●		●	●	0.94	0.1	0.98	0.09	0.8	0.25	0.92	0.19	0.91	0.16
	●	●	●	0.97	0.08	0.92	0.11	0.86	0.18	0.91	0.19	0.92	0.14
●	●	●	●	0.98	0.06	0.98	0.09	0.90	0.13	0.93	0.14	0.95	0.11

Table 5. Results for the second experiment using different set of signals. (HR) Heart Rate; (BR) Breath Rate; (SC) Skin Conductance; (ST) Skin Temperature

As it happened in the previous experiment, skin temperature is the one that showed higher relevance when used separately obtaining an AUC=0.80 in average. In this case, all the signals performed better than the previous case when used separately (0.68 for heart rate, 0.69 for breath rate and 0.75 for skin conductance). However the results across users are less regular in this case, especially when using the BR signal. When using signals in pairs, all the combinations performed an AUC>0.75, being the best pair the one that include skin temperature and heart rate with an AUC=0.88. When combining the signals three by three the results are very similar for each combination in this case with

an average $AUC=0.91$ and $AUC=90.92$ for the four users. In fact, the performance of the fusion of any three signals is very similar with the one obtained for the set of four signals. The influence of each signal in this experiment is also depicted in the right side of Fig. 3 in the same way as previously. In this case, the average accuracy for each signal across all the combinations when it is present is very similar. In summary, a high performance, $AUC>0.75$, can be obtained to discriminate one user from the rest by how concentration affects their physiological patterns:

- When the the skin conductance and skin temperature signals are used independently.
- When any possible pair of the selected signals is used.
- When any possible trio of the selected signals is used.
- When the four signals are used as it was already confirmed in previous sections.

5. Discussion of Results

HMM intra-subject models have shown to be extremely powerful at detecting concentration. Results reported in the first set of experiments are encouraging, and endorse the potential use of HMM models to detect concentration from physiological signals that can be captured without the need for expensive equipment. An average EER of 0.09 when using the four selected signals and the consistency of the results across all users open the way for further development of low-cost devices to be integrated into adaptive learning systems, by using individual models that allow the system to personalise content delivery to increase the student's engagement.

Regarding the relevance of each signal in the detection process, results reported are also encouraging in some respects. For instance, the use of the skin temperature sensor separately has shown a quite high accuracy when detecting concentration. This sensor is located in the wrist of the subject so it is not annoying at all when developing any task. In some scenarios the accuracy obtained with this sensor could be enough, and, although further studies need to be done, results reported here endorse the potential of this signal to detect concentration in a very unobtrusive manner. For scenarios with higher requirements of accuracy, results showed that any combination of three of the selected signals offer a minimum of $AUC=81\%$ with a maximum ERR of 0.26. The best result was obtained for the multi-fusion of the four signals, with an AUC of 0.95%. A very similar accuracy, 0.98% was reported in [52]. However, it is important to note that the scope of such work is different than the present one since the aim there was to detect basic emotions in spite of a complex one as it is the case of the concentrated state. Furthermore, while other approaches focus on using more signals to maximise accuracy [46,51,56], in this work we have focused on maximising practicality and applicability. Thus, although we collected the four mentioned signals from each subject, our findings translate to a smaller number of sensors needed for discriminating user engagement. This suppose consequently, lower cost in the experimentation setups, higher user comfort, lower subject reactivity to the measurement context, and thus higher reliability. However, it is also important to point that results within the first experiment have also shown how the use of more physiological signals make results more homogeneous across users.

The second experiment indicates that concentration does not manifest in the same way in different individuals so each individual can be recognised among others from how physiological signals are affected by the concentration state. Two signals, skin temperature and skin temperature could be used separately to discriminate subjects during their concentrated state in not very demanding scenarios. In this case, the fusion of any pair of the selected sensors offer high accuracy, $AUC>0.84$, and a maximum ERR of 0.16. While a lot of previous studies have focused on the recognition of the concentrated state of the students, none of them deal with the recognition of the subject from how concentration affects their physiological signals to the best of our knowledge. Thus, none comparison with previous results can be provided for the second experiment.

Finally, our third experiment has outlined the need for intra-subject models to better represent how concentration manifests on physiological signals; and showed that the same model construction process fails when using an inter-subject approach. The findings are directly in line with previous findings in [20]. However, when comparing our results to those of older studies that used an inter-subject

approach to recognise engaged concentration, those works with a larger dataset obtained better results as it could be expected. This is the case of [56], where they managed a 63% of accuracy for a subject-independent model with the fusion of the EEG and other peripheral information as GSR, ST, BR and BVP. It is important to pointed out that this result was obtained under different conditions compared with the present work. Regarding the size of the dataset, 20 participants were involved in the experiment. Regarding intrusiveness, the use of EEG signals is intrusive pose strong limitations on the movement of the subjects in real-world scenarios because of face-mounted electrodes.

We shall remark that our experiments are very sensitive to potential labelling mistakes and that a few incorrect labels may have a relatively high impact on the result. Despite that special care has been taken to avoid such mistakes, there is still room for improvement in the labelling methodology by increasing the number of experts and discarding samples where a full agreement is not achieved. We believe that such improvement would yield a further reduction of the EER and even better results.

In addition, results reported in the paper should be understood as a reference to i)justify the advantages of using intra-subject models, and ii)discover the performance of different multi-fusion of several physiological signals towards the development of low-cost and unobtrusive acquisition systems. However this should not be understood as representative of the performance of this type of models in a practical setting. In such a context, frames should be analysed as part of a sequence, rather than as an independent entity, i.e. considering the scores in consecutive signal frames and setting a threshold based of the proportion of positive judgements on the last frames. Such a more informative approach will indeed yield more accurate and consistent results than the ones reported in the experimental section.

6. Conclusions and future work

One current challenge in user-centred feedback and adaptive systems is the accurate detection of relevant mental states that contribute to improving adaptation capabilities. For practical reasons, and in order to achieve solutions that can be widely adopted, it is mandatory that this is achieved by using low-cost and unobtrusive devices and also desirable that the required signals can be captured by using wearables.

Some previous works have identified behavioural patterns that are valid across different individuals and can be detected by cameras or eye trackers, e.g. facial muscle activation [66]. However, psychological signals are more affected by subject traits, and how emotions, affect, or cognitive states affect their values varies significantly across a population of subjects. In this work, we have proved the success of using physiological signals to detect concentration when they are combined with an intra-subject modelling approach; and also shown the inability of inter-subject models in this particular context. The combination of 4 easy-to-capture signals has yielded quite a relatively high performance on detecting the concentration state, and very good results have also been obtained by fusing 3 and even 2 of the selected signals. The benefits of the results presented in this paper extend beyond the AUC and EER values reported, as physiological signals and visual sources provide different type of information and may be used in combination to further improve the already successful rates reported in previous works based on cameras or eye-tracking devices [18]. Together, they allow for practical setups that are minimally intrusive and do not impose any important limitation to the mobility of the user.

From a practical point of view, there are three signals that were captured in the wrist's subjects within this work that provided high accuracy on detecting subject's concentration. Those signals are: heart rate, skin conductance and skin temperature. From this point of view, we can easily imagine the possibility of implementation of a low-cost wristworn device that capture these three signals. The potential advantage of this application is promising: to afford better affect-aware experience in learning scenarios by predicting the student's engagement. It may be particularly useful in big data studies if devices with an acceptable error rate is achieved, by offering a cheap, mobile, non-intrusive, and scalable solution as compared with the expensive medical equipment that is commonly used in many works.

Despite the positive results reported in this paper, there are still a number of ways in which this work can and will be extended. First, there is a need to improve labelling methodologies, to be able to work with mistake-free data that allows for a better estimation of accuracy measures. Second, the already available labels in the same data set can be used to build predicting tools for other mental/affective states. Third, despite the negative results at using inter-subject models, we consider that they have to be further explored using other alternatives such as the subject-based normalisation proposed in [20] for EEG signals or for mouse and keyboard signals [23]. The development of inter-subject models that can be used on previously unseen subjects is a key issue from a practical perspective and would open the way to a seamless integration of this type of technology on today's learning applications.

In order to further study these aspects and advance the methodological approach and developments that open the way towards the development of affect-aware user-centred adaptive systems in real-world educational scenarios, we have started two new projects funded by the Spanish Ministry of Education. These are ITS-MathPS and INT2AFF. The first of these projects attempts to improve the learning of word problem solving by using, in between other data, the student's personal cognitive and affective characteristics as a solver. The second aims to advance the methodological and practical developments required to address the intertwined relationship between the learner's affective and cognitive states, as the centre and the target of a multisensorial affect-aware user-centred adaptive learning system, which considers the given context in order to provide the most appropriate response to a particular learner in a given situation.

Author Contributions: Conceptualization, M.A. and J.B.; methodology, M.A.; software, A.S, M.A. and G.C.; validation, A.S, M.A. and G.C.; formal analysis, M.A.; investigation, A.S, M.A.; resources, J.B.; data curation, M.A. and J.B.; writing—original draft preparation, A.S.; writing—review and editing, A.S, M.A., G.C. and J.B.; visualization, A.S.; supervision, M.A. and J.B.; project administration, J.B.; funding acquisition, J.B.

Funding: This research was partly supported by the Spanish Ministry of Economy and Competitiveness through projects TIN2014-59641-C2-1-P, PGC2018-096463-B-I00, TIN2014-59641-C2-2-P and the Spanish Ministry of Science, Innovation and Universities through the project PGC2018-102279-B-I00.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Picard, R.W.; Papert, S.; Bender, W.; Blumberg, B.; Breazeal, C.; Cavallo, D.; Machover, T.; Resnick, M.; Roy, D.; Strohecker, C. Affective learning - a manifesto. *BT Technology Journal* **2004**, *22*, 253–269. doi:10.1023/B:BTTJ.0000047603.37042.33.
2. Pardos, Z.A.; Baker, R.S.; San Pedro, M.; Gowda, S.M.; Gowda, S.M. Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics* **2014**, *1*, 107–128.
3. Pekrun, R.; Goetz, T.; Daniels, L.M.; Stupnisky, R.H.; Perry, R.P. Boredom in Achievement Settings: Exploring Control-Value Antecedents and Performance Outcomes of a Neglected Emotion. *Journal of Educational Psychology* **2010**, *102*, 531–549.
4. Ainley, M. Connecting with learning: Motivation, affect and cognition in interest processes. *Educational Psychology Review* **2006**, *18*, 391–405.
5. Baker, R.S.; D'Mello, S.K.; Rodrigo, M.M.T.; Graesser, A.C. Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human Computer Studies* **2010**, *68*, 223–241.
6. JAMES, W. II.—WHAT IS AN EMOTION ? *Mind* **1884**, *os-IX*, 188–205. doi:10.1093/mind/os-ix.34.188.
7. Andreassi. *Psychophysiology*; Psychology Press, 2010. doi:10.4324/9780203880340.
8. Shu, L.; Xie, J.; Yang, M.; Li, Z.; Li, Z.; Liao, D.; Xu, X.; Yang, X. A review of emotion recognition using physiological signals, 2018. doi:10.3390/s18072074.
9. Lane, H.C.; D'Mello, S.K. Uses of Physiological Monitoring in Intelligent Learning Environments: A Review of Research, Evidence, and Technologies; Springer, Cham, 2019; pp. 67–86. doi:10.1007/978-3-030-02631-8_5.

10. Kamišalić, A.; Fister, I.; Turkanović, M.; Karakatić, S. Sensors and functionalities of non-invasive wrist-wearable devices: A review. *Sensors (Switzerland)* **2018**, *18*. doi:10.3390/s18061714.
11. Taj-Eldin, M.; Ryan, C.; O'flynn, B.; Galvin, P. A review of wearable solutions for physiological and emotional monitoring for use by people with autism spectrum disorder and their caregivers, 2018. doi:10.3390/s18124271.
12. De Arriba-Pérez, F.; Caeiro-Rodríguez, M.; Santos-Gago, J.M. Towards the use of commercial wrist wearables in education. Proceedings of 2017 4th Experiment at International Conference: Online Experimentation, exp.at 2017. Institute of Electrical and Electronics Engineers Inc., 2017, pp. 323–328. doi:10.1109/EXPAT.2017.7984354.
13. Lohani, M.; Payne, B.R.; Strayer, D.L. A review of psychophysiological measures to assess cognitive states in real-world driving, 2019. doi:10.3389/fnhum.2019.00057.
14. Nelson, B.W.; Allen, N.B. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: Intraindividual validation study. *Journal of Medical Internet Research* **2019**, *21*. doi:10.2196/10828.
15. Uria-rivas, R.; Rodriguez-sanchez, M.C.; Santos, O.C.; Vaquero, J.; Boticario, J.G. Impact of physiological signals acquisition in the emotional support provided in learning scenarios. *Sensors (Switzerland)* **2019**, *19*, 4520.
16. Salmeron-Majadas, S.; Arevalillo-Herráez, M.; Santos, O.C.; Saneiro, M.; Cabestrero, R.; Quirós, P.; Arnau, D.; Boticario, J.G. Filtering of spontaneous and low intensity emotions in educational contexts. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, 2015, Vol. 9112, pp. 429–438.
17. Purnamasari, P.D.; Junika, T.W. Frequency-based EEG human concentration detection system methods with SVM classification. Proceedings: CYBERNETICSCOM 2019 - 2019 IEEE International Conference on Cybernetics and Computational Intelligence: Towards a Smart and Human-Centered Cyber World. Institute of Electrical and Electronics Engineers Inc., 2019, pp. 29–34.
18. Nizetha Daniel, K.; Kamioka, E.; Daniel, K.N.; Kamioka, E. Detection of Learner's Concentration in Distance Learning System with Multiple Biological Information. *Journal of Computer and Communications* **2017**, *5*, 1–15.
19. Reyes, F.M.; Bolivar, C.B.; Olivas, V.C.A.; Serna, J.G.G. KAPEAN: A supportive tool for observing performance and concentration of children with learning difficulties. Proceedings of 2015 International Conference on Interactive Collaborative and Blended Learning, ICBL 2015. Institute of Electrical and Electronics Engineers Inc., 2016, pp. 52–56.
20. Arevalillo-Herráez, M.; Cobos, M.; Roger, S.; García-Pineda, M. Combining inter-subject modeling with a subject-based data transformation to improve affect recognition from EEG signals. *Sensors (Switzerland)* **2019**, *19*.
21. Arevalillo-Herráez, M.; Chicote-Huete, G.; Ferri, F.J.; Ayesh, A.; Boticario, J.G.; Katsigiannis, S.; Ramzan, N.; González, P.A. On using EEG signals for emotion modeling and biometry. 33rd European Simulation and Modelling Conference. European Multidisciplinary Society for Modelling and Simulation Technology, 2019, pp. 229–233.
22. Arnau-Gonzalez, P.; Arevalillo-Herraez, M.; Katsigiannis, S.; Ramzan, N. On the influence of affect in EEG-based subject identification. *IEEE Transactions on Affective Computing* **2020**, *Early Access*. doi:10.1109/TAFFC.2018.2877986.
23. Salmeron-Majadas, S.; Baker, R.S.; Santos, O.C.; Boticario, J.G. A Machine Learning Approach to Leverage Individual Keyboard and Mouse Interaction Behavior from Multiple Users in Real-World Learning Scenarios. *IEEE Access* **2018**, *6*, 39154–39179.
24. Maiorana, E.; Campisi, P. Longitudinal Evaluation of EEG-Based Biometric Recognition. *IEEE Transactions on Information Forensics and Security* **2018**, *13*, 1123–1138. doi:10.1109/TIFS.2017.2778010.
25. Torres-Valencia, C.A.; Garcia-Arias, H.F.; Lopez, M.A.; Orozco-Gutierrez, A.A. Comparative analysis of physiological signals and electroencephalogram (EEG) for multimodal emotion recognition using generative models. 2014 19th Symposium on Image, Signal Processing and Artificial Vision, STSIVA 2014. Institute of Electrical and Electronics Engineers Inc., 2015. doi:10.1109/STSIVA.2014.7010181.
26. Schmidt, P.; Reiss, A.; Dürichen, R.; Laerhoven, K.V. Wearable-based affect recognition—a review, 2019. doi:10.3390/s19194079.

27. Samadiani, N.; Huang, G.; Cai, B.; Luo, W.; Chi, C.H.; Xiang, Y.; He, J. A review on automatic facial expression recognition systems assisted by multimodal sensor data, 2019. doi:10.3390/s19081863.
28. Dzedzickis, A.; Kaklauskas, A.; Bucinskas, V. Human emotion recognition: Review of sensors and methods, 2020. doi:10.3390/s20030592.
29. Lim, J.Z.; Mountstephens, J.; Teo, J. Emotion Recognition Using Eye-Tracking: Taxonomy, Review and Current Challenges. *Sensors* **2020**, *20*, 2384. doi:10.3390/s20082384.
30. Krishna, A. An integrative review of sensory marketing: Engaging the senses to affect perception, judgment and behavior. *Journal of Consumer Psychology* **2012**, *22*, 332–351. doi:10.1016/j.jcps.2011.08.003.
31. Garbas, J.U.; Ruf, T.; Unfried, M.; Dieckmann, A. Towards robust real-time valence recognition from facial expressions for market research applications. Proceedings - 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII 2013, 2013, pp. 570–575. doi:10.1109/ACII.2013.100.
32. Tokuno, S.; Tsumatori, G.; Shono, S.; Takei, E.; Yamamoto, T.; Suzuki, G.; Mituyoshi, S.; Shimura, M. Usage of emotion recognition in military health care. 2011 Defense Science Research Conference and Expo, DSR 2011, 2011. doi:10.1109/DSR.2011.6026823.
33. Ben Moussa, M.; Magnenat-Thalmann, N. Applying Affect Recognition in Serious Games: The PlayMancer Project. Motion in Games; Egges, A.; Geraerts, R.; Overmars, M., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; Vol. 5884, pp. 53–62.
34. Ghaleb, E.; Popa, M.; Hortal, E.; Asteriadis, S.; Weiss, G. Towards Affect Recognition through Interactions with Learning Materials. Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018. Institute of Electrical and Electronics Engineers Inc., 2019, pp. 372–379. doi:10.1109/ICMLA.2018.00062.
35. Farzaneh, A.H.; Kim, Y.; Zhou, M.; Qi, X. Developing a deep learning-based affect recognition system for young children. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, 2019, Vol. 11626 LNAI, pp. 73–78. doi:10.1007/978-3-030-23207-8_14.
36. Whitehill, J.; Serpell, Z.; Lin, Y.C.; Foster, A.; Movellan, J.R. The faces of engagement: Automatic recognition of student engagement from facial expressions. *IEEE Transactions on Affective Computing* **2014**, *5*, 86–98.
37. Baker, R.S.J.D.; Kalka, J.; Aleven, V.; Rossi, L.; Gowda, S.M.; Wagner, A.Z.; Kusbit, G.W.; Wixon, M.; Salvi, A.; Ocumpaugh, J. Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. Technical report, 2012.
38. Botelho, A.F.; Baker, R.S.; Heffernan, N.T. Improving Sensor-Free Affect Detection Using Deep Learning. Technical report, 2017.
39. Ocumpaugh, J.; Baker, R.; Gowda, S.; Heffernan, N.; Heffernan, C. Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology* **2014**, *45*, 487–501.
40. Arevalillo-Herráez, M.; Marco-Giménez, L.; Arnau, D.; González-Calero, J.A. Adding sensor-free intention-based affective support to an Intelligent Tutoring System. *Knowl.-Based Syst.* **2017**, *132*, 85–93. doi:10.1016/j.knosys.2017.06.024.
41. Cunha-Perez, C.; Arevalillo-Herráez, M.; Marco-Giménez, L.; Arnau, D. On Incorporating Affective Support to an Intelligent Tutoring System: an Empirical Study. *IEEE-RITA* **2018**, *13*, 63–69. doi:10.1109/RITA.2018.2831760.
42. Krithika, L.B.; Lakshmi Priya, G.G. Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric. *Procedia Computer Science*. Elsevier B.V., 2016, Vol. 85, pp. 767–776.
43. Sharma, P.; Joshi, S.; Gautam, S.; Filipe, V.; Reis, M.; Reis, M.C. IET Computer Vision Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning. Student Engagement Detection Using Emotion Analysis, Eye Tracking and Head Movement with Machine Learning. Technical report, 2019.
44. D'Mello, S.K.; Kory, J. A review and meta-analysis of multimodal affect detection systems, 2015. doi:10.1145/2682899.
45. Marín-Morales, J.; Higuera-Trujillo, J.L.; Greco, A.; Guixeres, J.; Llinares, C.; Scilingo, E.P.; Alcañiz, M.; Valenza, G. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific Reports* **2018**, *8*, 1–15. doi:10.1038/s41598-018-32063-4.

46. Kim, J.; André, E. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2008**, *30*, 2067–2083.
47. Valenza, G.; Lanata, A.; Scilingo, E.P. The role of nonlinear dynamics in affective valence and arousal recognition. *IEEE Transactions on Affective Computing* **2012**, *3*, 237–249. doi:10.1109/T-AFFC.2011.30.
48. Russell, J.A. A circumplex model of affect. *Journal of Personality and Social Psychology* **1980**, *39*, 1161–1178. doi:10.1037/h0077714.
49. Ekman, P. An Argument for Basic Emotions. *Cognition and Emotion* **1992**, *6*, 169–200. doi:10.1080/02699939208411068.
50. Petrantonakis, P.C.; Hadjileontiadis, L.J. Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Transactions on Affective Computing* **2010**, *1*, 81–97. doi:10.1109/T-AFFC.2010.7.
51. Gong, P.; Ma, H.T.; Wang, Y. Emotion recognition based on the multiple physiological signals. 2016 IEEE International Conference on Real-Time Computing and Robotics, RCAR 2016. Institute of Electrical and Electronics Engineers Inc., 2016, pp. 140–143. doi:10.1109/RCAR.2016.7784015.
52. Shin, D.; Shin, D.; Shin, D. Development of emotion recognition interface using complex EEG/ECG bio-signal for interactive contents. *Multimedia Tools and Applications* **2017**, *76*, 11449–11470. doi:10.1007/s11042-016-4203-7.
53. Wen, W.; Liu, G.; Cheng, N.; Wei, J.; Shangguan, P.; Huang, W. Emotion recognition based on multi-variant correlation of physiological signals. *IEEE Transactions on Affective Computing* **2014**, *5*, 126–140. doi:10.1109/T-AFFC.2014.2327617.
54. D'Mello, S. A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology* **2013**, *105*, 1082–1099. doi:10.1037/a0032674.
55. Alzoubi, O.; D'Mello, S.K.; Calvo, R.A. Detecting naturalistic expressions of nonbasic affect using physiological signals. *IEEE Transactions on Affective Computing* **2012**, *3*, 298–310. doi:10.1109/T-AFFC.2012.4.
56. Chanel, G.; Rebetez, C.; Bétrancourt, M.; Pun, T. Emotion assessment from physiological signals for adaptation of game difficulty. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans* **2011**, *41*, 1052–1063. doi:10.1109/TSMCA.2011.2116000.
57. Lascio, E.D.; Gashi, S.; Santini, S. Unobtrusive Assessment of Students' Emotional Engagement during Lectures Using Electrodermal Activity Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol* **2018**, *2*, 21. doi:10.1145/3264913.
58. Santos, O.C.; Saneiro, M.; Boticario, J.G.; Rodriguez-Sanchez, M.C. Toward interactive context-aware affective educational recommendations in computer-assisted language learning. *New Review of Hypermedia and Multimedia* **2016**, *22*, 27–57. doi:10.1080/13614568.2015.1058428.
59. Healey, J.A.; Picard, R.W.; Healey, J.A. Detecting Stress During Real-World Driving Tasks Using Physiological Sensors. Technical report.
60. Santos, O.C.; Uria-Rivas, R.; Rodriguez-Sanchez, M.C.; Boticario, J.G. An open sensing and acting platform for context-aware affective support in ambient intelligent educational settings. *IEEE Sensors Journal* **2016**, *16*, 3865–3874.
61. Declaration of Helsinki – WMA – The World Medical Association. <https://www.wma.net/what-we-do/medical-ethics/declaration-of-helsinki/>.
62. Schwarzer, R., & Jerusalem, M. Generalized Self-Efficacy scale. In *Measures in health psychology: A user's portfolio. Causal and control beliefs*; Weinman, J.; Wright, S.; Johnston, M., Eds.; NFER-NELSON: Windsor, England, 1995; pp. 35–37.
63. John, O.P.; Srivastava, S. The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. Technical report.
64. Brooke, J. SUS - a quick and dirty usability scale. In *Usability Evaluation In Industry*; P.W.Jordan.; Thomas, B.; Weerdmeester, B.; McClelland, I., Eds.; Taylor and Francis: London, 1996; pp. 189–194.
65. Saneiro, M.; Santos, O.C.; Salmeron-Majadas, S.; Boticario, J.G. Towards emotion detection in educational scenarios from facial expressions and body movements through multimodal approaches. *Scientific World Journal* **2014**, *2014*.
66. Ekman, R. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*; Oxford University Press, USA, 1997.

