Towards an Effective Arousal Detection System for Virtual Reality

Ifigeneia Mavridou
Centre of Digital Entertainment
Bournemouth University
Bournemouth, UK
imavridou@bournemouth.ac.uk

Ellen Seiss
Department of Psychology
Bournemouth University
Bournemouth, UK
eseiss@bournemouth.ac.uk

Theodoros Kostoulas
Dept. of Computing and Informatics
Bournemouth University
Bournemouth, UK
tkostoulas@bournemouth.ac.uk

Charles Nduka
Emteq Ltd.
Science Park Square,
Brighton, UK
charles@emteq.net

Emili Balaguer-Ballester
Dept. of Computing and Informatics
Bournemouth University
Bournemouth, UK
eb-ballester@bournemouth.ac.uk

ABSTRACT

Immersive technologies offer the potential to drive engagement and create exciting experiences. A better understanding of the emotional state of the user within immersive experiences can assist in healthcare interventions and the evaluation of entertainment technologies. This work describes a feasibility study to explore the effect of affective video content on heart-rate recordings for Virtual Reality applications. A low-cost reflected-mode photoplethysmographic sensor and an electrocardiographic chest-belt sensor were attached on a novel non-invasive wearable interface specially designed for this study. 11 participants responses were analysed, and heart-rate metrics were used for arousal classification. The reported results demonstrate that the fusion of physiological signals yields to significant performance improvement; and hence the feasibility of our new approach.

CCS CONCEPTS

•Human-centered computing→ Interaction paradigms: Virtual reality**; •Information systems**→ Sentiment analysis; • **Human-centered computing→** Interactive systems and tools

KEYWORDS

Virtual Reality; Arousal, Classification; PPG; ECG; C-SVM;

ACM Reference format:

Ifigeneia Mavridou, Ellen Seiss, Theodoros Kostoulas, Charles Nduka, Emili Balaguer-Ballester. 2018. Towards an Effective Arousal Detection System for Virtual Reality. In Proc. *of ACM Human-Habitat for Health (H3'18). ACM, Boulder, CO, USA, October 2018, 6 pages.*
DOI: 10.1145/3279963.3279969

Human-Habitat for Health (H3'18), October 16, 2018, Boulder, CO, USA

ACM ISBN 978-1-4503-6075-3/18/10⋯$15.00
https://doi.org/10.1145/3279963.3279969

1 INTRODUCTION

The increasingly evolving Virtual Reality (VR) technologies permit the adaptation of experimental protocols for their use with VR. Crucially, experiment design utilising VR can offer controlled laboratory conditions while granting a wealth of content resources and ecological validity [1]. User input and interface sensory modalities are currently integrated with VR, as they monitor the user's actions. These systems use various haptic and wearable user-interfaces to track head and body movements, eye gaze and speech patterns [2]. Such metrics can describe useful information related to the user's behaviour, preferences and actions within VR. As such, they can improve automatic emotion recognition, which is important to enhance VR user interactions. Previous research on affective computing offers a wealth of emotion detection solutions ranging from physiological and speech signals, to monitoring facial expressions, and movement analysis [3]. Understanding the user's emotions and behaviour within VR experiences could not only assist experience-designers to evaluate their content [4, 5] but also in healthcare interventions such as VR exposure therapy [6].

There are two basic challenges for emotion recognition in VR. Firstly, the Head Mounted Displays (HMDs) commonly used during VR experiences cover a significant part of the face which renders the detection of facial expressions difficult. Secondly, commercial immersive experiences require often intense head and limb movements, which could result in noise artefacts on potential wearable sensors. To overcome the first challenge, our team developed a novel prototype for facial expression recognition, Faceteq™ [7] with surface physiological sensors. This interface can be incorporated on a commercial HMD, acting as non-invasive, soft medium between the user's skin and the HMD.

In this work, we propose a system for the detection of high and low arousal in VR settings via capturing multimodal heart-rate responses (from low cost, custom-made photoplethysmographic (PPG) and electrocardiographic (ECG) sensors) and continuous self-ratings of HMD users.

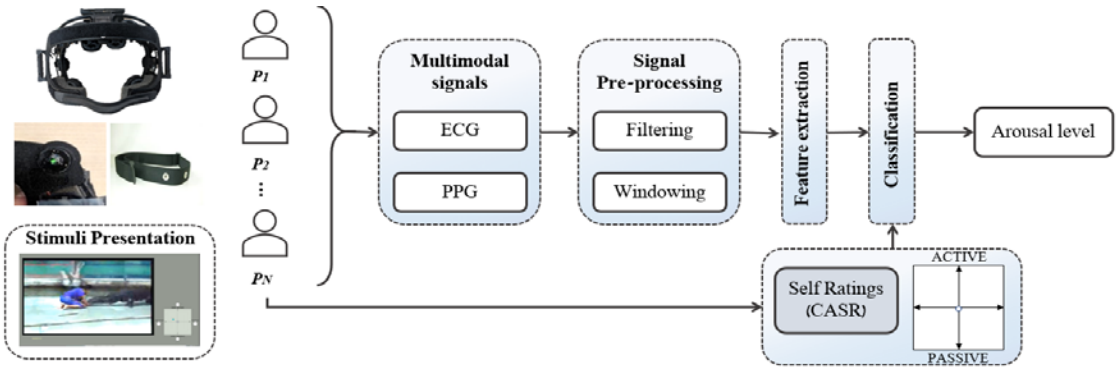


Figure 1: High-level overview of the proposed system. The Faceteq™ prototype is depicted on the upper left side.

2 RELATED WORK

Heart-rate (HR) metrics from ECG, plethysmographs and PPG have been previously used for stress and arousal detection [8]. These measures were correlated with measures of mental effort, as well as defensive (fear) and offensive (anger) emotions [9,10]. Commonly used HR measures include heart-rate variability (HRV), Inter-Beat-Interval (IBI), standard deviation (SD) of heart rate and blood pressure. Changes in the HRV have been associated with changes in attention and emotional states, e.g. during a stressful high HRV related to orienting responses and faster habituation, while low HRV with hypervigilance and defensive behaviour [11, 12].

Although ECG recordings have been primarily used for HRV monitoring due to its distinct profile of R peaks, there might be many advantages of measuring Pulse Rate Variability (PRV) from the PPG. The PPG is easy to use, non-invasive, cost-effective, it involves less sensors and it enables continuous, long-term recordings [13,14]. However, PPG signals are easily susceptible to movement artefacts and the detection of R-R intervals from arterial pulses from distant sources (e.g. fingertips or legs) could potentially be erroneous [15]. PPG sensors have been utilised in a large variety of experimental studies and on numerous body locations, including fingers, hands, forearms, earlobes, wrists, auditory canal, legs, buttocks, and the back [13]. Researchers have also recorded reflected PPG signals from the forehead [14,16]. PPG forehead placement showed advantages over other peripheral body location because it offered greater sensitivity to pulse changes during low blood flow [17], and because it was less susceptible to motion artefacts during certain body movements [18].

PPG sensors have also been utilised in VR research. Besides placing the sensors on common body-locations e.g. fingers [19], several attempts have been made towards facial placement and HMD incorporation. As such, [20] placed PPG sensors on the temple area (middle of the forehead) using a headband, and [21] placed the sensor directly on the face plate of the HMD; however, the performance of this approach was not described. As PPG measurements could be susceptible to changes in light perfusions and movement artefacts, we envisaged that by incorporating PPGs on an interface between the HMD and the user's skin, we could obtain a clear pulsative reading for reliable arousal detection in VR. The objective of this study was to test the feasibility of arousal detection via PPG on the superficial temporal vein; and to explore its performance efficiency when compared to an ECG (conventional method) and to the combination of both modalities.

3 SYSTEM DESCRIPTION

In 2016, we developed a novel interface prototype '*Faceteq™*' [7]. The interface was designed to work as an intermediate layer between the HMD and the face of the wearer, consisting of eight electromyography (EMG) sensors, two PPG sensors and one inertial measurement unit (IMU) including gyroscope and accelerometer. We hypothesised that by placing biometric sensors on the facial areas where the HMD covers already, could prove to be an easy-to-use, unobtrusive or user motion non-constraining solution for affect monitoring in VR. For the purposes of this paper, we focus exclusively on arousal detection via ECG and PPG sensors.

We designed a system where heart-rate responses were recorded using Faceteq™, while participants were watching videos with affective content and self-rating their level of arousal. As shown in Figure 1, recordings were sequentially becoming subject to signal processing where they were denoised and divided into segments, and feature extraction. The features, together with the participants' self-ratings, were then fed to a classifier in order to estimate the arousal levels from the participants' heart-rate responses to the stimuli.

4 EXPERIMENTAL SET UP

In this section, we describe the components of the proposed system as seen in Figure 1; the audio-visual stimuli, the participants recruited, the hardware and software utilised, the experiment procedure in which the participants' physiological signals were collected, the signal processing steps performed, and the classification experiments conducted.

4.1 Audio-visual Stimuli

A selection of the short videos from the affective film library

[22] was used for this study following the protocol introduced in [7]. The selected videos were intended to provoke one of two levels of valence and one of two levels of arousal, corresponding to each of the four quadrants of the dimensional model [23]. 5 videos were selected per each one of the four categories (High-Arousing Positive, High-Arousing Negative, Low-Arousing Positive and Low-Arousing Negative), as well as 20 neutral videos. The order of the videos was counterbalanced across participants and presented as follows: five videos from one category, followed by five neutral videos, then followed by five videos from another category and so on, until all videos have been played. Each video had a fixed duration of 25 seconds. Grey images lasting 8 seconds were added as 'breaks' after every video. The videos sequence within a category was randomised for each participant.

4.2 Participants

For this study, *n*=11 participants were recruited (P1-P11 in this paper; 5 female and 6 male) with a mean age of 21.5 years (SD: 2.6; range: 18-35 years). Prior to taking part in the study, we asked participants to avoid caffeinated drinks on the study day. After signing informed consent, basic demographic information was obtained from the participants and questionnaires verified that they were not suffering from anxiety, depression or any disorders of cardiovascular nature which could affect their heart-rate metrics at the time of the study. The study was reviewed and approved by the Science, Technology & Health Research Ethics Panel, Anonymous University (Ref. 13994).

4.3 Apparatus

The software solutions and hardware devices developed for this study are described below.

4.*3.1 The stimuli presentation*. For the study, an application with three environments was developed: (1) a self-rating training environment where participants were introduced to the terms of arousal and valence. During this training period (duration: 15-30 min) we asked the participants to get acquainted with rating these two dimensions using their mouse's pointer on our Continuous Affect Self-rating (CASR) interface (based on the 'FeelTrace' tool [24]), (2) a grey scene where the participants were asked to relax while neutral baseline data were recorded, and (3) a semi-dark cinema environment where the videos were presented next to the CASR interface, as suggested by [25, 26]. All participants were asked to perform minimal head movements during the recording.

4.3.2 Monitoring equipment and sensors. The interface prototype was equipped with a custom-made PPG sensor (reflection mode) on the upper left side of the mask (see Figure 1), corresponding to the area over the superficial temporal vein and artery. Additionally, a custom-made ECG chest-belt was developed comprising two ECG sensors, which were connected to the Faceteq™ interface. Both data streams were recorded simultaneously.

4.4 Experiment procedure

The study took approx. 50-60 minutes. During the study, every participant was wearing the Faceteq™ device and the ECG belt. Video capture of the participant's face and all physiological sensor data streams were recorded and synchronised with the video presentation via the Faceteq API. After setting up the equipment, each participant watched a sequence of 40 movie clips for 22 minutes. While watching the video, each participant was asked to rate their felt emotions in terms of arousal and valence using the *CASR* interface. They were advised to start rating as soon as a video commenced. After each video and during the grey images or 'breaks', participants were asked to return their rating pointer to the centre of the CASR interface (the neutral area).

4.5 Signal Preprocessing & Feature Extraction

PPG and ECG recordings from all participants were recorded using Faceteq API (sampling rate: 1000Hz). The analysis steps were as follows: First, the recorded raw data were filtered (Notch filter: 50Hz; band-pass Butterworth filter: 0.5Hz and 6 Hz for the PPG, 5Hz and 25Hz for the ECG; order: 2). Subsequently, the filtered recordings were divided in 25 seconds long time-window epochs corresponding to each video stimulus. Each epoch was further subdivided into 4.5 seconds overlap 5 second windows. Next, a peak detection method was applied on the PPG and ECG epochs.

The mean peak distance (IBImean) and the Root-Mean Square (RMS) of successive R-R interval distance (RMSIBI) per epoch were calculated. The whole feature vector was transformed based on the Minimum-Maximum normalisation [27]. The total number of processed samples per video-length was 48 per metric, resulting for 20 videos to a total number of 960 samples per participant per metric.

4.6 Classification experiments

As the automatic state recognition can be constrained by individual user differences, two scenarios were explored; a user-dependent and a user-independent approach. The following experiments were performed using: (1) the PPG derived metrics, (2) the ECG derived metrics and (3) the combination of both PPG and ECG derived metrics.

The two outputs (IBImean and RMSIBI) per modality were used as input to train a C-Support Vector Machine (SVM) using a gaussian kernel. The open-source libSVM framework [28] was adopted to train the binary C-SVM. In the *user-dependent* classification scenario, a 10-fold cross-validation was applied for each participant separately. In the *user-independent* scenario, we applied leave-one-participant-out cross validation by pooling all 11 datasets and predicting the rating of each participant in turn based on the remaining 10. The two free parameters of the method (the regularization penalty C and the standard deviation of the kernel function) were optimized exclusively on the training data for both dependent and user-independent scenarios.

The corresponding arousal CASR ratings per participant were used as the ground truth definition. Self-reported scores across users in terms of arousal for each video typically showed low dispersion (coefficient of variation, CV, Table 1), indicating high rating agreement per video across participants. Videos which presented rating disagreement were part of the positive emotion-inducing categories (videos 1-10), indicating higher rating variability during positive content.

Table 1: Agreement scores across users per video (mean value, standard deviation and Coefficient of Variation)

Video	M (Std), CV	Video	M (Std), CV	Video	M (Std), CV
1	0.7 (.73), 0 .9	8	1.0 (.49), 0.5	15	1.1 (.88), 0.8
2	0.4 (1.00), **2.5**	9	0.6 (.74), **1.2**	16	1.4 (.45), 0.3
3	1.2 (1.24), **1.0**	10	0.5 (.94), **1.8**	17	1.5 (.85), 0.5
4	0.2 (.65), **1.0**	11	1.6 (.81), 0.5	18	1.3 (.92), 0.7
5	0.7 (.86), **1.2**	12	2.0 (.90), 0.4	19	1.7 (.86), 0.5
6	0.8 (.62), 0.7	13	1.2 (.73), 0.6	20	1.4 (.84), 0.6
7	0.7 (1.0), **1.4**	14	0.8 (.66), 0.8		

The mean CASR value per participant for the user-dependent, and the mean value across all participants' ratings for the user-independent scenario were utilised as the division point for the of high and low arousal classes. The total number of samples per participant was 960 (total: 10560s). The mean number of samples for high arousal levels for the user-dependent scenario is 532±20 (mean ± SEM; ranging from 368 to 602 samples); and 5852 for all participants (P1-P11). Thus, classes are largely balanced.

5 EXPERIMENTAL RESULTS

We tested the feasibility of arousal detection via PPG sensor from the superficial temporal vein in VR. The C-SVM enabled us to map the level of arousal with the metrics calculated from the PPG and ECG recordings (IBImean and RMSIBI) during the presentation of four audio-visual stimuli categories.

User-dependent scenario – In Figure 2 we illustrate the receiver operating characteristic (ROC) curves per experiment performed. Each line represents a participant. The areas under curve (AUC) are included on the bottom right corner of each plot. Despite the variations in performance between participants, the system's capability for detecting changes in arousal is higher for the fusion versus unimodal approaches for most of the participants (Figure 2c)

To evaluate the performance of each classification experiment, we compared the AUC per metric in recording modality pairs (e.g. PPG against ECG) using the Bradley's test approach [29, 30]. The results from this test are reported on Table 2. We denote in bold when their AUC means are significantly different (see details in [29]). Detection performances between PPG and ECG modalities are generally significantly different except for participants 1, 3, 5, 8 and 9 (p<.05). Moreover, the combined PPG-EEG metric (termed here the fusion approach), outperforms the PPG and ECG modalities individually for 9 out of 11 subjects (Table 2).

User-independent scenario –The system shows a similar ability to identify high and low arousal levels from ECG, PPG and from the fusion of the two modalities (Figure 3). The AUC values for each experiment are included on the bottom right corner of the plot. The performance of the system when using only the PPG metrics is significantly lower than ECG and fusion. Finally, the best overall performance is achieved via the fusion approach (significance level at 5%, [29,30]), Table 3.

Table 2: Bradley scores between experiments per subject

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11
PPG-**E**CG	-1.5	**-1.8**	0.3	**2.5**	-0.7	**2.1**	**7.2**	1.3	0.2	**-3.3**	**4.6**
PPG-**F**usion	**-5.6**	**-7.0**	**-3.9**	**-5.7**	0.0	**-2.4**	**7.2**	-1.1	**-4.3**	**-5.7**	**-4**
ECG-**F**usion	**-4.0**	**-5.3**	**-4.2**	**-8.3**	0.7	**-4.6**	0.0	**-2.4**	**-4.5**	**-2.3**	**-8.8**

Significance level at 5% | *significant values are indicated in bold

Table 3: Bradley scores between experiments performed

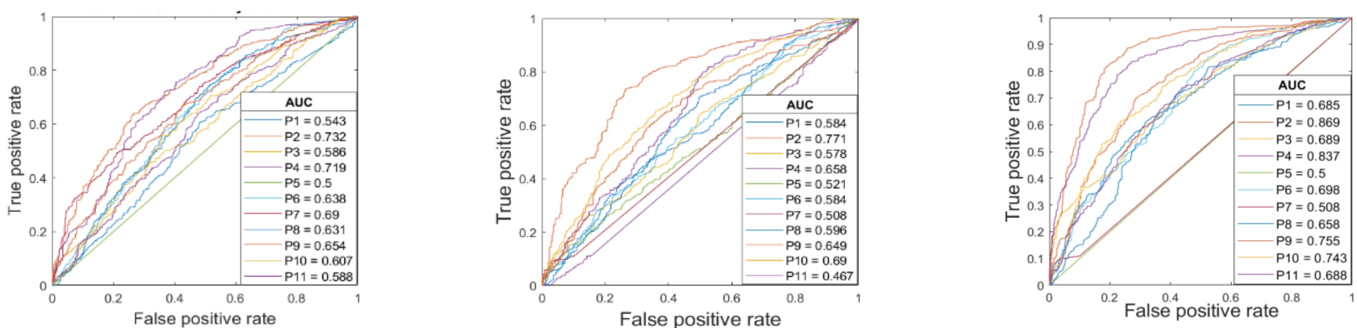
PPG-**E**CG	**-4.3**	**P**PG-**F**usion	**-9.1**	**E**CG-**F**usion	**-4.7**
Significance at 5% |		*significant values in bold			

6 DISCUSSION AND CONCLUDING REMARKS

We proposed a system for arousal detection in VR settings, by designing a novel interface which incorporates PPG and ECG sensors. During the study, participants facially expressed their emotions in response to video stimuli that resulted into limited head, and therefore sensor movements. Nonetheless, using the PPG metrics, our system yielded a similar detection performance to the ECG one for 5 out of 11 participants in the user-dependent scenario. This result supports our assumption regarding the system's capability to detect arousal via PPG recordings from the superficial temporal vessels, subject to individuals' variability. Moreover, the fusion of both methods provides an enhanced performance overall.

The arousal detection issues that occurred in all three experiments for participant 5 suggest that changes in heart-rate during audio-visual stimulation are elicited in different intensities among individuals. Thus, the detection capacity of the system was less reliable. Additionally, detection issues could have resulted from wrong sensor placement, sensor's quality, intense movements (e.g. laughter) which could also reposition the sensors, or to skin sweatiness. In this feasibility study the annotation interface was designed to minimize the cognitive effort. The use of alternative interfaces (e.g. [26]) will be explored in our future work.

The system's detection using PPG in the user-independent scenario performed slightly worse than using ECG (Figure 3). However, PPG sensors are affordable and easy to use, making them strong candidates for wearable integration in practice. Likewise, although ECG sensors are difficult to integrate at the moment, we envisage that improved ECG sensors will be readily available for integration with wearable devices and clothes in future. Thus, given the enhanced performance for the fusion set-up demonstrated in this feasibility study, the combination of both sensors for arousal detection seems a robust approach for multiple applications incorporating immersive technologies. Further research will focus on (1) recruiting a larger sample size, (2) exploring additional modalities for interface integration such as electrodermal activity sensors, (3) investigating supplementary physiological data metrics, and (4) applying regression approaches to detect finer-grained levels of arousal and valence in VR.



(a) PPG (b) ECG (c) Fusion (PPG & ECG)

Figure 2: ROC curves per participant per experiment performed (user-dependent).

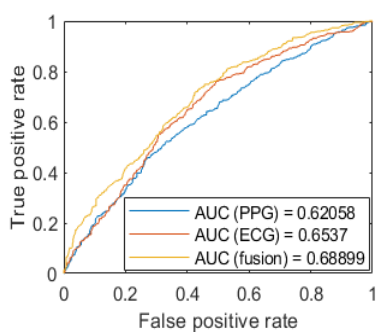


Figure 3: ROC curves per experiment (user-independent)

ACKNOWLEDGMENTS

This work is supported by the UK EPSRC and Emteq Ltd., via the Centre of Digital Entertainment (Grant No. EP/L016540/1).

REFERENCES

[1] Thomas D. Parsons. 2015. Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, 9, (Dec. 2015) 660. DOI: https://doi.org/ 10.3389/fnhum.2015.00660

[2] Jason Jerald. 2015. The VR Book: Human-Centered Design for Virtual Reality. Association for Computing Machinery and Morgan & Claypool, New York, NY, USA.

[3] Rafael A. Calvo and Sidney D'Mello. 2010. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Trans. Affect. Comput.* 1, 1 (Jan. 2010), 18-37. DOI: https://doi.org/10.1109/T-AFFC.2010.1

[4] Ricardo Baeza-Yates and Mounia Lalmas. 2012. User engagement: the network effect matters. *In 21st ACM international conference on Information and knowledge management* (CIKM '12). ACM, New York, NY, USA, 1-2. DOI= https://doi.org/10.1145/2396761.2396763

[5] Regan L. Mandryk, Kori M. Inkpen, and Thomas W. Calvert. 2006. Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & information technology* 25, 2, (Mar. 2016) 141-158. DOI: https://doi.org/10.1080/01449290500331156

[6] Botella, Cristina, Javier Fernández-Álvarez, Verónica Guillén, Azucena García-Palacios, and Rosa Baños. 2017. Recent progress in virtual reality exposure therapy for phobias: a systematic review. *Current psychiatry reports* 19, 7 (Jul 2017), 42.DOI: https://doi.org/10.1007/s11920-017-0788-4

[7] Ifigeneia Mavridou, James T. McGhee, Mahyar Hamedi, Mohsen Fatoorechi, Andrew Cleal, Emili Balaguer-Ballester, Ellen Seiss, Graeme Cox and Charles Nduka. 2017. FACETEQ interface demo for emotion expression in VR. *In Proc. of IEEE Virtual Reality (VR)*, IEEE , Los Angeles, CA, USA, 441-442. .DOI: https://doi.org/10.1109/VR.2017.7892369

[8] Foteini Agrafioti, Dimitris Hatzinakos, and Adam K. Anderson. 2012. ECG Pattern Analysis for Emotion Detection. IEEE Trans. Affect. Comput. 3, 1 (January 2012), 102-115. DOI: https://doi.org/10.1109/T-AFFC.2011.28.

[9] Daniel Kahneman. 1973. Attention and effort. New Jersey: Prentice-Hall.

[10] Paul Ekman, Robert W. Levenson, and Wallace V. Friesen.1983. Autonomic nervous system activity distinguishes among emotions. *Science* 221, 4616, 1208-1210. DOI: https://doi.org/10.1126/science.6612338

[11] Julian F.Thayer, Bruce H. Friedman, Thomas D. Borkovec, Bjorn H. Johnsen, and Silvia Molina. 2000. Phasic heart period reactions to cued threat and nonthreat stimuli in generalized anxiety disorder. *Psychophysiology*, 37, 3, (May 2000), 361-368. DOI: https://doi.org/10.1111/1469-8986.3730361

[12] Jennifer A. Healey. 2000. Wearable and automotive systems for affect recognition from physiology Ph.D. dissertation. Massachusetts Institute of Technology, Cambridge, MA.

[13] Fred Shaffer, Rollin McCraty, and Christopher L. Zerr. 2014. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. *Frontiers in Psychology,* 5, (Sep. 2014) 1040. DOI: https://doi.org/10.3389/fpsyg.2014.01040..

[14] Lena T. Nilsson, Goscinski, S. Kalman, L-G. Lindberg, Anders Johansson. 2007. Combined photoplethysmographic monitoring of respiration rate and pulse: a comparison between different measurement sites in spontaneously breathing subjects. *Acta Anaesthesiologica Scandinavica*, 51, 9 (Aug. 2007) 1250-1257. DOI: https://doi.org/10.1111/j.1399-6576.2007.01375.x

[15] Nicholas D. Giardino, Paul M. Lehrer, and Robert Edelberg. 2002. Comparison of finger plethysmograph to ECG in the measurement of heart rate variability. *Psychophysiology*, 39, 2, (Mar. 2003) 246-253. DOI: https://doi.org/10.1111/1469-8986.3920246.

[16] Nandakumar Selvaraj, Ashok Jaryal, Jayashree Santhosh, Kishore K. Deepak, and Sneh Anand. 2008. Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. *J. of medical engineering & technology*, 32, 6 (Nov. 2008) 479-84. DOI: https://doi.org/10.1080/03091900701781317

[17] Mendelson Y. and Pujary C., 2003. Measurement site and photodetector size considerations in optimizing power consumption of a wearable reflectance pulse oximeter. In *25th Annual International Conference on Engineering in Medicine and Biology Society,* 4, Mexico, IEEE, 3016-3019. DOI: https://doi.org/10.1109/IEMBS.2003.1280775

[18] Toshiyo Tamura, Tuka Maeda, Masaki Sekine, Masaki Yoshida. 2014. Wearable photoplethysmographic sensors—past and present. *Electronics,* 3, 2, (April 2014) 282-302. DOI: https://doi.org/10.3390/electronics3020282.

[19] Young Youn Kim, Eun Nam Kim, Min Jae Park, Kwang Suk Park, Hee Dong Ko, and Hyun Taek Kim. 2008. The application of biosignal feedback for reducing cybersickness from exposure to a virtual environment. *Presence: Teleoper. Virtual Environ.* 17, 1 (Feb. 2008), 1-16. DOI: https://doi.org/10.1162/pres.17.1.1.

[20] Seiya Otsuka, Kanami Kurosaki, Mitsuhiro Ogawa. 2017. Physiological measurements on a gaming virtual reality headset using photoplethysmography: A preliminary attempt at incorporating physiological measurement with gaming. In *Region 10 Conference, TENCON* 2017. IEEE, Penang, 1251-1256. DOI: https://doi.org/10.1109/TENCON.2017.8228049

[21] Guillermo Bernal and Pattie Maes. 2017. Emotional Beasts: Visually Expressing Emotions through Avatars in VR. *In 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, NY, USA, 2395-2402. DOI: https://doi.org/10.1145/3027063.3053207

[22] Andrea C. Samson, Sylvia D. Kreibig, Blake Soderstrom, A. Ayanna Wade, and James J. Gross. 2016. Eliciting positive, negative and mixed emotional states: A film library for affective scientists. *Cognition and Emotion,* 30, 5, (Jul. 2016) 827-856. DOI: https://doi.org/10.1080/02699931.2015.1031089.

[23] James A. Russell. 1980. A circumplex model of affect. *J. of personality and social psychology*, 39, 6, (Dec. 1980)1161-1178. DOI: https://doi.org/10.1037/h0077714

[24] Roddy Cowie, Ellen Douglas-Cowie, Susie Savvidou*, Edelle McMahon, Martin Sawey, and Marc Schröder. 2000. 'FEELTRACE': An instrument for recording perceived emotion in real time. In *ISCA Speech and Emotion 2, ISCA, Newcastle, UK.* DOI: https://doi.org/10.1037/a0024109.

[25] Angeliki Metallinou and Shrikanth Narayanan. 2013. Annotation and processing of continuous emotional attributes: Challenges and opportunities. *In Automatic* Face *and Gesture Recognition* (FG), 2013 10th IEEE International Conference and Workshops on. IEEE, Shanghai, China, 8 pages. DOI: https://doi.org/10.1109/FG.2013.6553804

[26] Georgios N. Yannakakis & Hector P. Martinez. 2015. Grounding truth via ordinal annotation. *In International Conference Affective Computing and Intelligent Interaction (ACII),* IEEE, Xi'an, China, 574-580. DOI: https://doi.org/10.1109/ACII.2015.7344627

[27] Slobodan Ribaric and Ivan Fratric. 2016. Experimental Evaluation of Matching-Score Normalization Techniques on Different Multimodal Biometric Systems. In *Proc. of Mediterranean Electrotechnical Conference MELECON 2006*. IEEE, Malaga, Spain, 498-501. Doi: https://doi.org/10.1109/MELCON.2006.1653147

[28] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 3, Article 27 (May 2011), 27 pages. DOI: https://doi.org/10.1145/1961189.1961199

[29] Andrew P. Bradley. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition,* 30, 7 (Jul 1997), 1145-1159. DOI: https://doi.org/10.1016/S0031-3203(96)00142-2

[30] Theodoros Kostoulas, Guillaume Chanel, Michal Muszynski, Patrizia Lombardo, and Thierry Pun. 2015. Dynamic Time Warping of Multimodal Signals for Detecting Highlights in Movies. In *Proc. of the 1st Workshop on Modeling INTERPERsonal SynchrONy And infLuence (INTERPERSONAL '15).* ACM, New York, NY, USA, 35-40. DOI: https://doi.org/10.1145/2823513.2823515