This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3273303

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.0092316

The effect of demanding mental tasks on electrodermal activity and heart rate during physical activity: a pilot study

JUAN A. CASTRO-GARCÍA[1,2], (Member, IEEE), ALBERTO J. MOLINA-CANTERO[1], MANUEL MERINO-MONGE[1], CLARA LEBRATO-VÁZQUEZ [1], ISABEL M. GÓMEZ-GÓNZALEZ[1,2],(Senior, IEEE)

[1]Departamento de Tecnología Electrónica, Universidad de Sevilla, 41012 Sevilla, Spain (e-mail: jacastro@us.es; almolina@us.es; manmermon@dte.us.es; clebrato@us.es; igomez@us.es).

[2]Instituto Universitario de Investigación en Ingeniería Informática, Universidad de Sevilla, 41012 Sevilla, Spain.

Corresponding author: Juan A. Castro-García (e-mail: jacastro@us.es).

This research was funded by project AAI, Grant PID2019-104323RB-C32 funded by MCIN/AEI/ 10.13039/501100011033 of Spain's Ministry of Science and Innovation.

ABSTRACT This paper presents the development of a virtual rehabilitation game and mental load with different difficulty blocks. The game was controlled with body tracking and physiological signals - electrocardiography and electrodermal activity - were recorded throughout the session. Several parameters -heart rate (HR), heart rate variability (HRV), skin conductance level (SCL), skin conductance response (SCR), energy expenditure ... - were extracted from these signals to check mental load influence on them. Mental load was found to affect the variation in kinetic power and instantaneous heart rate; a Support Vector Machine with linear kernel was trained with these two variables and an 82.3% accuracy rate was obtained. Furthermore, the mental load was reflected in the number of errors made by the volunteers, in the selection time and in the number of rounds in the game.

INDEX TERMS body tracker, ECG, EDA, mental load, rehabgame

I. INTRODUCTION

According to the World Health Organization (WHO), when people do exercise frequently, there is a reduction in the risk of suffering from cardiopathies, diabetes, high blood pressure, breast and colon cancers [1]. WHO guidelines state that young people and adults need to do 150 min. a week of moderate to vigorous aerobic exercise [2]. However, some physical activity can be negatively affected by certain factors. One of these factors is stress. Some people, used to doing sports, use physical exercise to cope with stress, although, in the following study [3], its authors concluded that, stress actually has a negative effect, as it reduces efforts to perform Physical Activity (PA). Stress elevates the concentration of cortisol in the blood and its negative effects on health [4] might be magnified when paired with a sedentary lifestyle, also implying an important risk factor related to obesity [5]. Stress can be defined as the *perceived information overload as the feeling of being overwhelmed when environmental demands exceed the perceived capacities to cope with them* [6], [7]. The most well-known cause is the work environment [8], but it also occurs during the performance of other daily activities such as driving a vehicle in urban environments [9].

To motivate and increase the frequency with which PA is carried out, virtual assistants can be used to guide the type of exercise and the number of repetitions to be performed [10]. These can be based on devices that monitor the subject, such as *body trackers* which have proven their efficiency in many additional applications, such as video games [11], [12], interaction with computers [13], [14], cardiopulmonary resuscitation (CPR) [15], estimate of caloric expenditure [16]–[19] or rehabilitation [20] in general. Additionally, the use of robots as tangible assistants to promote physical activity in the elderly [21] - also known as active aging - has also been tested, proving to be more motivating than virtual assistants [22]. While virtual assistants manage the rehabilitation session, other technologies, such as virtual reality [23] or the use of video games [24], can help with physical exercise.

Numerous studies have shown that stress, as well as the performance of physical activities, are reflected in several

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3273303

physiological variables. Stress can generate variations similar to those associated with PA, such as blood pressure (BP), heart rate (HR) [25]–[28], skin conductivity [29] and skin temperature [30]. This is why, the subject's monitoring could be used with a dual objective: the first one being to ensure that s/he is performing the exercise; and the second one being to extract information about his/her general state and the level of stress in particular that the proposed activity is generating, and in order to adapt it to a level of effort better suited to the individual's capabilities.

This pilot study was carried out to determine the effect that stress or increased mental demand has on certain physiological variables during light PA -below 39% of the heart rate reserve [31]-, and to thereby identify an excess of mental load during its execution. A Kinect-based video game was therefore adapted in which different geometric shapes have to be reached on the screen by only moving both arms. The skeleton provided by Kinect is used to estimate the kinetic power (P_k) of the movements and consequently the energy expenditure. The electrocardiogram (ECG) and electrodermal activity (EDA) of both hands are recorded simultaneously. PA should be light, with intensity in the control phase being similar to the stressful phase, to avoid increases in heart rate due to energy demand and possible sweating due to the effect of thermoregulation. The further goal of this study is to verify if there are physiological variables that indicate the stress level of a user with physical disability during a gamified rehabilitation session; however, as they are a particularly sensitive population, this study was performed with a healthy population.

II. STATE OF THE ART

A state-of-the-art review focused on finding the answer to "stress during physical activity" was carried out following the guidelines described in [32]. The PRISMA flow graph, which summarizes the results obtained in different phases, is depicted in Figure 1. As an initial approach, we used the text "stress AND physical AND activity" in two databases: IEEE Xplore and Scopus, which returned 343 items in IEEE Xplore and 36859 in Scopus. Given the large number of results in the second database, the search was refined by adding the following keywords: "heart AND rate AND (gsr OR eda OR sc OR skc)". This more restrictive search returned 170 items in Scopus.

We found an total of 13 duplicated papers that were removed, while the remaining 500 articles continued to the following two phases. In the first one, only the title and abstract were screened. This phase removed many studies about stress in physical materials.

In the second phase, the remaining 112 articles went through a full-text reading process, in which papers that met one of the following excluding criteria were discarded: 1) The study does not include physiological signals. 2) It uses very few samples of a physiological signal (*e.g.*, one single HR sample per day). 3) It uses animal models. 4) It is a review; and 5) It is a previous work from a selected paper.

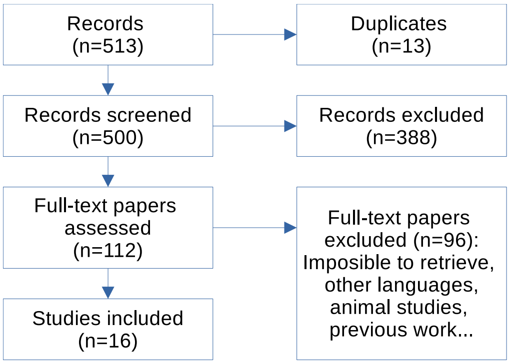

FIGURE 1: Number of selected articles throughout the review process.

Finally, we selected 16 related papers whose main features are shown in Table 1.

In the review carried out, we found that the most commonly used physiological variables for stress detection are derived from the recording of cardiac activity -87.5% of the cases-, in any of its variants: ECG [34], [38], [39], photoplethysmogram (PPG) [54] or blood volume pulse (BVP) [42], from which we obtain parameters based on the time or frequency domain, and in which Standard deviation of NN intervals (SDNN) is one of the most significant according to [41]. Secondly, we find EDA [42], [51], in 56.25% of cases, which has shown its efficacy when used alone or in conjunction with other variables, such as ECG [53], [57]. Less frequently, other authors have included physiological signals such as electroencephalography (EEG) [46], electromyography (EMG) for the measurement of muscle loading in the upper trapezius [46], BP [47], breathing rate (BR) [40], [41], maximal oxygen consumption ($V0_2$) [34] or skin temperature (ST) [41], [54]. To complement the physiological information, it may be advisable to include environmental variables, such as weather conditions [54], or the speed, position and state of the road, captured with a cell phone [42] while driving on a stretch of road; or those derived from the subject's own PA, which can be measured with accelerometers [39], [53] or markers placed on the human body that are then captured by a camera to analyze the subject's movement [52]. Accelerometers have also been used to limit the extent to which physiological information used to detect stress is corrupted by motion artifacts [33]. To complement the physiological information and to thus be able to classify it properly, validation tests of the state or level of stress to which the subject is subjected during the different phases of the experiment are needed. It is therefore known that, physiologically, a hormone, cortisol, is released in stressful situations. Some authors use periodic measurements of this hormone in saliva to check the stress level of the subject [33], [50]. However, the use of questionnaires to measure the emotional state is more widespread [34], [39], [42].

To induce stress many researchers have used a methodology in a controlled environment that seeks to exceed the

TABLE 1: Main features of the selected manuscripts obtained through the scoping review.

Ref.	#Subjects	Variables	Stress induction	Analysis technique	Performances / Results	Comments
[33]	30	EDA and Acc, salivary cortisol	Tier Social Stress Test	Several classifiers. Best results with kNN (k=3) (stress, no-stress, segment corrupted)	93% of participants correctly classified in stress phase, 63% during post-stress	The Acc. excludes EDA segments for classification. Cortisol tests every 15min.
[34]	14	ECG, VO2, questionnaire	Arithmetic operations: MIST [35]; PA: walking, ascending and running on a treadmill	Reactivity and recovery assess HRV differences between previous or post rest periods	Resting HRV correlated with reactivity and recovery for mental and physical stress.	V02 measures physical fitness. SAM test [36] to assess the valence and DT [37] for perceived stress
[38]	28	ECG	PA: rest and shoulder abduction of 45°; mental stressor (MENSA test), and both	Wilcoxon test applied to HRV features between conditions	Discrimination between all conditions. Mental load along with PA had an extra effect on HRV	HF was significant for all tasks at minutes 1 and 4, LF only being significant at minute 1.
[39]	8	ECG, Acc, questionnaire logs	Dialy life (2 weeks)	HRV time and frequency features; Acc (max, min, activity...) Several classifiers.	The best accuracy was obtained with Bagging 85.7%	Two stress levels: low and high. Acc data significantly complemented HRV data for stress modelling
[40]	21	ECG, BR, Acc, four-level questionnaire with 5 emotional words (cheerful, happy,...)	Mental arithmetic, cold pressure, talk in public	ECG and BR features and J48, AdaBoost, SVM. HMM for perceived stress model.	Stress (lab), best result J48 with AdaBoost 90.17%, model correlation of 0.74	Lab and daily life. Acc is used to remove data. *Physiological stress* classifier and model for *perceived stress*
[41]	48	ECG, BR, EDA, BVP, ST, EEG, Test to rate the subjective stress level	PA: 3 levels of cycling; two stress levels induced by videogames.	Ridge logistic regression (7 predictive models). Binary model's output	67.9% (all features) or 74.5% for only ECG and BR with high model confidence data	Based on the features' weights, the authors found that the most significant is SDNN.
[42]	21	ECG, PPG, Acc, (STAXI 2) [43] and UMACL [44] tests, photos, car position and speed	Driving task	RELIEFF for feature reduction. LDA, DT, kNN and an Ensemble classifier with Hill-Climbing [45]	Accuracy of 86.86% with physiological and driving features with the Ensemble classifier	Two Data Collection Scenarios with different driving features
[46]	19	ECG, EEG, EMG on the upper trapezius, questionnaire with a list of emotional words	Three types of Human-horse interaction: looking, grooming and leading	Classifiers: kNN, LSVM, SVM-RBF, DT, LDA	Valence: F1 = 78.3 with 1NN and EEG - PSDavg, Arousal: F1= 65.5 with LDA and EEG - MFCC	Binary classification (Negative vs Positive Valence) and (Negative vs Positive Arousal)
[47]	12	BP, EDA, ST, Heat flux and ECG	Stroop test, Arithmetic, Physical stress test evaluated through Borg RPE scale [48]	Statistical differences between stress phases	Significant differences in all parameters. BP and heat-flux correlated with stress	Mental stress tests had weaker effect compared to physical activity.
[49]	29	ECG, EEG and periodic RPE questionnaires	Cycling (resistance was adjusted for each participant)	Correlation between some indexes (CSI, AMHRR and EEG features) regarding RPE	CSI and HA_C4 (75% of people with significant correlation)	
[50]	34	EDA, ECG, ST, salivary cortisol	Stroop test, Tier Social Stress test, PA (walking at 3 speeds on a treadmill)	LDA, QDA, SVM and kNN and features selection	Best single modality (EDA), best accuracy of 97.13% with EDA + ST	Detect psychological stress and distinguish it from physical stress.
[51]	15	EDA	Simulated driving tests with stressors (accidents and sudden random sounds)	Smoothed Non-linear Energy Operator (SNEO) and peaks detection	Sensitivity of 95%	Filters for removing motion artifacts.
[52]	19	ECG, EEG-ERP, EDA, motion capture markers on legs and feet, survey, tone generation	Virtual reality high heights exposure during three balance exercises	Statistical comparison of conditions (Wilcoxon rsigned-rank test)	Increase in step-offs, HR, EDA and response time. Decreased EEG peak amplitude	Decreased peak amplitude in anterior cingulate (balance sensing)
[53]	20	ECG, EDA, Acc	Stroop test and mental arithmetic. Activities: sitting down, standing up and walking	Several classifiers: J48, Bayes Net, SVM	92.4% with J48 and all features. 80.9% for subjects classified.	Acc data improves classification accuracy
[54]	16	PPG, EDA, ST, Acc, weather, questionnaires: PSS [55], NASA-TLX [56],...	Exams with a jury	RF, MLP, kNN, LDA, SVM and features reduction by correlation and PCA	Accuracy of 80% with MLP for two output classes	By including weather features, classification improves. Two and three output classes
[57]	9	ECG, EDA	Stroop and PA: cycling and rest	BN, SVM, k-NN, J48, RF y AB. CFS, PCA, ... for feature reduction. Wilcoxon tests	Accuracy between 70 and 85%.	Stroop test is only performed during rest periods

subject's normal processing load in different ways: by using arithmetic calculation tests [34], [40], [53], with time-varying restrictions; with an intelligence quotient [38]; seeking conflict in the identification of a word's color that semantically expresses a color that is different from the color used to display it - Stroop test [53], [57]; with videogames [41], [51] where surprising elements are introduced that alter their course; by using virtual reality to simulate environments that induce extreme situations such as walking along a bar between two skyscrapers [52] or subjecting the subject to a stressful real environment such as speaking in public or defending an exam before a jury [42], [50]. There are also studies that seek experimentation in daily life environments [39], while driving [42] or with animal interaction [46].

The performance of PAs and how they affect physiological variables, in combination with stressful situations, is the main purpose of this study. In all the studies found, these experiments were carried out in controlled environments using a treadmill and involving walking at different speeds [50], running, uphill walking [34], among others ; balancing on a barbell [52], ergometer cycling [41], [49], [54] or just maintaining a posture - such as raising your arms 45^o and holding them up for several minutes [38].

Most of the studies analyzed seek to detect the existence - or non-existence - of stress, or different stress levels, through the use of classifiers. Among the most commonly used are k-nearest neighbors (kNN) [33], support vector machine (SVM), J48 [53], linear discriminant analysis (LDA) [46], artificial neural networks (ANN) [54], random forests (RF), decision tree (DT), or Bagging [39]. Several authors have studied the classifiers that generate the best results with the same dataset, while others combine them to improve prediction [42]. Feature space dimensionality reduction is a common preclassification task itself [42], [45] that can be used to determine which features are the most relevant. This is especially relevant for multi-modal systems that include several sensors that can extract multiple features. They do [41] in particular use the ridge logistic regression method to achieve this objective.

The use of statistical techniques, such as the Wilcoxon test, allows us to determine the statistical significance between different stress states -or levels- according to selected variables or characteristics. In the literature reviewed, several authors have been found to use this analysis type [38], [57] or another analysis type aimed at finding correlations between the level of stress measured by questionnaires and these physiological variables [34], [49].

Classification accuracies range between 70% and 97.13%. Better results are generally obtained when EDA is added to the physiological variables. In [51], the authors obtained a sensitivity of 95%; in [50] an accuracy of 97.13% by combining EDA with ST; in [33] a correct stress classification was obtained in 93% of the participants; whereas a slightly lower accuracy of 92.4% was obtained in [53]. It was only in [57] that the accuracy obtained was between 70% and 85%. Stress generally produces an increase in skin potential response (SPR) [51] or skin conductivity [33], [49], [52].

Heart rate-derived information permits correlations to be established between stressful phases and pre- and post-stress phases [34], allowing us to differentiate between all combinations of exercise and induced stress [38]. Furthermore, if BR data is added, the accuracy of the system is improved [41]. Some studies affirm that adding other non-physiological inputs improves prediction [42], [53], [54], [58]. Not many studies have used EEG among the signals analyzed, but, in the existing ones, spectral characteristics such as high alpha power at the C4 position are highly correlated in more than 75% of the participants with the different stress phases [49]. Furthermore, the average EEG power and Mel Frequency Cepstral Coefficients (MFCC) were found to be the most significant features for classifying valence and arousal dimensions in human-horse interaction [49]. A decrease in evoked potentials has also been found during stressful exercise performance, suggesting an increase in cortical demand for primary task performance [52].

Combining PA and stress produces an increase in HR compared to cases when these situations occurred separately [38]. This means that it is possible to detect stress in a subject, regardless of whether they are resting or at work, with mental stress having a weaker effect than physical stress [47]. In [41], the authors obtained an accuracy of 74.5% in stress discrimination while exercising with an indoor cycle at different intensities, based on features derived from HR and BR. Moreover, significant differences between the PA phases and stress have been proven when other variables, such as EDA, ST, BP or Heat flux, are used [47].

III. METHODOLOGY

This experiment is inspired in the idea of using games for rehabilitation (*rehabgame*). It is based on the use of Kinect v1 and a modification of *ShapeGame*, an application included in the Kinect package for Windows SDK v1.8. The participant's skeleton is shown on screen along with different spinning geometric shapes. According to a cue shown on the screen, one of these shapes has to be reached by performing upper limb movements, with the lower limbs having to remain still. Some physiological variables were recorded during the experiment as we will explain below.

Each participant attended a single session consisting of three activity blocks (B_n) plus a final rest period (R_4) (Figure 2). Each block is also divided into three parts: An initial rest period (R_n), a gentle activity level (RL_n) and a stressful period (SL_n). During rest periods, volunteers were asked to remain calm and it was suggested that they sat in a chair. If they sat down for 30 seconds before the start of the next part, they were instructed to stand up to prevent heart variability in the following part being influenced by the postural change [59]. Rest periods (R_n), as borders between consecutive activity parts, took 5 minutes - 300 sec. - , because EDA is a slow recovery physiological signal. [60].

In the gentle activity level, (RL_n) two shapes are shown on the screen: the cue, visible at the top; and the target, that

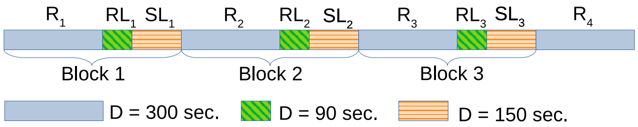

FIGURE 2: Experimental timing.

can appear randomly in different positions, but always within the reach of the skeleton's arms. Once a target is reached, a new one, with its cue, appears on the screen one second later. This delay allows the player to return to the initial position and perform similar movements for each target. The RL_n part takes 90 seconds. In the stressful level (SL_n) ten shapes appear distributed around the avatar and this takes 150 sec. Here, participants must search before selecting and be careful not to accidentally hit a non-cued shape. There are several geometric shapes that are identical to the cue, in a quantity that decreases as the experiment progresses. The total of PA per block times is 4 minutes, to avoid fatigue and keep it as light.

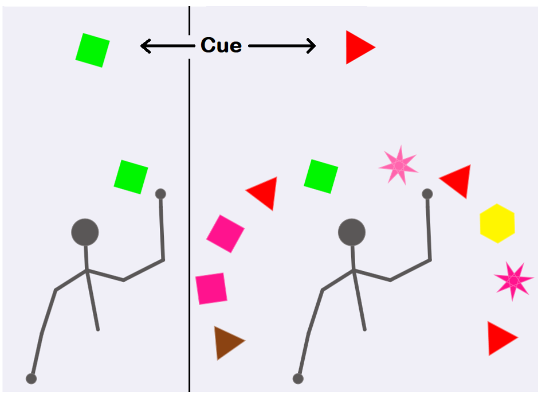

FIGURE 3: Game pictures: RL_i Level on the left and SL_1 on the right.

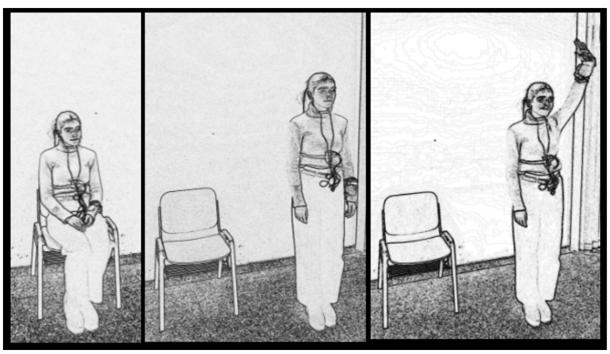

FIGURE 4: Volunteer playing: left in rest period, center and right, during the activity.

All blocks are basically similar, although with some subtle differences existing in two main points: 1) The maximum dwell time (DWT) in selecting a target; and 2) The number of available targets in the SL_n part. Both decrease as n increases, guaranteeing that mental demand progressively increases. As reaching a target on screen takes 2 sec on average, this is the minimum DWT value configured for the ending block, and a new target is shown on the screen when the previous one is selected, a similar level of PA being expected throughout the blocks, irrespective of the DWT established. If the participant does not select any shape before the DWT expires, or they select an incorrect one, an annoying sound will be played (negative feedback) and a new cue and target are then generated (a new round); there is a minimum number of rounds (N_R^{min}) even if the volunteer does not make any selections. The game also makes a warning sound if the player has correctly selected a target or approached a shape but not sufficiently to select it. This information is summarized in Table 2. Before the experiment, the volunteers played a demo in which, multiple valid targets identical to the cue, appeared at the same time. This training session lasted 1 minute.

TABLE 2: Description of the main features of each level for all blocks: duration (D) in seconds, number of geometric shapes shown (Sh), number of identical shape targets (NT), and dwell time (DWT) in seconds, minimum number of rounds (N_R^{min}).

	R_1	RL_1	SL_1	R_2	RL_2	SL_2	R_3	RL_3	SL_3	R_4
D (s)	300	90	150	300	90	150	300	90	150	300
Sh	0	1	10	0	1	10	0	1	10	0
NT	0	1	3	0	1	2	0	1	1	0
DWT (s)	-	6	6	-	4	4	-	2	2	-
N_R^{min}	-	12	21	-	18	30	-	30	50	-

The underlying idea of this experimental division is to establish comparisons between periods of similar physical activity (RL_n-SL_n) but with a different mental load. Differences between blocks stem from a progressive increase in mental load by means of reducing completion time and hindering the target search. We hypothesized that all RL_n periods have a similar level of PA and mental load, which are both higher than in the rest periods R_n. Significant differences must appear between SL_n periods with respect to their previous RL_n periods and between SL segments when increasing the mental demand. We also hypothesized that an increase in that cognitive load may slightly reduce PA, especially when the time constraints are more severe. Figure 5 shows the expected effect of experimental blocks and periods on PA and cognitive load.

A. DATA ACQUISITION

Physical activity was obtained by recording the joint coordinates that Kinect v1 offers (Figure 6). Two physiological signals were also acquired during the experiment: ECG and EDA. We used Ag/AgCl electrodes for ECG (Ambu WhiteSensor 4200), placed according to Einthoven II lead positions and specific wet electrodes for EDA (Biopac EL507), opposed to dry ones [29], [61], which were placed on the thenar and hypothenar eminences of both hands (see Figure 7); however, we only used the EDA channel for the participant's dominant hand. We also used PSM [62], an open-

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3273303

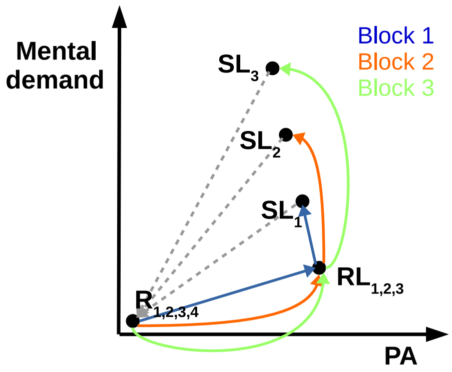

FIGURE 5: Illustration of the expected effects on PA and cognitive demand throughout the experimental blocks.

hardware shield placed on an Arduino Uno board, that was programmed with LBSP [63], an open-source library for real-time applications, in order to acquire data at a sampling frequency of 256 Hz per channel, and sending them wirelessly via Bluetooth.

Data was acquired using LSLRec software based on the Lab-Streaming Layer (LSL) library [64], [65]. This software makes it possible to save the data together with different synchronization marks, thereby facilitating data identification and segmentation throughout the different phases of an experiment.

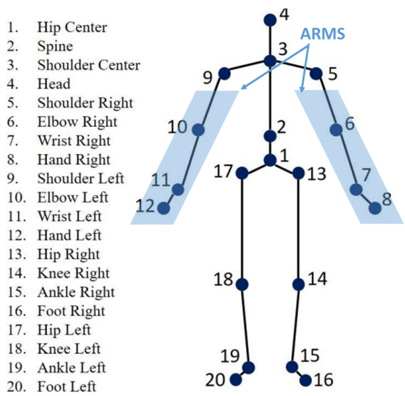

FIGURE 6: Kinect v1 joint list [66] and coordinate group used.

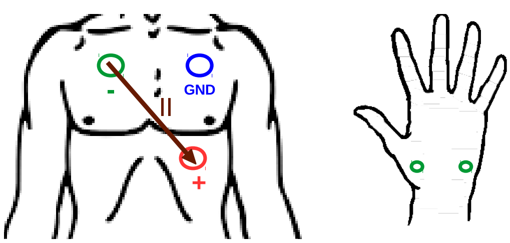

FIGURE 7: Electrode placement diagram of the ECG (left) and EDA (right) electrodes. Two EDA channels were used, one in each hand.

B. PARTICIPANTS

We initially gathered together a group of 29 healthy volunteers (16 men, 13 women), aged between 18 and 67 (mean 39.9 and standard deviation - SD - 14.2), and weighing between 50 and 120 kg (mean 74.7 and SD 16.4). However, two men were excluded because some signals were not properly registered. This sample size is within the range of pilot studies [67], moreover, using Cochran's equation with this population, a precision level of 15.3% with a confidence level of 90% was obtained. All participants agreed to take part in the experiment and signed the consent form. The Ethics committee at the Regional Government of Andalusia (Spain) approved the experiment (protocol code C.P. TAIS-C.I. 1130-N-17, 2018).

C. PROCESSING

This section describes the characteristics set analyzed. The sliding window technique was used for the time-series analysis at each block and level. The SL_n levels were divided into two segments of the same length as the RL_n level in order to obtain the same number of samples. These two segments are placed at the beginning and at the end of the SL_n period. In this way it is possible to analyze if any latency exists regarding the stressor effect onset, as well as avoiding possible biases. The first 90 seconds of the SL_n can be considered the transient time, or the time of major changes in level, while the last 90 seconds can be understood to be when the physiological variables are relatively stabilized or in a steady state. Furthermore, in order to analyze the evolution in more detail, we study the variation in each of the following variables between the RL_n level and its corresponding SL_n level.

To avoid the effect of outliers in every set of data associated to all the variables obtained, the method based on the interquartile range (IQR) was used, which considers outliers to be those values in the set that are out of the following range: $[Q_1 - 1.5IQR, Q_3 + 1.5IQR]$, with $IQR = Q_3 - Q_1$ and Q_i the i-th quartile. The Kruskal-Wallis test (KW Test) will be used to find statistical differences among sets. This is a non-parametric method that tests whether the samples originate from the same distribution and it does not assume a normal distribution of the residuals.

1) Game statistics

The number of rounds (N_R), number hit (H), missing targets (MT), wrong targets (WT) and selection time (T_s) are extracted from the game to compare the performance during the periods RL_n and SL_n in all difficulty blocks.

2) Kinetic power estimate P_k

Power expenditure can be estimated by evaluating the kinetic power, P_k, of the arm movements - joints 5-12 in the skeleton provided by Kinect (see Figure 6). To do this, the mass and velocity of the joint's coordinates are required. Velocity is easily obtained from the joint's positions at two consecutive time points, while for the mass distribution in the body, we used the results published in [19], and summarized in Table 3.

TABLE 3: Percentage of body weight [19] and associated clusters for kinetic power estimate.

Cluster	Body zone	Weight percentage (%)	IDs
ARM	Upper arms	4	$s_{\{5,6\}}$ $s_{\{9,10\}}$
	Forearms	3	$s_{\{6,7\}}$ $s_{\{10,11\}}$
	Hands	2.5	$s_{\{8\}}$ $s_{\{12\}}$

So, let $s_{\{.\}}$ be a set of joints that delimit an arm segment (forearm, and upper arm) or define the hand (see Table 3), and let $\vec{r}_{s_{\{.\}}}(t)$ be the position of its geometric center at time point t. The velocity, $\vec{v_s}(t)$, of a body segment s at time t is given by Eq. 1, and let m_s be the mass of segment s, obtained by applying the associated percentage to the whole body weight. Kinetic power, P_k, is therefore defined by Eq. 2, which includes the energy expenditure of all the arm segments.

$$\vec{v_s}(t) = \frac{\vec{r_{s_{\{.\}}}}(t) - \vec{r_{s_{\{.\}}}}(t-\Delta t)}{\Delta t} \tag{1}$$

$$P_k = \sum_s \frac{1}{2} m_s |\vec{v_s}(t)|^2 / \Delta t \tag{2}$$

$$s \in [s_{\{8\}}\ s_{\{12\}}\ s_{\{6,7\}}\ s_{\{5,6\}}\ s_{\{9,10\}}\ s_{\{10,11\}}]$$

The joint's coordinates are low-pass filtered by means of an exponential smoothing filter that Kinect implements. Data were then segmented by a sliding window with a 50% overlap.

3) EDA features

To process the EDA data, we used Ledalab tools [68], configured with *Continuous Decomposition Analysis (CDA)* to recover the characteristics of the underlying signal in the sudomotor nerve; with *Standard trough-to-peak (TTP)*, which analyzes data window maximums and miminums; and with *Global* which offers general data values. Table 4 - accessible via ledalab.de/documentation.htm - summarizes these variables. Recorded data were adapted to Ledalab input format by adding *events* at the time points when the first geometric shape of each set is shown. A two-second window (the smallest DWT) and a sensitivity of 1 μS were used.

4) ECG features

ECG was used to extract the instantaneous heart rate (IHR) for each level of the game and the heart rate variability (HRV), using 30-sec. windows with an overlap of 50% (ultra-short-term recording times [69]), which would allow us to obtain 5 windows at the shorter duration levels, (RL_n), while for the IHR we used 6-sec. windows with the same overlap as in the previous case. The HRV has special interest given that it allows us to evaluate the activity of the autonomic nervous system (ANS) [70], and specifically the parasympathetic nervous system (PNS) and the sympathetic nervous system (SNS) whose activity can be affected in situations of high mental demand such as stress. From HRV we observed time characteristics such as Standard deviation of NN intervals (SDNN), the root mean square of successive differences between normal heartbeats (RMSSD) and the percentage of successive RR intervals that differ by more than 50 ms (pNN50); frequenty characteristics were not studied since the windows are very small [69].

TABLE 4: EDA-variables' list.

Method	Variable	Description
CDA	nSCR	Number of significant SCR
	Latency	Latency of first significant SCR
	AmpSum	Sum of SCR-amplitudes of significant SCRs
	SCR	Average phasic drive
	ISCR	Area of phasic driver in the window
	PhasicMax	Maximum value of phasic activity
	Tonic	Mean tonic activity
TTP	TTP.nSCR	Number of significant SCR
	TTP.AmpSum	Sum of SCR-amplitudes of significant SCRs
	TTP.Latency	Response latency of first significant SCR
Global	Mean	Mean SC value within response window
	MaxDeflection	Maximum positive deflection

IV. RESULTS

A. GAME STATISTICS

Table 5 shows the averages of the four variables described in the previous section related to the activity performed by all participants. With respect to wrong targets (WTs), it can be observed during the SL_n periods that about 6.3% of failures occur in the first two levels while it almost doubles in SL_3. There is a statistically significant difference between this level and previous levels ($p < 0.0011$). Furthermore, WT cannot occur at RL_n levels, since a single target is displayed on the screen.

The missing target (MT) obtained low values at all levels in the first blocks ($< 1\%$). Only in the last block did this error reach 10.78%, standing out from the other levels and blocks. This difference was statistically significant ($p < 0.0015$). The MT number is non-zero at RL_n levels and somewhat higher for the last block, although without statistically significant differences. These results may be due to the Kinect skeleton not reaching the target properly some times, and requiring additional stretching that may not be done correctly in the pre-set time. In RL_3, moreover, we have a DWT that is at the limit of how long it takes a subject to reach a target on average, so it seems reasonable that the MT number increased. As previously mentioned, there are no statistical differences between the different levels and blocks, with the exception of SL_3.

As for the number of rounds (N_R), Table 5 shows the average for each level and block. Considering that the duration of the SL_n level is longer than the RL_n level, some normalization is required, dividing N_R by the duration of the level. The normalized values obtained are specifically[0.46 0.40 0.49 0.42 0.51 0.46] following the same order of RL_1-SL_3, as shown in the table below. One may expect N_R to decrease as the difficulty increases - the number of blocks advances - but the differences are not enough to make the regression slope statistically significant. However, note that a decrease in N_R can be observed when moving from the RL_n level to its SL_n, these differences having a significant difference - $p < 0.001$.

Finally, in focusing on the selection time (T_s), an increase in time is noted between each SL_n level and its respective RL_n, making the difference between them increase as the block number advances, *i.e.* as cognitive load rises. All mean times show statistically significant differences between them, as well as showing a positive slope with increasing block number.

TABLE 5: Wrong targets (WT) and missing targets (MT) in %, rounds, number of rounds (N_R) and selection time (T_s) for all participants.

	$\overline{WT}$ (%)		$\overline{MT}$ (%)		$\overline{N_R}$		$\overline{T_S}$ (s)	
	mean	SD	mean	SD	mean	SD	mean	SD
RL_1	0	0	0.21	0.76	41.73	4.77	1.12	0.66
SL_1	6.36	5.24	0.24	0.70	60.35	9.75	1.50	0.93
RL_2	0	0	0.10	0.49	44.85	3.39	0.97	0.38
SL_2	6.28	4.58	0.30	0.89	62.85	8.01	1.38	0.73
RL_3	0	0	1.57	2.90	45.69	3.04	0.94	0.32
SL_3	12.08	6.33	10.78	7.09	69.27	2.97	1.60	0.54

B. KINETIC POWER ESTIMATE

As an example, Figure 8 shows the behavior of variable P_k for participant S09, while Table 6 shows the values obtained by all the volunteers. Note that the motion is poor at the R_n levels. For each block, the SL_n level shows less motion with respect to its corresponding RL_n, probably due to searching for the target among all the available ones. This difference was statistically negative in all blocks and subjects, $\Delta P_k = P_k^{SL_n} - P_k^{RL_n} < 0$, yet, no significant correlation was found with the number of blocks, *i.e.*, with increased cognitive load, so that no change in this difference can be confirmed with task difficulty. These results are in accordance with those obtained from N_R.

TABLE 6: Mean and SD of P_K for all subjects and blocks.

	R1	RL1	SL1	R2	RL2	SL2	R3	RL3	SL3	R4
Mean	0.11	0.86	0.64	0.09	0.88	0.62	0.11	0.88	0.69	0.09
SD	0.11	0.30	0.29	0.07	0.30	0.28	0.09	0.36	0.28	0.05

C. PHYSIOLOGICAL FEATURES

As a first analysis, we compared the statistical significance of each proposed variable within each block R_n vs. RL_n vs. SL_n, and across the SL_n segments of the three blocks, in order to know if there is any effect on physical activity and/or cognitive demand in the variables. From the proposed characteristics, we selected those that were statistically significant for at least 90% of the total comparisons made between these levels and across all volunteers.

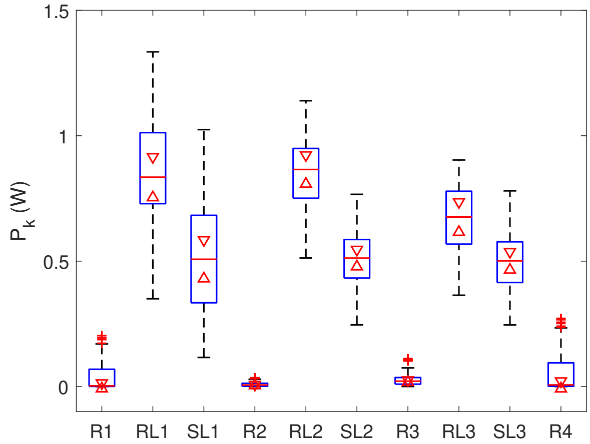

FIGURE 8: Subject S09's P_k at each level using a 6 sec. data window.

Table 7 shows these percentages by level comparison using the Kruskal Wallis test (KW Test) with a significance value ($p < 0.05$) for all EDA variables. The first and last 90 sec of the SL_n segment were also compared. As previously explained, the first 90 sec may reflect some transience, while greater stationarity is expected in the last 90 sec. Only the CDA.Tonic variable shows significant variations throughout the experiment, although another variable, Global.Mean, is close to the 90% limit.

TABLE 7: Percentage of times each EDA variable obtained significant p-values at full levels, in the first 90 seconds (F90EDA) and the last 90 seconds (L90EDA) for all subjets.

	EDA		F90EDA		L90EDA	
Variable	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
CDA.AmpSum	37.8	13.0	30.2	3.8	34.6	13.5
CDA.ISCR	39.7	7.9	37.2	5.3	40.4	5.3
CDA.Latency	35.2	7.0	31.4	14.1	28.2	7.6
CDA.nSCR	11.5	5.4	5.4	2.1	8.5	5.0
CDA.PhasicMax	48.7	7.2	44.2	5.3	41.0	7.9
CDA.SCR	39.7	7.9	37.2	5.2	40.4	5.3
CDA.Tonic	**91.7**	**6.6**	**86.5**	**7.6**	**96.1**	**3.4**
TTP.AmpSum	29.5	14.5	28.2	12.5	25.0	10.5
TTP.Latency	21.8	9.6	15.4	6.8	21.2	9.3
TTP.nSCR	15.4	2.4	15.4	6.8	13.5	3.2
Global.Mean	88.5	10.8	86.5	6.3	89.7	7.1
Global.MaxDeflection	30.8	16.0	30.1	12.0	25.0	10.0

CDA.Tonic and Global.Mean, show an accumulation effect as the experiment progresses, as has been shown in Figure 9, and this can be observed more clearly in Figure 10 on the same participant. There is no return to the prior rest level when moving across the blocks, even though the rest period is 5 min. It would therefore be interesting to analyze how these variables change in the SL_n segment with respect to the previous one (ΔEDA_n), and to determine whether these changes increase as the cognitive load grows.

The ΔCDA.Tonic$_n$ and ΔGobal.Mean$_n$ difference is positive for all subjects and blocks applying the bootstrap test [71] for a 5% significance level. These results were the same when we included the first 90 sec, the last 90 sec, or

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2023.3273303

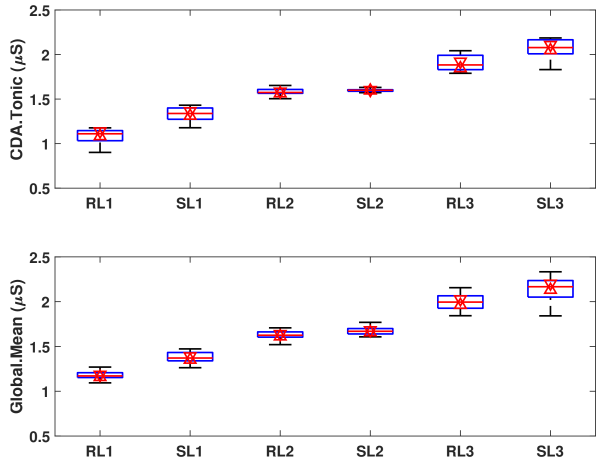

FIGURE 9: Subject S09's CDA.Tonic and Global.Mean at each level with PA.

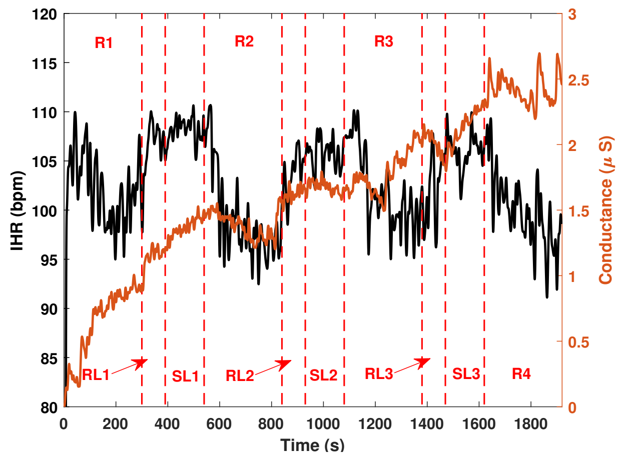

FIGURE 10: IHR -black- and EDA -orange- for volunteer S09 during the experiment. Level transitions are marked in red.

the whole segment of each level in the analysis. The linear regression shows that a positive slope exist for these variables with respect to the block number. This slope is statistically significant, which implies that as the difficulty increases, so does the Δs. This analysis of differences between segments and blocks was extended to the other EDA characteristics (Table 4). Only the ΔTTP.Latency regression shows a significant negative tendency. This parameter measures the delay time between a new target appearing on the screen, and the detection of a pulse in the EDA, the so-called event-related skin conductance response (SCR). As difficulty increases, latency decreases.

Regarding the features associated with HR, group differences have been observed between resting states R_n and any of the activity states - RL_n, SL_n. This was expected. Figures 10 and 11 show the S09 participant values for IHR. An increase in cardiac activity was observed during the PA segments with respect to those related to inactivity. Furthermore, unlike the EDA signal, the HR recovers in resting levels. Although Figure 11 shows a slight increase in HR values in the SL_n segments compared to their previous RL_n, that difference was not significant for the majority of the participants and blocks. Such a significance was only observed in ΔIHR between Block 1 and Block 3; and between Block 2 and Block 3 for all participants. No statistical differences were found in the HRV characteristics analyzed for the group.

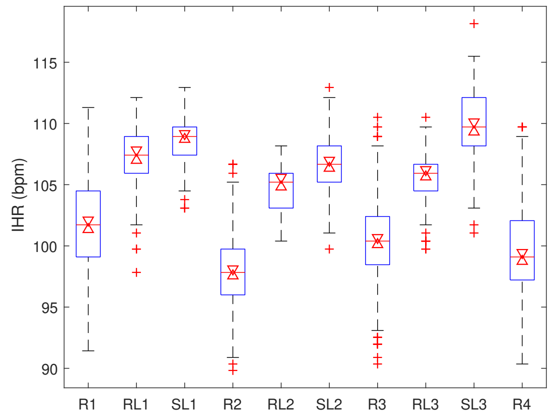

FIGURE 11: Subject S09's HR at each level

V. DISCUSSION

Indicators have been found to vary significantly during the different experimental phases. These include both physiological - ΔP_k, CDA.Tonic, ΔCDA.Tonic, ΔGlobal.Mean and ΔIHR - and non-physiological types - WT, missing target (MT), ΔRounds, Ts -; see Tables 5 and 8.

TABLE 8: Selected variable values between RL_n and SL_n for the three experimental blocks and all subjets.

		ΔP_k	ΔCDA.Tonic	ΔGlobal.Mean	ΔIHR
B_1	Mean	-0.22	0.13	0.13	0.51
	SD	0.21	0.23	0.26	4.13
B_2	Mean	-0.26	0.14	0.14	-0.26
	SD	0.18	0.38	0.44	3.77
B_3	Mean	-0.19	0.09	0.12	5.96
	SD	0.21	0.32	0.36	12.04

In analyzing the game variables, all of them show significant differences between the SL_n and its previous RL_n. However, only selection time (T_s) shows a more game dynamic behavior, increasing in value as the difficulty of the task at SL_n levels grows (rising n). This was expected, as reducing the number of correct targets increases the search time. And, as the available time to perform this task in the SL_3 state was significantly limited, selection errors WT and omission errors MT grow, while the number of rounds achieved (or targets achieved) was lower compared to the previous levels. An analysis among all these variables allows us to observe the existence of a high and significant correlation (Pearson's coefficient $\rho = 0.97$, p=0.001) just for T_s and WT. During RL_n periods the T_s is low and WT is zero. However, in the SL_n this time increases, together with the errors, and especially in the SL_3 level.

At RL_n levels, a decrease in T_s is observed as n grows and, consequently, an increase in N_R. These levels are characterized by displaying a single target on the screen. This variation seems to reflect a learn-to-play effect after the first block, however, the statistical regression analysis shows that such an effect is not significant. T_s is a particularly interesting variable, given that it can serve as a discriminator between situations with different cognitive loads. This means you could differentiate between the periods SL_3 and RL_3 with an accuracy of 96.2% by using a threshold of 1.28s, which is the block with the highest cognitive load. With a lower threshold of 1.1s, which is the optimal one to also differentiate blocks 1 and 2, the increase in cognitive load can be detected in 80.8% of the cases for all subjects. Figure 12 shows the ROC curve for this parameter within the three experimental blocks.

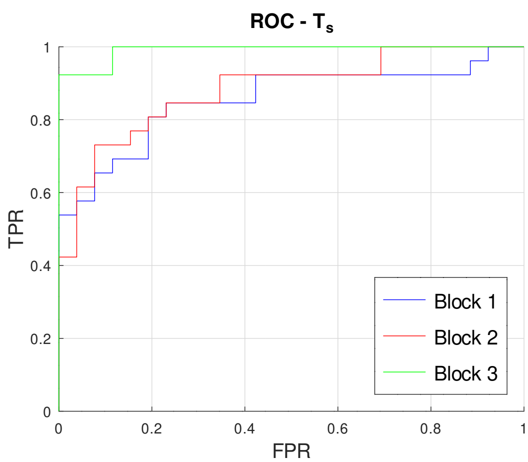

FIGURE 12: ROC curves of T_s for each of the blocks and with all participants.

With respect to the variable P_k, a reduction in the amount of movement in the SL_n periods had been observed with respect to the previous RL_n period. These decrements caused a ΔP_k negative and was statistically significant for all subjects. There were significant correlations between this variable and T_s (ρ= -0.9, p=0.01), N_R (ρ= 0.89, p=0.02) and WT (ρ= -0.81, p=0.05). This shows that an increase in the mean selection time results in a reduction in PA and, therefore, in a lower number of targets achieved or rounds.

Only 2 of the 12 EDA variables were useful, and they are similar since both measure the EDA's low frequency signal: Global.Mean is the average of the raw data, while CDA.Tonic is mainly composed of the low-frequency components of sweating. However, the behavior of this signal varied greatly from subject to subject. Results showed that skin conductance is higher at the SL_3 level than at the RL_3 one for most participants - 76.9%. That percentage decreased to 46% if the criterion is extended to all-phases-SL_n conductivity with respect to the previous RL_n. A possible reason is that not every level had an increased cognitive load which led to increased sympathetic activity. There are cases - 7.7% - where skin conductance was lower at the SL_n level than at its previous RL_n. This anomalous behavior for some subjects generalized to others if we observe a R_n with a RL_n. Sometimes the conductivity increases when we went from R_n to RL_n (34.6% of the cases), whereas in 26.9% it was the opposite. For the remaining volunteers, conductivity increases at some levels, but decreases at others. Following the same analysis as with T_s, we found, in this case, that the best accuracy that could be achieved with these variables would be 67%.

The ΔIHR variable was statistically significant. No other parameter derived from cardiac activity showed statistical significance. In contrast to the EDA variables, more homogeneous behavior was observed between subjects here, resulting in an increase in HR in the transitions from R_n to RL_n (change from rest to PA) and from RL_n to SL_n (increase in cognitive load), particularly in Block 3. This higher HR in the last level of each block coincides with a decrease in the amount of movement performed by the player (see Table 8), so it should be associated with the effects of a higher cognitive load.

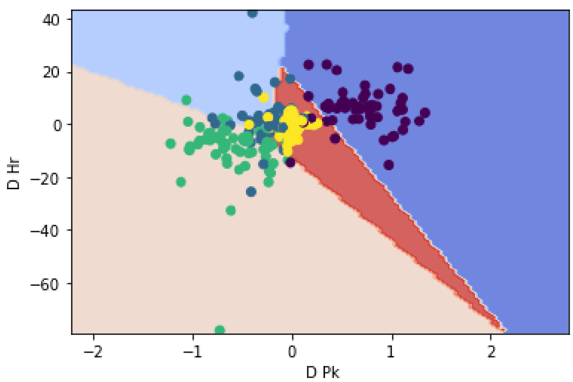

FIGURE 13: Feature space representation of ΔP_k and ΔHR of all subjets. In light green are the points associated with the SL_n-R_{n+1} change; those associated with RL_n-SL_n being in dark green; those with R_n-SL_n in magenta and the differences between the same level being in yellow.

Joining together the variables ΔP_k and ΔIHR could be used to identify situations of stress or higher cognitive load during PA. Let's suppose we use sliding windows with a 90 sec duration - the size used in the experiment. The purpose is to compare the PA and HR values in the current window with respect to the previous one. Figure 13 shows the scatter plot for four possible classes in the detection of physiological changes associated with intra and interlevel transitions. By using a classifier based on SVM with linear kernel, C=1.0, using 75% for training and the remainder for testing, we obtained an accuracy of 84.3% in the training set and 82.3% in the test set on average over 50 random iterations of the data set (26 volunteers multiplied by the number of blocks -3- and by the number of classes -4; the classes are associated with the change from R_n to RL_n; from RL_n to SL_n; from SL_n

to R_{n+1}; and with the continuity of the SL_n and R_n level).

Although, the window size used was large, it does not appear in reality that 90 sec would be a problem in detecting stressful situations in daily life. However, smaller window sizes, other classification techniques or other physiological variables that may enrich the classifier and improve its accuracy could be studied in further research.

Due to the limited existence of scientific literature in the field of stress or mental load detection during PA, there is a single paper of those selected in the scoping review with which we can compare our results. In [52], the authors try to determine whether people with vertigo maintain it in virtual reality tests in which the volunteers had to walk on a balance beam, they found that there was an increase in HR and response time to sound stimuli while decreasing the number of steps taken just by modifying the height in the virtual environment, which is consistent with the increase in HR and T_s variables and the decrease in P_k that has been detected in our trials.

A. LIMITATIONS

As this is a pilot study, the participant population was small. To achieve a precision level of 5% with a confidence level of 95% according to Cochran's equation, a sample of 385 people would be necessary.

VI. CONCLUSIONS

A *rehabgame* had been developed to check how mental load in light intensity exercises affects different variables, both physiological and non-physiological ones, and to try to determine whether the volunteer is -or is not- in a state of mental load while playing the game. With respect to the physiological variables analyzed, the experiment results show that a variation in the instantaneous heart rate (ΔIHR) and a variation in the kinetic power (ΔP_k) were the most significant ones, achieving an accuracy of 82.3% by using a SVM with a linear kernel. On the other hand, the selection time (T_s) and errors - both wrong target (WT) and MT - were also indicative of mental load, since they increased as the difficulty of the game grew. Thus, these variables could be used in adaptive difficulty rehabgames to keep people within the target zones for a greater amount of time.

As a future study, the effect of a decrease in the window size used for the classifier could be studied, and the results between different types of classifiers could be checked. However, the selected variables should also be checked to ensure that they are still viable in moderate or vigorous physical exercise, since in more intense activities, it may be difficult for physical fatigue itself or the increase in cognitive load during exercise to produce any measurable variations in heart rate. Furthermore, it would be useful to study what physiological changes occur when invalid targets are selected; to study the effect of figure color and shape on hit and miss rates; and to analyze the similarities and differences between the right and left hemispheres of the body in the EDA channels, using the second EDA channel that has not been analyzed in the present study.

ACKNOWLEDGMENT

The authors would like to thank the volunteers, for actively cooperating in the development of this work.

REFERENCES

[1] W. H. Organization, *A guide for population-based approaches to increasing levels of physical activity: implementation of the WHO global strategy on diet, physical activity and health*. World Health Organization, 2007.

[2] ——, "Global priority research agenda for improving access to high-quality affordable assistive technology," in *Global priority research agenda for improving access to high-quality affordable assistive technology*. World Health Organization, 2017.

[3] M. A. Stults-Kolehmainen and R. Sinha, "The effects of stress on physical activity and exercise," *Sports Medicine*, vol. 44, no. 1, pp. 81–121, Sep. 2013.

[4] W. H. Organization, *Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks*. World Health Organization, 2010.

[5] A. J. Tomiyama, "Stress and obesity," *Annual Review of Psychology*, vol. 70, no. 1, pp. 703–718, 2019.

[6] S. Misra and D. Stokols, "Psychological and health outcomes of perceived information overload," *Environment and Behavior*, vol. 44, no. 6, pp. 737–759, Apr. 2011.

[7] M. Merino, I. Gómez, and A. J. Molina, "Eeg feature variations under stress situations," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015, pp. 6700–6703.

[8] E. Garcia-Ceja, V. Osmani, and O. Mayora, "Automatic stress detection in working environments from smartphones' accelerometer data: A first step," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1053–1060, July 2016.

[9] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 2, pp. 156–166, June 2005.

[10] R. Komatireddy, A. Chokshi, J. Basnett, M. Casale, D. Goble, and T. Shubert, "Quality and quantity of rehabilitation exercises delivered by a 3-d motion controlled camera: A pilot study," *International Journal of Physical Medicine & Rehabilitation*, vol. 02, no. 04, 2014.

[11] N. Kang, J. Bai, J. Pan, and H. Qin, "Real-time animation and motion retargeting of virtual characters based on single rgb-d camera," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, March 2019, pp. 1006–1007.

[12] J. Barkman, K. Pfeiffer, A. Diltz, and W. Peng, "Examining energy expenditure in youth using XBOX kinect: Differences by player mode," *Journal of Physical Activity and Health*, vol. 13, no. s1, pp. 41–43, Jun. 2016.

[13] R. Cabrera, A. Molina, I. Gómez, and J. García-Heras, "Kinect as an access device for people with cerebral palsy: A preliminary study," *International Journal of Human-Computer Studies*, vol. 108, pp. 62–69, Dec. 2017.

[14] A. Molina-Cantero, J. Guerrero-Cubero, I. Gómez-González, M. Merino-Monge, and J. Silva-Silva, "Characterizing computer access using a one-channel EEG wireless sensor," *Sensors*, vol. 17, no. 7, p. 1525, Jun. 2017.

[15] J.-C. Wang, S.-H. Tsai, Y.-H. Chen, Y.-L. Chen, S.-J. Chu, and W.-I. Liao, "Kinect-based real-time audiovisual feedback device improves cpr quality of lower-body-weight rescuers," *The American Journal of Emergency Medicine*, vol. 36, no. 4, pp. 577 – 582, 2018.

[16] N. Komiya, M. Tokuoka, R. Egusa, S. Inagaki, H. Mizoguchi, M. Namatame, and F. Kusunoki, "Novel application of 3d range image sensor to caloric expenditure estimation based on human body measurement," in *2018 12th International Conference on Sensing Technology (ICST)*, Dec 2018, pp. 371–374.

[17] G. Brown, K. C Scheer, S. M Siebrandt, I. Shaw, and B. Shaw, "Wii, kinect, & move. heart rate, oxygen consumption, energy expenditure, and ventilation due to different physically active video game systems in college students," *International Journal of Exercise Science*, vol. 7, pp. 22–32, 01 2014.

[18] D. Nathan, D. Q. Huynh, J. Rubenson, and M. Rosenberg, "Estimating physical activity energy expenditure with the kinect sensor in an exergaming environment," *PLOS ONE*, vol. 10, no. 5, pp. 1–22, 05 2015.

[19] P. Tsou and C. Wu, "Estimation of calories consumption for aerobics using kinect based skeleton tracking," in *2015 IEEE International Conference on Systems, Man, and Cybernetics*, Oct 2015, pp. 1221–1226.

[20] B. Lange, C. Chang, E. Suma, B. Newman, A. S. Rizzo, and M. Bolas, "Development and evaluation of low cost game-based balance rehabilitation tool using the microsoft kinect sensor," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Aug 2011, pp. 1831–1834.

[21] J. Fasola and M. J. Matarić, "Socially assistive robot exercise coach: Motivating older adults to engage in physical exercise," in *Experimental Robotics*. Springer International Publishing, 2013, pp. 463–479.

[22] N. Jafari, K. Adams, M. Tavakoli, S. Wiebe, and H. Janz, "Usability testing of a developed assistive robotic system with virtual assistance for individuals with cerebral palsy: a case study," *Disability and Rehabilitation: Assistive Technology*, vol. 13, no. 6, pp. 517–522, Jul. 2017.

[23] R. Riener and M. Harders, *Virtual Reality in Medicine*. Springer, apr 2012.

[24] H. A. Hernandez, Z. Ye, T. N. Graham, D. Fehlings, and L. Switzer, "Designing action-based exergames for children with cerebral palsy," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2013, pp. 1261–1270.

[25] S. Hijazi, A. Page, B. Kantarci, and T. Soyata, "Machine learning in cardiac health monitoring and decision support," *Computer*, vol. 49, no. 11, pp. 38–48, Nov 2016.

[26] C. Thirumalai, A. Duba, and R. Reddy, "Decision making system using machine learning and pearson for heart attack," in *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, vol. 2, April 2017, pp. 206–210.

[27] N. Hjortskov, D. Rissén, A. K. Blangsted, N. Fallentin, U. Lundberg, and K. Søgaard, "The effect of mental stress on heart rate variability and blood pressure during computer work," *European Journal of Applied Physiology*, vol. 92, no. 1, pp. 84–89, Jun 2004.

[28] J. Cool and D. Zappetti, *The Physiology of Stress*. Cham: Springer International Publishing, 2019, pp. 1–15.

[29] W. Boucsein, *Electrodermal Activity*. Boston, MA: Springer US, 2012.

[30] S. Galán-Cuevas and E. Camacho-Gutiérrez, *Estrés y salud. Investigación básica y aplicada 1a.ed.* México D.F.: Manual Moderno, 2014.

[31] G. Liguori and American College of Sports Medicine (ACSM), *ACSM's guidelines for exercise testing and prescription*, 11th ed., ser. American College of Sports Medicine. Baltimore, MD: Wolters Kluwer Health, Apr. 2021.

[32] M. Peters, C. Godfrey, P. Mcinerney, C. Soares, H. Khalil, and D. Parker, *Methodology for JBI Scoping Reviews*. Joanna Briggs Institute, 08 2015, pp. 1–24.

[33] S. Amalan, A. Shyam, A. Anusha, S. Preejith, A. Tony, J. Jayaraj, and S. Mohanasankar, "Electrodermal activity based classification of induced stress in a controlled setting," in *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2018.

[34] S. Dong, M. Lee, H. Park, and I. Youn, "Stress resilience measurement with heart-rate variability during mental and physical stress," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 5290–5293.

[35] K. Dedovic, R. Renwick, N. K. Mahani, V. Engert, S. J. Lupien, and J. C. Pruessner, "The montreal imaging stress task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain," *Journal of Psychiatry and Neuroscience*, vol. 30, no. 5, p. 319, 2005.

[36] M. M. Bradley and P. J. Lang, "Measuring emotion: the self-assessment manikin and the semantic differential," *Journal of behavior therapy and experimental psychiatry*, vol. 25, no. 1, pp. 49–59, 1994.

[37] M. T. Hegel, E. D. Collins, S. Kearing, K. L. Gillock, C. P. Moore, and T. A. Ahles, "Sensitivity and specificity of the distress thermometer for depression in newly diagnosed breast cancer patients," *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer*, vol. 17, no. 6, pp. 556–560, 2008.

[38] J. Taelman, S. Vandeput, I. Gligorijević, A. Spaepen, and S. Van Huffel, "Time-frequency heart rate variability characteristics of young adults during physical, mental and combined stress in laboratory environment," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 1973–1976.

[39] M. Wu, H. Cao, H. Nguyen, K. Surmacz, and C. Hargrove, "Modeling perceived stress via hrv and accelerometer sensor streams," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 1625–1628.

[40] K. Plarre, A. Raij, S. M. Hossain, A. A. Ali, M. Nakajima, M. Al'absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, D. Siewiorek, A. Smailagic, and L. E. Wittmers, "Continuous inference of psychological stress from sensory measurements collected in the natural environment," in *Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks*, 2011, pp. 97–108.

[41] M. Parent, A. Tiwari, I. Albuquerque, J. Gagnon, D. Lafond, S. Tremblay, and T. H. Falk, "A multimodal approach to improve the robustness of physiological stress prediction during physical activity," in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Oct 2019, pp. 4131–4136.

[42] C. Dobbins and S. Fairclough, "Signal processing of multimodal mobile lifelogging data towards detecting stress in real-world driving," *IEEE Transactions on Mobile Computing*, vol. 18, no. 3, pp. 632–644, 2019.

[43] C. Spielberger, "The state-trait anger expression inventory-2 (staxi-2): Professional manual. lutz, fl: Psychological assessment resources," 1999.

[44] G. Matthews, D. M. Jones, and A. G. Chamberlain, "Refining the measurement of mood: The uwist mood adjective checklist," *British journal of psychology*, vol. 81, no. 1, pp. 17–42, 1990.

[45] K. Li and Y. Han, "Study of selective ensemble learning method and its diversity based on decision tree and neural network," in *2010 Chinese Control and Decision Conference*. IEEE, 2010, pp. 1310–1315.

[46] T. Althobaiti, S. Katsigiannis, D. West, and N. Ramzan, "Examining human-horse interaction by means of affect recognition via physiological signals," *IEEE Access*, vol. 7, pp. 77 857–77 867, 2019.

[47] J. Ogorevc, A. Podlesek, G. Geršak, and J. Drnovšek, "The effect of mental stress on psychophysiological parameters," in *2011 IEEE International Symposium on Medical Measurements and Applications*, 2011, pp. 294–299.

[48] G. Borg, "Perceived exertion as an indicator of somatic stress," *Scandinavian journal of rehabilitation medicine*, 1970.

[49] C. Jao, Y. Chen, T. H. Huang, C. Tseng, C. Yang, C. Y. Lin, S. J. Tsai, P. Wang, and Y. Wu, "Status change revealed by electrocardiography (ecg) and electroencephalography (eeg) during cycling exercise," in *2017 International Automatic Control Conference (CACS)*, 2017, pp. 1–5.

[50] A. Anusha, J. Jose, S. Preejith, J. Jayaraj, and S. Mohanasankar, "Physiological signal based work stress detection using unobtrusive sensors," *Biomedical Physics and Engineering Express*, vol. 4, no. 6, 2018.

[51] A. Affanni, R. Bernardini, A. Piras, R. Rinaldo, and P. Zontone, "Driver's stress detection using skin potential response signals," *Measurement: Journal of the International Measurement Confederation*, vol. 122, pp. 264–274, 2018.

[52] S. M. Peterson, E. Furuichi, and D. P. Ferris, "Effects of virtual reality high heights exposure during beam-walking on physiological stress and cognitive loading," *PLOS ONE*, vol. 13, no. 7, pp. 1–17, 07 2018.

[53] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss, "Activity-aware mental stress detection using physiological sensors," *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, vol. 76 LNICST, pp. 211–230, 2012.

[54] Y. S. Can, N. Chalabianloo, D. Ekiz, J. Fernández-Álvarez, C. Repetto, G. Riva, H. Iles-Smith, and C. Ersoy, "Real-life stress level monitoring using smart bands in the light of contextual information," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8721–8730, 2020.

[55] S. Cohen, T. Kamarck, and R. Mermelstein, "A global measure of perceived stress," *Journal of health and social behavior*, pp. 385–396, 1983.

[56] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.

[57] F. Delmastro, F. D. Martino, and C. Dolciotti, "Cognitive training and stress detection in mci frail older people through wearable sensors and machine learning," *IEEE Access*, vol. 8, pp. 65 573–65 590, 2020.

[58] A. A. S., S. P., S. V., S. K. S., S. A., T. J. Akl, P. S. P., and M. Sivaprakasam, "Electrodermal activity based pre-surgery stress detection using a wrist wearable," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 92–100, 2020.

[59] B. B. Gibbs, R. J. Kowalsky, S. J. Perdomo, M. Grier, and J. M. Jakicic, "Energy expenditure of deskwork when sitting, standing or alternating positions," *Occupational Medicine*, vol. 67, no. 2, pp. 121–127, 08 2016.

[60] C. Setz, B. Arnrich, J. Schumm, R. La Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable eda device," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 410–417, March 2010.

[61] W. Boucsein, D. C. Fowles, S. Grimnes, G. Ben-Shakhar, W. T. Roth, M. E. Dawson, and D. L. Filion, "Publication recommendations for electrodermal measurements," *Psychophysiology*, vol. 49, no. 8, pp. 1017–1034, 2012.

[62] J. A. Castro-García, A. J. Molina-Cantero, M. Merino-Monge, and I. M. Gómez-González, "An open-source hardware acquisition platform for physiological measurements," *IEEE Sensors Journal*, vol. 19, no. 23, pp. 11 526–11 534, Dec 2019.

[63] A. Molina-Cantero, J. Castro-García, C. Lebrato-Vázquez, I. Gómez-González, and M. Merino-Monge, "Real-Time Processing Library for Open-Source Hardware Biomedical Sensors," *Sensors 2018, Vol. 18, Page 1033*, vol. 18, no. 4, p. 1033, mar 2018.

[64] M. Merino-Monge, A. J. Molina-Cantero, J. A. Castro-García, and I. M. Gómez-González, "An easy-to-use multi-source recording and synchronization software for experimental trials," *IEEE Access*, vol. 8, pp. 200 618–200 634, 2020.

[65] C. Kothe, "Lab streaming layer (lsl)," https://github.com/sccn/labstreaminglayer, [Accessed: 2018-07-18].

[66] P. V. Kishore, P. S. Kameswari, K. Niharika, M. Tanuja, M. Bindu, D. A. Kumar, E. K. Kumar, and M. T. Kiran, "Spatial joint features for 3d human skeletal action recognition system using spatial graph kernels," *International Journal of Engineering & Technology*, vol. 7, no. 1.1, p. 489, Dec. 2017.

[67] M. A. Hertzog, "Considerations in determining sample size for pilot studies," *Research in Nursing & Health*, vol. 31, no. 2, pp. 180–191, 2008.

[68] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *Journal of Neuroscience Methods*, vol. 190, no. 1, pp. 80–91, 2010.

[69] F. Shaffer and J. P. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in Public Health*, vol. 5, p. 258, 2017.

[70] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Medical and Biological Engineering and Computing*, vol. 44, no. 12, pp. 1031–1051, Dec 2006.

[71] S. Carey, "Bootstrapping & the origin of concepts," *Daedalus*, vol. 133, no. 1, pp. 59–68, 2004.

JUAN A. CASTRO-GARCÍA (Member, IEEE) was born in Lora del Río, Seville, in 1986. He received the master's degree in computers and networks engineering from the Universidad de Sevilla, in 2011, and the Ph.D. degree from the TAIS Research Group (Technologies for Care, Inclusion and Health), University of Seville, in 2019. His research interests include signal processing, sensors, and human–computer interfaces.

ALBERTO J. MOLINA-CANTERO was born in Lucena, Córdoba, in 1967. He received the B.S. degree in physics with intensification in electronics, in Seville, in 1990, and the Ph.D. degree from the University of Seville, in 2010. He has been working as a Lecturer with the University of Seville, since 1990. He is the author of five books, more than 30 research articles, holds three inventions, and took part in more than ten projects. His research interests include signal processing, sensors, and the development of new devices or techniques for helping people to access a computer.

MANUEL MERINO-MONGE was born in Seville, Spain, in 1983. He received the master's degree in computer engineering from the University of Seville, in 2010, and the Ph.D. degree from the TAIS Research Group (Technologies for Care, Inclusion and Health), University of Seville, in 2015. His research interests include biomedical signal processing, affective computing, human–computer interface, augmentative and alternative communication, and assistive technology.

CLARA LEBRATO-VÁZQUEZ was born in Seville, Spain, in 1993. She received the B.Sc. degree in health engineering, the M.Sc. in water engineering and the M.Sc. in biomedic engineering from the Universidad de Sevilla, in 2015 and 2020 respectively. Since 2021 she is with the Department of Electronic Technology of the University of Seville as adjunct professor. She has been an author of a number of peer-reviewed journal publications and her work has been presented at several international conferences. Her research interests include the development and testing of sensors and interfaces for computer accessibility.

ISABEL M. GÓMEZ-GONZÁLEZ (Senior Member, IEEE) has been a Senior Researcher and a full-time Ph.D. Lecturer with the Electronic Technology Department, Universidad de Sevilla, since 1990. Her teaching is centered on digital electronic and biomedical technology disciplines in bachelor's and master's degrees of computer science. She has directed numerous final career projects about the design and implementation of serious games and toys adaptation in order to improve the cognitive and motor skills of adults and children with cerebral palsy. She is also the Head of the TAIS Research Group (Technologies for Care, Inclusion and Health), where she has been managing several Ph.D. thesis and national research projects related to human–computer interfaces, biosignal processing, and ambient assisted living.