Effects of Presence and Challenge Variations on Emotional Engagement in Immersive Virtual Environments

Oscar I. Caldas, Oscar F. Aviles, and Carlos Rodriguez-Guerrero

***Abstract*—Serious games and immersive virtual reality promote emotional engagement during learning tasks, mostly by providing (1) skill-adapted challenges with performance feedback (for trial and error learning) and (2) enhanced presence (further reactions to multimodal stimuli), respectively. However, it is still unclear how each of these two strategies independently influence emotional states to engage subjects to a task. This study assessed the dimensions of emotion (valence-arousal-dominance) of 87 healthy subjects in a virtual game, assigned to 2 groups that were exposed to a different set of 5 trials: Group A experienced game variations by virtual factors affecting user's presence, whereas group B experienced levels of difficulty, affecting challenge. Emotional reports and 26 features extracted from physiological signals were statistically analyzed. Results showed that presence-based experimental conditions were able to modify the sense of arousal, whereas valence and dominance responded to challenge variation, i.e. were positively correlated with game score. Arousal is likely to increase with low sense of coexistence (social presence) and decrease with low scenario realism (physical presence). Faster breathing and higher skin conductance (SC) were detected at high challenge, whereas heart rate variability and SC increased with higher arousal. The evidence from this study suggests that both strategies can be used to separately influence dimensions of emotion, pointing out the customization of presence-based factors as a promising method to adjust emotional engagement by impacting arousal. Further research should be undertaken to identify the independent effect of single presence factors on emotional states.**

***Index Terms*—Emotion recognition, engagement, presence, psychophysiology, virtual reality.**

Manuscript received November 8, 2019; revised March 3, 2020; accepted March 29, 2020. Date of publication April 3, 2020; date of current version May 8, 2020. This work was supported in part by the UMNG's Research Vice-Rectorate under Grant ING-IMP-2657. *(Corresponding author: Oscar I. Caldas.)*

Oscar I. Caldas is with the Faculty of Engineering, Universidad Militar Nueva Granada, Bogota 110111, Colombia, and also with the Department of Mechanical Engineering, Vrije Universiteit Brussel, 1050 Brussels, Belgium (e-mail: oscar.caldas@unimilitar.edu.co).

Oscar F. Aviles is with the Davinci Research Group, Faculty of Engineering, Universidad Militar Nueva Granada, Bogota 110111, Colombia (e-mail: oscar.aviles@unimilitar.edu.co).

Carlos Rodriguez-Guerrero is with the Department of Mechanical Engineering, Vrije Universiteit Brussel, 1050 Brussels, Belgium, and also with Flanders Make, 3920 Lommel, Belgium (e-mail: carlos.rodriguez.guerrero@vub.be).

Digital Object Identifier 10.1109/TNSRE.2020.2985308

I. Introduction

VIRTUAL reality (VR) and serious games have emerged together in recent years as powerful platforms for cognitive and psychomotor learning processes, i.e. education [1], training and rehabilitation [2], [3], essentially because serious games have proven to promote motivation and engagement [4], but also because VR contributes by enhancing them with repetition and feedback [5].

Studies such as those conducted by Fraser *et al.* and Lohse *et al.* have shown that repetition is necessary to produce skill learning and long-term retention [6], [7], and when used on task-specific exercises with feedback about performance, it can also promote neuroplasticity [5], [8]. Moreover, positive psychological states such as motivation and engagement are key factors to avoid boredom and loss of concentration after exposure to repetitions [9], [10], especially in motor tasks that need to be relatively simplistic (i.e. barely challenging) [11].

Surprisingly, the intrinsic potential of VR to provide individualized and enriched environments and its effects into the emotional states have not been closely examined yet [12]. However, we hypothesize that responses to immersive and coherent environments can be separated from responses to different kind of challenges, and therefore a further understanding of their independent effect should be addressed.

The main aim of this study is to separately investigate the effects of challenge-based and presence-based stimuli in VR environments, since it is unclear how each of these strategies influence in the subject's emotional states. Since reaching extremely low or high values of emotional states might cause a disharmonious experience and thus an impact on the willingness to carry on (i.e. engagement) [13]–[15], the primary goal was to identify significant changes of emotions after exposure to separated stimuli. The study used the dimensional approach for emotions: valence (from unpleasant to pleasant), arousal (from calmed to excited), and dominance (from dominated to dominant), which helps to reduce self-reports to only 3 graphical factors (easier for participants, as suggested by Nielsen and Kaszniak [16]). Later, significant positive or negative changes in emotions were analyzed based on the stimulus provided.

In order to address this hypothesis, a VR setup was designed to expose subjects to an immersive game-based task and to self-report their experience after one out of two sets of game scenarios: a) 5 trials with different difficulty levels, and b) 5 trials with different VR primitives (key factors) whose inclusion/exclusion is capable to change the sense of presence.

We also investigated how psychophysiological responses (i.e. signals associated with emotional states reflected by the central and autonomic nervous systems) correlate with subjects' reported emotions in both experimental sets of trials, as suggested on similar recent studies [17]–[20].

II. Background

A considerable amount of studies in VR recognizes the association between psychological changes and subjective responses to different types of environmental interactions and multimodal stimuli that are able to place physical and cognitive challenges [21], [22]. To provide adequate challenges in VR, theories of motivation and flow (i.e. difficulty modulation to retain a desired challenge/skill ratio [20], [23]) have been used extensively, but as claimed by Hocine *et al.* these strategies could still fail to integrate subjects′ responses to the task, such as physical effort and reactions to external stimuli [24].

However, there is an increasing amount of literature on investigating the interrelation between performance and presence as well, being presence the subjective psychological response to a VR system [25]–[27]. Skarbez *et al.* argued that high presence is reflected on a realistic response to stimuli from the virtual scenario [28], which happens when the subject experiences both illusions of real spatiality (high immersion) and real behavior (high coherence), or as proposed by Slater, Place Illusion (PI) and Plausibility Illusion (PSI), respectively [29]. Since PI depends on immersion, it could be obtained via tech allocation, or as previous researchers have established, by using technology to provide good sensory stimulation (image and sound quality) and ability to navigate (tracking level, stereoscopic vision, field of view, and refreshing rate) [30]–[32]. On the other hand, PSI could be modified by means of different environmental factors, generally agreed but called differently by authors, such as the 4 "coherence factors" introduced by Skarbez *et al.* (virtual-human behavior, subject′s virtual body, scenario coherence, and physical interaction) or the equivalent attributes proposed by Makranski *et al.* and classified into 3 dimensions of presence: sense of coexistence (social presence), bodily connectivity (self-presence), and both physical realism and control/act in VR (physical presence), [33]. If it is clear that PI and PSI are orthogonal components of presence, good VR technology could be provided to set fairly high levels of immersion and, consequently, allow presence to be adjusted in terms of coherence-factors/presence-attributes to modify the subjective response to the VR environment.

III. Methods

A. Hardware

The setup shown in Fig. 1 includes a rigid support structure fixed to the ground. The subject was attached to it to avoid falls and provided with hand grasps for self-sense of safety.

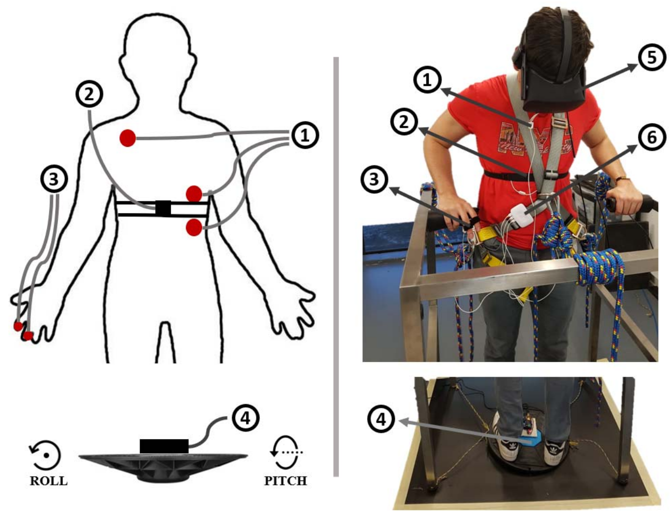

Fig. 1. Subject performing the task: schematic (left) and photograph (right) showing sensors placement. Sensors: 3 ECG electrodes on the chest (1), respiration strap (2), 2 GSR electrodes on index and middle fingers (3) and gyroscope over the balance board (4); devices: head-mounted display (5) and acquisition/communication device (6).

A Wobble board was used (diameter = 35 mm, capacity = 135 kg), but restricted to only two rotational degrees of freedom, so that subjects could stand over the balance board and perform part rotation about the transverse/lateral axis (i.e., pitching or backward-forward tilting) and part rotation about the longitudinal axis (i.e., rolling or left-right tilting) by following online instructions, similar to [34]. A 3-axis gyroscope was located to measure both rotations (range of motion: $-11°$ to $11°$).

Audiovisual stimuli and feedback were provided via an Oculus Rift HMD (Head-Mounted Display) with on-ear headphones, offering full immersion by position/orientation tracking, latency $<$ 10 ms, refreshment rate = 90 Hz, per-eye resolution = 1080x1200, and field-of-view = 110°.

Psychophysiological signals were acquired via a Biosignal-sPlux Explorer research kit (sampling rate = 1kHz, resolution = 16 bits) to measure electrocardiogram (ECG), Galvanic Skin Response (GSR), and Respiration (RSP). As in Fig. 1-left, ECG was sampled by 3 surface electrodes placed on the upper-right part of the chest, its downer-left part and over the floating ribs (ground). A GSR sensor measured skin conductance by 2 surface electrodes placed on the index and middle fingers (non-dominant hand). RSP was measured by a piezoelectric sensor in an elastic strap around the chest.

B. Virtual Environment Setup

The HMD-headphones-board setup allowed the subject to play a first-person skydiving game that asked to pass through colored rings before landing (see Fig. 2). The jump from the helicopter and the parachute deployment were automatically animated, and thus the player only had to tilt over the balance board to move the falling player and reach 1 ring in each of the 8 vertically-separated 4-ring sets (the game indicates the correct ring: yellow, red, blue or green), and to accurately land over a target placed in an island. The player could obtain points

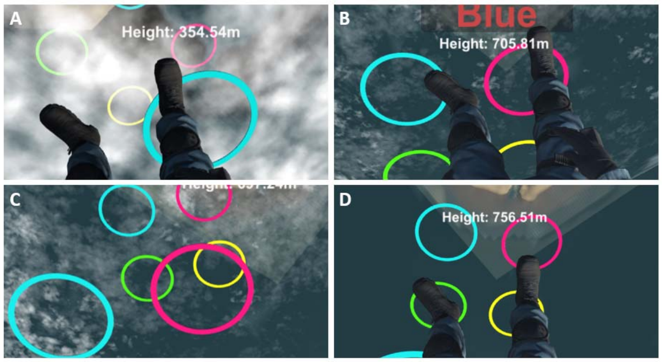

Fig. 2. First-person view of a subject performing the VR task (pass through the right ring) on different tests (modified scenarios): (A) Full presence-related assets presented to the subject, including animations, voice assistance, sounds, and embodiment. (B) Social presence diminished by replacing voice assistance by on-screen simple words as instructions. (C) Self-presence diminished by removing the body representation. (D) Physical presence diminished by removing environmental assets: clouds, trees, and sea animations, wind/sea/helicopter sounds and inertial effect of parachute deployment. A fifth option was B, C and D conditions at the same time.

at each of the 8 altitude levels (+10 for the right ring, 0 if missing and -10 points for any of the 3 wrong rings), as well as for landing right over the target (20 extra points), being 100 the maximum reachable score. The virtual environment was developed using Unity3D (version 2018.3.14f1).

C. Self-Report Questionnaires

Participants filled the Virtual Reality Sickness Questionnaire (VRSQ) proposed by Kim *et al.* [35] and similar to [34]. Nine oculomotor (general discomfort, fatigue, eyestrain and difficulty in focusing) or disorientation (headache, fullness of head, blurred vision, dizziness with eyes closed and vertigo) symptoms were assessed before and after the VR exposure.

After each trial, subjects were asked to answer the graphical 9-point scale test for emotional states given by Bradley and Lang called as Self-Assessment Manikin (SAM) [36]. This instrument allow to self-rate experiences in terms of arousal, valence, and dominance. It was easy to apply due to its non-verbal graphic depiction, but still reliable compared with other more complex self-report scales, as tested by Goljar *et al.* [37].

D. Feature Extraction

Twenty-six (26) features in time and frequency domain were extracted from the physiological signals during the task: ECG (9), GSR (11), and RSP (6). These signals were pre-processed by applying smoothing, notch, and band-pass filtering, and analyzed by morphology-based searching algorithms.

The ECG algorithm used adaptive heuristics to measure time between two normal heartbeats (N-N intervals) and thus to calculate a heart rate (HR) time series. Later, mathematical approaches helped to calculate its mean, median, maximum, and minimum values (μHR, $medianHR$, $maxHR$, and $minHR$); HR variability (HRV) was also calculated in both time domain: standard deviation (σHR) and root mean square of successive differences ($RMSSD$), and frequency domain: absolute Power Spectral Density for Low Frequencies [0.04-0.15 Hz] (aLF), for High Frequencies [0.15-0.4 Hz] (aHF) and ratio (LF/HF).

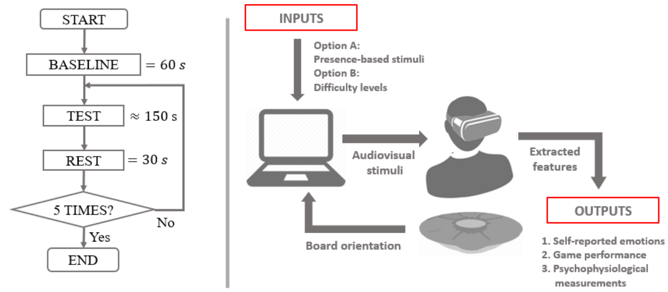

Fig. 3. Left: flowchart describing the 5-trials study. Trials duration changed if subject was assigned to group B. Right: General scheme describing the open-loop setup prepared for the study.

GSR analysis followed the well-defined GSR morphology to detect the signal phasic component (Skin conductance responses) and separate it from the tonic component (skin conductance level), by means of a state machine approach. Six (6) features were calculated from SCL: mean and standard deviation (μSCL and σSCL), global maximum, minimum and difference ($maxSCL$, $minSCL$ and $SCLrange$), and final-initial (ΔSCL); five (5) others were calculated after detecting each wave base and peak: number of SCRs/time ($nSCR$), mean amplitude and rise time (μSCR_A and μSCR_{tr}), and standard deviations (σSCR_A and σSCR_{tr}).

Finally, respiration features were mean respiratory rate and standard deviation (μRR and σRR), longest and shortest time between consecutive breaths ($maxTCB$ and $minTCB$), max deep breath ($Deep$), and min shallow breath ($Shallow$).

E. Study Design and Procedure

In order to objectively measure both emotions and psychophysiological signals, it was decided to compare how different versions of the virtual environment affect the subjects, similar to studies related to motivation assessment [38]. Therefore, participants were exposed to 5 different instances of the VR environment (approx. 150s each) to experience 5 experimental conditions (from now referred as Trials), after a 1-minute-long relaxing stage to acquire baseline data. Trials are presented randomly to avoid biases and subjects are also randomly assigned to either group A or B before starting, which defines what kind of variations will be deployed from trial to trial. The experimental procedure is explained by the scheme and flowchart shown in Fig. 3.

1) Group A - Affecting Virtual Presence: The VR environment experienced by participants was similar in all trials, except for the attributes that were included/excluded to affect presence, based on the Multimodal Presence Scale (MPS) [33] and the "coherence factors" of Skarbez *et al.* [28]. The following 3 sets of VR assets are proposed to affect presence if removed from the scene. As can be seen in Table I, trial 1 included all 3 sets, whereas in trials 2, 3, and 4 one set was removed, and trial 5 did not included any of the sets.

- Physical presence = Realism (scenario coherence).
  In VR: Audiovisual realistic animations, e.g. animated clouds, ocean, wind, and helicopter.

TABLE I
VARIABLES COMBINATION PER TRIAL. GROUP A: SETS OF PRESENCE-BASED VR ASSETS. GROUP B: DIFFICULTY LEVELS

	A - Presence			B - Difficulty		
No.	Social	Physical	Self	Rings	Speed	Quiz
1	Yes	Yes	Yes	Big	Slow	No
2	No	Yes	Yes	Medium	Slow	No
3	Yes	No	Yes	Medium	Fast	No
4	Yes	Yes	No	Small	Fast	No
5	No	No	No	Small	Fast	Yes

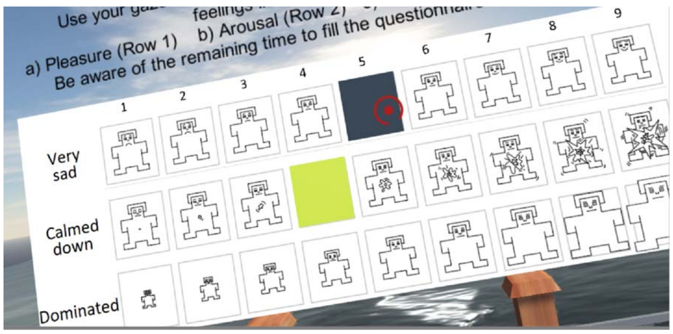

Fig. 4. VR Scene to fill the SAM test during the rest period. Subjects are asked to select pictures that reflect their recent experience in terms of valence, arousal, and dominance. Notice that the subject is able to select the desired options by gaze, with no external intervention.

- Social presence = Coexistence (virtual-human behavior). In VR: Simulated-radio instructions instead of plain text displayed on the screen.
- Self-presence = Bodily connectivity (virtual body). In VR: Avatar′s body is visible.

2) Group B - Affecting Game Difficulty: Subjects in Group B experienced difficulty changes in terms of a combinations of ring size, falling speed and inclusion of a general trivia quiz, as described in Table I. A trivia quiz was added as multiple-choice questions with 4 colored answers (right-answer's color indicated the ring to reach). Easy trivia questions were used to only elicit loss of time instead of high cognitive effort, which would have been out of this study's scope. As can be inferred, higher falling speed reduces the test duration.

3) Assessing Dimensions of Emotion: After each of the 5 trials, the 30-sec resting lapse shown in Fig. 3 was used to fill the SAM test. Koenig *et al.* warned that risk of disturbance on the physiological analysis is possible if distractions are permitted (e.g. speaking or writing) [39], hence a different VR scene was automatically rendered to perform hands-free gaze selection, without compromising the VR-setup (see Fig. 4).

F. Participants

n = 87 subjects were recruited, 36 females (41.38%) and 51 males (58.62%), mean age 24.64 (SD = 8.41). Inclusion criterion: healthy adults able to understand and sign informed consent. Exclusion criteria: 1) any pathology affecting body balance or mobility/strength in limbs and head, and thus impeding them to perform motion tasks; 2) any neurological condition or state not convenient for VR simulations (i.e. epilepsy, vertigo, seizure history, dizziness, etc.); 3) any cognitive impairment critical to understand and follow instructions.

One (1) male was removed after reporting eyestrain and headache during the VRSQ test, one (1) female presented extremely noisy ECG signal at Trial No.4, one (1) male had respiration sensor disconnected during the test and 7 participants had problems with the GSR sensor or electrodes, i.e. with either signal saturation or detachment. Subjects did not repeat the test and corresponding raw data were removed.

- Group A subjects: 38 (ECG) + 37 (GSR) + 38 (RSP)
- Group B subjects: 47 (ECG) + 42 (GSR) + 47 (RSP)

G. Statistical Analysis

A total of 30 features were collected from participants after the trials, which were organized as follows:

- Ordinal (4): 3 emotional (Valence, Arousal, Dominance); 1 Performance (game score)
- Continuous (26): 26 Psychophysiological (section III-D)

All analyzes were carried out using Matlab (2018b). Descriptive statistical analysis was firstly performed to graphically observe variations of emotions and performance along trials, i.e. central tendency (mean and median) and dispersion (inter-quartile range) from self reports and game-score data. The tested null hypothesis was having no significant differences in the mean value of each independent variable in all trials, as in (1), using a desired overall alpha level of $\alpha = 0.05$.

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5. \quad (1)$$

where μ_i is the mean of the observed feature for each trial.

Since emotions and score were ordinal categorical variables and deviated from normality (Kolmogrov-Smirnov tested), the non-parametric Friedman's analysis of variance (Friedman's ANOVA) was performed to assess the overall differences between experimental conditions, and Kendall's W coefficient to estimate effect sizes. The post-hoc test to perform pairwise comparisons was Wilcoxon signed-rank test, with a Bonferroni correction ($p < \alpha/n$; $n = 5$). On the other hand, Repeated Measures ANOVA (within-subjects) was performed to test significant differences in psychophysiological features, and partial eta-squared (η_p^2) was used to calculate effect sizes.

Paired sample t-test was used to evaluate differences between both groups for the VRSQ (before and after exposure).

Finally, Spearman rank-order correlation (ρ) was calculated to measure the strength of association between the dimensions of emotion and the extracted psychophysiological features.

IV. RESULTS

The database containing all gathered data, as well as basic signal processing algorithms, is available online at IEEE DataPort: http://dx.doi.org/10.21227/vj8w-v224 [40].

We first analyzed the results of the self-reports (SAM test) to verify that emotional changes were successfully induced with the different experimental conditions presented to all

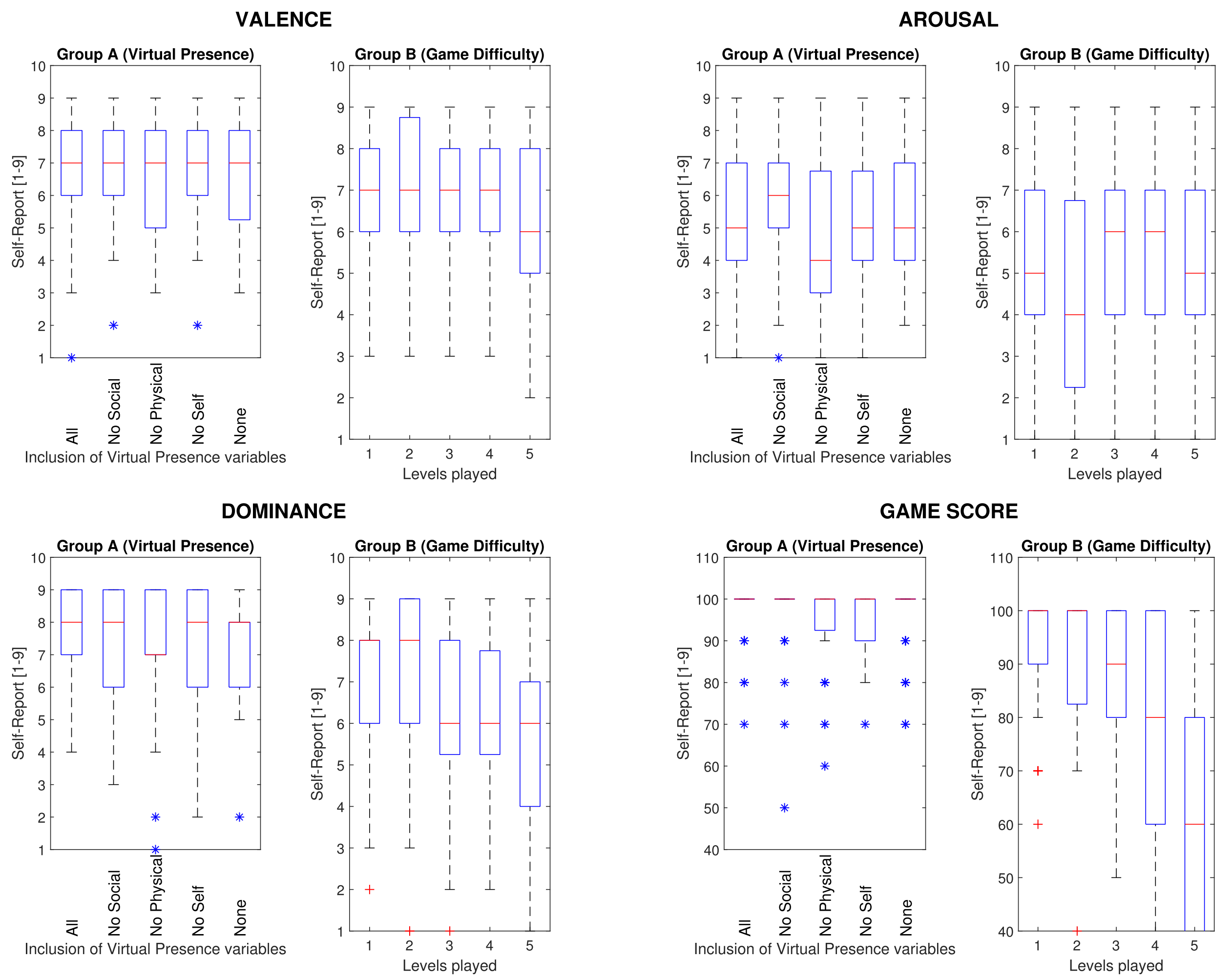

Fig. 5. Changes on emotions and performance between experimental conditions (trials). Group A: Virtual presence; Group B: Game difficulty.

participants (described in Section III-D), i.e. showed significant changes between trials. Then, pairwise comparisons were performed to identify the association between emotional changes and specific presence-based or challenge-based factors in trials. The same analysis was done with significantly changing psychophysiological features.

No significant differences were found between the VRSQ results before and after the VR exposure ($p = 0.47$).

A. Effects of Presence-Based Factors on Emotions

Table II presents the results obtained from the analysis of variance performed over the ordinal features (emotions and performance) between trials. As can be seen, only arousal changed significantly due to inclusion/exclusion of presence-related factors, i.e. reported by participants in group A. The post-hoc test reported significant pairwise differences between trials 2-3, 2-5, and 3-4 ($p < 0.05$). The direction of the differences can be identified by observing Fig. 5.

None of the psychophysiological features extracted from participants' signals in Group A showed significant differences between the 5 experimental conditions. Nonetheless, some of they showed significant correlation ($p = 0.00$) with arousal (the only emotion that changed between trials): aLF ($\rho = -0.20$), LF/HF ($\rho = -0.25$), and $nSCR$ ($\rho = 0.23$).

TABLE II
DIFFERENCES BETWEEN ORDINAL FEATURES AND EFFECT SIZES FOR THE 5 EXPERIMENTAL CONDITIONS

Feature	H_0	Effect size	H_0	Effect size
	Group A		Group B	
Score	$p = 0.71$	W = 0.65	**p = 0.00**	W = 0.44
Valence	$p = 0.40$	W = 0.72	**p = 0.01**	W = 0.59
Arousal	**p = 0.05**	W = 0.76	$p = 0.12$	W = 0.59
Dominance	$p = 0.71$	W = 0.67	**p = 0.00**	W = 0.55

* In bold: $p < 0.05$ (rejecting the null hypothesis)

B. Effects of Challenge-Based Factors on Emotions

Table II also provides evidence of significant differences in score, valence, and dominance between trials with different challenge factors, i.e. reported by participants in Group B (also corroborated by Fig. 5). Pairwise differences ($p < 0.05$) were reported for valence between trials 1-5 and 3-5, for dominance between all trials excepting between 1-2, and for game score between all trials excepting 1-2 and 2-3.

10 psychophysiological features showed differences between trials (shown in Table III). What is interesting about the data is that significant changes (i.e. $p < 0.05$) in ECG were only detected in features related with frequency-domain

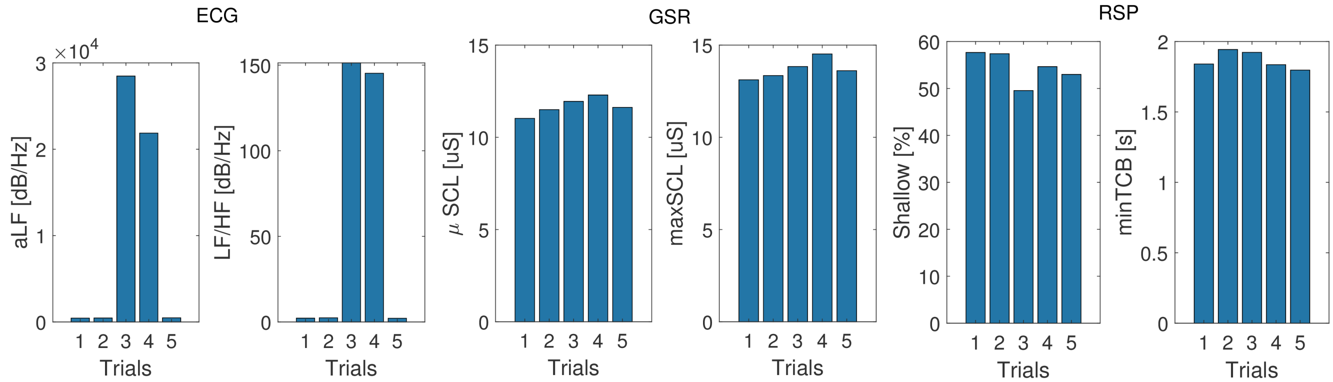

Fig. 6. Results for the two features per physiological signal that showed the highest differences between trials in group B (from Table III), i.e. main values of the feature in each of the 5 levels of difficulty defined in Table I.

TABLE III
PSYCHOPHYSIOLOGICAL FEATURES WITH SIGNIFICANT DIFFERENCES BETWEEN TRIALS (GROUP B ONLY)

	Signal	Feature	H_0	Effect size
7	ECG	aLF	**p = 0.00**	η_p^2 = 0.97
9		LF/HF	**p = 0.00**	η_p^2 = 0.99
10	GSR	μSCL	**p = 0.00**	η_p^2 = 0.81
13		$maxSCL$	**p = 0.01**	η_p^2 = 0.78
14		$minSCL$	**p = 0.00**	η_p^2 = 0.75
20		$nSCR$	**p = 0.00**	η_p^2 = 0.70
21	Resp	μRR	**p = 0.00**	η_p^2 = 0.88
23		$Deep$	**p = 0.04**	η_p^2 = 0.65
24		$Shallow$	**p = 0.00**	η_p^2 = 0.86
25		$minTCB$	**p = 0.00**	η_p^2 = 0.90

* In bold: $p < 0.05$ (rejecting the null hypothesis)

HRV, whereas both SCL and SCR features were significant from the GSR signal. Besides, 3 of these features showed significant correlation ($p = 0.00$) with the ordinal features, although with moderate effect sizes (Spearman's ρ): μRR ($\rho = 0.24$) and $minTCB$ ($\rho = 0.22$) showed positive correlation with dominance, whereas $nSCR$ showed negative correlation with score ($\rho = -0.26$) and positive with arousal ($\rho = 0.28$).

The mean values for six of these features are shown in Figure 6. It is apparent from this bar graphs that the highest values of these ECG and GSR features were on trials 3 and 4, and that there is a fair decreasing trend of RSP.

Despite of the fact that findings were consistent with the literature and that effect sizes of 0.2 are not small as to be trivial [41], caution must be applied since significance might be due to big sample sizes (Group A = 195; Group B = 235), i.e. results do not rule out the influence of other factors.

V. DISCUSSION

It was hypothesized that participants would report different kind of changes in emotional states (via valence, arousal, and dominance) after experiencing 5 trials whose differences were defined as sets of VR factors that followed either presence or challenge criteria, i.e. reported by participants assigned to groups A and B, respectively.

A. Effects of Presence-Based Factors on Emotions

There was no significant evidence that the overall perception of valence and dominance could be affected by changing the VR environment in terms of social presence (virtual-human behavior), physical presence (scenario realism), and self-presence (virtual body). However, a wider and lower interquartile range in valence and a lower median in dominance can be noticed when not providing good scenario coherence ("No Physical" trial in Figure 5) might suggest rejection of unattractive environments. In fact, this behavior was also mentioned by Wulf and Lewthwaite as a less interesting environment leading subjects to focus in the self instead of in the task goal, and thus impeding self-efficacy [42].

Arousal was significantly higher when replacing radio commands with written instructions ("No Social" trial), i.e. arousal increased as a consequence of loss of coexistence. Contrary to expectations, subjects were less aroused with a non-coherence scenario ("No Physical" trial), which also suggests the importance of the focus of attention. Finally, removing the body from the scene did not influence arousal and therefore, it is not possible to discuss about the role of self-presence.

B. Effects of Challenge-Based Factors on Emotions

Results indicate that changing the trial by different challenging conditions mostly affects the participant′s performance (difficulty-score correlation: $p = 0.00$, $\rho = -0.47$) and consequently their dominance, due to the positive score-dominance correlation (see Figure 5) or a confidence deterioration [20], which suggests that subjects tend to report perception of challenging tasks by changes in dominance. In addition, a closer inspection on Figure 5 and the results of the pairwise comparisons reported let to suggest that the greater change on valence (decreasing) was produced when involving cognitive effort (quiz trivia in trial 5). However, this must be approached with some caution because it was the only cognitive-demanding trial and it could also indicate displeasure leaded by deviation of expectation [13].

The most surprising aspect of the data from group B is the absence of evidence for arousal changes, and thus it is almost certain that it does not follow the rising challenge trend. However, it can be noticed a greater influence of reaction time reduction (falling faster) compared with precision and cognitive tasks (rings size and trivia), which can be seen as higher arousal at trials 3 and 4 in Figure 5. This might require further studies on the individual effect of challenging factors.

Regarding to the psychophysiological signals, the results of this study indicates that the tonic level of GSR (SCL) increases with higher challenge, whereas time between breaths decreases (faster breathing), not to mention that this happens along with a fall in dominance and score. Besides, HRV and SCL seem to increase when the participants were overwhelmed by falling faster (when arousal increased), which might suggest that this game characteristic could release arousal-related behaviors, such as stress and fear, which are typical in high challenging tasks. In fact, these results are in line with those of previous studies, where GSR was found to be positively correlated with arousal, anxiety and stress, whereas respiratory time tend to decrease in the same situations, as well that HRV was associated with mental workload [17], [19].

C. Study Limitations and Future Work

Full combination of the 3 presence factors (physical realism, coexistence, and body connectivity) could not be analyzed by this setup, since it would have demanded participants to perform eight 2.5-minute-long tests (about 20 minutes of VR exposure), which was considered as exhausting and risky in terms of VR simulation sickness [43]. Future studies should focus in assessing single hypotheses regarding these factors for VR environments.

Some measurements showed low effect sizes when calculating differences and correlations among trials and thus further experimental designs should consider using fine-grained or continuous units to assess emotions and to try a higher dosage of the independent variable (more disruptive presence and challenge related factors). In fact, having a setup intended to find “awareness thresholds” after external presence-based stimuli would be a fruitful area for further work, i.e. identifying magnitudes at which presence attributes could trigger the changes on emotional states, evident enough to be self-reported and detected by features extracted from physiologically signals.

Additionally, a more representative population and a normally distributed sample would allow to investigate the effect of demographic characteristics on emotional responses.

Finally, only game score and psychophysiological features were measured. Future setups could consider measuring biomechanical measurements (applied forces and torques, position and velocity error, etc.) and behavioral measures, such as reaction time, postures, and facial expressions.

VI. Conclusions

Keeping participants out of extremely high or low levels of valence, dominance, and arousal could be considered as a fair goal to assure a harmonious experience and promote emotional engagement. However, it is important to identify how different stimuli-delivering strategies could affect such emotions and how can they be used to promote changes in the subjective perception of VR experiences. In this study, an experimental setup was designed to expose subjects to 5 trials of a motion task in a serious immersive virtual game.

Overall, this study strengthened the idea that the emotions changed according to the challenge provided in the task, but only in terms of valence and dominance, whereas arousal suggested to be depending of the nature of the factors selected to define Presence in the VR environment. In accordance to the literature, both dominance and valence were confirmed to be positively correlated to game score, i.e. decreased when challenge was increasing between trials. The lack of significance in arousal changes could be possibly explained by understanding that subjects tend to focus on accomplishing the challenging task, and thus dominance was an interpretation of self-assessment that overshadowed arousal as the key emotion to define pleasantness. However, arousal did increase along with some psychophysiological features (HRV and SCL) at higher falling speed (originally selected as a challenge factor), which suggests its influence as a factor of scenario realism (presence). Additionally, a faster breathing and higher SCL were detected when dominance and score decreased due to higher challenges.

Oppositely, with a non-challenging task (score did not change significantly) and different audiovisual stimuli (presence-based changes between trials), valence and dominance were stable, and arousal was the emotion reported to be changing. Results suggested that cutting down the sense of coexistence (social presence) increased arousal, but removing scenario realism (physical presence) decreased it, whereas removing the virtual body (self-presence) did not show any effect. These remarks suggested that arousal could be increased/decreased by means of well-identified presence-based factors even more than by difficulty changing.

We authors consider that these findings should be used as a more solid base of knowledge to understand the nature of emotional responses to different kind of variables commonly used to modify experimental conditions in immersive VR tasks, but particularly in terms of the separated effect of such modifications. Therefore, further studies must be addressed to gather detailed data about the use of fine-grained values of those variables.

Acknowledgment

The authors would like to acknowledge all subjects who voluntarily decided to participate in this study. They would also thank master’s students Juan D. Abril and Oswaldo Rivera for their contributions on the setup, as well as to researchers of UMNG and VUB for their technical support and advisory.

Ethics and Disclosure Statement

This study protocol was approved by UMNG’s Ethics Committee. All participants were provided with informed consent and decided to participate voluntarily and anonymously after knowing the purpose of the study and their freedom to withdraw. Authors also report no conflicts of interest.

REFERENCES

[1] P. Wouters, E. D. V. D. Spek, and H. V. Oostendorp, "Current practices in serious game research: A review from a learning outcomes perspective," in *Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices*. Philadelphia, PA, USA: IGI Global, 2009, ch. 14, pp. 232–250.

[2] T. M. Connolly, E. A. Boyle, E. MacArthur, T. Hainey, and J. M. Boyle, "A systematic literature review of empirical evidence on computer games and serious games," *Comput. Edu.*, vol. 59, no. 2, pp. 661–686, Sep. 2012.

[3] M. Graafland, J. M. Schraagen, and M. P. Schijven, "Systematic review of serious games for medical education and surgical skills training," *Brit. J. Surgery*, vol. 99, no. 10, pp. 1322–1330, Oct. 2012.

[4] R. Sigrist, G. Rauter, R. Riener, and P. Wolf, "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review," *Psychonomic Bull. Rev.*, vol. 20, no. 1, pp. 21–53, Feb. 2013.

[5] M. K. Holden, "Virtual environments for motor rehabilitation: Review," *CyberPsychol. Behav.*, vol. 8, no. 3, pp. 187–211, Jun. 2005.

[6] C. Fraser *et al.*, "Driving plasticity in human adult motor cortex is associated with improved motor function after brain injury," *Neuron*, vol. 34, no. 5, pp. 831–840, May 2002.

[7] K. R. Lohse, C. G. E. Hilderman, K. L. Cheung, S. Tatla, and H. F. M. Van der Loos, "Virtual reality therapy for adults post-stroke: A systematic review and meta-analysis exploring virtual environments and commercial games in therapy," *PLoS ONE*, vol. 9, no. 3, 2014, Art. no. e93318.

[8] M. E. Hackney, J. Nocera, T. Creel, M. D. Riebesell, and T. Kesar, "Exercise and balance in older adults with movement disorders," in *Locomotion and Posture in Older Adults: The Role of Aging and Movement Disorders*. Cham, Switzerland: Springer, 2017, ch. 21, pp. 323–346.

[9] J. Laut, F. Cappa, O. Nov, and M. Porfiri, "Increasing patient engagement in rehabilitation exercises using computer-based citizen science," *PLoS ONE*, vol. 10, no. 3, 2015, Art. no. e0117013.

[10] M. Widmer, J. P. Held, F. Wittmann, O. Lambercy, K. Lutz, and A. R. Luft, "Does motivation matter in upper-limb rehabilitation after stroke? ArmeoSenso-reward: Study protocol for a randomized controlled trial," *Trials*, vol. 18, no. 1, pp. 1–9, Dec. 2017.

[11] M. Mihelj, D. Novak, M. Milavec, J. Ziherl, A. Olenšek, and M. Munih, "Virtual rehabilitation environment using principles of intrinsic motivation and game design," *Presence, Teleoperators Virtual Environ.*, vol. 21, no. 1, pp. 1–15, Feb. 2012.

[12] M. F. Levin, "Can virtual reality offer enriched environments for rehabilitation?" *Expert Rev. Neurotherapeutics*, vol. 11, no. 2, pp. 153–155, Feb. 2011.

[13] I. Bakker, T. van der Voordt, P. Vink, and J. de Boon, "Pleasure, arousal, dominance: Mehrabian and russell revisited," *Current Psychol.*, vol. 33, no. 3, pp. 405–421, Sep. 2014.

[14] A. H. Lequerica and K. Kortte, "Therapeutic engagement," *Amer. J. Phys. Med. Rehabilitation*, vol. 89, no. 5, pp. 415–422, 2010.

[15] M. S. Cameirão, A. L. Faria, T. Paulino, J. Alves, and S. B. I Badia, "The impact of positive, negative and neutral stimuli in a virtual reality cognitive-motor rehabilitation task: A pilot study with stroke patients," *J. NeuroEng. Rehabil.*, vol. 13, no. 1, pp. 13–70, Dec. 2016.

[16] L. Nielsen and A. W. Kaszniak, "Conceptual, theoretical, and methodological issues in inferring subjective emotion experience recommendations for researchers," in *Handbook of Emotion Elicitation and Assessment*. New York, NY, USA: Oxford Press, 2007, ch. 22, pp. 361–375.

[17] S. D. Kreibig, "Autonomic nervous system activity in emotion: A review," *Biol. Psychol.*, vol. 84, no. 3, pp. 394–421, Jul. 2010.

[18] C. L. B. Maia and E. S. Furtado, "Using psychophysiological measures to estimate dimensions of emotion in hedonic experiences," *Comput. Electr. Eng.*, vol. 71, pp. 431–439, Oct. 2018.

[19] A. C. Dirican and M. Göktürk, "Psychophysiological measures of human cognitive states applied in human computer interaction," *Procedia Comput. Sci.*, vol. 3, pp. 1361–1367, 2011.

[20] C. Rodriguez-Guerrero, K. Knaepen, J. C. Fraile-Marinero, J. Perez-Turiel, V. Gonzalez-de-Garibay, and D. Lefeber, "Improving Challenge/Skill ratio in a multimodal interface by simultaneously adapting game difficulty and haptic assistance through psychophysiological and performance feedback," *Frontiers Neurosci.*, vol. 11, p. 242, May 2017.

[21] D. Novak, A. Nagle, U. Keller, and R. Riener, "Increasing motivation in robot-aided arm rehabilitation with competitive and cooperative gameplay," *J. NeuroEng. Rehabil.*, vol. 11, no. 1, p. 64, 2014.

[22] S. Sarabadani, L. C. Schudlo, A.-A. Samadani, and A. Kushki, "Physiological detection of affective states in children with autism spectrum disorder," *IEEE Trans. Affect. Comput.*, early access, Mar. 27, 2018, doi: 10.1109/TAFFC.2018.2820049.

[23] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. Palatine, IL, USA: Harper, 1990.

[24] N. Hocine, A. Gouaïch, S. A. Cerri, D. Mottet, J. Froger, and I. Laffont, "Adaptation in serious games for upper-limb rehabilitation: An approach to improve training outcomes," *User Model. User-Adapted Interact.*, vol. 25, no. 1, pp. 65–98, Mar. 2015.

[25] S. Poeschl, K. Wall, and N. Doering, "Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2013, pp. 129–130.

[26] E. B. Nash, G. W. Edwards, J. A. Thompson, and W. Barfield, "A review of presence and performance in virtual environments," *Int. J. Hum.-Comput. Interact.*, vol. 12, no. 1, pp. 1–41, 2000.

[27] M. Slater, P. Khanna, J. Mortensen, and I. Yu, "Visual realism enhances realistic response in an immersive virtual environment," *IEEE Comput. Graph. Appl.*, vol. 29, no. 3, pp. 76–84, May/Jun. 2012.

[28] R. Skarbez, S. Neyret, F. P. Brooks, M. Slater, and M. C. Whitton, "A psychophysical experiment regarding components of the plausibility illusion," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 4, pp. 1369–1378, Apr. 2017.

[29] M. Slater, "Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments," *Phil. Trans. Roy. Soc. B, Biol. Sci.*, vol. 364, no. 1535, pp. 3549–3557, Dec. 2009.

[30] J. J. Cummings and J. N. Bailenson, "How immersive is enough? A meta-analysis of the effect of immersive technology on user presence," *Media Psychol.*, vol. 19, no. 2, pp. 272–309, Apr. 2016.

[31] B. Balakrishnan and S. S. Sundar, "Where am I? How can I get there? Impact of navigability and narrative transportation on spatial presence," *Hum.-Comput. Interact.*, vol. 26, no. 3, pp. 161–204, 2011.

[32] W. Wirth *et al.*, "A process model of the formation of spatial presence experiences," *Media Psychol.*, vol. 9, no. 3, pp. 493–525, May 2007.

[33] G. Makransky, L. Lilleholt, and A. Aaby, "Development and validation of the Multimodal Presence Scale for virtual reality environments approach," *Comput. Hum. Behav.*, vol. 72, no. 1, pp. 276–285, 2017.

[34] O. I. Caldas, J. D. Abril, O. Rivera, C. Rodriguez-Guerrero, and O. F. Avilés, "Contribution of virtual environments to the perception of balance rehabilitation tasks: A psychophysiological study," in *Proc. CLAIB-CNIB*, Cancún, Mexico: Springer, 2019, pp. 1–7.

[35] H. K. Kim, J. Park, Y. Choi, and M. Choe, "Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment," *Appl. Ergonom.*, vol. 69, pp. 66–73, May 2018.

[36] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *J. Behav. Therapy Experim. Psychiatry*, vol. 25, no. 1, pp. 49–59, Mar. 1994.

[37] N. Goljar *et al.*, "Psychophysiological responses to robot training in different recovery phases after stroke," in *Proc. IEEE Int. Conf. Rehabil. Robot.*, Zurich, Switzerland, Jun. 2011, pp. 1–6.

[38] D. Novak, "Promoting motivation during robot-assisted rehabilitation," in *Rehabilitation Robotics*. Amsterdam, The Netherlands: Elsevier, 2018, ch. 11, pp. 149–158.

[39] A. Koenig *et al.*, "Psychological state estimation from physiological recordings during robot-assisted gait rehabilitation," *J. Rehabil. Res. Develop.*, vol. 48, no. 4, p. 367, 2011.

[40] O. I. Caldas, O. F. Aviles, and C. Rodriguez-Guerrero, "Experimental results to compare the effects of presence and challenge modulation on immersive virtual environments," *IEEE DataPort*, to be published, doi: 10.21227/vj8w-v224.

[41] G. M. Sullivan and R. Feinn, "Using effect size—Or why the p value is not enough," *J. Graduate Med. Edu.*, vol. 4, no. 3, pp. 279–282, 9 2012.

[42] G. Wulf and R. Lewthwaite, "Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning," *Psychonomic Bull. Rev.*, vol. 23, no. 5, pp. 1382–1414, Oct. 2016.

[43] J. Munafo, M. Diedrick, and T. A. Stoffregen, "The virtual reality head-mounted display oculus rift induces motion sickness and is sexist in its effects," *Exp. Brain Res.*, vol. 235, no. 3, pp. 889–901, Mar. 2017.