HOW VISUAL DISCOMFORT AFFECTS 3DTV VIEWERS' EMOTIONAL AROUSAL

Miguel Barreda-Ángeles[1], Romuald Pépion[2], Emilie Bosc[2], Patrick Le Callet[2], and Alexandre Pereda-Baños[1]

[1]Barcelona Media – Innovation Center, Spain
[2]IRCCYN/IVC, Polytech Nantes/Université de Nantes, France

ABSTRACT

We present the results of a research on the effects of image distortions known to induce visual discomfort (namely, the introduction of Gaussian blur on one visual field) on the emotional reactions of 3DTV viewers. Emotional reactions were explored by means of self-reported as well as psychophysiological methods. The results showed that, despite visual discomfort not having an effect on self-reported emotions, the level of electrodermal activity (a physiological correlate of emotional arousal) was clearly affected, suggesting the existence of seemingly complex effects of visual discomfort over viewers' emotional arousal.

Index Terms **—** 3DTV, Quality of Experience, emotions, entertainment, psychophysiology, electrodermal activity

1. INTRODUCTION

The goal of many entertainment consumption situations is to obtain arousing emotional experiences. Viewers want dramas to move them, comedies to make them crack up, or horror movies to send shivers down their spine. Although the kind and intensity of the emotions experienced during entertainment depend to a large extent on semantic aspects of the contents (e.g. events in the story, actions of the characters, etc.), formal and presentation attributes can also contribute to determine viewers' emotional reactions [1]. In the case of 3DTV, stereoscopic images may introduce issues related to image distortions known to elicit visual discomfort, which have consequences on the overall visual experience [2]. However, how those distortions may affect viewer's emotions beyond visual experience has been hardly addressed to date.

Since Quality of Experience (QoE) is related to the satisfaction of users' expectations with a certain service or application [3], and most users seek for emotional experiences when consuming entertainment contents, understanding the impact of image features on emotions can be considered a key issue in 3DTV QoE research. Here we present some results of a Short Term Scientific Mission conducted within the frame of the 3DConTourNet COST Action, whose objective was to initiate an exploration of how image distortions affect users' emotions. The results presented here focus on the arousal dimension of emotions.

2. VISUAL DISCOMFORT AND EMOTIONAL AROUSAL

Emotions are commonly defined in terms of valence (or the extent to which the emotion is positive or negative) and arousal (meaning the intensity of the emotion) [4]. When watching entertainment contents the extent to which viewer's attention is focused on the contents, and not on the context of consumption, has been related to the experienced arousal [5-6]: the more *involved* is the viewer with the contents, the more easily will experience arousing emotions.

Visual discomfort produced as a consequence of image distortions (such as parallax problems or video coding artifacts) is likely to impact on viewers' involvement with contents, and, hence, to have consequences in their emotional reactions. In this sense, two hypotheses may be elaborated. First (H1), visual discomfort may disrupt viewer's involvement with contents, leading to a lesser emotional arousal. On the other hand (H2), experiencing visual discomfort may be perceived as an arousing negative emotion itself, therefore the emotional arousal would be higher in contents with image distortions. In the next section, we describe the experiment we conducted in order to test these two hypotheses.

3. METHOD AND PROCEDURE

3.1 Design and materials

Participants watched a series of nine 3DTV sequences presenting different levels of image distortions known to elicit visual discomfort, as well as contents able to produce different degrees of emotional arousal on viewers. Thus, the experimental design included two independent variables: content and visual discomfort (VD), with three levels each. The three types of contents (Type 1, 2 and 3) varied in the level of arousal, from very calm contents in the case of Type 1, to highly arousing contents in the case of Type 3. Three sequences for each type of contents were selected, so creating the series of nine sequences. Type 1 contents were 3D images of still shots of objects or landscapes, with low level of movement within the scene, selected from a public available dataset [7]. Type 2 contents were segments of a synthetic image short-film, also publicly available [8], which included different shots with moderate to high level of movement within shots and a mildly arousing story. Type 3 contents were segments of a commercial action and horror film presenting natural images, several different shots, moderate to high movement within the scenes, and very arousing moments. The level of arousal was initially set in conformance with the researchers' criterion, and afterwards confirmed by participants' ratings and physiological responses (see Results). The duration of the sequences was about two minutes each; in the case of shorter Type 1 contents, they were edited to reach duration of two minutes.

The three levels of VD were low, middle, and high, and were manipulated by including Gaussian blur on the visual field corresponding to the observer's dominant eye [9], from no-distortion to mild distortion (30 pixels Gaussian blur) to high distortion (60 pixels Gaussian blur). Thus, three versions of each of the three sequences of each type were created.

A mixed design was applied to the viewing conditions in a way such each participant watched the nine sequences including the three types of contents, and for each type one of the sequences

978-1-4799-4758-4/14/$31.00 ©2014 IEEE

included one of the levels of image distortion. The relationship between levels of VD and sequences within each type was balanced between participants. Therefore all the participants watched three sequences of each type including the three levels of VD, but the sequence that presented a particular level of VD was not the same for all participants. This kind of design is often used in media psychology research because it allows permuting all the contents by all the conditions, so that participants do not have to watch the same clip in different VD conditions.

On the other hand, to determine the most arousing moments for each of the three Type 3 sequences (which was useful for analyzing the effects of VD in the specifically arousing moments, see Results) a pretest was conducted in which six participants (two women) with ages between 29 and 38 years old (M = 32.00; SD = 3.16) watched a 2D version of all three Type 3 videos and marked the more arousing moments of each one of them. In order to correct for inaccuracy in the temporal positions given by the participants, the temporal marks differing less than four seconds were considered as referring to the same moment, and a posterior viewing by one of the researchers served to determine the exact second of each arousing moment. Then participants' selections were rated by assigning three points to the moment selected by each observer as the most arousing, two points to the second most arousing moment, and one point to the third. The four moments with higher ratings for each Type 3 clip were considered as the most arousing moments.

3.2 Measurements

Emotional arousal of contents was assessed by means of self-reported and physiological methods. A nine points Self-Assessment Manikin (SAM) questionnaire [10] was used to measure self-reported valence and arousal. Meanwhile, since increases in the electrodermal activity (EDA) are related to increases in emotional arousal [11], EDA was taken as a physiological index of emotional arousal. To record EDA data, a TMSi equipment and a Matlab in-house application were used, with a sampling rate of 2048 Hz.

3.3 Participants and procedure

Eighteen volunteers (including 12 women) with ages between 18 and 33 years (M = 22.72, SD = 4.27) participated in the experiment. The experimental protocol included a visual acuity test in a Monoyer chart (to guarantee participants' visual acuity), and a test to determine eye-dominance, consisting in selecting their preferred ones among pairs of 3D videos that included image distortions in either the left of the right visual field, considering that if a participant showed preference for the videos with the distortions in one visual field (e.g. left field), his or her dominant eye was the opposite (e.g. right eye). Then participants signed informed consent, received instructions to fill the SAM scales, and the electrodes for the physiological measurement were placed, including two electrodes on the fingers of the non-dominant hand to measure EDA.

The tests were conducted in an ITU conforming laboratory, and the stimuli were displayed following ITU-T BT.500 specifications by means of an Acer GD245HQ associated to a NVidia 3D Vision system. The order of presentation of the sequences was randomized and different for each participant. A black screen was displayed during 30 seconds between the videos, in order to separate sequences and allowing participants to fill in the SAM questionnaire after each sequence.

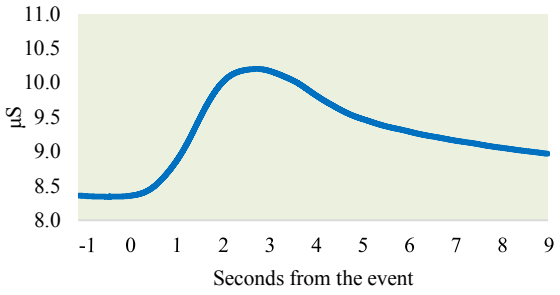


Figure 1. Example of phasic EDA response to an emotional event

4. RESULTS

Data were analyzed following two strategies. The first analysis regarded the viewers' emotional reactions to contents and VD for the whole set of nine video sequences. Then, in a second analysis, we examined the results concerning the subset of the most arousing and entertainment-like contents (Type 3 videos). This makes sense since those videos were the more similar to entertainment contents usually found in real-life consumption. In particular, we considered that the effects of VD over emotional reactions could be more evident in the more emotional events of the contents. Therefore we focused on the viewers' EDA responses at the most arousing moments, in which phasic increases in skin conductance may be expected (see Figure 1).

4.1 Analysis of questionnaires data

A series of ANOVAs were conducted to analyze participants' ratings of valence and arousal on SAM questionnaires. A significant difference was found in arousal ratings related to content type, $F(2,34) = 64.62$, $p < .001$, $\eta^2_p = .79$. Paired comparisons with Bonferroni adjustment showed that the average ratings for arousal were different for the three types of contents, being Type 1 contents (M = 1.54, SE = 0.18) less arousing than Type 2 (M = 3.72, SE = 0.41), and those less arousing than Type 3(M = 5.44, SE = 0.36) ($p < .001$ in the three cases), so confirming the manipulation of the level of arousal between content types. However, no differences were found in the ratings of arousal related to VD, $F(2,34) = 1.46$, $p < .25$, $\eta^2_p = .08$, nor to the interaction between VD and content type, $F(2,34) = 1.61$, $p < .18$, $\eta^2_p = .09$.

A second analysis focused just in the three Type 3 sequences (which will be called T3.1, T3.2, and T3.3). Among these the ratings showed no significant differences in the level of arousal by content, $F(2,34) = 1.47$; $p = .24$; $\eta^2_p = .08$. Neither significant effects of VD were found on the ratings of arousal, $F(2,34) = 1.65$; $p = .21$; $\eta^2_p = .08$.

4.2 Analysis of EDA data

The EDA signal was smoothed by averaging each sample with the 200 ms previous and posterior samples. Then data were averaged for each three seconds period, resulting in a temporal series of 40 points for each video sequence of each participant. Multilevel mixed models (see [11]) were applied to analyze the data. Model construction followed a progressive *step-by-step* approach in which simpler model including the least number of elements were constructed first and then random and fixed factors were added to the model progressively, comparing more complex models to simpler ones by likelihood tests to ensure that more

	Model 1	
Fixed effects	Coefficients (Vx10^{-6})	
Intercept	0.568	***
Content Type 2	-0.007	
Content Type 3	0.011	**
VD	-0.010	***
Time	0.002	***
Order	-0.032	*
Type 2 x VD	0.025	***
Type 3 x VD	0.013	***
VD x Time	0.0002	**
Type 2 x Time	-0.02	
Type 3 x Time	-0.13	
Type 2 x VD x Time	-0.0005	***
Type 3 x VD x Time	-0.0003	**
Random effects	*SD*	
Intercept	0.433	
Time	0.002	
Order	0.043	
Time x Order	0.0002	
Residuals		
Phi estimate	0.999	
-2LogLikelihood	-28842.5	

*$p < .05$; **$p < .01$; ***$p < .001$

Table 1. Coefficients of the Model 1 (Model for the three types of contents).

complex models fitted data better than simpler ones. Random intercepts for subjects, random slopes for the effect of time, order of presentation and their interaction, and functions considering autocorrelation of residuals were included in the cases that they improved model fitting. Then fixed factors were added to the model: type of content, VD, time within each sequence (from 0 to 40), the interaction between those three factors, and the effects of the order of presentation of the sequences. Since the different levels of VD keep an arithmetical relationship (a distortion of 0, 30 and 60 pixels Gaussian blur is related to low, middle, and high VD levels, respectively), VD was considered as a continuous variable.

The model that best fitted the data is summarized in Table 1. As shown in the model, Type 3 contents increased EDA levels compared to Type 1 contents (which were taken as baseline), while there was not significant effects of Type 2 contents. The main effect of VD was a reduction of EDA levels; nevertheless, the interaction between VD and time suggests that, as time elapses, this effect is moderated. However, in the case of Type 2 and 3 contents (more arousing than Type 1, which is consistent with SAM ratings), those effects are reversed: the interactions between VD and Type 2, and between VD and Type 3, suggest that VD increases EDA in those kinds of contents, while the triple interaction between these contents, VD and time suggests that this effect is reduced and even reversed over time. Summarizing, in the case of Type 1 contents VD reduces arousal, but this reduction decreases over time, while in Type 2 and 3 contents VD increases arousal, but this increase in reduced over time.

The next step in the analysis was to assess how VD affected emotional arousal in the more arousing moments of Type 3 videos. A new model was calculated following the approach applied for Model 1, but in this case, in order to preserve temporal resolution of the signal that facilitates the presence of phasic skin conductance responses, EDA signal was averaged every one second period (instead every three seconds as in the previous case). Then, the model was calculated considering only the EDA values between two seconds before and seven seconds after the four more emotional events for each one of the three Type 3 videos for each subject. Model 2 (summarized in Table 2) is consistent with Model 1 in showing a reduction of EDA values associated to the presence of VD, but significant interactions between VD and T3.2 and T3.3 contents show that the reduction of emotional arousal in lower in T3.3 sequences, and even that VD increases arousal in the case of T3.2 (since the coefficient associated to the interaction is higher than the coefficient associated to the main effect of VD).

	Model 2	
Fixed effects	Coefficients (V x 10^{-6})	
Intercept	0.453	***
T3.2	-0.057	***
T3.3	-0.046	**
VD	-0.028	***
Time	0.0005	***
Order	-0.012	*
T3.2 x VD	0.042	***
T3.3 x VD	0.008	***
Random effects	*SD*	
Intercept	0.219	
Time	0.001	
Residual	0.098	
Phi estimate	0.989	
-2LogLikelihood	-6399.6	

*$p < .05$; **$p < .01$; ***$p < .001$

Table 2. Coefficients of the Model 2 (Model for Type 3 videos).

5. DISCUSSION

The goal of this experiment was to explore two hypotheses that seem a priori opposed. H1 forecasted that, by reducing viewers' involvement, VD would reduce emotional arousal, while H2 forecasted that the unpleasant feeling of VD would be perceived as a negative emotion that would increase arousal. The results show that, even if the main effect of VD in both models is a reduction in arousal, VD is related to increases or decreases in EDA levels depending on the interactions between VD and type of contents and time, which suggest that both hypothesized effects of VD may be present at a time.

Regarding Model 1, VD is associated to a decrease in arousal for the baseline contents (Type 1), which is moderated over time. It thus can be interpreted that, for those contents, the involvement of the viewer is lower when VD is present. Meanwhile, the unpleasant emotion associated to VD could be responsible for the increase on arousal over time. Therefore in the case of Type 1 contents, the reduction in emotional arousal as a consequence of lower involvement would be higher than the increase on it predicted by H2. By contrast, Type 2 and 3 contents show the opposite pattern: VD is associated to increased arousal, moderated over time. Since Type 2 and 3 contain moving images and many changes of shots, it is possible that in that case VD would be more evident and disturbing, so the unpleasant emotion associated to VD could have been stronger in those cases, compared to Type 1 contents. Moreover, the more emotional events of Type 2 and 3 contents concentrate at the final seconds of the clips, which is consistent with the fact that involvement had been lower when VD is present, and consequently the reduction on arousal associated to VD was higher when the variable time had higher values in Type 2 and 3 sequences (i.e. at the final seconds of the sequences).

On the other hand, if we compare the coefficients associated to VD and the coefficient of the interactions of VD and Type 2 and 3, and also the coefficients of the interaction between VD and time and those of the interaction between VD, time and Type 2 and 3 contents, we can see that in all cases the absolute values of the coefficients are larger when Type 2 or Type 3 contents are present. This suggests that the absolute effects of VD are stronger for Type 2 and 3 contents than for Type 1. This makes sense if we

consider that, compared to Type 1, contents of Type 2 and 3 present both more arousing stories and more visual components (movement, scene changes, etc.). Therefore, the lower level of emotional arousal and visual features may have produced ceiling effects that limit the extent to which the both hypothesized effects of VD had been present in Type 1 contents.

As can be seen in Model 2, the reduction of arousal related to VD is also present in the more arousing moments of the contents. In this case, the model also shows differences between the three groups of sequences, being the baseline sequences (T3.1) more arousing than the other two groups (T3.2 and T3.3), and the effects of VD over arousal were not the same for the three groups of sequences. The reduction of VD is lower for T3.3 compared to T3.1, and even VD is related to an increase in arousal in the case of T3.2. Maybe the different features (both in content and in visual aspects) of the three groups of sequences might explain this fact; nevertheless, since those differences between contents are much more subtle and less evident between the three Type 3 groups of sequences than between the three different types of sequences (Type 1, 2, and 3), it is difficult to figure out which factors could be responsible of them.

The mentioned interactions between VD, content aspects, and time highlight the need to take into account content (semantic) aspects within QoE research. While the use of relatively simple stimuli (such as Type 1 contents) in QoE research is necessary and useful to carry out controlled and detailed explorations of the effects of technological and visual features on QoE, when higher-level psychological factors (such as emotions) are considered, it is also necessary to have in mind their specificities as psychological processes. For example, in our experiment we needed to use longer segments than the ones commonly used in research on visual aspects of QoE, in order to allow for some narrative development of the stories which could induce emotions on viewers. The inclusion of variables related to semantic aspects involves a large increase in the possible factors related to the contents and their interaction with user's features (preferences, background, context of consumption, etc.) that brings important challenges for researchers on the field.

Another important aspect of these results is that self-reported and physiological measures of arousal showed very different patterns of responses to VD. While any effect of VD was evident on SAM ratings of arousal, EDA results indicated complex effects of VD over emotional arousal. This illustrates the utility of physiological measures in QoE research, especially for addressing questions that can fall beyond users' awareness, as, for example, subtle changes in emotional responses. The combined application of self-reported and psychophysiological methods may thus be useful to obtain a detailed picture of the psychological processes that determine QoE, and of which factors act in levels under and above user's awareness.

This research has some limitations; one of the main of which is the lack of control for many visual and content factors that could be impacting on the perceived emotions, which should be considered in future explorations of this issue. Other limitations refer to the limited size of the contents sample, the limited number of participants, and especially, the fact that the experienced VD was not controlled. By controlling this factor (for example, by asking observers to rate the experienced VD) future experiments could address how different aspects of both content and technical aspects interact in the overall perceived QoE.

6. CONCLUSIONS

Despite its limitations, this experiment shows that image distortions eliciting visual discomfort on viewers also have consequences on their emotional reactions. Specifically, the found evidences support the idea that visual discomfort has not a straight relationship to viewer's emotions, but that different effects may interact between them and with content and time factors. Future research should carry out more fine-grained explorations on how visual discomfort interacts with different content-related factors to determine users' emotions, without forgetting the important role of time that, through habituation or cumulative processes, may moderate the effects found with short stimuli. Taken together, all these questions make 3DTV QoE research a challenging field in which the collaboration of multidisciplinary teams will build up as the focus of the research moves from technological and low-level perceptual questions, to higher-level cognitive and emotional processes, and, eventually, even social (e.g. referring to the context of consumption) aspects of the experience.

7. ACKNOWLEDGMENTS

This research was partially supported by the EU COST Action IC1105 (3D-ConTourNet).

8. REFERENCES

[1] B. H. Detenber and A. Lang, "The influence of form and presentation attributes of media on emotion", in K. Döveling, C. von Scheve, and E.A. Konijn (Eds.), *The Routledge handbook of emotions and mass media* (pp. 275-293). New York: Routledge, 2011.

[2] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "Quality of experience model for 3DTV", in *Proc. SPIE 8288 Stereosc. Disp. Appl. XXIII,* San Francisco, USA, Jan. 2012, 82881P.

[3] P. Le Callet, S. Möller, and A. Perkins (eds.), "Qualinet white paper on definition of quality of experience". March, 2013.

[4] M. Bradley and P. J. Lang, P.J., "Emotion and motivation", in J.T. Cacioppo, L.G. Tassinary, and G.G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 581-607). New York: Cambridge University Press, 2007.

[5] D. Zillmann, "Dramaturgy of emotions from fictional narration", in J. Bryant and P. Vorderer (Eds.), *Psychology of entertainment* (pp. 215-238), Mahwah, NJ: Laurence Erlbaum, 2006.

[6] P. Vorderer. F. F. Steen, and E. Chan, "Motivation", in J. Bryant and P. Vorderer (Eds.), *Psychology of entertainment* (pp. 3-17). Mahwah, BJ: Lawrence Erlbaum, 2006.

[7] E. Chen, "RMIT3DV: Pre-announcement of a creative commons uncompressed HD 3D video database", in *Proc. 4th Int. Workshop Qual. Multimed. Exp. (QoMEX 2012)*, Yarra Valley, Australia, Jul. 2012, 212-217.

[8] Orange Open Movie Team, "Elephants dream", 2006. Retrieved from: http://www.elephantsdream.org/

[9] M. Lambooij, M. Fortuin, I. Heynderickx and W. A. IJsselsteijn, "Visual discomfort and visual fatigue of stereoscopic displays: A review", *J. Imaging Sci. Technol.*, vol. 53, pp. 30201-1-30201-14, May 2009.

[10] M. Bradley and P. J. Lang. "Measuring emotion: The self-assessment manikin and the semantic differential", *J. Behav. Ther. Exp. Psychiatry,* vol. 25, pp. 49-59, Mar. 1994.

[11] M. E. Dawson, A. M. Schell, and D. L. Fillion, "The electrodermal system", in J.T. Cacioppo, L.G. Tassinary, and G.G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 159-181). New York: Cambridge University Press, 2007

[12] P. Bliese, "Multilevel modeling in R (2.5): A brief introduction to R, the multilevel package and the nmle package". Retrieved from: http://cran.r-project.org/doc/contrib/Bliese_Multilevel.pdf