Automatic recognition of Virtual Reality sickness based on physiological signals

Nicolas Martin, Nicolas Mathieu, Nico Pallamin, Martin Ragot, Jean-Marc Diverrez

AUTOMATIC RECOGNITION OF VIRTUAL REALITY SICKNESS BASED ON PHYSIOLOGICAL SIGNALS

Nicolas Martin[1], Nicolas Mathieu[2], Nico Pallamin[1], Martin Ragot[1], Jean-Marc Diverrez[1]

[1]b<>com, France; [2]Editorial user research lab, Ubisoft Paris, France

ABSTRACT

Virtual Reality (VR) sickness seems one of the main limitations to the large-scale adoption of VR technologies. This disturbance seems to induce physiological changes that affect the sympathetic and parasympathetic activities of the users. Thereby, it seems relevant to measure users' physiological data in order to prevent and reduce VR sickness. This paper presents the results of an initial real-life experiment of VR sickness detection based on physiological data. The electrodermal, cardiac and subjective data of 27 participants was recorded during VR sessions. Machine Learning algorithms were trained and the best model (Gradient Boosting) explained 48% of the VR sickness variance. These results demonstrate the opportunity to develop an automatic and continuous tool to detect the appearance of VR sickness based on physiological signals. This tool will prove very valuable to the VR industry.

INTRODUCTION

Virtual Reality (VR) appears as a major technological breakthrough and a main business opportunity for the entertainment industry. The VR market is expected to expand exponentially with worldwide revenues for the AR/VR growing to more than $162 billion in 2020 [1]. However, one main limitation to its large-scale adoption is VR sickness especially because of health, ethical, legal and acceptability aspects. VR sickness is a common problem that could affect up to 60% of adult users [2].

The necessity to better detect and prevent the appearance of VR sickness is at the origin of this research cooperation between b<>com and Editorial user research lab, Ubisoft Paris aiming at the development of an automatic VR sickness detection tool.

VIRTUAL REALITY SICKNESS: CAUSES AND SYMPTOMS

Previous research proposed several hypotheses to explain VR Sickness. The most common theory is the sensory conflict theory [3]. According to this theory, VR sickness is the result of conflicts and inconsistency between the different sensorial information sent to the brain when the user evolves in the virtual world. In VR, the most common conflict is a discrepancy between the motion information coming from the vestibular system and the motion information coming from the visual system. Indeed, this information comes from two separate systems: the vestibular system located in the inner ear and the visual system

[4]. This discrepancy will be detected by the brain and will induce, in sensitive participants, the symptoms of VR sickness.

VR sickness is a complex phenomenon and while motion cues play a primary role, multiple factors are known to contribute to the appearance of the sickness. Three main categories of factors can be identified:

1) The factors related to the characteristics of the stimuli: the spatial frequency [5], the reactivity of the system [6], the wideness of the Field Of View (FOV) [7], etc.
2) The factors related to the predispositions of the users: gender [8], age [9], the predisposition to suffer from migraine attack [10], etc.
3) The users' past experiences [11], according to the sensory conflict theory. In fact VR sickness appears only when a present set of vestibular and visual information is not congruent with what is expected from previous experiences [12].

The complex nature of VR sickness is not limited to its different causes, but it is reflected in its symptomatology as well. While the most evident and detrimental symptom of VR sickness is the nausea, the complete symptomatology of VR sickness includes other elements like general discomfort, headache, disorientation and eye strain. The intensity as well as the duration of the symptoms are quite variable. They depend on the characteristic of the stimulus and the user predisposition to VR sickness. In the majority of cases, the symptoms disappear some minutes after the end of the stimulation. Nevertheless, there are documented cases when the symptoms were still present 6 hours after the VR experience [13].

THE ASSESSMENT OF VIRTUAL REALITY SICKNESS

The traditional way to evaluate VR sickness is based on subjective questionnaires. Various questionnaires have been previously developed to assess a sickness level. The most popular is probably the "Simulator Sickness Questionnaire" (SSQ) proposed by Kennedy in 1993 [14]. It is constituted of 16 questions to evaluate 3 categories of symptoms: nausea related symptoms, oculomotor symptoms and disorientation symptoms.

While still widely used for the simplicity of its deployment and analysis, the SSQ has several limitations. Indeed, this measure is global, gathered a posteriori, intrusive and punctual. These limitations make the SSQ (and any other questionnaire) inadequate to be integrated in an automatic tool to evaluate VR sickness. Another method was proposed by Keshavarz and collaborators [15] which consists in requesting the participants to verbally give their evaluation of sickness level on a scale ranging from 0 to 20, with zero representing no discomfort and twenty representing barely supportable sickness. This approach seems to be a very good alternative to the SSQ since very strong correlation (r = .79) between verbal rating and SSQ scores were reported in this study. Nevertheless, this method is cognitively disruptive for the user experience in a context of VR-video games.

To be effective and valuable for the VR industry, the assessment method must ideally be continuous, automatic and real time. Considering these constraints, physiological data related to the perceived VR sickness was used in this paper.

The physiological measures of VR sickness

The physiological response associated to VR sickness is due to the connections between the vestibular and the autonomic nervous system. The conflicting inputs from visual, vestibular and somatosensory afferents induce a vestibular autonomic response. This response involves both the sympathetic and the parasympathetic systems and affects for instance the heart rate variability and the skin conductance [16].

Various researches support the hypothesis of a measurable VR sickness by physiological response. For instance, Gavgani and collaborators [17] underline a correlation between phasic skin conductance activity and the reported nausea ratings. While Dennison [18] found relations between subscale scores of the SSQ and bradygastric power, breathing rate, pulse amplitude and blinking. Moreover, Ohyama and collaborators [19] investigated the cardiac responses associated to VR sickness. But, cardiac responses didn't statically correlate with the user subjective evaluation.

These findings suggest the existence of a relation between the physiological responses and VR sickness. Nevertheless, such relation seems too complex to be exploited using classical analyses. To handle such complex data, Machine Learning approach has spread in recent years to go beyond the limits of classical statistical approaches. Nam and collaborators [20] use such method to detect nausea in VR using various bio-signals (EEG, ECG, PPG, SCL, EOG, SKT). Results seem promising as they are able to partially detect nausea in real time [20].

Automatic recognition of VR sickness: the Machine Learning approach

To automatically detect VR sickness using physiological data, the use of Machine Learning algorithms seem relevant. In this approach, physiological and declared VR sickness are mapped using Machine Learning algorithms. In others words, the training of Machine Learning algorithms, using the supervised techniques [21], aims to infer the function between the input data (i.e., physiological data) and the output data (i.e., subjective labels). Indeed, Machine Learning algorithms have the ability to learn without being explicitly programmed [22]. After training, the models should be able to automatically recognize in real time, for any new user, the VR sickness related to the physiological data without requesting a subjective response.

Physiological signals and subjective VR sickness level were collected to investigate the possibility of detecting VR sickness automatically. A large and relevant labeled dataset (i.e., physiological data labeled with subjective evaluation of VR sickness) to train models is fundamental to obtain an efficient system of VR sickness detection. So, Editorial user research lab, Ubisoft Paris and b<>com decided to cooperate with the intent to collect and analyze consistent amounts of physiological data of test users viewing VR content. Contrary to previous studies [19], [20] the aim of the current paper is to develop and evaluate a new solution more adapted to the industrial context than previous approaches based on invasive and/or expensive sensors (e.g., EEG or eye tracking).

METHOD

27 participants (22 men and 5 women - average age: 25.93 years; standard deviation: 4.39) were recruited by Ubisoft. Participants were requested to test three VR game prototypes during 30 minutes. The prototypes were interactive games currently under

development and they implied a relevant amount of camera's movements. The first game was a space simulation using a third-person point of view. The second was an arcade car racing game (in first-person and third-person point of view) and the third was a space first-person shooter. To increase the probability of VR sickness we didn't implement in these prototypes any of the typical countermeasures (i.e. Reduction of the peripheral vision) applied in the gaming industry to prevent VR sickness. The participants were informed that they were free to stop the experiment at any time.

In order to induce various levels of VR sickness, the game's levels tested by the participants were designed to increase progressively the VR sickness induction. VR content was presented using the "Oculus Rift"[1] ® or "HTC Vive" [2] ® devices.

Measures

Two types of measures were collected: the physiological data and subjective data of the users' sickness.

The physiological data was collected using the "Shimmer GSR+"®[3]. This sensor measures the electrodermal activity (EDA) and the Blood Pulse Volume (BVP) using PhotoPlethysmoGraphy (PPG) as the measure of cardiac activity. EDA sensors were placed on the middle phalanx of the non-dominant hand two first fingers. The data was collected in a continuous way during the experimental session and the sampling frequency was set at 128 Hz.

The subjective data was collected following the method proposed by Keshavarz and collaborators [15] presented in the previous section. Participants were instructed to express their sickness level evaluation on a scale from 0 to 20 in response of auditory stimuli presented every 40 seconds. To grant the accuracy of the reply and the correct time alignment, the audio of the responses was recorded during each session and transcribed at the end of each experimental session.

Features extraction

The raw physiological signals appear very complex to interpret directly. So, it is generally necessary to extract specific features from raw signals. To extract these features, a toolbox was specially developed in Python. For this purpose, the raw signals were divided into subsamples of 30 seconds. For each of the 715 reported subjective responses, the last 30 seconds of physiological data (BVP and EDA) are used to extract the physiological features.

Concerning the cardiac activity, two types of features are extracted [23]: time-domain features (HR, AVNN, SDNN, nn50, pNN50) and frequency-domain features (VLF, LF, LF/HF, Total Power).

Concerning the EDA, the signal is composed of two components: a phasic and a tonic part [24]. The phasic part (also called Skin Conductance Level - SCL) corresponds to slow change in the EDA while the tonic part (also called Skin Conductance Responses - SCR)

---

[1] https://www.oculus.com/rift/
[2] https://www.vive.com/eu/
[3] http://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor

corresponds to the rapid physiological responses to a stimulus. The process of extracting those two components from the raw signal is constituted of different steps. A low-pass filter (cutoff frequency = 1Hz, order = 3) is firstly applied to reduce noise in the signal. Then, the tonic part of EDA is extracted from this filtered signal using a low-pass filter (cutoff frequency = 0.05Hz, order = 3) [25] and averaged over the whole filtered signal (mean EDA). The phasic part is obtained by subtracting the tonic signal to the filtered signal. On the phasic part, the SCR are manifested by peaks on signals. On these peaks[4], the amplitude (mean, standard deviation, minimum and maximum) and half of recovery time (mean, standard deviation, minimum and maximum) are measured. Moreover, an Empirical Mode Decomposition [26] was applied to EDA signal to extract Intrinsic Mode Functions (IMF).

Table 1.List of extracted physiological features

BVP features	**EDA features**
HR (Heart Rate) AVNN (Average of all NN intervals) SDNN (Standard deviation of all NN intervals) rMMSD (Root-mean square differences of successive of all NN intervals) nn50 (The number of interval differences of successive NN intervals greater than 50 ms) pNN50 (Percentage of differences between adjacent NN intervals that are greater than 50 ms) VLF (Total spectral power of all NN intervals between 0.003 and 0.04 Hz) LF (Total spectral power of all NN intervals between 0.04 and 0.15 Hz) HF (Total spectral power of all NN intervals between 0.15 and 0.4 Hz) LF/HF (Ratio of low to high frequency power) Total Power (Total spectral power)	Mean EDA Amplitude: mean, standard deviation, minimum and maximum Half of recovery: mean, standard deviation, minimum and maximum IMF(1 to 3): mean, standard deviation, minimum and maximum

RESULTS

Analysis of correlation

Based on the extracted features, Pearson correlations between the physiological features and the subjective data were computed. Only five correlations appear as significant. Of

[4] An EDA peak is especially characterized by the amplitude (the height of the peak) and the recovery time (time to return to the level EDA before the peak).

these five correlations, three concern the EDA features: the Mean Amplitude of the EDA (r = .12), the Minimal Amplitude of the EDA (r = -.14) and the Maximal Amplitude of the EDA (r = -.07). The two remaining correlations concern cardiac features: the Heart Rate (r = .22) and the NN50 (r = .18).

The majority of the 33 computed correlations appear as not significant. This can be the consequence of the multidimensional and nonlinear relationships between the physiological features and the subjective evaluation of VR. So, to handle such complex data and recognize VR sickness, Machine Learning techniques were used.

Machine Learning

In order to create an automatic tool to recognize VR sickness, Machine Learning models were trained. To test the performance of trained model, the dataset was randomly split in two independent parts: a training dataset and a testing dataset corresponding respectively to the 80% (572 samples) and the 20% (143 samples) of the original dataset. Then, three Machine Learning models were trained to select the most efficient model for our task. The three selected models were: 1) Random Forest, 2) Gradient Bosting and 3) Support Vector Machine.

Regression models were chosen instead of classification models. Indeed, a continuous measure of VR sickness seems more adapted to represent the human perception compared to exclusive classes. Three metrics were computed for each method: RMSE (Root Mean Square Error), MAE (Mean Absolute Error) and R-Square (i.e., coefficient of determination)[5].

Table 2. Results for Random Forest, Gradient Boosting and Support Vector Machine

Model	RMSE	MAE	R-Squared
Random Forest	1.52	1.15	.45
Gradient Boosting	1.55	1.05	.48
Support Vector Machine	1.57	1.31	.03

According to these results, Gradient Boosting method seems to offer the best result with more variance explained and smaller error.

CONCLUSION

VR is potentially a major business opportunity with an exponential growth. But serious acceptability issues, in particular VR sickness, could interfere with the uptake of this technological breakthrough by the mass market. In this paper, a solution to detect the development of VR sickness based on physiological signals was tested. Models were trained to recognize VR sickness using Machine Learning methods and showed promising results (the best model explained 48% of the VR sickness variance). Indeed, the ability to evaluate in a continuous and automatic way the appearance of the users' sickness seems

[5] RMSE and MAE evaluate the distance between the predicted data and the ground truth (0 corresponds to a perfect prediction). R-Square corresponds to the explained variance (it ranges from 0 to 1, with 1 indicates that the model explains all the variability of the output variable).

a necessary stage to prevent VR sickness. Further research work will be conducted to confirm the results and evaluate the performance of the tool in real time.

Lastly, this type of automatic system able to assess the level of VR sickness of the end user seems extremely valuable to all the stakeholders in the VR ecosystem. As a standalone application, this tool can in fact be used to assess the acceptability of the content (both videogames and 360° videos) in the early stages of the production chain or can be used to validate the efficiency of sickness reduction strategies (e.g., reduction of the FOV).

REFERENCES

[1] IDC, "Worldwide Semiannual Augmented and Virtual Reality Spending Guide," *IDC: The premier global market intelligence company*, 2016. [Online]. Available: http://www.idc.com/getdoc.jsp?containerId=prUS41676216.

[2] C.-H. Chang, W.-W. Pan, L.-Y. Tseng, and T. A. Stoffregen, "Postural activity and motion sickness during video game play in children and adults," Experimental Brain Research, vol. 217, no. 2, pp. 299 to 309, March, 2012.

[3] J. T. Reason, "Motion sickness adaptation: a neural mismatch model.," Journal of the Royal Society of Medicine, vol. 71, no. 11, pp. 819 to 829, November, 1978.

[4] M. Mousavi, Y. H. Jen, and S. N. B. Musa, "A Review on Cybersickness and Usability in Virtual Environments," Advanced Engineering Forum, vol. 10, pp. 34 to 39, December, 2013.

[5] C. Diels and P. A. Howarth, "Frequency characteristics of visually induced motion sickness," Human Factors, vol. 55, no. 3, pp. 595 to 604, June, 2013.

[6] M. E. St. Pierre, S. Banerjee, A. W. Hoover, and E. R. Muth, "The effects of 0.2Hz varying latency with 20–100ms varying amplitude on simulator sickness in a helmet mounted display," Displays, vol. 36, pp. 1 to 8, January, 2015.

[7] 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, pp. 201 to 210.

[8] A. C. Paillard *et al.*, "Motion sickness susceptibility in healthy subjects and vestibular patients: effects of gender, age and trait-anxiety," Journal of Vestibular Research: Equilibrium & Orientation, vol. 23, no. 4–5, pp. 203 to 209, 2013.

[9] L. L. Arns and M. M. Cerney, "The relationship between age and incidence of cybersickness among immersive environment users," in IEEE Proceedings. VR 2005. Virtual Reality, 2005., 2005, pp. 267 to 268.

[10] A. Cuomo-Granston and P. D. Drummond, "Migraine and motion sickness: what is the link?," Progress in Neurobiology, vol. 91, no. 4, pp. 300 to 312, August, 2010.

[11] G. Bertolini and D. Straumann, "Moving in a Moving World: A Review on Vestibular Motion Sickness," Frontiers in Neurology, vol. 7, February, 2016.

[12] C. M. Oman, "A heuristic mathematical model for the dynamics of sensory conflict and motion sickness," Acta Oto-Laryngologica, vol. 392, pp. 1 to 44, 1982.

[13] E. M. Kolasinski and U. S. A. R. I. for the B. and S. Sciences, *Simulator sickness in virtual environments*. U.S. Army Research Institute for the Behavioral and Social Sciences, 1995.

[14] 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, vol. 3, no. 3, pp. 203 to 220, July, 1993.

[15] B. Keshavarz and H. Hecht, “Validating an efficient method to quantify motion sickness,” Human Factors, vol. 53, no. 4, pp. 415 to 426, August, 2011.

[16] 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.

[17] P. J. Gianaros, K. S. Quigley, E. R. Muth, M. E. Levine, R. C. Vasko, and R. M. Stern, “Relationship between temporal changes in cardiac parasympathetic activity and motion sickness severity,” Psychophysiology, vol. 40, no. 1, pp. 39 to 44, January, 2003.

[18] M. S. Dennison, A. Z. Wisti, and M. D’Zmura, “Use of physiological signals to predict cybersickness,” Displays, vol. 44, pp. 42 to 52, September, 2016.

[19] S. Ohyama *et al.*, “Autonomic responses during motion sickness induced by virtual reality,” Auris Nasus Larynx, vol. 34, no. 3, pp. 303 to 306, September, 2007.

[20] 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, 2001, vol. 2, pp. 2013 to 2015 vol.2.

[21] 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, Amsterdam, The Netherlands, The Netherlands, 2007, pp. 3 to 24.

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

[23] S. D. Kreibig, “Autonomic nervous system activity in emotion: a review,” Biological Psychology, vol. 84, no. 3, pp. 394 to 421, July, 2010.

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

[25] 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,” Psychophysiology, vol. 49, no. 1, pp. 1017 to 1034, 2013.

[26] 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, New York, NY, USA, 2015, pp. 73 to 80.