Towards valence detection from EMG for Virtual Reality applications

I Mavridou[1], E Seiss[2], M Hamedi[4], E Balaguer-Ballester[3], C Nduka[4]

[1] Centre of Digital Entertainment, Bournemouth University, Fern Barrow, Poole, UK

[2] SC2AN Research Group, Department of Psychology, Bournemouth University, Fern Barrow, Poole, UK

[3] Department of Computing and Informatics, Bournemouth University, Fern Barrow, Poole, UK

[4] Emteq Ltd., Science Park Square, Brighton, UK

[1]imavridou@bournemouth.ac.uk, [2]charles@emteq.net,

ABSTRACT

The current practical restraints for facial expression recognition in Virtual Reality (VR) led to the development of a novel wearable interface called *Faceteq*. Our team designed a pilot feasibility study to explore the effect of spontaneous facial expressions on eight EMG sensors, incorporated on the Faceteq interface. Thirty-four participants took part in the study where they watched a sequence of video stimuli while self-rating their emotional state. After a specifically designed signal pre-processing, we aimed to classify the responses into three classes (negative, neutral, positive). A C-SVM classifier was cross-validated for each participant, reaching an out-of-sample average accuracy of 82.5%. These preliminary results have encouraged us to enlarge our dataset and incorporate data from different physiological signals to achieve automatic detection of combined arousal and valence states for VR applications.

1. INTRODUCTION

Multiple computational systems and interfaces are available to enhance Human Computer Interaction (HCI). For emotion recognition, various methods are currently under investigation to assess human affective states and expressions (see review by Calvo & D'Mello, 2010). Typically, emotions are measured in a two-dimensional space consisting of their arousal (activation or excitement levels) and valence (positive/ negative polarity levels) which are considered as the basic quantitative dimensions of emotions (dimensional model by Russel, 1980). Multiple studies have explored the link between self-rated levels of arousal and valence, and physiological and/or behavioural changes, and their classification for HCI applications (e.g. Healey & Picard, 2005). Although, researchers in interactive technologies tend to focus on the identification of arousal through physiological measures (e.g. Wu et al, 2010; Yannakakis & Paiva, 2014), we believe that the coalescence of both dimensions is required to evaluate the nature of the emotion elicited during an immersive experience. The majority of research in valence detection emphasises the interpretation and analysis of facial expressions; either via computer vision (see Zeng et al. 2009 review) or facial EMG signals (e.g. Fridlund & Cacioppo, 1986; Cheng & Liu, 2008). The majority of these measures are today often combined for multimodal affective and physiological computing.

Commercial low-cost first and second generation Virtual Reality (VR)technologies became largely available after 2014 (with the introduction of Oculus Rift DK2), enabling a larger number of researchers to investigate user-experience effects in VR. VR provides a platform for the design of controlled experimental conditions while granting ecological validity and an abundance of content resources. Research from diverse disciplines gradually adapt VR for their experimental designs and executions (e.g. Bekele et al, 2007; Yang et al,2017; Burke, 2018). As the number of VR-related projects is increasing, emotion detection in VR is anticipated to become a vital piece for future research. The study of emotion elicitation and detection in VR however is still in its infancy (Diemer et al, 2015). Research in VR often combines different measures such as behavioural observations and physiological measures as e.g. heart-rate and galvanic skin response (Slater et al, 2006; Giakoumis et al, 2009) to report arousal levels; although emotion-related studies in VR such as Riva et al (2007) often utilise questionnaires to identify psychological states and valence levels. The area of the face which is rich with valence information is usually left unexplored since the Head-Mounted Display (HMD; commonly used for immersive VR) is covering almost 2/3 of the face, including the most informative facial muscles. As VR technologies are improving and their uses are increasing, it is fundamental to propose alternative, novel approaches to measure emotions. This will enable a better understanding of the emotional states induced in VR and the latter's effect on human feelings and elicited responses.

In 2016, we proposed a novel hardware solution, 'Faceteq', for facial muscle activation monitoring (Mavridou et al, 2016) in VR. Faceteq was designed to work as an extra intermediate layer between the HMD and the face of the wearer, consisting of eight electromyography (EMG) sensors, two Photoplethysmogram sensors (PPG) and one inertial measurement unit (IMU) including gyroscope and accelerometer. We hypothesise that such an interface can track the valence information needed for continuous emotion assessment in VR. To our current best knowledge, this is the first study where integrated surface EMG sensors have been used for valence detection in VR. Therefore, we investigated the feasibility of this approach in controlled conditions, using audio-visual stimuli on a monitor. This feasibility study has been specifically designed to induce three valence levels to test the sensitivity of the metrics derived from the wearable sensors (EMG, PPG, IMU) employed on the Faceteq interface. For the purpose of this paper, we included only the surface EMG data and valence ratings in our analysis.

2. EXPERIMENT SET-UP

2.1 Stimuli

A selection of the videos with affective content from the affective film library by Samson et al. (2016) was used for this study. The selection of videos was intended to elicit two levels of valence and two levels of arousal, in order to obtain physiological responses for each of the 4 quadrants of the dimensional model. For this reason, 5 videos per category (four categories) with the highest and lowest rating of valence and arousal were selected, as well as 20 neutral videos. The order of videos categories was counterbalanced across participants and presented as follows: five affective videos from one category, followed by five neutral videos, followed by five affective videos from another category etc. Each video had a fixed duration of 25 seconds. Grey images were added after every video for 8 seconds each. The video order within a category was newly randomised for each participant.

2.2 Participants

For this preliminary study, 35 participants were recruited (20 female), from 18 to 40 years old (M: 22.8, SD: 5.2). Participants were told that we were monitoring the electrical conductivity of their skin and were instructed how to rate their felt emotions in terms of arousal and valence prior the video presentation. Questionnaires verified that the participants were not suffering from anxiety, depression or any disorder that can affect their facial movements at the time that the study took place.

Figure 1. *Video presentation environment*

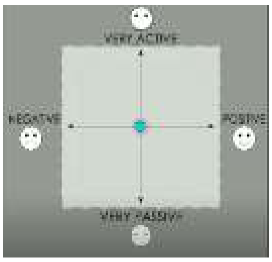

Figure 2. *CASR interface*

Figure 3. *Faceteq interface*

2.3 Apparatus

The stimuli presentation. We designed and developed an application with three environments/scenes: (1) a training environment where we asked participants to get familiar with the Continuous Affect Self-rating (CASR) interface (based on the *'FeelTrace'* tool for affect rating; Cowie et al, 2000) (Figure 1 & 2), (2) a grey scene where the participant were asked to relax while we recorded neutral (baseline) data, and (3) a semi-dark cinema environment where the videos were presented next to the CASR interface which was placed on the bottom right corner of the screen (Figure 1). The participants were requested to face the monitor and perform minimum head movements during the recording.

Monitoring Equipment and sensors. The interface prototype was equipped with eight surface dry electromyography (f-EMG) sensors on the right and the left side of the face (Channels; 1 & 2 on Zygomaticus major, 3 & 4 on Frontalis, 5 & 6 on Orbitocularis oculi and 7 & 8 Corrugator muscles) using an adapted protocol described by van Boxtel (2010).

2.4 Experiment procedure

The study took approximately 50-60 minutes and was conducted at our laboratory at Innovation Centre at Sussex University. During the study, each participant watched a randomised sequence of movie clips for approx. 22 mins. During this time, they were asked to rate their felt emotions using the CASR interface. During the video presentation, video capture of the participant's face and physiological responses were recorded. All sensor data streams were synchronised with the video presentation via the Faceteq API to ensure ease in analysis and efficient event detection. Each participant was compensated by a £5 voucher for their time. The study was reviewed and approved by the Science, Technology & Health Research Ethics Panel, Bournemouth University (Ref. 13994 on 26/01/2017).

3. DATA ANALYSIS

EMG recordings from 8 channels were recorded using Faceteq API (sampling rate: 1000Hz) and afterwards analysed offline in Matlab. Firstly, a baseline correction was applied, by subtracting the mean EMG values. We then removed 50Hz and their harmonics up to 350Hz using Notch filters. The signals were band-pass filtered from 30 to 450 Hz. Extreme outliers caused by motion artefacts were removed using a Hampel filter. Next, the EMG recordings from the 8 channels were divided in epochs of 22 seconds which is corresponding to the video duration minus the first 3 seconds of each video. Next, the Root-Mean Square value per 512 samples window was calculated. As EMG are highly variable between wearers, and since we are interested in detecting valence states (negative, neutral, positive) we applied a Maximum-Minimum normalization function. The data were used as input to train a C-Support Vector Machine (SVM), using the libSVM (Chang & Lin, 2011). For each video and for each participant from the data set, the ground truth was defined by the corresponding participant's CASR valence-only scores. The data and labels were sent into an SVM (RBF kernel) for classification, using 10-fold cross validation for each participant separately. The low computational cost of the implementation enabled the approach to provide a cross-validated readout in less than 0.5 seconds per participant.

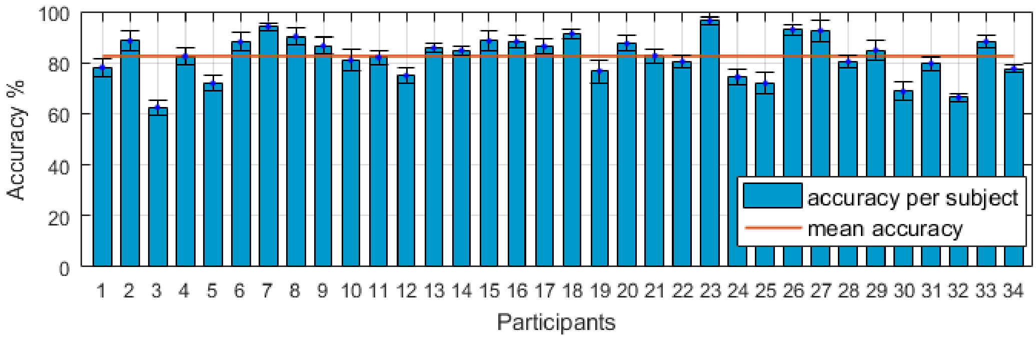

Figure 4. *Classification accuracies for all 34 participants*

4. RESULTS AND DISCUSSION

Overall, we tested the feasibility of our prototype for valence detection in VR. The C-SVM enables us to map the levels of activation of EMG channels with the spontaneous expressions during the stimuli categories. Each model, trained with data from each of 34 female and male participants achieved an accuracy ranging from 62.4% to 96.4%, with an average accuracy across the group of 82.5% (Std: 8.2) (Figure 4).The results of this initial study confirmed the feasibility of our approach. We are currently analysing the remaining recorded physiological responses related to heart rate, head movement and electrical conductivity of the skin, which are expected to provide further insights, as well as assist on the refinement of our valence and arousal detection for VR applications. We envisage that a larger sample size will enable the system to achieve higher accuracy on valence and arousal detection and a more robust model. This model may be able to discover common facial patterns in clusters of participants, and more fine-grained levels of valence and arousal i.e. additional levels of classes.

Acknowledgements: We would like to thank Dr. Theodoros Kostoulas for the scientific guidance. This material is based upon work supported by the UK EPSRC and Emteq Ltd., via the Centre of Digital Entertainment under Grant No. EP/L016540/1).

5. REFERENCES

Bekele, E, Bian, D, Peterman, J, Park, S, & Sarkar, N, (2017), Design of a Virtual Reality System for Affect Analysis in Facial Expressions (VR-SAAFE); Application to Schizophrenia. *IEEE Trans. on Neural Systems and Rehabilitation Engineering,* 25(6), pp. 739-749.

Bradley, MM, & Lang, PJ, (1994), Measuring emotion: the self-assessment manikin and the semantic differential. *J. of behavior therapy and experimental psychiatry*, 25(1), pp. 49-59.

Burke, RR, (2018), Virtual Reality for Marketing Research. In *Innovative Research Methodologies in Management* (pp. 63-82). Palgrave Macmillan, Cham.

Calvo, RA, & D'Mello, S, (2010), Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Trans. on affective computing*, 1(1), pp. 18-37.

Chang, CC, & Lin, CJ, (2011), LIBSVM: a library for support vector machines. *ACM Trans. on intelligent systems and technology (TIST)*, 2(3), pp. 27.

Cheng, B, & Liu, G, (2008, May), Emotion recognition from surface EMG signal using wavelet transform and neural network. In *Proc. of the 2nd international conference on bioinformatics and biomedical engineering (ICBBE)* (pp. 1363-1366).

Cowie, R, Douglas-Cowie, E, Savvidou, S, McMahon, E, Sawey, M, & Schröder, M, (2000), 'FEELTRACE': An instrument for recording perceived emotion in real time. In *ISCA tutorial and research workshop (ITRW) on speech and emotion.*

Fridlund, AJ, & Cacioppo, JT, (1986), Guidelines for human electromyographic research. *Psychophysiology*, 23(5), pp. 567-589.

Healey, JA, & Picard, RW, (2005), Detecting stress during real-world driving tasks using physiological sensors. *IEEE Trans. on intelligent transportation systems*, 6(2), pp. 156-166.

Mavridou, I, McGhee, JT, Hamedi, M, Fatoorechi, M, Cleal, A, Ballaguer-Balester, E, Seiss, E, Cox, G, & Nduka, C, (2017, March), FACETEQ interface demo for emotion expression in VR. In *Virtual Reality (VR), 2017 IEEE* (pp. 441-442). IEEE.

Riva, G, Mantovani, F, Capideville, CS, Preziosa, A, Morganti, F, Villani, D, Gaggioli, A, Botella, C, & Alcañiz, M, (2007), Affective interactions using virtual reality: the link between presence and emotions. *CyberPsychology & Behavior*, 10(1), pp. 45-56.

Russell, JA, (1980), A circumplex model of affect. *J. of personality and social psychology*, 39(6), pp. 1161.

Samson, AC, Kreibig, SD, Soderstrom, B, Wade, AA, & Gross, JJ, (2016), Eliciting positive, negative and mixed emotional states: A film library for affective scientists. *Cognition and Emotion*, 30(5), pp. 827-856.

Slater, M, Guger, C, Edlinger, G, Leeb, R, Pfurtscheller, G, Antley, A, Garau, M, Brogni, A, & Friedman, D, (2006), Analysis of physiological responses to a social situation in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, 15(5), pp. 553-569.

Van Boxtel, A, (2010, August), Facial EMG as a tool for inferring affective states. In *Proc. of measuring behavior* (pp. 104-108). Wageningen: Noldus Information Technology.

Wu, D, Courtney, CG, Lance, BJ, Narayanan, SS, Dawson, ME, Oie, KS, & Parsons, TD, (2010), Optimal arousal identification and classification for affective computing using physiological signals: Virtual reality stroop task. *IEEE Trans. on Affective Computing*, 1(2), pp. 109-118.

Yannakakis, GN, & Paiva, A, (2014), Emotion in games. *Handbook on affective computing*, pp. 459-471.

Zeng, Z, Pantic, M, Roisman, GI, & Huang, TS, (2009), A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Trans. on pattern analysis and machine intelligence*, 31(1), pp. 39-58.