Virtual reality sickness detection: an approach based on physiological signals and machine learning

Nicolas Martin* Nicolas Mathieu† Nico Pallamin* Martin Ragot* Jean-Marc Diverrez*

*IRT b<>com, Cesson-Sevigne, France
†Ubisoft, Montreuil, France

ABSTRACT

Virtual Reality (VR) is spreading to the general public but still has a major issue: VR sickness. To take it into consideration and minimize its occurrence, evaluation methods are required. The current methods are mainly based on subjective measurements and therefore have several drawbacks (e.g., non-continuous, intrusive). Physiological signals combined with Machine Learning (ML) methods seem an interesting approach to go beyond these limits. In this paper, we present a large-scale experimentation (103 participants) where physiological data (cardiac and electrodermal activities) and subjective data (perceived VR sickness) were gathered during 30-minute VR video game sessions. Using ML methods, models were trained to predict VR sickness level (based on the physiological data labeled with the subjective data). Results showed an explained variance up to 75% (in a regression approach) and an accuracy up to 91% (in a classification approach). Despite generalization issues, this method seems promising and valuable for a real time, automatic and continuous evaluation of VR sickness, based on physiological signals and ML models.

Index Terms: H.1.2 [Models and principles]: User/Machine Systems—Human factors; I.3.6 [Computer graphics]: Methodology and Techniques—Ergonomics

1 INTRODUCTION

Consumer-grade Virtual Reality (VR) equipment is now broadly distributed on the public market. However, one of the remaining pitfalls to its usage is cybersickness in VR (i.e., "sickness or general feelings of malaise experienced due to exposure to VR" [119]). Consequently, it is still necessary to propose some solutions to reduce and, if it is possible, eliminate the occurrence of cybersickness. For that, be able to measure it precisely seems an essential preliminary step. The current methods are only based on subjective evaluations with some limitations: they are offline, non-continuous, temporally imprecise and non-automatic. An automatic cybersickness recognition could be useful to adapt, potentially in real time, the content to provide the most comfortable and adapted VR experience to the user. Indeed, the current countermeasures proposed to reduce the occurrence of cybersickness, such as reduction of the Field Of View (FOV) [34, 40], are almost always done regardless of the current state of the user. Consequently, they could reduce the immersion and therefore feeling of presence [75] even if the user does not experience any trouble. Ideally, these methods must be tailored to the actual feelings of the user. For this purpose, in this paper, we present a large-scale experiment (103 participants) to detect VR sickness based on physiological signals and Machine Learning (ML) algorithms.

*e-mail: firstname.lastname@b-com.com
†e-mail: firstname.lastname@ubisoft.fr

In this paper, we present a solution to detect in real-time the perceived cybersickness of users in VR. The Section 2 presents an overview of related work. The Section 3 describes the study, which was conducted to assess the physiological activity (i.e., electrodermal and cardiac) of 103 participants in the context of VR games. The processing chain of the collected data are presented including the feature extraction and the normalization of physiological data to reduce the inter-individual variability. The Section 4 presents the recognition performance of trained models to detect cybersickness in VR, using machine learning algorithms. Inferential statistics on extracted features and several data splitting are compared. The best configurations reached a recognition performance of 91% for two level of cybersickness and a variance explained of 75% in regression approach. Moreover, to fill the gap of prior works, the generalization to new users is explored and showed limitations of the trained models. Finally, these results are further discussed in Section 5; and Section 6 provides the concluding remarks. Finally, 3 main contributions are provided by this work: (1) a machine learning approach for VR sickness recognition based on only two physiological activity (cardiac and electrodermal); (2) training and evaluation on a large dataset (103 participants); (3) evaluation of the generalization of trained models to new users. The experimental results also provide insights on the methods of normalization.

2 RELATED WORK

2.1 Cybersickness in VR

In this paper, we will use indistinctly the terms "cybersickness" and "VR sickness".

2.1.1 Definitions and theories

Cybersickness in VR can broadly be defined as the feeling of discomfort that users can experience while using VR devices [119]. In VR, the most common conflict is a discrepancy between the motion information coming from two separate systems [90]: the vestibular system (i.e., sensory system related to the sense of balance, self-motion and spatial orientation located in the inner ear) and visual system (i.e., central nervous system related to the process of visual details [54]). This discrepancy will be detected by the brain [20, 63] and will induce, within sensitive participants, the symptoms of VR sickness. Several theories have been proposed to explain the causes of VR sickness: the poison theory [12], postural instability theory [71], rest frame theory [103, 104] and sensory conflict theory [109] (see Davis [30] or Martirosov [85] for a review on cybersickness). The most widely accepted theory is the sensory conflict theory [109].

2.1.2 Cybersickness induction

Cybersickness in VR is a complex phenomenon in which motion cues play a primary role [10, 33]. However, several other factors are involved in the VR sickness etiology, grouped into 3 main categories:

- Characteristics of the stimuli (e.g., FOV [41, 130], reactivity of the system [64], spatial frequency [10, 33])

978-1-7281-8508-8/20/$31.00 ©2020 IEEE
DOI 10.1109/ISMAR50242.2020.00065

- Predispositions of the users (e.g., gender [97], age [4], migraine predisposition [29])
- Users' past experiences [10]

2.1.3 Cybersickness symptomatology

Nausea and vomiting are the most evident and detrimental symptoms of cybersickness [86]. Other symptoms like headache, disorientation and eye strain can occur [11, 56, 110]. The intensity as well as the duration of the symptoms are quite variable. In the majority of cases, the symptoms disappear some minutes after the end of the stimulation. However, some studies report the persistence of the symptoms several hours after the VR experience [89, 91].

2.1.4 Cybersickness reduction methods

Over the years, multiple methods to reduce cybersickness has been proposed such as virtual nose [128], dynamic blur [17], or reduction of the FOV [18, 130]. Recently, eight methods from the prior research have been implemented in an open-source software package, called GingerVR, for the Unity game engine [3]. Nevertheless, these methods are almost always applied regardless of the actual state of the user.

2.2 Cybersickness in VR: measuring methods

2.2.1 Subjective methods

The usual way to evaluate VR sickness is based on subjective questionnaires. This approach seems relevant because it is based on actual and declared user perception of sickness level. Thereby, various questionnaires have been previously proposed to assess VR sickness. The most popular is probably the "Simulator Sickness Questionnaire" (SSQ) [56]. Some adaptations to VR have been recently proposed [59, 112]. While still widely used for the simplicity of its deployment and analysis, the subjective methods are not flawless. Indeed, subjective evaluations are generally done after the experiment. Thus, only a punctual, a posteriori and global measure of the user perception can be gathered [52]. Moreover, the participants have to explicitly perform a self-assessment of their state. To cope with these limitations, some online subjective evaluation of VR sickness (i.e., evaluation during the task) have been tested [57, 58, 133] but they can lead to intrusiveness. Indeed, the explicit self-assessment can interfere with the task that the participant was performing and reduce the feeling of presence [42]. In summary, subjective measurements are easy to pick up [94], but do not allow continuous, automatic and real time measurement of VR sickness. For these reasons, various researchers have investigated the possibility to exploit the users' physiological data to assess VR sickness in real time.

2.2.2 Physiological methods

The conflicting inputs from visual, vestibular and somatosensory afferents at the origin of VR sickness induce a vestibular autonomic response [115, 131]. Such response, due to the connections between vestibular and autonomic nervous systems [99], implies both the sympathetic and parasympathetic activities [28, 38, 45, 63]. These activities could consequently be detected and exploited to assess the occurrence of VR sickness or motion sickness.

Various physiological parameters, requiring more or less intrusive sensors, have been studied to assess their reliability as indicators of VR sickness. Studies adopting intrusive sensors concern, for instance, the gastric activity and, in particular, the gastric tachyarrhythmia (increased electrogastrography activity in the 4-to 9-cycles/min frequency) are correlated with motion sickness according to the prior studies [51, 65]. Using respiratory sensors, some authors like Dennison [31] found a significant negative interaction between breathing rate and cybersickness symptom severity adopting a regression analysis. Lastly, Chen [23] found five brain processes (for instance the augmentation of the theta and delta bands in the occipital areas) related to motion sickness using a 32-channel EEG system.

Using less intrusive sensors, some research have been conducted on the electrodermal and cardiac activity [9, 80, 81]. For instance, concerning the electodermal studies, Golding [47] founded that phasic skin conductance activity on the forehead correlates with motion sickness onset and recovery. The study of Meusel [87] showed a weak but significant correlation between EDA and nausea subscale of the SSQ. More recently, Plouzeau et al. [102] demonstrated that the EDA variation over time could be used for estimating the VR sickness. Finally, concerning the works conducted on cardiac activity, for example, Yokota [132] emphasized a significant effect of motion sickness over the cardiac activity manifested by an increase of Low-Frequency power and a decrease of the High-Frequency power of the Heart Rate Variability (HRV). This result is coherent with previous research (e.g., [36]) that found a significant reduction in the power spectral density of the R-R interval (at the medium and high frequencies) during a vestibular disorientation test.

2.2.3 VR sickness recognition using physiological data: an application of Machine Learning approach

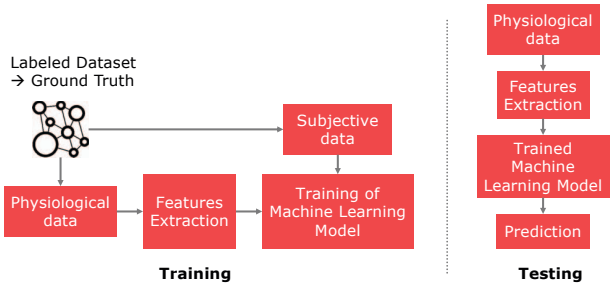

Figure 1: Machine Learning approach: Training and Testing

As previously presented, numerous studies (e.g., [31, 47, 51, 65]) supported the hypothesis of an existing relationship between the physiological responses and VR sickness. However, the exploitation of physiological data and its relation with VR sickness is complex especially due to their non-linear relationships [32, 84] and the large inter-individual variability (e.g., [95]). Thus, to handle this complex data and respect constraints of real time automatic measurement, the classical statistical methods of data analysis (e.g., linear regression) are not suitable. In this context, the use of Machine Learning (ML) models seems promising and has spread in recent years [73], including the field of cybersickness recognition [74]. Using supervised techniques [66], the training of ML models intends to automatically infer the function between the input data (i.e., extracted physiological features) and output data (i.e., subjective responses) [94]. This way, after training, these models could provide a relevant method to automatically recognize VR sickness, in real time, using the physiological data without requesting a subjective response: see figure 1.

3 Experiment

In line with the previous research, the possibility to recognize VR sickness using Machine Learning and physiological signals was evaluated in the current study. For this purpose, an experiment was conducted in order to collect physiological and subjective data during VR video game sessions. The main long term goal is to facilitate the study of cybersickness and the real-time in-game adaptation using only few intrusive physiological sensors.

While physiological measures have several advantages over the subjective methods (e.g., like their ability to provide information in real time and continuously [27]), all the physiological measures do

not satisfy the criteria of low intrusiveness required in the frame of our study. For this reason, only two physiological responses were gathered: the cardiac and the electrodermal activities. Thereby, as previously explained, various studies showed a relationship between the Heart Rate Variability (HRV) (e.g., [36, 44, 132]) as well as the electrodermal activity (EDA) (e.g., [47, 80, 87, 102]) and the VR or motion sickness. These physiological signals can be easily gathered [108] and are suitable for consumer or industrial use. For instance, other studies integrate EEG (e.g., [92]), electrogastrogram, electrooculogram and respiratory devices (e.g., [31]). The use of such laboratory-grade devices increase the diversity of physiological data collected but they can be intrusive and complex to deploy out of a laboratory. In the same way, a majority of physiological studies are conducted using devices like the Biopac MP150 [31] or Biopac MP100 [92] that grant an accurate collection and synchronization of the different physiological data. Nevertheless, due to their cost and complexity, they are usually reserved to specialized laboratories.

Even if ML models have already been exploited to recognize VR sickness (e.g., [31,53,92,96]), our research differs from the previous studies in various aspects. Indeed, to the best of our knowledge, no study has been conducted at large scale using only cardiac and electrodermal activities with the objective to recognize VR sickness using ML algorithms. Another peculiarity, our study explicitly aims to obtain non-intrusive, online and automatic recognition of the VR sickness level while using interactive VR content in natural settings. This choice imposed a series of constraints concerning the content, the data used for the ML training and the adopted physiological sensors, these constraints were absent in the other studies.

Regarding the content presented to the user, only few researchers (e.g., [53]) opted to use interactive content (i.e., video games) while the majority of studies preferred to use non interactive content like 360-degree videos (e.g., [96]) or ad-hoc and simplified environments (e.g., [31]). Our choice to use various prototypes of commercial games instead of ad-hoc environment should increase the ecological validity of our study and the generalization of our recognition model.

Considering the data exploited by the algorithms, only the physiological responses of the participant, easily measurable, were considered without exploiting any information concerning the content. Exploiting information concerning the content like the optical flow could be used to increase the accuracy of the predictions [96], but it requires access to the content itself. In our study, the choice to focus only on the physiological data was also dictated by the need to be able to apply our method to any VR environment (e.g., HMD, CAVE), without having access to the content itself. This could be particularly valuable to assess commercial products in case of proprietary and/or confidential contents. Moreover, it may improve the generalization of the approach across contents.

All these constraints, introduced by the willingness to propose an operational and easily deployable VR sickness assessment solution, suitable for commercial and industrial use, have strongly influenced the protocol details.

3.1 Participants

One hundred and three (103) participants (86 men and 17 women - mean age = 26.12; standard deviation = 6.31) were involved in the study. All declared a normal or corrected-to-normal vision (contact lenses). The selected participants could already have experienced VR but should not be professional gamers or regular users of VR (i.e., no more than few VR experiences). To ensure of the VR sickness induction, they may be susceptible to the cybersickness according to Motion Sickness History Questionnaire [55]. They declared to be free of any medical treatment likely to affect cognitive state or proprioceptive systems. Participants who reported illness likely to influence cognitive state like migraine sensitivity, mental disorders, vestibular disorders were excluded from the recruitment. All the experimental procedures followed the guidelines of the Declaration of Helsinki [5]. The participants were informed that they were free to stop the experiment at any time. They also completed and signed an informed consent form at the beginning of the session.

3.2 Measures

Two types of measures were collected: the physiological data and the users' Subjective Responses (SR). The physiological data was collected using the Shimmer GSR+[1]. This device measures the Blood Volume Pulse (BVP) using PhotoPlethysmoGraphy (PPG) as the measure of cardiac activity and electrodermal activity (EDA). Based on literature (e.g., [124]) and preliminary studies, the position of sensors have been evaluated to select the ones that maximize the sensor sensibility and minimize the noise, without increase the intrusiveness. The BVP sensor was placed at the extremity of the little finger [79, 121]. The EDA sensors were placed on the middle phalanx of the two first fingers of the non-dominant hand [13, 124]. The data was gathered in a continuous way (sampling frequency = 128 Hz) during the experimental session. The subjective data (i.e., perceived cybersickness) was collected following the method proposed by Keshavarz and collaborators [57]. Participants were instructed to express orally their perceived VR sickness level on a scale from 0 to 20 in response of auditory stimuli presented every 45 seconds (regardless the current content of the game). A score of 0 corresponds to the lowest level and a total absence of VR sickness symptoms. A score of 20 corresponds to the highest level of VR sickness state in which participants lose completely their autonomy. According to Keshavarz et al. [57], this scale is strongly correlated (r = .79) with the gold standard SSQ [56]. All data were referenced to the same clock to avoid asynchrony (participant PC clock).

3.3 Experimental Protocol

The participants were requested to test two games selected among five VR game prototypes during 30 minutes. Since our final goal was to use the developed solution during real VR customer experiences like VR game experiences, our strategy was to build a training database representative of the different games experiences and the various hardware setups to increase the generalization of our trained models. Thus, the VR devices and game experiences were voluntarily diversified to grant the effectiveness of the algorithms independently from the content and the VR device adopted by the end-user. The participants were seated throughout the experience. They were equipped by an Oculus Rift VR (93 participants) or HTC Vive helmet (20 participants). The VR game prototypes were launched on a high-end VR ready computer (Nvidia GTX 1070, Intel Xeon E2100, 16 Go of RAM). The VR game prototypes were interactive games currently under development. They have been selected as they offer a variety in gaming situations and potential occurrences of cybersickness that users can encounter in the VR gaming. Thus, they imply a relevant variety of motion commonly present in video game, especially in terms of amount of camera and avatar movements: walk, run, drive and flight, as well as first and third points of view (POV). The games were a space simulation, an arcade car racing game, a space first-person shooter, a horror exploration puzzle game and an astronaut space simulation. None of the typical countermeasures to prevent VR sickness (e.g., reduction of the peripheral vision, fake nose or camera rotation constraints) were implemented in these prototypes. All games have similar framerates and lag times. Xbox 360 gamepad was used for the arcade car racing game and the Oculus Touch was used for the remained games. The distribution of participants per game is provided in Table 1.

A double audio "beep" played every 45 seconds was used to indicate to the participant to give their subjective VR sickness level. The soundtrack was played in addition to the VR game sounds in the audio headset. The subjective VR sickness levels orally given by

[1] http://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor

the participants were recorded with the microphone integrated in the Head-Mounted Display (HMD). After the presentation and setup of the different devices and sensors (see figure 2), the participants were instructed to give orally their current subjective VR sickness level (scale from 0 to 20) each time they heard the sound signal. Next, participants were instructed to play "as if they were at home". In this way, they were instructed to do what they wanted into the game prototypes, to follow their natural way of play and not to interact with the experimenters except if they needed it. Each participant played 2 games during 30 minutes (15 minutes for each game). The second game was launched by an experimenter without removing the participants VR helmet. After each experimental session, the SRs were transcribed with the corresponding timestamps.

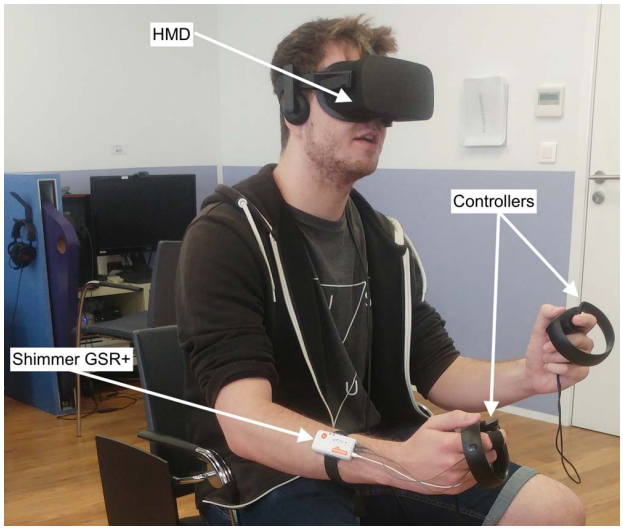

Figure 2: Participant with full devices and sensors

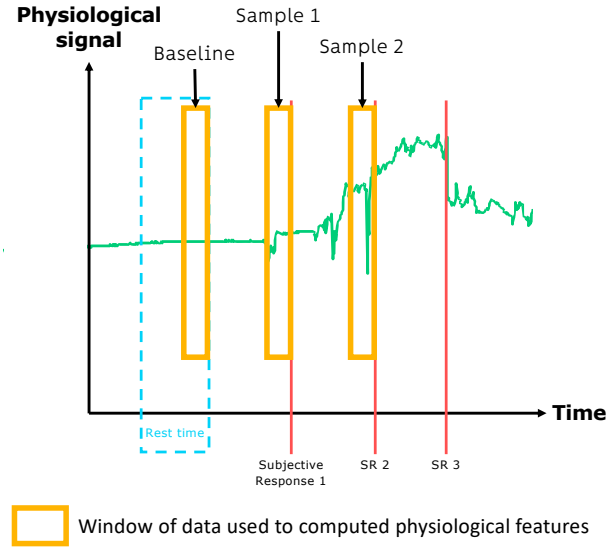

Figure 3: Rest time, Baseline and Subjective Responses

Table 1: Descriptive statistics on game

Game	Space simulation	Arcade car racing game	Space first-person shooter	Horror exploration puzzle game	Astronaut space simulation
Female	1	15	15	0	1
Male	14	62	57	11	9
Total	15	77	72	11	9

3.4 Features extraction

To fully exploit the raw physiological signal, it is necessary to extract specific features [1]. For this purpose, physiological responses related to perceived sickness need to be extracted. Thus, for each of the reported SR, a fixed-size window (several window sizes have been explored: 10, 30, 60, 90, 120 seconds) of the raw physiological data (BVP and EDA) was extracted. Thus, the X seconds of physiological data before the SR are extracted to estimate the physiological features (X corresponding to the evaluated window size) [42]: see figure 3. The number of labeled data corresponds to the number of collected SRs. Thus, for the 10 and 30 seconds window sizes, the physiological windows are not overlapped (i.e., physiological data are used only for 1 SR). For the 60, 90 and 120 seconds, physiological windows are overlapped (i.e., same physiological data are used for several SRs).

Based on these windows of physiological signals, the common features from the relevant state of the art were used.

Concerning the **cardiac activity**, time-domain and frequency-domain features based on the InterBeat Interval (IBI), related to the HRV, were extracted [68]. To estimate the IBI, a bandpass filter was firstly applied (cutoff frequency = [0.66; 3.33] Hz, order = 3) to only select the normal cardiac activity between 40 and 200 beats per minute and reduce noise such as user's motions [1]. Then, the peaks on the BVP signal are detected using a threshold (arbitrarily set) and an estimation of local minima/maxima [122]: see figure 4. Based on these IBI, 6 time-domain and 5 frequency-domain features were computed: see table 2.

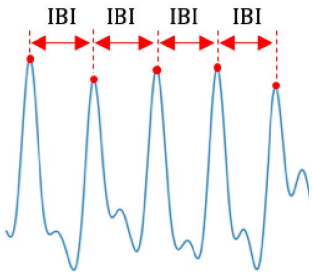

Figure 4: InterBeat Interval (IBI) on BVP

Concerning the **electrodermal activity**, the signal is composed of two parts: the phasic and tonic part [13]. The phasic part (also called Skin Conductance Level - SCL) corresponds to slow changes in the EDA while the tonic part (also called Skin Conductance Responses - SCR) corresponds to the rapid physiological responses to a stimulus. Different steps compose the process of extracting those two components from the raw signal. A low-pass filter (cutoff frequency = 1Hz, order = 3) is firstly applied to reduce noise in the signal [13]. Then, the tonic part of EDA is extracted from this filtered signal using a low-pass filter (cutoff frequency = 0.05Hz, order = 3) and averaged over the whole filtered signal [14]. The phasic part is

obtained by subtracting the tonic signal to the filtered signal. On the phasic part, 9 features were calculated on the estimated EDA peaks[2]. On the tonic part, 29 features were extracted: see table 2. A part of these features are based on previous research from other domains suggesting the introduction of features related to the frequency-domain of EDA [77, 111, 113], or also the use of a data-driven signal decomposition called Empirical Mode Decomposition (EMD) [49].

In summary, 50 features were computed based on the raw BVP and EDA signals: see table 2.

Table 2: Extracted physiological features

Cardiac features
Time-domain features
Heart Rate
Average of NN intervals
Standard deviation of NN intervals
Root mean square of successive differences between NN intervals
Number of interval differences of successive NN intervals greater than 50 ms
Percentage of interval differences of successive NN intervals greater than 50 ms
Frequency-domain features
Very Low Frequency (0.003 to 0.004 Hz)
Low Frequency (0.04 to 0.15 Hz)
High Frequency (0.15 to 0.4 Hz)
Ratio of Low Frequency and High Frequency
Total Spectral Power
EDA features
Tonic EDA features
Mean, Standard deviation, Maximum, Range, Inter-quartile range and Root mean square error of EDA signal
Mean absolute value of 1st differences of the EDA signal
Mean absolute value of 2nd differences of the EDA signal
Mean absolute value of the 1st differences of the standardized EDA signal
Mean absolute value of the 2nd differences of the standardized EDA signal
Skewness and Kurtosis of EDA signal
Mean, Standard deviation, Min and Max of 3 Intrinsic Mode Functions (IMFs)
Very low frequency (0 to 0.1 Hz)
Low frequency (0.1 to 0.2 Hz)
Middle frequency (0.2 to 0.3 Hz)
High frequency (0.3 to 0.4 Hz)
Very high frequency (0.4 to 0.5 Hz)
Phasic EDA features
Number of EDA peaks
Mean, SD, min and max of EDA peak amplitude
Mean, SD, min and max of half of recovery time of EDA peaks

3.5 Baseline and Normalization

Previous research in other fields showed an important variability between participants concerning the physiological responses (e.g., [8, 42]) which could be an important limitation to the recognition performance (e.g., [129]). So, several approaches were proposed in the literature in order to reduce the influence of the inter-individual variability. Concerning the EDA, the most common approach is to collect data during a rest time (i.e., baseline; see figure 3) and subtract the mean value of EDA during this rest time on the whole signal [14]. This method aims to provide comparable data between participants even if their EDA levels are different (inter-individual variability). Nevertheless, it cannot be applied to periodic signal such as BVP[3]. Conversely, deal with inter-individual variability at feature level seems a more interesting approach [94] as it can be applied to all signals. Thus, two types of approaches have been previously proposed:

- **Normalization by substracting at feature level (NFL)**: For each participant, subtracting the features estimated for each subjective response with the features estimated during the baseline [24, 61, 118];
- **Normalization by adding to feature map (NAF)**: For each participant, adding the features estimated during the baseline to the dataset (i.e., the feature space) [101, 123].

In the current study, the following approach is applied for each participant: (1) extraction of the physiological features; (2) selection of the values of the physiological features during the baseline (window of physiological data collected during the rest time); (3) subtraction of these values to all other samples of the participant (NFL method) or addition of these values to the feature space to all other samples of the participant (NAF method). The NSF and NAF methods have been evaluated on the same data. For this purpose, a window of physiological data collected during the rest time was used at baseline (the window size is equal to the window used for extracting the physiological features). Otherwise, rest-phase corresponds to the phase where participants are listening instructions for the experiment. They do not have to move and they do not wear the VR helmet.

3.6 Machine Learning

Based on the extracted physiological features and related subjective responses, models were trained to detect VR sickness level using supervised Machine Learning (ML) algorithms. Thus, the training were done in the way to infer the function between the input data (i.e., extracted physiological features) and output data (i.e., subjective responses) [94] without explicitly defining the relationship [7]. In this way, trained models could provide a VR sickness estimation only based on physiological features without requesting any user evaluation (see figure 1). For this purpose, three ML algorithms were evaluated: Support Vector Machine (SVM) [50], Gradient Boosting (GB) [43] and Random Forest (RF) [15]. All training were realized using R [106] and the library caret [69].

In summary, the following parameters were explored and evaluated:

- **Input:** 50 extracted physiological features for each SR
  - **Normalization:** No normalization, NFL, NAF
  - **Window size:** 0, 30, 60, 90, 120 seconds
- **Output/Label:** Subjective Responses
  - **Regression approach:** SRs are considered as continuous values

[2] An EDA peak is especially characterized by the amplitude (the height of the peak) and recovery time (time to return to the level EDA before the peak) [13, 117]

[3] The BVP signal is a periodic signal unlike the EDA which is a non-period (aperiodic) signal. Thus, subtract the mean value of BVP (estimated during the baseline) will only bring the BVP signals to the same level (roughly the same mean). But, the features related to the BVP signal are based on the IBI which is not influenced by the mean level.

– **Classification approach:** SRs are divided into classes

- **ML models:** Support Vector Machine (SVM) [50], Gradient Boosting (GB) [43] and Random Forest (RF) [15] were evaluated.
- **Training strategies:** Person-dependent models (one model per participant), Group models (10-fold cross-validation on all data), and person-independent models (LOOCV).

4 RESULTS

In total, 3022 SRs were collected (see table 3 for the data distribution). The participants are asked to provided their SR every 45 seconds, theoretically corresponding to more than 3022 SRs (103 participants * ≈ 30 min * 60 seconds / 45 seconds per SR). However, some participants continued to play and forgot to provide some SRs in response to the beep sound (e.g., during a complex action in the game, beep sounds were masked by game sound). Moreover, the duration of VR game can vary according to the selected games.

Based on the gathered SRs, regression and classification approaches were evaluated. Indeed, the SR are numeric values between 0 (i.e., no discomfort) and 20 (i.e., barely supportable sickness). They can be considered as a continuous variable (regression approach) or divided into classes (classification approach) [94].

Table 3: Contingency table of SRs

SR	0	1	2	3	4	5	6	7	8	9	10	11	12
N	1887	291	373	151	122	103	27	29	14	6	13	5	1

4.1 Statistical analysis

Before training models to detect VR sickness, some statistical analysis have been conducted. The extracted physiological features were compared between two classes: C1: SR =<1 and C2: SR>= 2 (see table 4). As data are repeated, mixed models were used to assess statistical effects using the R statistical software [107] with the lme4 library [6]. The participants were considered as a random factor, and the SRs were considered as a within-subject factor. Analyses showed significant ($p < .05$) differences between the two classes on the following features: mean, max and range of EDA signal, number of EDA peaks, mean, SD, min, and max of amplitude of EDA peaks, Mean absolute value of 1st differences and of the 2nd different of the EDA signal and also on very low, low, middle, high and very high frequency of EDA. Similar results were found with the data splitting C1: SR = 0 and C2: SR >= 1.

Table 4: Data splitting on SRs

2 classes (see table 6)	2 classes (see table 7)	3 classes (see table 8)
C1: SR =< 1	C1: SR = 0	C1: SR = 0
C2: SR >= 2	C2: SR >= 1	C2: SR >= 1 & =< 2
		C3: SR >= 3

As previously explained, several ML algorithms were tested but, due to space limitation[4], only the best combination of parameters are listed. Thus, Random Forest (RF) presented almost always the best performance on the dataset as in similar contexts (e.g., [37]). Consequently, only the results estimated using this algorithm will be presented. The number of trees as well as the number of randomly selected predictors at each cut in the tree were tuned during the training of RF. As the evaluation of feature selection (i.e., principal component analysis) showed no improvement of the recognition accuracy and as the tree-based models are generally robust against unhelpful features, all the features per sensor were considered.

[4]Please contact the first author for details

4.2 Group models - Cross-Validation 10-fold

To evaluate the performance of the models obtained, all results were computed using a 10-fold cross-validation method [94, 114]. Technically, the initial dataset (i.e., the 3022 labeled physiological data window) is divided into 10 sub-samples. One sub-sample is used as the testing dataset. The other 9 samples are used for training. It is repeated 10 times (i.e., for each sub-sample). The average result on the 10 sub-samples is used as the final result.

4.2.1 Regression approach

Using the regression approach, the SRs are considered as a continuous variable. Models were trained to produce a numeric value between 0 and 12 (i.e., range of collected SRs; see table 3) based on the labeled physiological data. Using several sizes of window and type of normalization, regression models were trained to recognize VR sickness: see table 5. To evaluate the recognition performance, three metrics were used: Root Mean Square Error (RMSE)[5], Mean Absolute Error (MAE)[6] and R-squared (R^2)[7]. To facilitate the evaluation of the results, performance of a naive model corresponding to the mean of SRs was presented. The variable importance calculated on the model showing the best performance are the following top-10 features [8]: mean of IMF3 during the baseline (100.00), pNN50 (73.87), mean absolute value of 1st differences of the EDA signal (71.27), HR (65.91), mean absolute value of 2nd differences of the EDA signal (64.98), min of peak amplitude (62.45), Mean absolute value of 2nd and 1st differences of the standardized of EDA signal during the baseline (respectively 62.40 and 60.70), min of half recovery of EDA peaks (59.28), nn50 (57.97), Mean absolute value of 2nd and 1st differences of the standardized of EDA signal (57.07).

Table 5: Influence of window size and type of normalization (Norm) on recognition performance

Norm	**Metrics**	**Window Size**				
		10	30	60	90	120
Naive model	*RMSE*	*1.83*				
	R^2	*NA*				
	MAE	*1.34*				
No Norm	RMSE	1.66	1.59	1.52	1.48	1.42
	R^2	.19	.23	.32	.37	.42
	MAE	1.17	1.11	1.07	1.01	0.96
NFL	RMSE	1.24	1.15	1.18	1.15	1.14
	R^2	.56	.62	.60	.62	.63
	MAE	0.81	0.78	0.76	0.75	0.75
NAF	RMSE	1.03	0.99	0.97	0.95	**0.94**
	R^2	.69	.72	.73	.74	**.75**
	MAE	0.60	0.59	0.56	0.55	**0.53**

[5]$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{y_i}-y_i)^2}{n}}$ where y_i corresponds to the predicted value and $\widehat{y_i}$ to the correct value.

[6]$MAE = \sqrt{\frac{\sum_{i=1}^{n}|\widehat{y_i}-y_i|}{n}}$ where y_i corresponds to the predicted value and $\widehat{y_i}$ to the correct value.

[7]$R^2 = \frac{\text{Residual sum of squares}}{\text{Total sum of squares}}$

[8]The feature importance was estimated based on the impurity decrease. It corresponds to the mean decrease in impurity averaged over all nodes where that feature was used to split the node [105]

4.2.2 Classification approach

The SR can be also classed by grouping the SR values into subclasses (e.g., two classes: SR =< 1 corresponding to the class 1 and SR >= 2 corresponding to the class 2). Based on the data distribution and over-representation of SR close to 0 (see figure 3), several data splitting on SR were evaluated: see table 4. The results are presented in the tables 6, 7 and 8.

Only the best configuration in the regression approach (i.e., window of 120 seconds and NAF normalization) were presented in the classification approach. Several metrics were used to measure the recognition performance: Accuracy, Kappa, Sensitivity, Specificity, Precision and F1. The Accuracy[9] varies between 0 and 1 (0 = all data are misclassified and 1 = all data are correctly classified). The Kappa[10] metric [70] is based on the comparison of the observed accuracy with the expected accuracy (corresponding to the random chance). The Sensitivity measures the proportion of positive classes (e.g., SR = 1 identified as SR = 1) that are correctly identified. The Specificity measures the proportion of negative classes (e.g., SR != 1 identified as SR != 1) that are correctly identified. The Precision measures the ratio of true positives to combined true and false positives. The F1 score is the harmonic mean of Precision and Sensitivity.

Table 6: Classification performance (confusion matrix) on 2 classes problem (SR =< 1 and SR >= 2)

		Reference	
		C1	C2
		(SR =<1; 71.3%)	(SR >=2; 28.7%)
Prediction	C1	*68.0%*	5.4%
	C2	3.4%	*23.3%*

Accuracy (average) = .91 (Kappa = .78); Sensitivity = .95; Specificity = .81; Precision = .93; F1 = .94

Table 7: Classification performance (confusion matrix) on 2 classes problem (SR = 0 and SR >= 1)

		Reference	
		C1	C2
		(SR = 0; 61.6%)	(SR >=1; 38.4%)
Prediction	C1	*57.4%*	5.7%
	C2	4.2%	*32.7%*

Accuracy (average) = .90 (Kappa = .79); Sensitivity = .93; Specificity = .85; Precision = .91; F1 = .92

Table 8: Classification performance (confusion matrix) on 3 classes problem (SR = 0; SR >= 1 & =< 2 and SR >= 3)

		Reference		
		C1	C2	C3
		(SR=0;	(SR>=1 & =<2;	(SR>=3;
		61.6%)	22.0%)	15.6 %)
Prediction	C1	*58.3%*	4.7%	1.4%
	C2	2.5%	*15.7%*	2.2%
	C3	0.8%	1.9%	*12.4%*

Accuracy (average) = .87 (Kappa = .75); Sensitivity = .81; Specificity = .92; Precision = .83; F1 = .82

[9] $\text{Accuracy} = \frac{\text{True Positive + True Negative}}{\text{True Positive+True Negative+False Positive+False Negative}}$

[10] $\text{Kappa} = \frac{\text{Observed Accuracy-Expected Accuracy}}{\text{1 - Expected Accuracy}}$

Table 9: Generalization - Classification performance (confusion matrix) on 2 classes problem (SR =< 1 and SR >= 2)

		Reference	
		C1	C2
		(SR =<1; 71.3%)	(SR >=2; 28.7%)
Prediction	C1	*67.5%*	27.8%
	C2	3.8%	*0.9%*

Accuracy (average) = .67 (Kappa = .02); Sensitivity = .96; Specificity = .06; Precision = .70; F1 = .78

Similarly to the regression approach, the top-10 of variable importance for each model have been extracted: see table 11.

Moreover, the models were trained on an unbalanced dataset (i.e., one or several classes are over-represented compared to the others). Even if RF are suitable for this type of situation [16,98], it can lead to weak performances. Several methods to handle unbalanced datasets were evaluated (i.e., random up- and down-sampling [48], SMOTE [22] and weights proportional to the class imbalance [35]), without providing any increase on the ML performances and generalization.

4.3 Generalization

In the context of VR sickness recognition, a person-independent recognition system seems desirable to provide an efficient tool. It requires to be able to recognize cybersickness levels for new users (i.e., unseen participant during the model training). For this purpose, the generalization of our models was evaluated. This step is generally neglected in majority of studies implying physiological data and, generally lead to weak accuracy [114, 125]. In the results presented above, cross-validation was used during the models training. Thus, some data of each participant could be seen by the algorithms during the training. Indeed, using this method, the dataset is shuffled and 10-fold of data are created. Although no data point could be in both the training and the testing datasets, some data gathered from the same participant could be on the training and testing datasets. Consequently, the results may be biased by this approach. In case of weak generalization, for each new user, the model should be retrained. That means, it could be necessary to collect physiological data on the new user under sickness context and retrain the model. So, in order to evaluate the generalization of trained models, another method of validation was used: Leave-One-Out Cross-Validation (LOOCV). This method consists of training a model on all participants except one [114]. Data of the discarded participant are used to test the model. This process is repeated along all participants (As 103 participants were involved in the current study, so 103 models were trained). This approach offers the possibility to evaluate the generalization of the model by testing it on unseen participants. Thereby, the parameters presenting the best performance using the cross-validation method were used to evaluate the generalization of our models to detect VR sickness level: RF algorithm, NAF normalization, and window size of 120 seconds. Under regression approach, the following results have been found: RMSE = 1.71, MAE = 1.52 and R^2 = .14. Under classification approach, an accuracy of 67 % with a Kappa of .02 has been found: see Table 9.

Lastly, to facilitate interpretation of performance, the previous training strategies have been compared to person-dependent and naive models. The results are presented in table 10.

4.4 Real-time Recognition

Based on the processing chain (i.e., filtering and features extraction) and trained models described above, a Proof of Concept was implemented in python using numpy, scipy and scikit-learn. It allows to estimate VR sickness level every second: see figure 5. Preliminary evaluation showed that it is suitable for real-time applications (detection provided in less than 30 ms).

Table 10: Summary of results according to the training strategy (R^2 for regression and Accuracy for classification)

Training strategy	Regression	2 classes	2 classes	3 classes
Naive models[a]	NA	62.4%	72.1%	62.4%
Person-dependent models[b]	.40	88.3% (11.1)	89.0% (10.7)	85.9% (12.3)
Group models[c]	.75	90.1%	91.7%	86.2%
Person-independent models[d]	.14	57.3%	68.3%	58.5%

[a] Model predicts the mean (regression approach) or the most represented class (classification approach)
[b] One model for each participant (models trained and test on data of each participant)
[c] Models trained on all participants
[d] Models trained on all participants except one and tested on this remained participant. The process is repeated along all participants.

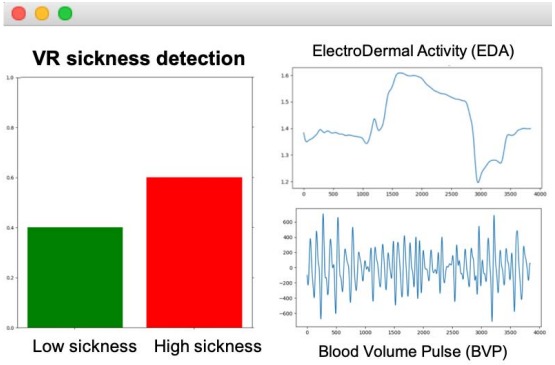

Figure 5: Interface of real-time VR sickness detection

5 Discussion

Cybersickness appears as an huge barrier for the mass adoption of VR. Despite some recent technological equipment improvements to reduce the effects of VR sickness (e.g., reduced latency, refresh improvement, adjustment of FOV [18, 130]), this drawback is still present. Previous research showed that physiological measurements labeled with subjective data can be an efficient approach to recognize cognitive states (e.g., [78, 82]). Thus, sensors whose physiological data has been mapped to subjective data may represent a relevant approach for measuring cybersickness. The current study was conducted in order to train models to recognize VR sickness based on physiological signals. Results showed an interesting recognition performance of up to 75 % of explained variance using regression approach and up to 91 % of accuracy using classification approach. First, it can be noted that the features significantly related to increased cybersickness (see Section Statistical analysis) are similar to those having the most importance in RF models (see Table 11). This confirms the relevance of the trained models and the selected features. Second, several previous studies have attempted to set up physiological measurements for cybersickness. For example, Kim et al. [62] presented a study aggregating several physiological responses (i.e., cardiac, electrodermal and gastrointestinal activities). They found a correlation between some physiological data and the severity of cybersickness. Based on electroencephalogram and electrocardiogram measures using classical statistical modeling, Lin and collaborators [76] distinguished three classes of cybersickness. Dennison and collaborators [31] obtained an accuracy of 77.8% on 2 classes based on 8 sensors. Nevertheless, the comparison with prior research is tricky due to the huge differences in terms of cybersickness induction, sensors or methods of recognition between papers. Li et al. [74] obtained a binary accuracy of 76.3% using EEG, center of pressure, and the head and waist motion trajectories. These approaches seem offer satisfactory performance but require multiple sensors which can be particularly intrusive and difficult to use (e.g., Electrooculography or Electrogastrogram). To the best of our knowledge, there is no study with this type of protocol related to this goal. First, only BVP and EDA were used to predict cybersickness level. Moreover, the sensors used are relatively non-intrusive and easily deployed in industrial and commercial context [108]. Second, interactive contents (i.e., VR video games) and relatively ecological context were used to induce VR sickness which can lead to noise in the data (e.g., motion artifact) and imply a series of difficulties. The

Table 11: Features importance according to data splitting[a]

Rank	C1=<1; C2>=2	C1=0; C2>=1
1	EDA - MAV of the 1st differences of the stzd signal during baseline (100.00)	EDA - Min amplitude during baseline (100.00)
2	EDA - Number of peaks during baseline (98.42)	EDA - SD of IMF1 during baseline (54.93)
3	EDA - Mean of IMF2 (93.29)	MAV of the 2nd differences of the stzd signal (53.35)
4	EDA - MAV of the 2nd differences of the stzd signal during baseline (91.79)	EDA - Mean of half recovery during baseline (51.91)
5	Cardiac - Heart rate during baseline (88.20)	Cardiac - Heart rate (51.80)
6	Cardiac - pNN50 during baseline (85.17)	EDA - Nb of peaks during baseline (50.36)
7	Cardiac - NN50 during baseline (81.71)	EDA - Mean of IMF3 during baseline (48.48)
8	EDA - Mean of half recovery time during baseline (80.09)	Cardiac - pNN50 (47.35)
9	EDA - MAV of the 1st differences of the signal (76.08)	EDA - Min of half recovery time (43.84)
10	EDA - Inter-quartile range during baseline (70.36)	Cardiac - NN50 during baseline (43.25)
Rank	**C1=0; C2>=1 & =<2; C3>=3)[b]**	
1	EDA - Min of peak amplitude during baseline (92.70)	
2	EDA - Nb of peaks during baseline (56.79)	
3	EDA - MAV of the 1st differences of the stzd during baseline (55.68)	
4	EDA - Mean of peak half of recovery during baseline (54.21)	
5	Cardiac - pNN50 (51.86)	
6	Cardiac - Heart rate (51.34)	
7	EDA - MAV of the 2nd differences of the stzd signal during baseline (49.13)	
8	Cardiac - LF/HF (48.63)	
9	EDA - Max of signal (48.15)	
10	Cardiac - NN50 during baseline (47.07)	

[a] Acronyms: MAV = Mean Absolute Value; stzd = standardized
[b] Importance is averaged over the 3 classes

choice to use interactive video games was essential to grant the value of our recognition model for the entertainment industry, but such choice introduced a first drawback: the increase of motion artifacts in the physiological data [1] due to the fact that the participants were not necessarily static. A second drawback connected with the use of our content was that the used prototype video games were created to be engaging: as consequence, the user emotional response influenced the physiological activity of the participants [26]. Third, no information about the content (e.g., quantity of motion) was exploited in the recognition model. Furthermore, even if our results showed an improvement in some way compared to the state of the art, the diversity of experimental settings (e.g., sensors, protocols and ML algorithm) complicates the comparison between studies. Finally, questions may arise about the ability of electrodermal signals to predict cybersickness. Indeed, some authors like [126] found that the increases in skin conductance did not correlate with single indices of motion sickness.

All these constraints, introduced by the need to have an operational and easily deployable VR sickness assessment solution exploitable for commercial and industrial use, made our task particularly challenging and could explain the differences in the performances compared to the other studies in the VR sickness domain. Moreover, the problem of generalization remains a huge scientific limitation of physiological methods and is rarely explored in previous research especially due to weak accuracy [114, 125]. The current research and generalization tests tend to show that the system is quite efficient to identify cybersickness when the participant has been already seen by the model during the training. Beyond good recognition rates, comparison to naive and person-dependent models confirm the relevance of our approach. Conversely, the system shows weak performance when the player is unknown from the system. This weak generalization could be related to the inter-individual variability (e.g., [88]). Some methods of normalization were tested in order to reduce this variability. These approaches have improved the recognition performance without improving generalization. Although the sample of participants are relatively large, it may not be enough to train a powerful system of VR sickness recognition. Lastly, even if person-dependent models are less valuable than person-independent model, it can allow, after training, to adapt a video-game in real time according to the cybersickness state of the user. Indeed, current models can already perform cybersickness detection in real time and can be used for the implementation of cybersickness adaptive countermeasures. These contermeasures could be a simple adaptation such as FOV reduction (which can be applied regardless to the game) or more complex and game dependent adaptation (e.g., modify the scenario of the game to avoid the scenes most likely to induce cybersickness such as stairs or fast movements). A phase of calibration of the model (i.e., the model is adapted with user data), requiring labeling of the data by the user, must be carried out and integrated in the game.

5.1 Limitations and Future Works

Different limitations inherent to this field of research can be cited. First, the accuracy of the recognition model is perfectible. It is likely that other factors besides cybersickness influence physiological signals, making automatic recognition and generalization of results more complex (e.g., motion [120], inter-individual variability [129]). The proposed approach aims to deal with this point by training models to link physiological responses to cybersickness perception by the users. But, the work needs to be extended to evaluate the sensibility and specificity of the proposed detection models of cybersickness and other sickness. Moreover, cybersickness may generate delayed physiological responses [91], which would impact the efficiency of real time recognition. Subjective scale can also skew the data labeling. Indeed, this measure may be influenced by memory process (e.g., peak-and-end effect [46], primacy and recency effects [19]). Moreover, the subjective responses may also suffer from inter-individual variability (i.e., people's estimation of the severity of their sickness can also highly vary across participants. A score of '3' for person A may not be the same as a score of '3' given by person B). Nevertheless, as the individual sensitivity to cybersickness is quite variable [2, 95, 127], the subjective measures seem a more interesting solution than using the content (e.g., quantity of motion) to label the physiological data to be closer to the feeling of the individual. Lastly, the collected SRs are imbalanced with a low average level of cybersickness. It is primary related to the ethical and legal impossibility of voluntarily inducing too high level of cybersickness. More, VR game prototypes have been deliberately used to induce realistic levels of cybersickness present in the video game market. The data collection could be extended to include more extreme level of cybersickness and improve the generalization of the proposed approach. However, the slightest cybersickness is not acceptable for the user and the video game industry and should be avoided as much as possible.

Gender consideration remains an important limitation in VR research and women are underrepresented [100]. The current sample needs to be extended to consider more women and improve the generalization of our approach.

The main objective of the current paper is to build a detection system based on physiological signals. In this way, the relationship between the physiological responses and the subjective perception of cybersickness has been explored through machine learning algorithms. The protocol of induction aims to produce the greater variability in physiological responses under cybersickness condition, in a safe way (i.e., no high level of cybersickness). The way to induce cybersickness is an important aspect and needs to be explored in future studies to improve the generalization of the approach. In this way, the content of stimuli could be more controlled in future studies by exposing the participants exactly to the same stimulus. Nevertheless, it can be complicated in VR games as participants interact with the games resulting in differences in the content viewed.

Some perspectives and possible improvements can be also formulated. The electrodermal and cardiac activities gathering were based on the Shimmer GSR+ sensor. Even if it is relatively not intrusive, it can impose physical constraints on the participant. These physiological activities are now easily gathered using wearable sensors (e.g., Empatica E4) and could provide comparable information [108] while being less intrusive. Furthermore, in the current study, Shallow Learning approach was used. This approach involves the extraction of statistical features (e.g., heart rate from electrocardiogram) from raw data (based on extensive domain knowledge [1, 72]) followed by the training of a model using ML algorithms [1, 67, 72]. Using Deep Learning approach, the features extraction and the classifier are trainable at the same time (i.e., end-to-end trainable model) [83]. This approach drastically increased recognition rates in other areas (e.g., [25, 116]) and could be relevant for VR sickness recognition based on physiological data. For this purpose, it would be necessary to expand the present dataset. As mentioned previously, inter-individual physiological variability may restrain generalization. To solve this issue, some normalization methods have been tested, but are still inefficient. New normalization methods (e.g., [93]) could be evaluated to improve the generalization. Another approach, based on the training of more specialized models, could be used to deal with the inter-individual physiological variability [78]. First, in correspondence with their specific physiological responses, the users could be categorized into subgroups. Then, based on this first categorization, specifically trained ML models (specialized models) could be appropriate for each users groups. Some studies have used other physiological signals (e.g., ocular behavior) or other objective data (e.g., postural sway or optical flow) to classify cybersickness levels [39, 60, 62, 76, 96]. It will be relevant to test the improvement due to inclusion of new types of data. For example, some

correlations are established between cybersickness and ocular behavior [62]. Otherwise, eye tracking will be soon widely available in HMD offering potentially more information about physiological changes induced by cybersickness. Research based on postural instability theory also showed a link between postural sway and VR sickness [21]. Thereby, embedded accelerometer and gyroscope could provide a fully non-intrusive measure of the postural sway and consequently of the VR sickness. Moreover, as cybersickness is related to motion cues in the content (e.g., optical flow [96]), this information could be integrated in the extracted features to perform cybersickness classification. These areas for improvement could enhance the generalization of our models and could be the topic of future research.

6 Conclusion

The major issue of Virtual Reality is still the cybersickness despite the technological improvements. To minimize its occurrence, it is necessary to be able to detect it. The use of physiological data seems to be an interesting approach. For this purpose, physiological signals (electrodermal and cardiac activities) and perceived VR sickness were gathered on 103 participants during VR video game sessions. Using ML methods, models were trained to predict cybersickness intensity and showed an explained variance up to 75 % (in a regression approach) and an accuracy up to 91 % (in a classification approach). Although the results are hopeful for the development of a system requiring calibration steps for each participant, further studies are required to find solutions for a person-independent recognition model. To conclude, this study paves the way to a new method to automatically and continuously detect cybersickness in VR based on physiological signals and Machine Learning algorithms. This approach could help to better understand the VR sickness phenomenon and the characteristics of inductive stimuli. Cybersickness recognition could be useful to evaluate VR content acceptability and offer real time content adaptation.

Acknowledgments

This study was carried out within b<>com, an institute of research and technology dedicated to digital technologies. It received support from the Future Investments program of the French National Research Agency (grant no. ANR-07-A0-AIRT).

References

[1] M. Ali, A. H. Mosa, F. A. Machot, and K. Kyamakya. *Emotion Recognition Involving Physiological and Speech Signals: A Comprehensive Review*, vol. 109, pp. 287–302. Springer International Publishing, 2018. doi: 10.1007/978-3-319-58996-1_13

[2] B. Allen, T. Hanley, B. Rokers, and C. S. Green. Visual 3d motion acuity predicts discomfort in 3d stereoscopic environments. *Entertainment Computing*, 13:1–9, Mar. 2016. doi: 10.1016/j.entcom.2016.01.001

[3] S. Ang and J. Quarles. GingerVR: An Open Source Repository of Cybersickness Reduction Techniques for Unity. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 460–463. IEEE, Atlanta, GA, USA, Mar. 2020. doi: 10.1109/VRW50115.2020.00097

[4] L. Arns and M. Cerney. The relationship between age and incidence of cybersickness among immersive environment users. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, vol. 2005, pp. 267–268, 2005. doi: 10.1109/VR.2005.1492788

[5] W. M. Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA*, 310(20):2191, Nov. 2013. doi: 10.1001/jama.2013.281053

[6] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01

[7] J. Bell. *Machine Learning: Hands-On for Developers and Technical Professionals*. John Wiley & Sons, Inc, Nov 2014.

[8] M. Benedek and C. Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1):80–91, Jun 2010. doi: 10.1016/j.jneumeth.2010.04.028

[9] R. Bertin, C. Collet, S. Espié, and W. Graf. Objective Measurement of Simulator Sickness and the Role of Visual-Vestibular Conflict Situations. 2004.

[10] G. Bertolini and D. Straumann. Moving in a moving world: A review on vestibular motion sickness. *Frontiers in Neurology*, 7, Feb 2016. doi: 10.3389/fneur.2016.00014

[11] P. Bockelman and D. Lingum. *Factors of Cybersickness*, vol. 714, pp. 3–8. Springer International Publishing, 2017. doi: 10.1007/978-3-319-58753-0_1

[12] S. Bouchard, G. Robillard, P. Renaud, and F. Bernier. Exploring new dimensions in the assessment of virtual reality induced side effects. *Journal of Computer and Information Technology*, 1(3):20–32, 2011.

[13] W. Boucsein. *Electrodermal Activity*. Springer-Verlag New York Inc., 2nd ed. ed., Sep 2011.

[14] J. J. Braithwaite, D. G. Watson, R. Jones, and M. Rowe. *A Guide for Analysing Electrodermal Activity (EDA) & Skin Conductance Responses (SCRs) for Psychological Experiments*. 2015.

[15] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001. doi: 10.1023/A:1010933404324

[16] I. Brown and C. Mues. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3):3446–3453, Feb 2012. doi: 10.1016/j.eswa.2011.09.033

[17] P. Budhiraja, M. R. Miller, A. K. Modi, and D. Forsyth. Rotation Blurring: Use of Artificial Blurring to Reduce Cybersickness in Virtual Reality First Person Shooters. *arXiv:1710.02599 [cs]*, Oct. 2017. arXiv: 1710.02599.

[18] H. Buhler, S. Misztal, and J. Schild. Reducing vr sickness through peripheral visual effects. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 517–9. IEEE, Mar 2018. doi: 10.1109/VR.2018.8446346

[19] E. Canic and T. Pachur. Serial-position effects in preference construction: a sensitivity analysis of the pairwise-competition model. *Frontiers in Psychology*, 5, 2014. doi: 10.3389/fpsyg.2014.00902

[20] E. Chang, I. Hwang, H. Jeon, Y. Chun, H. T. Kim, and C. Park. Effects of rest frames on cybersickness and oscillatory brain activity. In *2013 International Winter Workshop on Brain-Computer Interface (BCI)*, pp. 62–64. IEEE, Feb 2013. doi: 10.1109/IWW-BCI.2013.6506631

[21] J.-R. Chardonnet, M. A. Mirzaei, and F. Mérienne. Features of the postural sway signal as indicators to estimate and predict visually induced motion sickness in virtual reality. *International Journal of Human-Computer Interaction*, 33(10):771–785, Oct 2017. doi: 10.1080/10447318.2017.1286767

[22] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, Jun 2002. doi: 10.1613/jair.953

[23] Y.-C. Chen, J.-R. Duann, S.-W. Chuang, C.-L. Lin, L.-W. Ko, T.-P. Jung, and C.-T. Lin. Spatial and temporal EEG dynamics of motion sickness. *NeuroImage*, 49(3):2862–2870, Feb. 2010. doi: 10.1016/j.neuroimage.2009.10.005

[24] I. C. Christie and B. H. Friedman. Autonomic specificity of discrete emotion and dimensions of affective space: a multivariate approach. *International Journal of Psychophysiology*, 51(2):143–153, Jan 2004. doi: 10.1016/j.ijpsycho.2003.08.002

[25] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber. Multi-column deep neural network for traffic sign classification. *Neural Networks*, 32:333–338, Aug 2012. doi: 10.1016/j.neunet.2012.02.023

[26] A. Clerico, C. Chamberland, M. Parent, P.-E. Michon, S. Tremblay, T. H. Falk, J.-C. Gagnon, and P. Jackson. Biometrics and classifier fusion to predict the fun-factor in video gaming. In *2016 IEEE Conference on Computational Intelligence and Games (CIG)*, pp. 1–8. IEEE, Santorini, Greece, Sept. 2016. doi: 10.1109/CIG.2016.7860418

[27] M. Coral. Analyzing cognitive workload through eye-related mea-

surements: A meta-analysis. *Browse all Theses and Dissertations*, Jan 2016.

[28] P. S. Cowings, S. Suter, W. B. Toscano, J. Kamiya, and K. Naifeh. General autonomic components of motion sickness. *Psychophysiology*, 23(5):542–551, Sep 1986. doi: 10.1111/j.1469-8986.1986.tb00671.x

[29] A. Cuomo-Granston and P. D. Drummond. Migraine and motion sickness: What is the link? *Progress in Neurobiology*, 91(4):300–312, 2010. doi: 10.1016/j.pneurobio.2010.04.001

[30] S. Davis, K. Nesbitt, and E. Nalivaiko. A systematic review of cybersickness. In *Proceedings of the 2014 Conference on Interactive Entertainment*, IE2014, pp. 8:1–8:9. ACM, 2014. doi: 10.1145/2677758.2677780

[31] M. S. Dennison, A. Z. Wisti, and M. D'Zmura. Use of physiological signals to predict cybersickness. *Displays*, 44:42–52, Sep 2016. doi: 10.1016/j.displa.2016.07.002

[32] A. Diab, M. Hassan, C. Marque, and B. Karlsson. Performance analysis of four nonlinearity analysis methods using a model with variable complexity and application to uterine EMG signals. *Medical Engineering & Physics*, 36(6):761–767, June 2014. doi: 10.1016/j.medengphy.2014.01.009

[33] C. Diels and P. A. Howarth. Frequency characteristics of visually induced motion sickness. *Human Factors*, 55(3):595–604, Jun 2013. doi: 10.1177/0018720812469046

[34] P. DiZio and J. R. Lackner. Circumventing Side Effects of Immersive Virtual Environments. In *HCI*, 1997.

[35] Q. Dong, S. Gong, and X. Zhu. Imbalanced deep learning by minority class incremental rectification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2018. doi: 10.1109/TPAMI.2018.2832629

[36] I. Doweck, C. R. Gordon, A. Shlitner, O. Spitzer, A. Gonen, O. Binah, Y. Melamed, and A. Shupak. Alterations in r-r variability associated with experimental motion sickness. *Journal of the Autonomic Nervous System*, 67(1-2):31–37, Dec 1997.

[37] C. Elkin and V. Devabhaktuni. Comparative Analysis of Machine Learning Techniques in Assessing Cognitive Workload. In H. Ayaz, ed., *Advances in Neuroergonomics and Cognitive Engineering*, Advances in Intelligent Systems and Computing, pp. 185–195. Springer International Publishing, Cham, 2020. doi: 10.1007/978-3-030-20473-0_19

[38] A. D. Farmer, Y. Al Omran, Q. Aziz, and P. L. Andrews. The role of the parasympathetic nervous system in visually induced motion sickness: systematic review and meta-analysis. *Experimental Brain Research*, 232(8):2665–2673, Aug. 2014. doi: 10.1007/s00221-014-3964-3

[39] T. Feigl, D. Roth, S. Gradl, M. Wirth, M. E. Latoschik, B. M. Eskofier, M. Philippsen, and C. Mutschler. Sick Moves! Motion Parameters as Indicators of Simulator Sickness. *IEEE Transactions on Visualization and Computer Graphics*, 25(11):3146–3157, Nov. 2019. doi: 10.1109/TVCG.2019.2932224

[40] A. S. Fernandes and S. K. Feiner. Combating VR sickness through subtle dynamic field-of-view modification. In *IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 201–210, Mar. 2016. doi: 10.1109/3DUI.2016.7460053

[41] A. S. Fernandes and S. K. Feiner. Combating vr sickness through subtle dynamic field-of-view modification. In *2016 IEEE Symposium on 3D User Interfaces, 3DUI 2016 - Proceedings*, pp. 201–210, 2016. doi: 10.1109/3DUI.2016.7460053

[42] A. Fortin-Côté, C. Chamberland, M. Parent, S. Tremblay, N. Beaudoin-Gagnon, A. Campeau-Lecours, J. Bergeron-Boucher, and L. Lefebvre. Predicting video game players' fun from physiological and behavioural data. 2018.

[43] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.

[44] A. Garcia-Agundez, C. Reuter, P. Caserman, R. Konrad, and S. Göbel. Identifying Cybersickness through Heart Rate Variability alterations. *International Journal of Virtual Reality (IJVR)*, 19(01):1–10, Mar. 2019.

[45] A. M. Gavgani, K. V. Nesbitt, K. L. Blackmore, and E. Nalivaiko. Profiling subjective symptoms and autonomic changes associated with cybersickness. *Autonomic Neuroscience*, 203:41–50, Mar 2017. doi: 10.1016/j.autneu.2016.12.004

[46] X. Geng, Z. Chen, W. Lam, and Q. Zheng. Hedonic evaluation over short and long retention intervals: The mechanism of the peak-end rule: Hedonic evaluation over retention intervals. *Journal of Behavioral Decision Making*, 26(3):225–236, Jul 2013. doi: 10.1002/bdm.1755

[47] J. F. Golding. Phasic skin conductance activity and motion sickness. *Aviation, Space, and Environmental Medicine*, 63(3):165–171, Mar 1992.

[48] H. He and E. Garcia. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284, Sep 2009. doi: 10.1109/TKDE.2008.239

[49] L. He, D. Jiang, L. Yang, E. Pei, P. Wu, and H. Sahli. Multimodal affective dimension prediction using deep bidirectional long short-term memory recurrent neural networks. In *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*, AVEC '15, pp. 73–80. ACM, 2015. doi: 10.1145/2808196.2811641

[50] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf. Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4):18–28, 1998.

[51] S. Hu, W. F. Grant, R. M. Stern, and K. L. Koch. Motion sickness severity and physiological correlates during repeated exposures to a rotating optokinetic drum. *Aviation, Space, and Environmental Medicine*, 62(4):308–314, Apr 1991.

[52] B. E. Insko. *Measuring presence: Subjective, behavioral and physiological methods*, pp. 109–119. Studies in new technologies and practices in communication. IOS Press, 2003.

[53] W. Jin, F. Jianyu, D. Gromala, and P. Pasquier. Automatic prediction of cybersickness for virtual reality games. 2018.

[54] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, eds. *Principles of neural science*. McGraw-Hill, Health Professions Division, New York, 4th ed ed., 2000.

[55] R. S. Kennedy, J. E. Fowlkes, K. S. Berbaum, and M. G. Lilienthal. Use of a motion sickness history questionnaire for prediction of simulator sickness. *Aviation, Space, and Environmental Medicine*, 63(7):588–593, July 1992.

[56] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, Jul 1993. doi: 10.1207/s15327108ijap0303_3

[57] B. Keshavarz and H. Hecht. Validating an efficient method to quantify motion sickness. *Human Factors*, 53(4):415–426, Aug 2011. doi: 10.1177/0018720811403736

[58] B. Keshavarz, H. Hecht, and L. Zschutschke. Intra-visual conflict in visually induced motion sickness. *Displays*, 32(4):181–188, Oct 2011. doi: 10.1016/j.displa.2011.05.009

[59] H. K. Kim, J. Park, Y. Choi, and M. Choe. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. *Applied Ergonomics*, 69:66–73, May 2018. doi: 10.1016/j.apergo.2017.12.016

[60] J. Kim, W. Kim, S. Ahn, J. Kim, and S. Lee. Virtual reality sickness predictor: Analysis of visual-vestibular conflict and vr contents. *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, 2018. doi: 10.1109/QoMEX.2018.8463413

[61] K. H. Kim, S. W. Bang, and S. R. Kim. Emotion recognition system using short-term monitoring of physiological signals. *Medical and Biological Engineering and Computing*, 42(3):419–427, May 2004. doi: 10.1007/BF02344719

[62] Y. Y. Kim, E. N. Kim, M. J. Park, K. S. Park, H. D. Ko, and H. T. Kim. The application of biosignal feedback for reducing cybersickness from exposure to a virtual environment. *Presence*, 17(1):1–16, Feb 2008. doi: 10.1162/pres.17.1.1

[63] Y. Y. Kim, H. J. Kim, E. N. Kim, H. D. Ko, and H. T. Kim. Characteristic changes in the physiological components of cybersickness. *Psychophysiology*, 42(5):616–625, Aug 2005. doi: 10.1111/j.1469-8986.2005.00349.x

[64] A. Kinsella. The effect of 0.2 hz and 1.0 hz frequency and 100 ms and 20 - 100 ms amplitude of latency on simulatory sickness in a head mounted display. 2014.

[65] K. L. Koch, R. M. Stern, M. W. Vasey, J. F. Seaton, L. M. Demers, and T. S. Harrison. Neuroendocrine and gastric myoelectrical responses to illusory self-motion in humans. *The American Journal of Physiology*, 258(2 Pt 1):E304–310, Feb 1990. doi: 10.1152/ajpendo.1990.258.2.E304

[66] S. B. Kotsiantis. Supervised machine learning: A review of classification techniques. In *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, pp. 3–24. IOS Press, 2007.

[67] A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M. R. Wróbel. Emotion recognition and its application in software engineering. In *2013 6th International Conference on Human System Interactions (HSI)*, pp. 532–539, Jun 2013. doi: 10.1109/HSI.2013.6577877

[68] S. D. Kreibig. Autonomic nervous system activity in emotion: a review. *Biological Psychology*, 84(3):394–421, Jul 2010. doi: 10.1016/j.biopsycho.2010.03.010

[69] M. Kuhn, J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, t. R. C. Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, and T. Hunt. *caret: Classification and Regression Training*. 2018.

[70] J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977. doi: 10.2307/2529310

[71] J. LaViola, Joseph J. A discussion of cybersickness in virtual environments. *SIGCHI Bull.*, 32(1):47–56, Jan 2000. doi: 10.1145/333329.333344

[72] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015. doi: 10.1038/nature14539

[73] J. H. Lee, J. Shin, and M. J. Realff. Machine learning: Overview of the recent progresses and implications for the process systems engineering field. *Computers & Chemical Engineering*, 114:111–121, Jun 2018. doi: 10.1016/j.compchemeng.2017.10.008

[74] Y. Li, A. Liu, and L. Ding. Machine learning assessment of visually induced motion sickness levels based on multiple biosignals. *Biomedical Signal Processing and Control*, 49:202–211, Mar. 2019. doi: 10.1016/j.bspc.2018.12.007

[75] J. J. Lin, H. B. L. Duh, D. E. Parker, H. Abi-Rached, and T. A. Furness. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proceedings of the IEEE Virtual Reality*, pp. 164–171. Orange County, CA, USA, Mar. 2002. doi: 10.1109/VR.2002.996519

[76] Y.-T. Lin, Y.-Y. Chien, H.-H. Wang, F.-C. Lin, and Y.-P. Huang. 65-3: The quantization of cybersickness level using eeg and ecg for virtual reality head-mounted display. *SID Symposium Digest of Technical Papers*, 49(1):862–865, May 2018. doi: 10.1002/sdtp.12267

[77] D. Lopez-Martinez and R. Picard. Multi-task neural networks for personalized pain recognition from physiological signals. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*, pp. 181–184. IEEE, Oct 2017. doi: 10.1109/ACIIW.2017.8272611

[78] D. Lopez-Martinez, O. Rudovic, and R. Picard. Physiological and behavioral profiling for nociceptive pain estimation using personalized multitask learning. *arXiv:1711.04036 [cs]*, Nov 2017. arXiv: 1711.04036.

[79] Y. Maeda, M. Sekine, and T. Tamura. Relationship between measurement site and motion artifacts in wearable reflected photoplethysmography. *Journal of Medical Systems*, 35(5):969–976, Oct 2011. doi: 10.1007/s10916-010-9505-0

[80] T. Magaki and M. Vallance. Developing an Accessible Evaluation Method of VR Cybersickness. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1072–1073. IEEE, Osaka, Japan, Mar. 2019. doi: 10.1109/VR.2019.8797748

[81] T. Magaki and M. Vallance. Real-time Monitoring Method for Cybersickness using Physiological Signals. In *25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–2. ACM, Parramatta NSW Australia, Nov. 2019. doi: 10.1145/3359996.3364700

[82] N. Martin and J.-M. Diverrez. *From Physiological Measures to an Automatic Recognition System of Stress*, vol. 618, pp. 172–176. Springer International Publishing, 2016. doi: 10.1007/978-3-319-40542-1_27

[83] N. Martin, J.-M. Diverrez, S. Em, N. Pallamin, and M. Ragot. How to Recognize Emotions Without Signal Processing. In *Neuroergonomics*, pp. 191–194. Elsevier, 2019. doi: 10.1016/B978-0-12-811926-6.00032-4

[84] N. Martin, N. Mathieu, N. Pallamin, M. Ragot, and J.-M. Diverrez. Automatic recognition of Virtual Reality Sickness based on physiological signals. In *International Broadcasting Convention (IBC)*. Amsterdam, The Netherlands, Sept. 2018.

[85] S. Martirosov and P. Kopecek. *Cyber Sickness in Virtual Reality - Literature Review*, vol. 1, pp. 0718–0726. DAAAM International Vienna, 1 ed., 2017. doi: 10.2507/28th.daaam.proceedings.101

[86] M. E. McCauley and T. J. Sharkey. Cybersickness: Perception of self-motion in virtual environments. *Presence: Teleoperators and Virtual Environments*, 1(3):311–318, Jan 1992. doi: 10.1162/pres.1992.1.3.311

[87] C. R. Meusel. *Exploring mental effort and nausea via electrodermal activity within scenario-based tasks*. Master of Science, Iowa State University, Digital Repository, Ames, 2014. doi: 10.31274/etd-180810-3757

[88] J. C. Miller, T. J. Sharkey, G. A. Graham, and M. E. McCauley. Autonomic physiological data associated with simulator discomfort. *Aviation, Space, and Environmental Medicine*, 64(9 Pt 1):813–819, Sep 1993.

[89] J. M. Mittelstaedt, J. Wacker, and D. Stelling. Vr aftereffect and the relation of cybersickness and cognitive performance. *Virtual Reality*, Nov 2018. doi: 10.1007/s10055-018-0370-3

[90] M. Mousavi, Y. H. Jen, and S. N. B. Musa. A review on cybersickness and usability in virtual environments. *Advanced Engineering Forum*, 10:34–39, Dec 2013. doi: 10.4028/www.scientific.net/AEF.10.34

[91] E. R. Muth. The challenge of uncoupled motion: duration of cognitive and physiological aftereffects. *Human Factors*, 51(5):752–761, Oct 2009. doi: 10.1177/0018720809353320

[92] Y. H. Nam, Y. Y. Kim, H. T. Kim, H. D. Ko, and K. S. Park. Automatic detection of nausea using bio-signals during immersion in a virtual reality environment. In *2001 Conference Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 2013–2015 vol.2, 2001. doi: 10.1109/IEMBS.2001.1020626

[93] N. Nourbakhsh, Y. Wang, F. Chen, and R. A. Calvo. Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks. In *Proceedings of the 24th Australian Computer-Human Interaction Conference on - OzCHI'12*, pp. 420–423. ACM Press, 2012. doi: 10.1145/2414536.2414602

[94] D. Novak, M. Mihelj, and M. Munih. A survey of methods for data fusion and system adaptation using autonomic nervous system responses in physiological computing. *Interacting with Computers*, 24(3):154–172, May 2012. doi: 10.1016/j.intcom.2012.04.003

[95] S. Ohyama, S. Nishiike, H. Watanabe, K. Matsuoka, H. Akizuki, N. Takeda, and T. Harada. Autonomic responses during motion sickness induced by virtual reality. *Auris, Nasus, Larynx*, 34(3):303–306, Sep 2007. doi: 10.1016/j.anl.2007.01.002

[96] N. Padmanaban, T. Ruban, V. Sitzmann, A. M. Norcia, and G. Wetzstein. Towards a machine-learning approach for sickness prediction in 360° stereoscopic videos. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1594–1603, Apr 2018. doi: 10.1109/TVCG.2018.2793560

[97] A. C. Paillard, G. Quarck, F. Paolino, P. Denise, M. Paolino, J. F. Golding, and V. Ghulyan-Bedikian. Motion sickness susceptibility in healthy subjects and vestibular patients: Effects of gender, age and trait-anxiety. *Journal of Vestibular Research: Equilibrium and Orientation*, 23(4-5):203–210, 2013. doi: 10.3233/VES-130501

[98] M. Pal. Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1):217–222, Jan 2005. doi: 10.1080/01431160412331269698

[99] B. Patrão, S. Pedro, and P. Menezes. How to Deal with Motion Sickness in Virtual Reality. In *Sciences and Technologies of Interaction (SciTecIN'15)*. Coimbra, Portugal, 2015.

[100] T. C. Peck, L. E. Sockol, and S. M. Hancock. Mind the Gap: The Underrepresentation of Female Participants and Authors in Virtual

Reality Research. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1945–1954, May 2020. doi: 10.1109/TVCG.2020.2973498

[101] R. W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(10):1175–1191, Oct 2001. doi: 10.1109/34.954607

[102] J. Plouzeau, J.-R. Chardonnet, and F. Merienne. Using Cybersickness Indicators to Adapt Navigation in Virtual Reality: A Pre-Study. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 661–662. IEEE, Reutlingen, Mar. 2018. doi: 10.1109/VR.2018.8446192

[103] J. Prothero and D. Parker. *A unified approach to presence and motion sickness*, pp. 47–66. Lawrence Erlbaum Associates Publishers, 2003. doi: 10.1201/9781410608888.ch3

[104] J. D. Prothero. *The Role of Rest Frames in Vection, Presence and Motion Sickness*. PhD thesis, University of Washington, 1998.

[105] Y. Qi. Random Forest for Bioinformatics. In C. Zhang and Y. Ma, eds., *Ensemble Machine Learning*, pp. 307–323. Springer US, Boston, MA, 2012. doi: 10.1007/978-1-4419-9326-7_11

[106] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

[107] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.

[108] M. Ragot, N. Martin, S. Em, N. Pallamin, and J.-M. Diverrez. *Emotion Recognition Using Physiological Signals: Laboratory vs. Wearable Sensors*, vol. 608, pp. 15–22. Springer International Publishing, 2018. doi: 10.1007/978-3-319-60639-2_2

[109] J. T. Reason. Motion sickness adaptation: a neural mismatch model. *Journal of the Royal Society of Medicine*, 71(11):819–829, Nov 1978.

[110] L. Rebenitsch and C. Owen. Review on cybersickness in applications and visual displays. *Virtual Reality*, 20(2):101–125, Jun 2016. doi: 10.1007/s10055-016-0285-9

[111] A. Sano, W. Chen, D. L. Martinez, S. Taylor, and R. W. Picard. Multimodal ambulatory sleep detection using lstm recurrent neural networks. *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2018. doi: 10.1109/JBHI.2018.2867619

[112] V. Sevinc and M. I. Berkman. Psychometric evaluation of Simulator Sickness Questionnaire and its variants as a measure of cybersickness in consumer virtual environments. *Applied Ergonomics*, 82:102958, Jan. 2020. doi: 10.1016/j.apergo.2019.102958

[113] Y. Shimomura, T. Yoda, K. Sugiura, A. Horiguchi, K. Iwanaga, and T. Katsuura. Use of frequency domain analysis of skin conductance for evaluation of mental workload. *Journal of Physiological Anthropology*, 27(4):173–177, June 2008. doi: 10.2114/jpa2.27.173

[114] L. Shu, J. Xie, M. Yang, Z. Li, Z. Li, D. Liao, X. Xu, and X. Yang. A review of emotion recognition using physiological signals. *Sensors (Basel, Switzerland)*, 18(7), Jun 2018. doi: 10.3390/s18072074

[115] A. Shupak and C. R. Gordon. Motion sickness: advances in pathogenesis, prediction, prevention, and treatment. *Aviation, Space, and Environmental Medicine*, 77(12):1213–1223, Dec. 2006.

[116] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, Jan. 2016. doi: 10.1038/nature16961

[117] Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures. Publication recommendations for electrodermal measurements: Publication standards for EDA. *Psychophysiology*, 49(8):1017–1034, Aug. 2012. doi: 10.1111/j.1469-8986.2012.01384.x

[118] C. L. Stephens, I. C. Christie, and B. H. Friedman. Autonomic specificity of basic emotions: Evidence from pattern classification and cluster analysis. *Biological Psychology*, 84(3):463–473, Jul 2010. doi: 10.1016/j.biopsycho.2010.03.014

[119] W. B. Stone Iii. *Psychometric evaluation of the Simulator Sickness Questionnaire as a measure of cybersickness*. PhD thesis, Iowa State University, 2017.

[120] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss. Activity-aware mental stress detection using physiological sensors. In M. Gris and G. Yang, eds., *Mobile Computing, Applications, and Services*, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, pp. 211–230. Springer Berlin Heidelberg, 2012.

[121] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida. Wearable photoplethysmographic sensors-past and present. *Electronics*, 3(2):282–302, Apr 2014. doi: 10.3390/electronics3020282

[122] B. S. Todd and D. C. Andrews. The identification of peaks in physiological signals. *Computers and Biomedical Research*, 32(4):322–335, Aug 1999. doi: 10.1006/cbmr.1999.1518

[123] E. L. van den Broek, V. Lisý, J. H. Janssen, J. H. D. M. Westerink, M. H. Schut, and K. Tuinenbreijer. Affective Man-Machine Interface: Unveiling Human Emotions through Biosignals. In A. Fred, J. Filipe, and H. Gamboa, eds., *Biomedical Engineering Systems and Technologies*, Communications in Computer and Information Science, pp. 21–47. Springer Berlin Heidelberg, 2010.

[124] M. van Dooren, J. G.-J. de Vries, and J. H. Janssen. Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology & Behavior*, 106(2):298–304, May 2012. doi: 10.1016/j.physbeh.2012.01.020

[125] P. van Gent, T. Melman, H. Farah, N. van Nes, and B. van Arem. Multi-Level Driver Workload Prediction using Machine Learning and Off-the-Shelf Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, 2672(37):141–152, Dec. 2018. doi: 10.1177/0361198118790372

[126] L. A. Warwick-Evans, R. E. Church, C. Hancock, D. Jochim, P. H. Morris, and F. Ward. Electrodermal activity as an index of motion sickness. *Aviation, Space, and Environmental Medicine*, 58(5):417–423, 1987.

[127] S. Weech, J. Parokaran Varghese, and M. Barnett-Cowan. Estimating the sensorimotor components of cybersickness. preprint, Neuroscience, May 2018. doi: 10.1101/318758

[128] C. Wienrich, C. K. Weidner, C. Schatto, D. Obremski, and J. H. Israel. A Virtual Nose as a Rest-Frame - The Impact on Simulator Sickness and Game Experience. In *2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)*, pp. 1–8. IEEE, Wurzburg, Sept. 2018. doi: 10.1109/VS-Games.2018.8493408

[129] D. Wu and T. D. Parsons. *Inductive Transfer Learning for Handling Individual Differences in Affective Computing*, vol. 6975, pp. 142–151. Springer Berlin Heidelberg, 2011. doi: 10.1007/978-3-642-24571-8_15

[130] H. K. Yaqub. *Reducing Head Mounted Display VR Sickness Through Dynamic Field of View Constriction*. PhD thesis, University of Bath, Jun 2018.

[131] B. J. Yates and A. M. Bronstein. The effects of vestibular system lesions on autonomic regulation: observations, mechanisms, and clinical implications. *Journal of Vestibular Research: Equilibrium & Orientation*, 15(3):119–129, 2005.

[132] Y. Yokota, M. Aoki, K. Mizuta, Y. Ito, and N. Isu. Motion sickness susceptibility associated with visually induced postural instability and cardiac autonomic responses in healthy subjects. *Acta Oto-Laryngologica*, 125(3):280–285, Mar 2005.

[133] D. Zielasko, A. Meißner, S. Freitag, B. Weyers, and T. W. Kuhlen. Dynamic field of view reduction related to subjective sickness measures in an hmd-based data analysis task. In *IEEE 4th Workshop on Everyday Virtual Reality (WEVR)*, 2018.