Investigating the Process of Emotion Recognition in Immersive and Non-Immersive Virtual Technological Setups

Abstract

This paper investigates the use of Immersive Virtual Environment (IVE) to evaluate the process of emotion recognition from faces (ERF). ERF has been mostly probed by using still photographs resembling universal expressions. However, this approach does not reflect the vividness of faces. Virtual Reality (VR) makes use of animated agents, trying to overcome this issue by reproducing the inherent dynamic of facial expressions, but outside a natural environment. We suggest that a setup using IVE technology simulating a real scene in combination with virtual agents (VAs) displaying dynamic facial expressions should improve the study of ERF. To support our claim we carried out an experiment in which two groups of subjects had to recognize VAs facial expression of universal and basic emotions in IVE and No-IVE condition. The goal was to evaluate the impact of the immersion in VE for ERF investigation. Results showed that the level of immersion in IVE does not interfere with the recognition task and a high level of accuracy in facial recognition suggests that IVE can be used to investigate the process of ERF.

Keywords: emotion recognition, facial expression, virtual environments, virtual reality, immersive virtual environment, emotional virtual agents, Ekman basic emotion

Concepts: •**Human-centered computing** → **Virtual reality;** *Empirical studies in HCI; Empirical studies in HCI;* **Virtual reality;** *Empirical studies in HCI;*

1 Introduction

Face is considered a primary system in social interaction and it plays a relevant role in the emotional communication. The main approach to study the process of emotion recognition in faces (ERF) came from the idea that there is a universal expression of emotion in both humans and animals [Darwin 1955]. This theory was investigated by the psychologist Paul Ekman who discovered six facial *basic* emotions, that are universally recognized: Anger, Disgust, Fear, Happiness, Sadness and Surprise [Ekman 1993]. On this approach, many experiments are carried out to investigate the underpinning mechanism of ERF by using both still photographs and dynamic 3D representations of virtual faces [Adolphs 2002; Spencer-Smith et al. 2001]. In fact, an Immersive Virtual Environment (IVE) combined with a VA provide a strong sense of presence[Slater and Wilbur 1997] . The applications that make use of VAs in IVE had a big impact in the field of psychotherapy, especially for studying the processes of social perception [Gaggioli et al. 2003] and for devising new treatments for emotional diseases [Vanni et al. 2013]. In this paper an exploratory experiment investigates if the *basic* emotions are well recognized in IVE to evaluate the effect of IVE on the study of ERF. For this purpose, we compared in a facial emotion recognition task a non-immersive virtual environment (N-IVE), displaying a static scene, and an Immersive Virtual Environment (IVE), displaying a 360° interactive VE, both populated with an animated VA. In the following section a review of previous related works is presented, then the Method used to conduct our experiment is illustrated. Finally, the salient results are discussed.

SIGGRAPH 2016 Posters, July 24-28, 2016, Anaheim, CA
ISBN: 978-1-4503-ABCD-E/16/07
DOI: http://doi.acm.org/10.1145/9999997.9999999

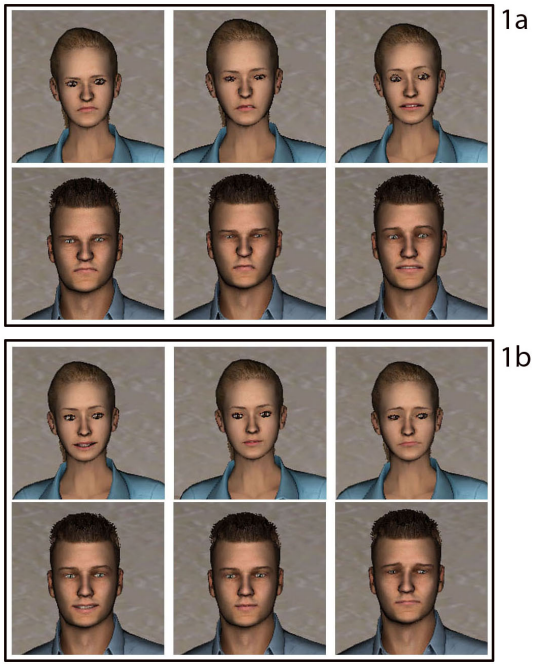

Figure 1: *Facial expressions of basic emotions in the male and female VAs used in the study.* ***Panel 1a****, from lef to right: Anger, Disgust, Fear.* ***Panel 1b****, from left to right: Happiness, Neutral and Sadness.*

2 Related Work

Most studies investigated ERF process using static pictures of Ekman *basic* emotions [Ekman 1993; Adolphs 2002]. These studies were limited by still faces not affording the liveliness of real-face emotional expressions. Subsequent research showed that faces in movement are better recognized than fixed faces, since the dynamism represents the three-dimensional structure of the face [Knight and Johnston 1997]. [Sato et al. 2004] showed that the perception of face in movement enhanced the neural activity in humans, opposite to static pictures. For this reason, the inherent dynamism of natural facial expressions has been created by animating virtual characters in a very realistic way and it has been evaluated in a lot of experiments[Spencer-Smith et al. 2001; Dyck et al. 2008; Yang and Bhanu 2011]. In particular, [Dyck et al. 2008] highlighted a significant correlation between the recognition of emotional expressions on virtual and natural faces. [Moser et al. 2007] showed that there is an activation increase in response to natural faces compared to virtual faces. In a recent experiment we have evaluated

the combination of dynamism and realism of virtual faces in reproducing facial *basic* emotions [Faita et al. 2015]. In parallel, other research showed that realistic 3D characters may be a useful tool to improve interpersonal communication in people suffering from emotional disorder[Gutiérrez-Maldonado et al. 2014]. However, none of these studies focused on the fact that the presentation of facial stimuli in isolation could be a limitation because the real facial encounter takes place in context. Recent works with photographs remarked that the environment largely influences the perception of facial emotions, hence stressing the need for a new paradigm that includes an environment to study ERF [Barrett et al. 2011]. In order to overcome this limitation, an IVE can be used to simulate a realistic scene in which a VA is animated with facial emotions. However, the correlation between emotional expressions in virtual faces and the virtual environment in which such faces came across has not yet been studied. In this research we have studied the ability of 24 healthy subjects in the emotion recognition task of two VAs by comparing Immersive and Non-Immersive technological setups. The novelty introduced by our research is the combination of the strong sense of immersion in a VE with the dynamism of VA facial emotions. We suggest that IVE can lead to improve the study of ERF, especially for clinical diseases.

3 Method

3.1 Participants

Twenty-four healthy subjects (6 females) were recruited for the experiment. They were aged between 23 and 36 (28.38 ± 4.20 y.o.), with normal or corrected-to-normal vision. The subjects presenting history of depression or anxiety disorders were excluded. All participants provided their informed consent to undertake the experiment by reading and signing a form explaining the background, objectives and procedures of the study, and the confidential handling of all the collected data.

3.2 Stimuli and Apparatus

The stimuli used in our study were a VE and two VAs to be used in an emotion recognition task. The VE presented a room with sofas, a television set and common furniture and it was accompanied with an environmental sound (Figure 2). The sounds were stored in separate *.MP3* files and were delivered through stereophonic wireless headphones. A male and a female human-like VAs were used for displaying the basic facial expressions: Anger, Disgust, Fear, Happiness, and Sadness (Figure1), plus Neutral according to the description of Ekman in the handbook of FACS [Ekman and Friesen 1978]. The faces were animated using *Faceshift 1.2* (Faceshift AG, Zurich, Switzerland), a motion-capture software able to perform marker-less facial tracking using Microsoft's Kinect cameras [Bouaziz et al. 2013]. The cleaned captured data were applied to the characters model and managed through the HALCA library [Gillies and Spanlang 2010] on which *XVR* (Extreme Virtual Reality, VR-Media S.r.l., Pisa, Italy) [Carrozzino et al. 2005] relies. We used *XVR* for the real-time rendering of the scene and to manage participant interaction. Halca allows to have different characters into which load the animations during the real-time execution of the experiment (Figure1). Stimuli for the emotion recognition task were arranged in two blocks of trials, each representing a single VA displaying one of the facial expressions in the VE. The dynamism of each trial is accomplished by combining 0.1sec. (6 frames) of Neutral expression and 0.9 sec. (54 frames) of transition from the Neutral to the emotional face to be recognized. Before disappearing, the face to be recognized remains for 4 sec. (240 frames), followed by the screen showing the choice options for the recognition task. In the IVE condition the VE was displayed to participants by using *Oculus Rift* (Oculus rift-virtual reality headset for 3d gaming, Oculus VR, Inc., Irvine, California, United States) with a resolution of 1280×800 ($16 : 10$ aspect ratio) split between each eye. It supports a 110° FOV, stereoscopic vision, accelerometers, gyroscopes and magnetometers. Conversely, in the N-IVE, the VE was visualized on a 15.6 inch computer screen with a resolution of 1920×1080 pixels. Oculus Rift provides immersion and interaction with the VE through low-latency head tracking while no kind of interaction with the VE was achieved in the N-IVE condition.

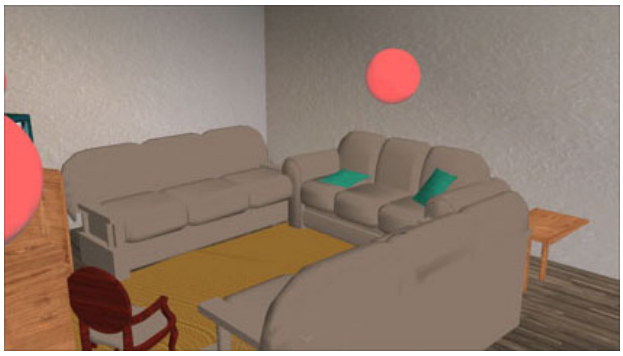

Figure 2: *Screenshot from the VE during the exploration phase in the IVE setup.*

3.3 Procedure

Prior to the experiment, participants self-rated their familiarity with desktop PC, videogames and VR systems and completed the Positive and Negative Affect Schedule (PANAS) [Watson et al. 1988]. Then, they were randomly divided in two groups: 12 participants (3 females) experimented the IVE condition, whereas the other 12 used the N-IVE setup. The task for both groups was to observe the VAs facial expressions arranged as one block of 24 trials, representing twice one of the 6 possible facial expressions in the face of a male or female VA in a randomized order (Figure1). After each trial participants choose what emotion had been shown from a list of six possible options: Anger, Disgust, Fear, Happiness, Neutral and Sadness (Figure3). There was a timeout of 7 seconds that assured the automatic triggering of next trial. The IVE group started the experiment with an exploration phase in which they stood up to foster a complete 360° FOV and interacted with the VE by turning their head. This phase lasted 30 seconds during which participants came across three red balls located at the VE boundaries, that became green after being displayed and disappeared. The interaction through ray-cast operation gave us the evidence that the participants explored the entire VE. Then the VA with a Neutral expression appeared in the VE, but the animation was activated by the participant with a ray cast operation. After all the trials, the IVE participants answered a questionnaire about their experience. The entire duration for IVE group was about 20 minutes included preparation, experiment and questionnaire session. The N-IVE group did not need the exploration phase because they sat in front of screen and visualized a static picture of the VE. Also, they did no fulfill any questionnaire. Therefore the entire duration for N-IVE group was around 10 minutes.

3.4 Measures

The primary response variable was the participants accuracy in the recognition of the basic emotions or neutrality shown on the VA's faces. It was calculated for each participant as the percent of correct identifications of the emotion types displayed by the VA across

the replicates of each facial emotion presentation in each block of trials. A no-response was counted as a wrong identification. The familiarity of participants with computers, videogames and VR systems was self-rated on a 5-point scale. PANAS questionnaire [Watson et al. 1988] was used to assess the participant's positive (PA) and negative (NA) affect state at the time of the experiment. A selection of questions from the Igroup Presence Questionnaire (IPQ) [Regenbrecht and Schubert 2002] with two additional questions on recognition difficulty and VAs expression engagement were used for obtaining information about the immersion in the VE.

Figure 3: ***Left panel****: a participant of IVE group during the experiment while speaking out the label corresponding to the emotion recognized.* ***Right panel****: screenshot of the options screen appearing to participants at the end of every trial.*

3.5 Data analysis

A mixed model ANOVA with Group as a between-subject factor with two levels (N-IVE and IVE) and Face as a within-subject factor with six levels (one for each different facial emotion displayed) was carried out on the collected data. The aim was to investigate whether a difference exists in the recognition accuracy between the N-IVE and IVE, and if there is an interaction between the potential effects of the visualization system used (Group) and the facial emotion to be identified (Face). A two-sample t-test for independent groups was carried out on PANAS PA and NA scores for both groups. A Wilcoxon rank sum test for independent samples was applied to the scores self-rating the familiarity of the N-IVE and IVE groups with personal computers, videogames and virtual environments. Statistical analyses were performed using the Matlab Statistics Toolbox (MATLAB and Statistics Toolbox Release 2011b, The MathWorks, Inc., Natick, Massachusetts, United States) and the language environment for statistical computing *R 3.1.0* [R Core Team 2014].

4 Results and Discussion

In this study we compared the emotion recognition accuracy of the N-IVE and IVE groups for assessing: *(i)* if the the level of immersion in VE invalidate the universality of *basic* emotions and *(ii)* if there is recognition accuracy of facial emotion in an immersive scenario simulating a real life situation. The Wilcoxon rank sum test revealed a homogeneity of the two groups in the use of computer (IVE=$4,417 \pm 0,260$; N-IVE=$4,583 \pm 0,2294,417 \pm 0,260$), videogames (IVE=$2,833 \pm 0,386$; N-IVE=$2,583 \pm 0,336$) and VEs (IVE=$2,583 \pm 0,468$; N-IVE=$2,5 \pm 0,289$). Moreover, t-test did not evidence any statistical difference in the emotional state of participants between the IVE (PA=$29,83 \pm 1,40$; NA=$13,58 \pm 1,40$) and N-IVE (PA=$31,92 \pm 1,12$; NA=$15,25 \pm 1,06$) groups.

The mixed ANOVA $Group \times Face$ (Figure 4) yielded no significant main effect of Group and no significant interaction between

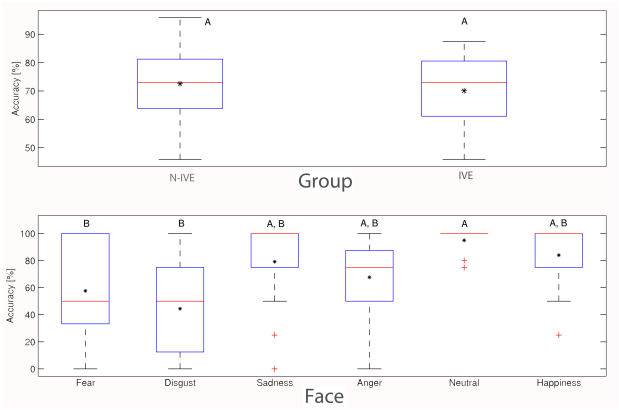

Figure 4: ***Upper:*** *Box plots of percent emotion recognition accuracy for N-IVE and IVE groups.* ***Lower:*** *Box plots of percent emotion recognition accuracy for every facial emotion displayed in both setups across either groups.The asterisk indicates the mean, while the solid line indicates the median. The couples of factor resulting significantly different after a Tukey post-hoc test are marked with different letters.*

Group and Face. In particular, the overall mean recognition accuracy for the two groups was 71.33%. Moreover, the trend of mean scores for IPQ questions reveled that the level of immersion in the IVE was very high and participants perceived the VE as a real life experience. These results confirm the idea that the the level of immersion does not invalidate the universality of *basic* emotions. However, ANOVA results revealed a main effect of Face ($F(5,110) = 11.18, p < 0.0001$). The *post-hoc* Tukey test contrasting between the Face levels returned only two significant differences between means of recognition accuracies. Specifically, a difference was found between the effects of Neutral and Disgust ($p < 0.001$) and Fear and Neutral ($p < 0.05$) facial expressions. Specifically, the mean accuracy of the Neutral face recognition (95.00%) is significantly different from the recognition accuracy of Fear (57.64%) and Disgust (44.44%). The main difficulty in recognition accuracy in both groups (mean=44.44%) was for Disgust. This result is in line with previous studies that compared emotion recognition on virtual and natural faces [Spencer-Smith et al. 2001; Moser et al. 2007; Dyck et al. 2008; Gutiérrez-Maldonado et al. 2014]. [Fabri et al. 2004] showed that Disgust was the only emotion that obtained a lower level of recognition accuracy on virtual than on natural face. This result can be explained by the fact that Disgust is a mixture of other expressions and it is more difficult to be replicated virtually. Moreover, Disgust is characterized by wrinkling at the base of the nose (the element that distinguishes it from Anger), an area very difficult to model in graphical terms, since it requires a large number of polygons [Spencer-Smith et al. 2001]. Differently to previous research we found the highest recognition accuracy for the Neutral expression (mean=95.00%) and not for Happiness (mean=84.03%) [Dyck et al. 2008; Gutiérrez-Maldonado et al. 2014; Kirita and Endo 1995]. However, [Dyck et al. 2008] showed that there is a recognition advantage of the Neutral expression on natural versus virtual faces. The results we obtained might depend on an enhanced consistency of our virtual Neutral with a natural Neutral face. Moreover, in [Dyck et al. 2008] and [Gutiérrez-Maldonado et al. 2014] all 3D facial expressions are static differently to our research in which Neutral is the only face without a blending function, therefore the easier to be recognize. As for the other expressions, Anger obtained a level of accuracy completely in line with previous findings, while Fear and Sadness were recognized by our participants with a lower and higher accuracy respec-

tively compared to the findings for virtual and natural faces reported by [Dyck et al. 2008] and [Gutiérrez-Maldonado et al. 2014].

In conclusion, we obtained a high level of accuracy in the emotion recognition of VA's faces in both groups. This result suggests that IVE can be used in the study of ERF and that the immersion in VE does not invalidate the recognition of basic emotions. Based on this assumption we suggest that IVE can be used in the field of therapy as a training tool to improve emotional skills. The major limitation of our study is the decrease of ecological validity that might result from the absence of a relevant interaction with the VA. This consideration becomes evident in the poor score obtained in an IPQ question: *How much engaging were the expressions displayed on the avatar's face?* However, the engagement with the VA is a natural way of the face-to-face encounter and we will consider this aspect in the next study. We are in fact planning to improve this study by comparing recognition accuracy of VA's faces in a VE and of human faces in a real environment, and to evaluate differences in perception of male and female VAs.

References

ADOLPHS, R. 2002. Recognizing emotion from facial expressions: Psychological and neurological mechanisms. *Behavioral and cognitive neuroscience reviews 1*, 1, 21–62.

BARRETT, L. F., MESQUITA, B., AND GENDRON, M. 2011. Context in emotion perception. *Current Directions in Psychological Science 20*, 5, 286–290.

BOUAZIZ, S., WANG, Y., AND PAULY, M. 2013. Online modeling for realtime facial animation. *ACM Transactions on Graphics (TOG) 32*, 4, 40.

CARROZZINO, M., TECCHIA, F., BACINELLI, S., CAPPELLETTI, C., AND BERGAMASCO, M. 2005. Lowering the development time of multimodal interactive application: the real-life experience of the xvr project. In *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology*, ACM, 270–273.

DARWIN, C. 1955. *The Expression of the Emotions in Man and Animals: With Photographic and Other Illus. With a Pref. by Margaret Mead*. Philosophical Library.

DYCK, M., WINBECK, M., LEIBERG, S., CHEN, Y., GUR, R. C., AND MATHIAK, K. 2008. Recognition profile of emotions in natural and virtual faces. *PLoS One 3*, 11, e3628.

EKMAN, P., AND FRIESEN, W. V., 1978. Facial action coding system: A technique for the measurement of facial movement. palo alto.

EKMAN, P. 1993. Facial expression and emotion. *American Psychologist 48*, 4, 384.

FABRI, M., MOORE, D., AND HOBBS, D. 2004. Mediating the expression of emotion in educational collaborative virtual environments: an experimental study. *Virtual reality 7*, 2, 66–81.

FAITA, C., VANNI, F., LORENZINI, C., CARROZZINO, M., TANCA, C., AND BERGAMASCO, M. 2015. Perception of basic emotions from facial expressions of dynamic virtual avatars. In *Augmented and Virtual Reality*. Springer, 409–419.

GAGGIOLI, A., MANTOVANI, F., CASTELNUOVO, G., WIEDERHOLD, B., AND RIVA, G. 2003. Avatars in clinical psychology: a framework for the clinical use of virtual humans. *Cyberpsychology & behavior 6*, 2, 117–125.

GILLIES, M., AND SPANLANG, B. 2010. Comparing and evaluating real time character engines for virtual environments. *Presence: Teleoperators and Virtual Environments 19*, 2, 95–117.

GUTIÉRREZ-MALDONADO, J., RUS-CALAFELL, M., AND GONZÁLEZ-CONDE, J. 2014. Creation of a new set of dynamic virtual reality faces for the assessment and training of facial emotion recognition ability. *Virtual Reality 18*, 1, 61–71.

KIRITA, T., AND ENDO, M. 1995. Happy face advantage in recognizing facial expressions. *Acta Psychologica 89*, 2, 149–163.

KNIGHT, B., AND JOHNSTON, A. 1997. The role of movement in face recognition. *Visual cognition 4*, 3, 265–273.

MOSER, E., DERNTL, B., ROBINSON, S., FINK, B., GUR, R. C., AND GRAMMER, K. 2007. Amygdala activation at 3t in response to human and avatar facial expressions of emotions. *Journal of neuroscience methods 161*, 1, 126–133.

R CORE TEAM. 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

REGENBRECHT, H., AND SCHUBERT, T. 2002. Real and illusory interactions enhance presence in virtual environments. *Presence: Teleoperators and virtual environments 11*, 4, 425–434.

SATO, W., KOCHIYAMA, T., YOSHIKAWA, S., NAITO, E., AND MATSUMURA, M. 2004. Enhanced neural activity in response to dynamic facial expressions of emotion: an fmri study. *Cognitive Brain Research 20*, 1, 81–91.

SLATER, M., AND WILBUR, S. 1997. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and virtual environments 6*, 6, 603–616.

SPENCER-SMITH, J., WILD, H., INNES-KER, Å. H., TOWNSEND, J., DUFFY, C., EDWARDS, C., ERVIN, K., MERRITT, N., AND PAIR, J. W. 2001. Making faces: Creating three-dimensional parameterized models of facial expression. *Behavior Research Methods, Instruments, & Computers 33*, 2, 115–123.

TRAUFFER, N. M., WIDEN, S. C., AND RUSSELL, J. A. 2013. Education and the attribution of emotion to facial expressions. *Psihologijske teme/Psychological Topics 22*, 2, 237–247.

VANNI, F., CONVERSANO, C., DEL DEBBIO, A., LANDI, P., CARLINI, M., FANCIULLACCI, C., BERGAMASCO, M., DI FIORINO, A., AND DELL'OSSO, L. 2013. A survey on virtual environment applications to fear of public speaking. *European review for medical and pharmacological sciences 17*, 12, 1561–1568.

WATSON, D., CLARK, L. A., AND TELLEGEN, A. 1988. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology 54*, 6, 1063.

YANG, S., AND BHANU, B. 2011. Facial expression recognition using emotion avatar image. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, IEEE, 866–871.