Design of an Emotion Elicitation Tool using VR for Human-Avatar Interaction Studies

P-H. Orefice[1] ✉, M. Ammi[2] M. Hafez[3], and A. Tapus[1]

[1] U2IS, ENSTA ParisTech, Univ. Paris-Saclay
[2] CNRS/LIMSI, University of Paris-Sud, Univ. Paris-Saclay
[3] CEA, List, Sensoriel and Ambient Interfaces Laboratory
pierre-henri.orefice@ensta-paristech.fr

Abstract. With the development of socially interacting machines, it is important to understand how people react depending on their emotional state. Research in this area require emotion elicitation devices. This paper presents such a tool using virtual reality (VR), that merges classical elicitation techniques to emphasize emotional response. The design choices are depicted for four emotions, and a performance analysis using questionnaires is achieved.

1 Introduction

Emotions are a fundamental component of human life. Indeed, they have an impact on human's psychology, physiology, and behavior, and they influence human social interactions. Several technology-based systems have been developed so as to socially interact with humans. However, they mostly lack of emotion embodiment in order to make interactions more realistic and natural. To address this issue, it is important to measure people's reaction caused by their emotional experience. That leads to design reliable tools that elicit a given emotion, in order to relate with the conducted measures. For instance, [6] explore the link between physiological data and the emotion elicited. Investigating how social interaction is altered by the emotional state of subjects has applications in human machine interaction in general. It would enable to detect the emotional state of a subject depending on its reaction during the social interaction. Our research project lays in this context and our approach is to compare social interactions before and after an emotion elicitation phase. This paper focuses on designing an emotion elicitation device that can generate an emotion long enough to remain during interaction phases.

Many methods enable to elicit emotions, using various modalities: through internal process or self-elicitation, by a passive reaction to discrete visual or audio stimuli displayed, or via an interactive tasks. In this study, we are interested in inserting discrete stimuli in a virtual reality (VR) scenario. Indeed, VR enables to use a succession of stimuli combining several modalities, to have the participant involved as he/she controls its mobility, and to increase emotional arousal while adjusting the sense of presence [1]. Several parameters can be balanced given the

emotion to be elicited, which gives higher control on elicitation [4]. [3] designed five environments to elicit joy, anger, anxiety, boredom, and sadness. Based on these examples and motivated by the idea that inserting in VR discrete stimuli from databases would magnify the emotional arousal and duration, the following describes an elicitation platform we designed and presents an evaluation study.

2 VR platform design

Elicitation design: The first steps we had to complete to create the system were to relevant emotions, the right stimuli, and the right timing and scenario. We selected 3 emotions, which are part of the 6 Ekman basic emotions and distant in the VAD space (**V**alence-**A**ctivation-**D**ominance), and a Neutral emotion. This makes 4 conditions. The hypothesis to answer are: *(a)* there is a significant difference in emotion felt depending on elicitation; *(b)* these emotions are close to the associated VAP values: Joy (high **V**, **A** and **D**), Fear (high **A**; low **V** and **D**), and Sadness (low **V**, **A**, and **D**). Neutral (medium **V** and **D**; low **A**). Elicitations of these emotions are respectively referred as **J**, **F**, **S**, and **N**.

The discrete stimuli were chosen from pre-evaluated databases including pictures, musics and video samples. Concerning images, the IAPS database [5] is well-known as it has been evaluated and tested by many studies, in VAD and qualitative [7] spaces. We extracted from this database 10 pictures per condition. For **J**, we picked all pictures noted with at least amusement, or contentment. We ordered them with highest scores for these dimensions, and selected the ones with high valence and arousal. We did the same for For **F** and For **S** with suitable categories and VAD. For **N**, we used the PAD space and selected pictures with valence just above the medium value and chose the ones with low arousal. Music brings a time dimension. The dataset we exploited [2] uses movies soundtracks. We decided to play, for each elicitation, two $1min$ musics and we selected the ones with high score for the given emotion, high arousal, and low level of confusion. Video content depicts more complex situations. We used a French database [8] and selected the samples with high arousal and high score for a given emotion. The VR scenario is as follow: the participant arrives in a corridor with pictures displayed on both sides. He/she has $1.5min$ to cross the corridor gazing at pictures with an average time of $6s$. During this step, a music is played. Then, the participant enters a theater room. A movie sample of $2min$ is displayed. When the video ends, the participant removes its VR helmet.

Elicitation evaluation tools: In order to evaluate our system measure the cognitive response of participants using questionnaires. This is done through two tools. The first one is the Self Assessment Manikin scale (SAM), which returns the emotional state on a 9 points three dimensional VAD scale. The second is composed of discrete adjectives to tick, several responses being accepted. The adjective list is from [8]. These questions are asked after each elicitation phase in VR. We also measure the variability of the emotion felt, and the modality to which the participant was the most attentive and receptive. After the experiment, participants had to watch again every stimulus separately on a screen,

evaluate the emotion felt, and say if it was more intense than in the VR elicitation.

Experimental protocol: The experiment was carried out thanks to 16 participants with scientific background, aged between 22 and 30 years old, mainly used to play video-games. The first phase of the experiment is to let the subject train to use VR and the joystick, playing a basic game. When the participant is comfortable, the main phase begins. 4 emotional conditions are presented, and each occur in four steps. First, a relaxation step with the VR helmet on and a relaxing music of $2min$. Then, the VR environment is displayed, it last between 3 and $4min$. After that a quick questionnaire is answered, and a cognitive game of $2min$ (a Sudoku) enable the emotion to vanish. After the 4 conditions, the participant could provide some comments about the experiment. The total duration was one hour.

3 Results of the platform evaluation

In order to validate our work, we used one-way ANOVA and Tukey post-hoc test when significant. The p-value ($\mathbf{p}$) used for significance is 0.05. The mean is noted μ and standard deviation sd.

Emotion elicitations The first step is to check significance in elicitation evaluation differences. Valence significantly depends on condition ($F(3;58) = 14.2$, $\mathbf{p} < 10^{-6}$). **J** has more valence than all the others (+2.7 from **N**, +3.5 from **F**, and +4.6 from **S**). However, the Tukey test is not significant between **N**, **F**, and **S**. The arousal is discriminating ($F(3;58) = 9.8, \mathbf{p} < 10^{-6}$) but just because **N** case has low arousal (-1.7 from **S**, -2.9 from **F**, and **J**). The Tukey test is not significant for other comparisons. The dominance component is not discriminant. In a second step, we check the correspondence with the expected VAD values. The **N** condition has neutral values: medium valence ($\mu : 4.9; sd : 2.5$) and low arousal ($\mu : 3.5; sd : 2.0$). **J** has high valence ($\mu : 7.6; sd : 1.6$), arousal ($\mu : 6.5; sd : 1.8$), and dominance ($\mu : 7.1; sd : 1.5$). **S** corresponds for valence which is low, but arousal and dominance remain medium. Finally, the **F** VAD responses lack of magnitude but go in the right tendency. If we count the number of good adjectives to describe the elicitation phases, **J** and **S** have good scores (79% and 77%) while the **N** case is noisy (56%) and the **F** case is often confused with "surprise" or "anger" (50%).

Comparison of modalities We also compare which modality is more efficient depending on the emotion elicited. We found that for **S**, subjects were more attentive and receptive to images. The same effect was seen for **F** with the video. In general, little attention was given to music. However, participants got the most receptive to this modality. Many people did not pay attention to a particular modality but they could choose an answer for receptivity. We finally compare the emotional intensity and category differences between viewing stimuli in the post-questionnaire (on a screen) and in VR. The images were not perceived significantly different from VR. However, videos were less intense during the post-questionnaire. Participant judged it more immersible in VR.

Music, however, was significantly perceived more intense in post-questionnaire, probably due to the higher attention level.

4 Discussion, conclusion, and perspectives

The experiment taught us a lot about how the stimuli we proposed are perceived and how VR alters this perception. Our system succeeded to elicit the right emotions, and more efficiently for **J**. The noise in **N** adjectives can be caused by the fact no adjective qualify precisely such an emotion. **S** had low valence, mainly thanks to images. However video and musics were evaluated as "calm" and mi-aroused in the post-questionnaire. Better stimuli may have strengthen the **S** condition. The **F** did not fit exactly the hypothesis. The individual stimuli evaluation are noisy, images have a lot of qualification errors and too low arousal, the video's valence is too random. The musics are better marked but subject payed less attention to this modality. This confirms that the attention given to a modality can change the emotion felt in the whole elicitation.

This paper proposed an example of emotion elicitation using virtual reality. Several benefits of VR, including the ability to merge several modalities, are shown. The proposed solution revealed that classical elicitation methods integrated in the VR setup improves immersion and emotional intensity. Future work will take advantage of the presented results in order to have a robust elicitation method, that will allow to do more complex evaluations, and study the link between tactile interaction with virtual agents or robots, and emotions.

References

1. Bouvier, P.: La présence en réalité virtuelle, une approche centrée utilisateur. Ph.D. thesis, Université Paris-Est (2009)
2. Eerola, T., Vuoskoski, J.: A comparison of the discrete and dimensional models of emotion in music. Psychology of Music (2010)
3. Felnhofer, A., Kothgassner, O., Schmidt, M., Heinzle, A.K., Beutl, L., Hlavacs, H., Kryspin-Exner, I.: Is virtual reality emotionally arousing? investigating five emotion inducing virtual park scenarios. International journal of human-computer studies 82, 48–56 (2015)
4. Geslin, E., Jégou, L., Beaudoin, D.: How color properties can be used to elicit emotions in video games. International Journal of Computer Games Technology 2016, 1 (2016)
5. Lang, P., Bradley, M., Cuthbert, B.: International affective picture system : Affective ratings of pictures and instruction manual. Technical report A-6 (2005)
6. Ménard, M., Richard, P., Hamdi, H., Dauce, B., Yamaguchi, T.: Emotion recognition based on heart rate and skin conductance. In: PhyCS. pp. 26–32 (2015)
7. Mikels, J., Fredrickson, B., Larkin, G., Lindberg, C., Maglio, S., Reuter-Lorenz, P.: Emotional category data on images from the international affective picture system. Behavior research methods 37(4), 626–630 (2005)
8. Schaefer, A., Nils, F., Sanchez, X., Philippot, P.: Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. Cognition and Emotion 24(7), 1153–1172 (2010)