PHYSIOLOGICAL MEASURES IN VIDEO GAMES USER EXPERIENCE MODELING

Ing. Martin Čertický
Technická Univerzita v Košiciach
Letná 9, Košice
e-mail: martin.certicky@tuke.sk

prof. Ing. Peter Sinčák, CSc.
Technická Univerzita v Košiciach
Letná 9, Košice
e-mail: peter.sincak@tuke.sk

Abstract
Analyses of user experience in electronic entertainment industry currently rely on self-reporting methods, such as surveys, ratings, focus group interviews, etc. We argue that self-reporting alone carries inherent problems - mainly the subject bias and interpretation difficulties - and therefore should not be used as a sole metric. To deal with this problem, we propose a possibility of creating a model of consumer experience based on psychophysiological measurements and describe how such model can be trained using machine learning methods. Models trained exclusively on real-time data produced by autonomic nervous system and involuntary physiological responses is not susceptible to subjective bias, misinterpretation and imprecision caused by the delay between the experience and the interview. This paper proposes a potentially promising direction for future research and presents an introductory analysis of available biological data sources, their relevance to user experience modeling and technical prerequisites for their collection. Multiple physiological measurements (such as heart rate, electrodermal activity or respiratory activity) should be used in combination with self-reporting methods to prepare training sets for machine learning models.
Key words: Physiological measurements, Enjoyment, Modeling, Machine learning, Time-series

INTRODUCTION

Electronic entertainment (EE), especially video-games, has undergone an extensive growth over the past decade. In the USA alone (a country with the biggest video-game market in the world) the sales went from $7.3 billion in 2006 to $23.5 billion spent in the game industry in 2015. The global value of video-game industry is estimated to reach $120 billion by 2019 [1].

Yet the assessment of user experience in video-games is still done using old-fashioned self-report techniques, such as questionnaires and interviews.

We propose an approach to assessing player experience more effectively while eliminating some disadvantages of self-report techniques, such as untruthfulness or emotional bias.

Based on the assumption that psychophysiological measurements, such as heart rate, electrodermal activity or respiratory activity, correlate with user experience, we suggest combining them with the player's self-reports to create machine learning-based model of the player's experience.

STATE OF THE ART

Kivikangas et al. reviewed many scientific research papers concerning the use of psychophysiological measures in video-games [2]. The review investigated various types of psychophysiological metrics, pros and cons of research and particular causes of changes in subject's psychophysiological states. Another study relevant to our research is the work of Drachen et al. who studied a few correlations between self-reporting and psychophysiological measurements [3]. In their research, they used iGEQ (In-Game Experience Questionnaire) as a self-reporting method in combination with heart rate and electrodermal activity measures [4].

They found statistically significant correlations between HR/EDA and iGEQ dimensions, although with different patterns of covariance. To prevent such covariance (considering more dimensions with a plan to use more psychophysiological measures) we plan to design our own questionnaire for future experiments. Drachen focused on the effects of particular game genres on the subject's physiological states [5]. Further research of psychophysiological responses in video-games has focused on studying various effects of game genre and specific situations on the subject's psychophysiological states. For instance, Ballard and Wiest studied in 1996 the effects of violent video-games on the hostility and cardiovascular responses of males [6].

One of our assumptions is that psychophysiological changes correlate with game events and the subject's emotions. Some work addressing this topic has already been published: Mandryk et al. used multiple psychophysiological measures in their experiments with players set in two different conditions, playing both against the computer and against another co-located player:

- EKG (Cardiovascular measures)

- HR (Heart Rate)
- GSR (Galvanic Skin Response)
- EMG (Electromyography)
- Resp Rate (Respiratory Rate)
- RespAmp (Respiratory Amplitude)

Afterwards, they were able to compile a list of possible correlations between particular psychophysiological measures and subjective responses [7]:

- **Fun** significantly correlated with **GSR**.
- **Boredom** correlated with **EMG**.
- **Challenge** correlated linearly with **RespAmp EMG**.
- **Ease** also correlated linearly with **RespAmp** and **EMG**.
- **Frustration** significantly correlated with **GSR** and **RespRate**.

Several other papers have addressed the correlation problem of psychophysiological measures with similar results. Drachen, Nacke, Yannanakis and Petersen have produced several papers in the field, indicating how various emotions affect the psychophysiological measures of subjects. In 2009, their results indicated that HR as a measure of arousal is a good correlator with self-report measures of player experience, both positive and negative [3], [8].

PROBLEM DOMAIN

User experience assessment in EE is currently done mostly by focus groups or by developer evaluation using subjective estimations. We feel that the development process might benefit largely from evaluating particular sections of the product (be it levels in video-games, or segments of movies) in a reliable way. As stated above, questionnaires in focus groups have several drawbacks regarding reliability, such as personal bias or untruthfulness of subjects.

We propose a solution taking advantage of the autonomous nervous system (ANS), which functions automatically.

PHYSIOLOGICAL MEASUREMENTS

Any research method in which the dependent variable is a physiological measure and the independent variable is behavioral or mental (such as memory) is a psychophysiological method.

Physiological measures take many forms and range from blood flow or neural activity in the brain to heart rate variability and eye movements. These measures can provide information about processes including emotion and cognition, as well as interactions between them.

Physiological measures thus offer a very flexible set of tools for researchers to answer questions about behavior, cognition, and health [9].

All the measurements used in our research are parts of the human nervous system, more specifically the autonomic nervous system (ANS), which controls involuntary physiological responses.

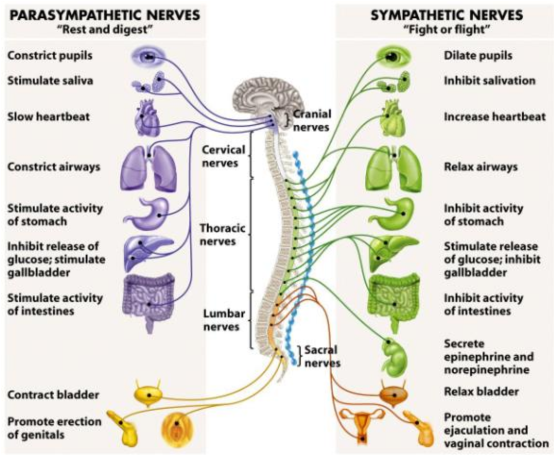

Figure 1. The human autonomic nervous system and its response stimulations [20].

Measures controlled by the ANS function automatically. Therefore, they can not be affected or controlled consciously, which reduces the risk of unsuccessful experiments or imprecise results. We believe that a combination of traditional techniques with psychophysiological measurements will provide a reliable insight into the player experience. Although there is no prior verification of reliability of such a combination, interviews and questionnaires are currently the only way to evaluate the player experience, and therefore the only way to create our model.

This section is a short introduction of how the ANS and psychophysiological measures used in our work function.

Heart rate + Heart rate variability:

Heart rate (HR) is a measure of cardiovascular activity which reflects the emotional state. It has been found to increase for a number of negative emotions (e.g. anger, anxiety, embarrassment, fear, sadness) as well as for some positive emotions (e.g. happiness, joy) and surprise [10].

The heart rate variability (HRV) is the time difference between each successive heartbeat, otherwise known as the R-R interval or the inter-beat interval. The time between each heartbeat is not fixed or consistent, but it varies with every beat - hence the term variability.

Historically, HRV has been measured using an electrocardiogram (ECG), but with the development of technology, it can now be reliability measured using smartphone applications combined with a heart rate strap [11] or a pulse sensor [12]. Whilst there are many other metrics used to measure HRV, the most common is known as the "root-mean square difference of successive normal R-R intervals (RMSSD)" [13].

Respiratory activity:
Respiration is measured as the rate of volume at which an individual exchanges air in their lungs. Previous research has found that respiration rate is increased by emotional arousal and decreased by rest and relaxation [14]. Overall, respiratory activity (RA) is rather easily measured during the experiments. There are several types of devices used for measuring respiratory activity, worn on the chest, torso, neck or even the wrist.

Electrodermal activity:
Electrodermal activity (EDA) measures the activity of the eccrine sweat glands and has been found to be a linear correlate to arousal [15]. Although room temperature, humidity, participants activities and the correct attachment of the electrodes have to be carefully considered, tonic EDA is a well researched and valid method to record arousal and has been used for measuring emotions for interaction with systems [16], [17].

EDA sensors are typically worn at the fingertips, but nowadays wrist-worn sensors and even ring sensors are becoming available.

EXPERIMENTS

This section describes our initial experiments along with apparatus description and some preliminary results.

Preliminary Experiments:
Before expanding the total number of measured physiological features to 10 (which we measure in the final experiments), we have conducted an initial round of experiments with one participants, measuring his heart rate using wrist-worn smartwatch sensor. The participant played ~20 hours of Dota 2 throughout several days. This experiment provided insight for future measurements described later in this section. In Figure 2., there is heart rate recorded from 4 games throughout 3 different days.

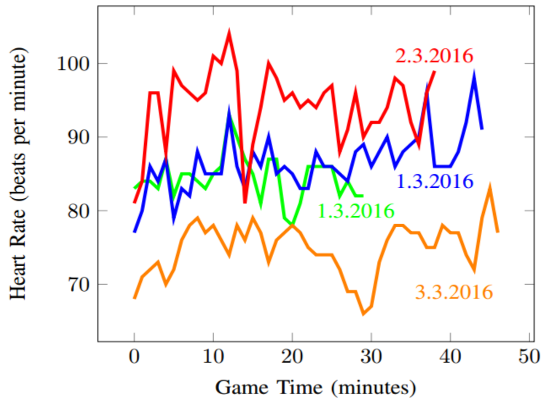

Figure 2. 4 games of Dota 2 played by the tested subject in 3 sessions.

Afterwards, we used different sensors and methodology in order to create more precise experiments for our research. Experiments using only 1 physiological measures proved unable to provide sufficient data to train any machine learning model.

APPARATUS

For data collection, we used apparatus consisting of three different sensors and a webcam:

- **Zephyr Bioharness 3**[1] for measuring heart rate, respiratory and physical activity.
- **Neurosky Mindwave Mobile**[2] for measuring EEG and blinking frequency.
- **Grove GSR**[3] Sensor in combination with **Arduino Uno** for measuring electrodermal skin activity.
- **Creative Senz3D**[4] for recording the experiment.

1 https://www.zephyranywhere.com/
2 http://neurosky.com/
3 http://wiki.seeedstudio.com/
4 https://us.creative.com/

Sensor	Measured Feature	Data Range	Unit
Zephyr Bioharness 3	Heart Rate	25 – 240	BPM (beats per minute)
Zephyr Bioharness 3	Heart Rate Amplitude	0.25 – 15	mV
Zephyr Bioharness 3	Breathing Rate	4 – 70	BMP (breaths per minute)
Zephyr Bioharness 3	Breathing Rate Amplitude	0 – 65534	16 bit unsigned integer
Zephyr Bioharness 3	Activity	+-16 (any axis)	VMU (vector magnitude unit)
Seeedstudio Grove GSR Skin Sensor Module	GSR Resistance	0 – 1023	Ohm
Seeedstudio Grove GSR Skin Sensor Module	GSR Conductance (1/Resistance)	0 – 1023	Siemens
Neurosky Mindwave Mobile	Attention	0 – 100	%
Neurosky Mindwave Mobile	Meditation	0 – 100	%
Neurosky Mindwave Mobile	Blinking Frequency	0 – 65534	16 bit unsigned integer

Table 1. List of measured physiological features and sensors used.

DATA COLLECTION

Using psychophysiology, it is important to consider carefully prior psychophysiological states of the subject.

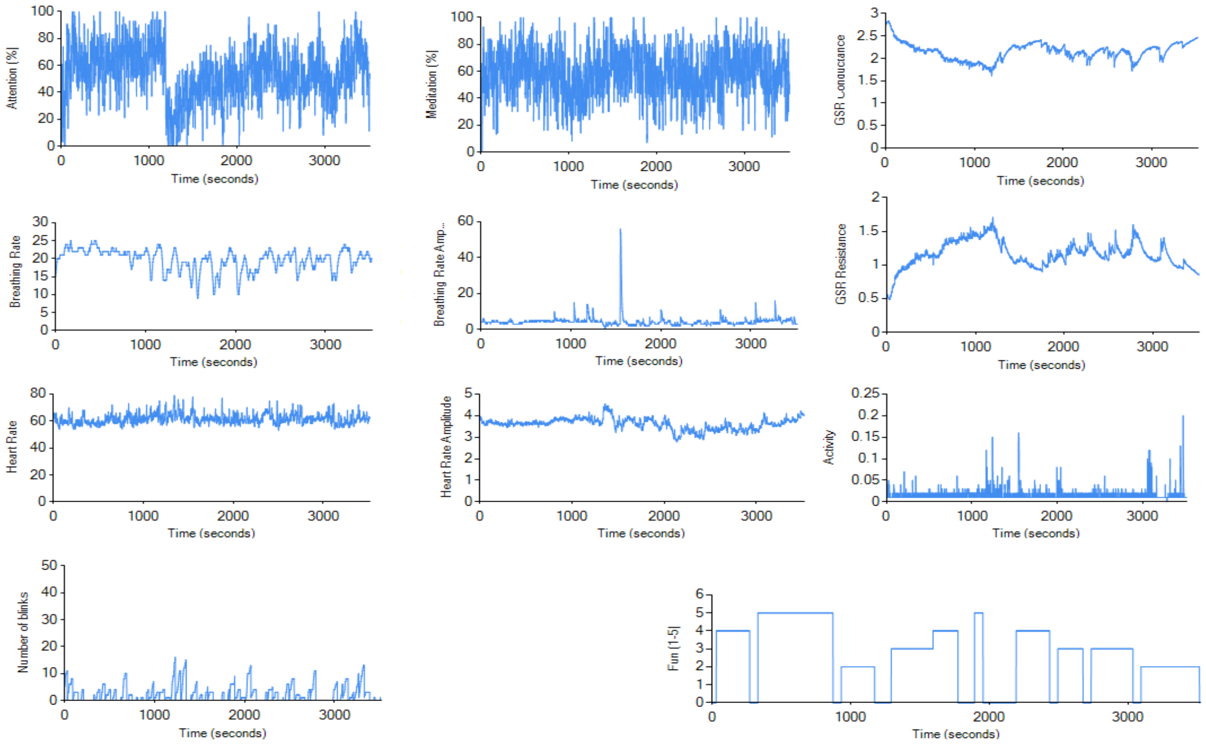

Figure 3. Sample data from data collection software

Chart of our experiments on Azure Machine Learning StudioAs described in [7] as well as found by our initial experiments, particular readings of subjects vary, depending on several factors. For example, in Figure 2. there are 4 games of the same subject player playing the same game in separate sessions during three days. This figure shows that average values of different measures vary significantly while taken in different days.

Implementing psychophysiological measurements in research has to be done with great care due to the many variables that can affect the result. Temperature, humidity, attachment of electrodes, individual differences, differences concerning gender, age, time of the day, the use of stimulants such as coffee or energy drinks, medicaments, etc. can cause different reactions in sensor readings [18], [19]. Therefore we decided to ignore absolute values of HR, RA, EEG and EDA measurements in the future and use only deviations from the average values.

Besides the psychophysiological state, the game genre and format also affect experimental readings considerably. While in Single-player games players are focused solely on game experience, in Multi-player format the social aspect comes into play. And while our initial findings show that HR is not affected by verbal communication with teammates and opponents, this does not have to apply to other measures that are used in experiments.

Data collection was done both real-time and post-experiment. With Bioharness sensor, we were not able to access the data real-time (no API provided), therefore the data was collected after experiments using simple csv log parser. Mindwave EEG sensor was able to provide data real-time, so our database was updated every second during the experiments: attention, meditation, blinking frequency were stored. EDA sensor we used was connected to Arduino and while able to send data real-time using a serial port, we have decided to store it into text files during the experiment. This was necessary due to data noise provided by used EDA sensor, which required some corrections to be done before updating the database. List of measured features with units is shown in Table 1.

As stated above, we have decided to use deviations from the average of all features instead of absolute values. We created new columns in our database for average values and replaced absolute values with the difference. This applied to all physiological measures except blinking frequency. Blinking frequency was saved at the end of every minute (number of blinks in last 60seconds). However, such data is not usable in modeling so we decided to calculate another features into our database: we use blink frequency (number of blinks) in last 10, 30 and 60 seconds. This means that every database entry (1 per second) has a value of how many times the subject blinked in said intervals.

The initial set of experiments included 15 participants with the average age of 30.7, 13 of them males. We store information about their age, gender, and frequency of play with three different levels (non-player, average player, frequent player). Every participant played approximately an hour of different games, mostly 2-3 games for ~20 minutes. After the gameplay, we conducted an interview in combination with screen capture video where participants labeled their enjoyment level in different parts of games. For example, with games with short levels, we let the participants label the level as one event (for example, NFS: Payback levels are 2-3 minutes long). If the levels were too long, we asked the subjects to determine which events had an effect on their enjoyment levels and labeled between those. Based on games and participant's ability to assess their enjoyment, the times between enjoyment alterations differed.

We realize that using subjective assessment in determining enjoyment in our experiments is susceptible to same questionnaire drawbacks that we plan to eliminate with our research. However, with sufficient amount of experiments and different participants, we believe we will be able to train accurate enjoyment model for later use.

Model Testing:

We used Microsoft Azure Machine Learning studio (AMLS) to model our initial data. In Figure 4. is a screenshot of the training process steps. We used the Decision Forest Regression block available on AMLS with following parameters:

- Resampling method: Bagging
- Trainer mode: Single Parameter
- Number of decision trees: 8
- Maximum depth of the trees: 32
- Number of random splits per node: 128
- Minimum number of samples per leaf node: 1

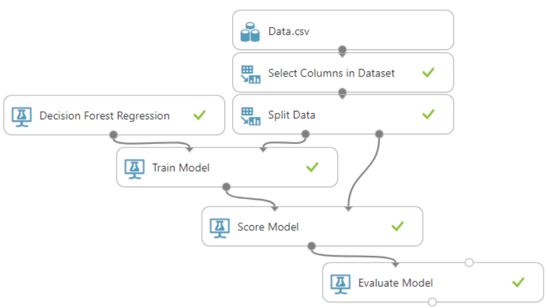

Figure 4. Chart of our experiments on Azure Machine Learning Studio

In our initial experiments, we used data samples from 15 subjects which in total provided ~35 000 examples. After training the model using data we have collected up until now, we obtained a model able to predict the enjoyment with a coefficient of determination (the proportion of the variance in the dependent variable that is predictable from the independent variable) of **0.88206**.

The root mean squared error of our model was **0.40712** with values of enjoyment scaling from 1 to 5. Therefore the initial model's prediction was roughly 0.40712 away from the actual value of enjoyment. We have also tested our data on different models available on AMLS, such as Neural Network Regression and Linear Regression, although with unsatisfying results.

CONCLUSION

In future research, we plan to increase the number of experiments substantially and use the findings for creating a model of user experience in correlation with the psychophysiological state of the EE consumer. However, as stated in Introduction section, simply measuring psychophysiological states of the subject is not sufficient for a successful training of user experience model. We plan to combine these measurements with interviews as well as video recordings of the subject's gameplay in order to determine the subject's enjoyment. As stated in Physiological Measurements section, to create the model we have to rely initially on the said questionnaires and interviews. With this information, experiments will pose as a training set for our model, with the goal to predict the experience of players in the future.

References

[1] D.E.S.A. Washington, E. Sales, demographics and usage data: Essential facts about the computer and video-game industry 2015.

[2] Kivikangas, J.M.; Chanel, G.; Cowley, B.; Ekman, I.; Salminen, M.; Järvelä, S.; Ravaja, N. A review of the use of psychophysiological methods in game research. Journal of Gaming & Virtual Worlds 2011 , 3, 181–199.

[3] Drachen, A.; Nacke, L.E.; Yannakakis, G.; Pedersen, A.L. Psychophysiological Correlations with Gameplay Experience Dimensions. arXiv preprint arXiv:1004.0243 2010.

[4] IJsselsteijn, W.; Van Den Hoogen, W.; Klimmt, C.; De Kort, Y.; Lindley, C.; Mathiak, K.; Poels, K.; Ravaja, N.; Turpeinen, M.; Vorderer, P. Measuring the experience of digital game enjoyment. Proceedings of Measuring Behavior. Noldus Information Tecnology Wageningen, Netherlands, 2008, pp. 88–89.

[5] Drachen, A.; Nacke, L.E.; Yannakakis, G.; Pedersen, A.L. Correlation between heart rate, electrodermal activity and player experience in first-person shooter games. Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games. ACM, 2010, pp. 49–54.
[6] Ballard, M.E.; Wiest, J.R. Mortal Kombat (tm): The Effects of Violent Videogame Play on Males' Hostility and Cardiovascular Responding1. Journal of Applied Social Psychology 1996, 26, 717–730.
[7] Mandryk, R.L.; Inkpen, K.M.; Calvert, T.W. Using psychophysiological techniques to measure user experience with entertainment technologies. Behaviour & information technology 2006, 25, 141–158.
[8] Yannakakis, G.N.; Hallam, J.; Lund, H.H. Entertainment capture through heart rate activity in physical interactive playgrounds. User Modeling and User-Adapted Interaction 2008, 18, 207–243.
[9] Infantolino, Z.; Miller, G.A. Psychophysiological Methods in Neuroscience. http://nobaproject.com/modules/psychophysiological-methods-in-neuroscience. Accessed: 2016-03-15.
[10] Kreibig, S.D. Autonomic nervous system activity in emotion: A review. Biological psychology 2010, 84, 394–421.
[11] Flatt, A.A.; Esco, M.R. Validity of the ithleteTM smart phone application for determining ultra-short-term heart rate variability. Journal of human kinetics 2013, 39, 85–92.
[12] Heathers, J.A. Smartphone-enabled pulse rate variability: an alternative methodology for the collection of heart rate variability in psychophysiological research. International Journal of Psychophysiology 2013, 89, 297–304.
[13] Aubert, A.E.; Seps, B.; Beckers, F. Heart rate variability in athletes. Sports medicine 2003, 33, 889–919.
[14] Stern, R.M.; Ray, W.J.; Quigley, K.S. Psychophysiological recording; Oxford University Press, USA, 2001.
[15] Schell, M.E.D.A.M.; Filion, D.L. The electrodermal system. Principles of Psychophysiology: Physical, social and inferential elements, pp. 295–324.
[16] McCarthy, J.; Wright, P. Technology as experience. interactions 2004, 11, 42–43.
[17] Ward, R.D.; Marsden, P.H. Physiological responses to different web page designs. International Journal of Human-Computer Studies 2003, 59, 199–212.
[18] Ganglbauer, E.; Schrammel, J.; Deutsch, S.; Tscheligi, M. Applying psychophysiological methods for measuring user experience: possibilities, challenges and feasibility. Workshop on user experience evaluation methods in product development. Citeseer, 2009.
[19] Certicky, M. Correlation between user experience in electronic entertainment and psychophysiological measurements. Proceedings of Scientific Conference of Young Researchers, 2016, pp. 80–83.
[20] Parasympathetic Nervous System Rest and Digest. http://ktmc.info/parasympathetic-nervous-system-rest-and-digest. Accessed: 2016-03-15.