Cognitive Computing and Internet of Things, Vol. 43, 2022, 93–101
https://doi.org/10.54941/ahfe1001843

Estimation of Change in Affective State Using Eye Tracking Features from Virtual Reality Technologies

Martin Pszeida[1], Amir Dini[1], Michael Schneeberger[1], Melanie Lenger[1], Lucas Paletta[1], Silvia Russegger[1], Sybille Reidl[1], Sarah Beranek[1], Sandra Schüssler[2], Alfred Häussl[2], Robert Hartmann[3], Martin Sighart[3], Sebastian Mayer[3], Patricia Papic[3], Beatrix Koch[4], and Hermine Fürli[4]

[1]Joanneum Research Forschungsgesellschaft mbH, Graz, Austria
[2]Medical University of Graz, Graz, Austria
[3]Netural GmbH, Linz, Austria
[4]Perisutti Pflegezentrum, Eibiswald, Austria

ABSTRACT

Affective states play a prominent role in the context of human activation and motivation. Immersive VR-based presence provides opportunities to activate elderly people in the context of preferred leisure activities (Häussl et al., 2021) or to apply mindfulness interventions for their cognitive reserve (Paletta et al., 2021). The appropriate design of positively activating content is pivotal for appropriate changes in users' affective states. The presented study provided insight into the potential of non-invasive VR-based eye tracking for automated estimation of affective state induced by video content, in an explorative pilot study with seven elderly persons living in a nursing home. The results indicate the feasibility of estimating mood change from characteristic eye movement features, such as, fixation duration and pupil diameter, as a promising future research topic.

Keywords: Virtual reality, Eye tracking, Affective state change

INTRODUCTION

Virtual Reality (VR) technologies offer great potential to provide immersive presence in artificial environments, in particular, to activate elderly people, such as, in the context of dementia (Häussl et al., 2021; Paletta et al., 2021). With the help of VR glasses that are adapted to the needs of the elderly, people can visit places that would be very difficult or even impossible to access in reality. Being able to experience something new within these virtual excursions can help to raise their emotional and motivational state, relieve stress and strengthen resilience.

Mental activation and emotion play a prominent role in the context of process- and outcome-oriented activation and motivation (Toure-Tillery & Fishbach, 2014). From this we deduce that affective states are capable to cause either a reinforcing an inhibiting effect on human behavior, especially

Figure 1: Resident of nursing home using the VR-based technology for activation in the explorative research study.

in mental processes associated with cognitive control and attentional processes. Cognitive reserve determines the capacity for emotional regulation, which is an important function for mental processes in the context of executive functions (Diamond, 2014), and regulatory effort largely determines the impact of negative emotions on mental health and physical well-being (Ochsner & Gross, 2005).

A key aspect in the appropriate design of positively activating and motivating VR technology is to understand the impact of presenting a specific audio-visual content on the affective state of the user. There exist several publications that describe how to recognise the affective user state when applying VR headsets, under application of EEG- (Liu et al., 2011), fNIRS- (Hu et al., 2019), or classical psychophysiological biosensor-based technologies providing EDA-, HR- and respiration rate-related data (Petrescu et al., 2020) for further interpretation. However, these technologies are either too expensive or too invasive to be frequently applied at home or within the nursing home and therefore lack practical relevance.

The aim of the presented explorative pilot study was therefore to get more insight into the potential of applying non-invasive VR-based eye tracking to provide evidence that changes in the affective state could be applied in an automated manner in a typical healthcare environment, such as, at home or in a nursing home.

VIRTUAL REALITY-BASED ACTIVATION

Related Work

The reduction of one's autonomy can have a greater impact on the mental health of residents than their physical limitations (Boyle, 2014). Autonomy is the strongest predictor of the absence of depression (Tester et al., 2004) and the perception of personal control correlates with long-term physical and emotional well-being. Virtual reality (VR) represents a great potential

Figure 2: Typical samples of video content-based semantic categories, superimposed with the gaze (red) of the observer: (a) 'global' and (b) 'personalized' category, (c) 'regional' category illustrated by a 360° video frame of "Styrian wine road", a well-known destination in the vicinity of the study participants' nursing home.

for creating realistic, interactive worlds of experience in which multisensory stimuli can be exploited for mental activation, for experiencing self-efficacy and for triggering positive emotions. VR supports the impression of physical presence (immersion), in a three-dimensional, potentially interactive environment (Benoit et al., 2015).

Semantic Video Content Categories

Following the user requirement analysis (Häussl et al., 2021), we selected those scenarios to represent the video content that were most interesting for male and female participants. The principal idea of clustering video content into semantic categories was motivated by the fact that the activation as well as the valence might be different for the users with respect to each particular semantic reference. In order to test the affective impact of presenting different semantic categories of video content, we clustered the videos into three basic categories, i.e., of 'regional', 'personalized' and 'global' content (see Figure 2). We evaluated then the usability as well as the affective impact of the persons living at the nursing home within specifically targeted studies.

Videos With Regional Content

Inhabitants of nursing homes sadly miss the typical environment of their daily activities, they are interested in watching their rural or urban surroundings, their home, their working place, or typical environments of leisure activities that they are not anymore capable to perform.

Videos With Personalized Content

In cooperation with the nursing home of Perisutti care centre, personalized videos were created for the participants. The videos reflect the interests of the participants and were created independently by the staff with a GoPro camera. For the video material, it was important to examine in advance which places and everyday situations are interesting for the respective persons. The videos should not have any negative psychological effects on the participants. The finished videos were then edited by the company Netural. The length of the videos was shortened to a maximum length of two minutes. The corresponding video length was defined by the sub-studies in the project as the optimal length for target persons. In addition, the brightness and sharpness of the video material were adapted to the VR glasses. Since the videos were recorded with the GoPro camera, it was unavoidable that wind noise was very dominant in the audio. These disturbing noises were adjusted so that they could be perceived as pleasant for the participants.

Videos With Global Content

Videos with content that does not have any obvious reference to regional or personalized aspects is classified as 'global' content category.

MOOD ANALYTICS

Measuring Mood

Insight into mood is mandatory to understand how design can influence user behavior. People's preferences are impacted by their momentary mood states (Maier et al., 2012). Well-being requires a positive mood balance (Morris, 1999), while a lasting disturbance of this balance is one of the main reasons for human ill-being (e.g., Peeters et al., 2006). The terms mood and emotion are often used interchangeably, however, they represent different phenomena. Emotions are acute, and exist only for a relatively short period of time, usually seconds, minutes or hours at most. In Contrast, moods are always present, they tend to have a relatively long-term nature, they can last for several hours or even for several days. This implies that our mood is the affective background state to what we do, whereas emotions are momentary 'perturbations' that are superimposed on this affective background. Furthermore, another substantial difference is that emotions are targeted while moods are global. Several mood questionnaires are available that obtain a reliable understanding of a respondent's mood state. Questionnaires are easy to administer and analyse, and can be used to measure subtle and nuanced mood patterns. Visual scales are promising because they are quick, easy, and (when properly validated) reliable. Several visual scales are available that measure basic dimensions of affect, the most famous of which is SAM (Lang, 1980), however, SAM is limited in that it requires considerable explanation before

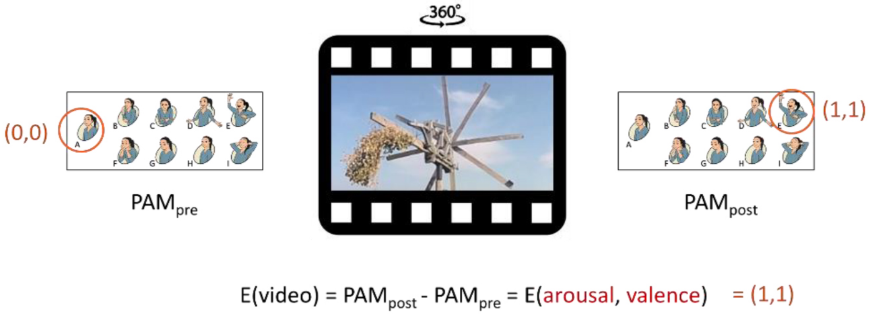

Figure 3: The mood state of a person was queried before and after each video presentation by means of Pick-A-Mood (PAM; Desmet et al., 2016; Desmet et al., 2019).

respondents can effectively report their feelings for each factor separately (Broekens and Brinkman, 2013).

Pick-A-Mood

(PAM; Desmet et al., 2016) is a validated cartoon-based pictorial instrument for reporting and expressing moods. It measures eight distinct mood states in a quick and intuitive way and can be used both for qualitative and quantitative research. Pick-A-Mood consists in principle of a female cartoon character (Figure 3). This character includes nine expressions that represent eight distinct mood states (and one neutral character). The sets are interchangeable; choice of character can be based on respondent demographics or on other pragmatic considerations. The PAM was applied in a specific research study in order to investigate the particular affective impact of the presentation of each video, and, in particular, in the context of the different semantic categories of the video content. For this purpose, PAM queries were applied before and after each video presentation and the effect of the video E(·) was computed based on the differences in both arousal and valence (see Figure 3).

Eye Tracking Analytics

Eye Tracking measurements were recorded at 90Hz via a HTC Vive Pro Eye headset. Each participant performed and successfully completed a 5-point calibration at the beginning of a session. Measurement data were segmented into the relevant sequences through the events triggered by the Unity application, for example, to represent observation of a video, PAM interaction, etc. Individual gaze points were analyzed and assigned to fixations by employing the dispersion-based fixation algorithm by Salvucci & Goldberg (Salvucci et al., 2000), with a dispersion threshold of 3° and a duration threshold of 100 ms. Furthermore, we applied blink detection by detecting events via the eye openness values provided by the Vive Pro Eye, triggering an event whenever eyes were closed about more than 50% and applying a duration constraint between 80 ms to 500 ms. We also calculated the mean pupil diameter of both eyes from the raw values provided. With the help of these measurements and calculated values, we acquired the median fixation duration, median saccade duration, the fixation rate (number of fixations

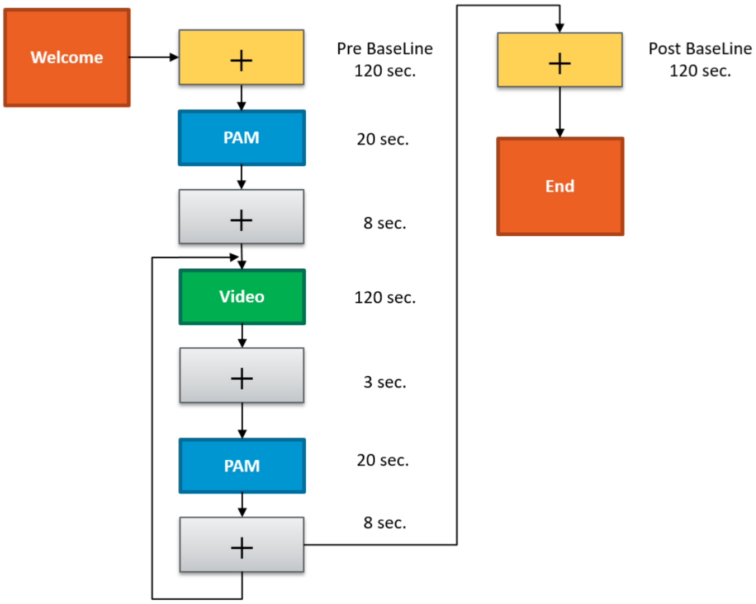

Figure 4: Study design: The participants were involved in a VR intervention of about 20 minutes in total, presenting six panoramic videos (each with a duration of 2 minutes) for exploration, two of them with personalized, two with regional and two with global content.

per second), as well as the mean pupil diameter and the number of blinks per minute. Having done this for every sequence of a session, we computed the relative increase & decrease of eye tracking features for the presentation of activation caused by a video.

EXPERIMENTAL RESULTS

Study Design and Descriptive Statistics

The study was implemented with N = 7 elderly persons (57.1% female) with M = 83.0 (SD 5.3) years of age, with MoCA cognitive assessment (Nasreddine et al., 1996) M = 19.0 (SD 7.3) and Geriatric Depression Scale (GDS) score (Yesavage & Brink, 1983) M = 2.0 (SD 1.9). The participants were involved in a VR intervention of about 20 minutes in total, presenting six panoramic videos (each with a duration of 2 minutes) for exploration, two of them with personalized, two with regional and two with global content, respectively (see *Figure 4*). A Pick-a-Mood (PAM) scale (Desmet et al., 2016) was used to query about the affective state of the user, eye movements were logged throughout the session. We investigated whether the difference between measurements of pre- and post-video PAM-based valence and arousal would correlate with descriptive statistics of eye movement features measured before and after the video exploration.

Inferential Statistics

To measure the activation of a participant, we compared the mean/median values of a an eye tracking feature in terms of a digital biomarker during

Table 1. Correlations between mood scores and eye tracking features for all videos (N = 42).

Mood dimension	Eye movement Feature	Type	Rho	p-val
valence	median fixation duration	Pearson	−0.3054	0.0492*
valence	median fixation duration	Spearman	−0.3497	0.0232*
valence	mean pupil diameter	Pearson	−0.3179	0.0402*
valence	mean saccade rate	Spearman	0.3774	0.0137*

Table 2. Correlation between valence and mean pupil diameter for videos with regional and global significance.

Mood dimension	Eye movement Feature	Type	Rho	p-val
Valence	mean pupil diameter	Pearson	−0.4810	0.009**

(a)

(b)

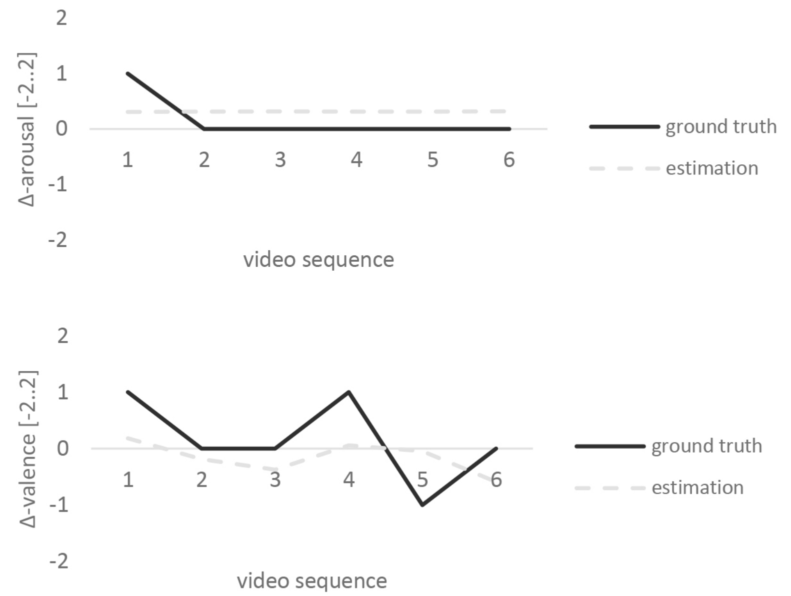

Figure 5: Prediction of mood based on linear regression model for the example of a selected study participant. (a) Prediction of change in arousal based on change of mean pupil diameter. (b) Prediction of change in valence based on change in median fixation duration.

the Pick-A-Mood (PAM) phase preceding a video and the mean/median values during a video. By comparing these activations with the subjective perceived mood provided before and after each video sequence, we can identify the following correlations. Based on the activation of all video classes (global, regional, personalized) of all participants, a statistically significant correlation was found between the change in median fixation duration and valence of the provided PAM responses. In particular, we found that the distribution of increments in mood correlate in valence with the median of fixation duration ($p = .023^*$, $rho = -0.350$; Spearman) and mean pupil diameter ($p = 0.040^*$, $rho = -0.318$; Pearson). Furthermore, specific video content (regional and global content) correlates with statistical significance (mean pupil diameter: $p = .009^{**}$, $rho = -0.481$; Pearson) with changes in valence.

To test how well a set of data points fits a normal distribution, the Anderson-Darling test (Anderson, 1952) was performed on all sets. The Anderson-Darling test is a statistical test, which is used to determine whether a sample of a population fits a specific distribution. Only the fixation duration features passed the null hypothesis at a significance level of 0.05.

The graphs below illustrate the results of a prediction of valence and arousal based on a linear regression model. While the discrete values do not align, a trend, in which increase & decrease align with the change in valence and arousal respectively. The mean deviation between ground truth and prediction of the discrete case below for arousal is 0.38 and for valence a mean deviation of 0.64 was observed.

CONCLUSION

These results emphasise the potential of measuring appropriate eye movement features to represent variation in affective state in real-time, i.e., with respect to short-time changes in visual content. The results of the presented exploratory study can help to understand how adaptive content could be implemented to serve the activation of elderly people, in particular, persons with high risk for dementia or even already diagnosed persons with dementia.

ACKNOWLEDGMENT

This work was partly funded from project VR4Care under grant No 873757 of the Austrian FFG femtech programme co-financed by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK) and from project MARA under grant No 115629 of programme Eurostars-2 CoD 15 (Eurostars is a joint programme between EUREKA and the European Commission).

REFERENCES

Anderson, T.W., & Darling, D.A., (1954). "A Test of Goodness of Fit", Journal of the American Statistical Association, 49:268, 765–769, DOI: 10.1080/01621459.1954.10501232

Benoit, M., Guerchouche, R., Petit, P.-D., Chapoulie, E., Manera, V., Chaurasia, G., et al. (2015). "Is it possible to use highly realistic virtual reality in the elderly? A feasibility study with image-based rendering", Neuropsychiatric Disease and Treatment, 11, 557–563.

Boyle, G.J., Helmes, E., Matthews, G., and Carroll, E.I. (2014). "Measures of Affect Dimensions", in Boyle, G.J., Saklofske, D., Matthews, G., eds., Measures of Personality and Social Psychological Constructs, Academic Press.

Diamond, A. (2013). "Executive functions", Annual Review of Psychology, 64, 135–168.

Häussl, A., Schuessler, S., Paletta, L., Fuerli, H., Koch, B., Binder, T., Schneeberger, M., Andreu, J.-P., Reidl, S., Beranek, S., Hartmann, R., Sighart, M. (2021). "Requirements Analysis on Emotional Preferences for Leisure Activities in Virtual Reality for Female Nursing Home Residents – A Mixed Method Approach". In: Ayaz H., Asgher, U., Paletta, L., eds., Advances in Neuroergonomics and Cognitive Engineering (AHFE 2021). Springer Nature, ISBN 978-3-030-80285-1

Hu X, Zhuang C, Wang F, Liu Y-J, Im C-H and Zhang D (2019). "fNIRS Evidence for Recognizably Different Positive Emotions", Front. Hum. Neurosci. 13:120. doi: 10.3389/fnhum.2019.00120

Liu Y., Sourina O., Nguyen M.K. (2011). "Real-Time EEG-Based Emotion Recognition and Its Applications", in: Gavrilova, M.L., Tan, C.J.K., Sourin, A., Sourina, O., eds., Transactions on Computational Science XII, Lecture Notes in Computer Science, vol 6670, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-22336-5_13.

Ochsner K N, & Gross J J. (2005). "The cognitive control of emotion", Trends Cogn Sci, 2005, 9: 242–249.

Paletta, L., Pszeida, M., Schüssler, S., Andreu, J.-P., Dini, A., Zweytik, E., Steiner, J., Grabher, A., Lodron, J. (2021). "Virtual Reality-based Sensory Triggers and Gaze-based Estimation for Mental Health Care", in: Ayaz H., Asgher, U., Paletta, L., eds., Advances in Neuroergonomics and Cognitive Engineering (AHFE 2021). Springer Nature, ISBN 978-3-030-80285-1

Petrescu L, Petrescu C, Mitruţ O, Moise G, Moldoveanu A, Moldoveanu F, Leordeanu M. (2020). "Integrating Biosignals Measurement in Virtual Reality Environments for Anxiety Detection", Sensors (Basel), Dec 10;20(24):7088. doi: 10.3390/s20247088.

Salvucci, Dario D., & Goldberg, Joseph H. (2000). "Identifying fixations and saccades in eye-tracking protocols", in Proceedings of the 2000 Symposium on Eye tracking Research & Applications (ETRA '00). Association for Computing Machinery, New York, NY, USA, 71–78. https://doi.org/10.1145/355017.355028.

Tester, S., Hubbard, G., Downs, M., MacDonald, C., Murphy, J. (2004). "Frailty and institutional life", in: Walker, A., Hagan Hennessy, C. eds., Growing older: quality of life in old age. Open University Press, Maidenhead, 209–224.

Touré-Tillery, M., and Fishbach, A. (2014). "How to Measure Motivation: A Guide for the Experimental Social Psychologist", Social and Personality Psychology Compass, 8(7).