The influence of the visualization task on the Simulator Sickness symptoms - a comparative SSQ study on 3DTV and 3D immersive glasses

Raluca Vlad, Olha Nahorna, Patricia Ladret, Anne Guérin-Dugué

THE INFLUENCE OF THE VISUALIZATION TASK ON THE SIMULATOR SICKNESS SYMPTOMS - A COMPARATIVE SSQ STUDY ON 3DTV AND 3D IMMERSIVE GLASSES

Raluca Vlad, Olha Nahorna, Patricia Ladret, Anne Guérin

GIPSA-lab, 11 rue des Mathématiques, Grenoble Campus,
BP46 F - 38402 Saint Martin d'Hères Cedex, France

ABSTRACT

The human factors are an essential aspect to take into consideration in order to explain the level of public acceptability of new stereoscopic devices. A study using the Simulator Sickness Questionnaire allowed us to illustrate the differences in symptoms after the visualization of 3D images on a 3DTV screen and on a pair of prototype immersive 3D glasses. Also, the results of our study showed that the visualization task influenced the exploration of the scenes, and therefore influenced the evolution of the simulator sickness symptoms.

Index Terms— SSQ, Simulator Sickness Questionnaire, simulator sickness, human factors, human evaluation, subjective evaluation, nausea, oculomotor effort, disorientation, stereoscopic display, 3DTV, 3D glasses, head-mounted display.

1. INTRODUCTION

In time, the appeal exerted by stereoscopy on the public proved to be strong. History shows that there existed several *waves* of substantial commercial success for stereoscopic 3D movies in the past [1] [2]. However, despite the fascination that these movies created, the technology did not succeed in being adopted neither at a large scale, nor on a long term. The main reason for this failure was the low quality of the stereoscopic content in terms of visual quality.

At present, new stereoscopic movies appeal the cinema goers. We can definitely say that we are *on a new 3D wave*, judging from the large diversity of stereoscopic 3D compatible devices we are surrounded by (television screens, head mounted displays, smartphones, game consoles, medical instruments *etc.*) and by their permanently improved performances.

However, research shows that a satisfying level of visualization comfort is still a distant goal for the stereoscopic displays, even today. Besides the large binocular disparities, factors like frequent changes in accommodation and vergence, 3D artifacts, and unnatural blur have been identified as sources of visual fatigue in the current stereoscopic systems [3].

In the introduction to his book [4], Mendiburu predicts that one day the stereoscopic technology will become fully integrated in our daily lives and will replace the existing two-dimensional technologies. Still, judging from the previous moments of failure for stereoscopy, it is certain that in order for this to happen, the viewing comfort of stereoscopic systems must be at least comparable to that of conventional television [5].

Subjective and objective methods exist for quantifying the effect that the stereoscopic content has on the physical symptoms of the viewers [3] and focusing on this human factors perspective is considered of essential value in the studies on the acceptance of stereoscopic technologies [5].

In the present paper, we focus our attention on the Simulator Sickness Questionnaire (SSQ), a subjective method that allows to estimate the physical symptoms that are likely to deteriorate after the immersion in a simulator-like experience. The SSQ was developed by Kennedy *et al.* as a method of evaluating the effects of aviation simulator displays on a series of physical symptoms [6]. It was however soon used, as-is or adapted, for evaluating human factor aspects related to various other consumer immersive technologies.

Howarth and Costello [7] used the SSQ in an adapted form in order to illustrate the differences in the human effects determined by the use of a head-mounted display (HMD) as a 'personal display system' (head-tracking disabled), in contrast to the effects determined by a classic visual display. The authors expected to find that the sensory conflict specific to the personal display system (for head movement, no visual response indicating this movement) would generate stronger symptoms than a classic display. Their results confirmed the hypothesis, with augmented scores in the case of the personal display system for general discomfort, fatigue, headache, nausea, dizziness, and stomach awareness.

Häkkinen *et al.* performed several SSQ studies. When analyzing the symptoms after watching a movie on a television, watching the same movie on a HMD, and playing a game on the same HMD, they could conclude that playing the game on a HMD produced the strongest levels of nausea and disorientation. However, surprisingly, more nausea was recorded after watching the movie on the classic television than on the head-mounted display (HMD) in the considered test conditions [8]. In a study where the participants played two video games (a fast-paced car racing game and a simple and stationary car racing game) on two different displays (a head-worn display and a CRT display), the authors observed that simulator sickness were perceived in all the four cases, showing that a captivating and audio-visually appealing game does not decrease the reported simulator sickness symptom levels. Since the levels of nausea and disorientation were larger for the fast-paced game than for the simple one in a similar way on the two displays, it was also concluded that the main difference in the context of the experiment was between the game types and not between the display types [9]. A third study on video games showed that for the same game played on a CRT display and on a head-worn virtual display with or without stereoscopy, the strongest symptoms were recorded for the stereoscopic mode on the head-worn display, with only small differences of symptoms between the two cases of 2D displays [10].

Jumisko-Pyykkö *et al.*, through a group of five experiments performed on auto-stereoscopic displays, during which the SSQ was administered several times, observed a slight and mainly short term increase of the simulator sickness symptoms [11]. Their findings

indicate that, for certain visualization conditions, an adaptation can occur that leads to a reduction in intensity of the symptoms for a longer duration of the immersion.

Previous experiments thus focused on the impact of watching stereoscopic video media on various types of screens and showed that the simulator sickness symptoms can be influenced by the type of the display, by the content watched, by the activity performed by the observer and by the duration of the visual immersion.

Our experiment concentrated on understanding the influence on the human symptoms of factors like the type of the display, the level of artifacts, or the visualization task. This in a context where the observers watched stereoscopic still images containing artifacts, displayed on two different devices: an ultra-wide 3DTV or a pair of prototype immersive 3D glasses. The influence of all the three factors under study was confirmed by the results.

2. EXPERIMENTAL APPROACH

2.1. Context and experiment structure

The experiment presented in this article was performed as an in-depth analysis of the findings in [12]. Its main purpose was to lead to the definition of a better model of the way the naïve observers perceive and judge the quality of stereoscopic content.

Therefore, the test was divided in four sub-experiments, each represented by a task of quantitative psycho-perceptual evaluation of an attribute related to the stereoscopic quality: (1) the *image quality*, (2) the *comfort*, (3) the *realism of the rendering*, and (4) the *overall 3D quality* perceived for a set of stereoscopic images. The four tasks were repeated with different participants on two stereoscopic displays: a 3DTV and a set of immersive 3D glasses.

The model-related results of the experiment will not be presented here, the article will only illustrate the human factors aspect related to the evolution of the simulator sickness symptoms of the participants after the visualization experiment.

2.2. Test stimuli

Our stereoscopic test database was made up of still images of four different test scenes: two outdoor and two indoor. In each indoor or outdoor group, one scene contained only still objects and the other scene a human character. They were of varying spatial information complexity. The scenes are illustrated in Table 1.

Table 1. The four different scenes at the basis of our stereoscopic database.

scene 1	scene 2	scene 3	scene 4

For capturing the database images (at the resolution of 1080p), we used a compact Fujifilm FinePix Real 3D W3 camera, having a lens separation of 7.5 cm. All the images were intended to be as natural as possible, as taken by ordinary users, thus placing the experiment in a consumer utilization context, from the point of view of the content to be visualized.

For each of the four scenes, we derived 27 test conditions, leading to a set of 108 distinct presentations. These 27=3x3x3 test conditions were represented by the combination of: 3 different levels of circular averaging blur for which the radius parameter is given, 3 different levels of maximum crossed disparities given as intervals expressed in minutes of arc, and 3 different levels of camera-scene distance at capture, conveniently chosen in function of each test scene

(details in Table 2). The 3 levels of crossed disparities were computed considering the observations on the maximum disparities supported by the human visual system in [13].

Table 2. The test conditions considered for our stereoscopic database.

blur	disparities	capture distance			
		scene 1	scene 2	scene 3	scene 4
0	[-21,-4]	6.26m	10.25m	28.5m	5.5 m
1	[-86,-53]	7.76m	15m	31.6m	10m
1.75	[-140,-104]	9.26m	20m	36m	15 m

The 3D database used in this study is available online [1].

2.3. Displays

The first display was a Panasonic TC-P50VT20 stereoscopic TV screen with active shutter glasses. The size of this screen was of 111 cm x 70 cm and its resolution was of 1080p. The visualization distance for the 3DTV was 1.5 m and the stereoscopic geometry of the images in our database was configured having as reference these visualization parameters. The database images were displayed in the side-by-side mode.

The second display was a prototype head-mounted device, in the form of a pair of 3D glasses with two micro OLED displays. The device did not include a head-tracking option, therefore the concept of *personal display system* used in [7] is compatible to the context of our experiment. The field of view of the micro screens of the prototype 3D glasses was of approximately 28° at a WVGA resolution of 854 x 480 and a system of lenses inside the glasses was used for the magnification of the image on the two screens and also for adjustments by each participant, in order to adapt the glasses to any different diopters. The maximum resolution accepted for the content was 720p in side-by-side mode, therefore our images were adapted to this format. This device was a prototype produced within the MOOV3D project.

2.4. Methodology

The experimental procedure was designed according to the ITU-R BT.500-13 Recommandation [14] for the subjective assessment of visual quality. The structure of each of the four sub-experiments corresponded to the SSMR method (single stimulus with multiple repetition). Thus, in one test session, all the 108 presentations of the database were shown twice, with a several minutes pause between the two repetitions. During each sub-experiment, the presentations had to be rated in function of the task given, using the five-rating ITU-R quality and impairment scale highlighted on a keyboard (the keyboard could be seen while wearing the 3D glasses). For a common interpretation by all the participants, the four concepts to rate were explained on the task sheets. Also, a learning phase was included before the first repetition, with 8 presentations representing the extreme test conditions. For each repetition, the images of the database were displayed in a random order. The display duration of each presentation was of 5 seconds and was always preceded by a fixation cross on a gray background. The voting was performed after the visualization of the stimulus, during the display of a gray voting screen which reminded the scale.

The whole experiment took place in a standard experiments room, where a daylight lamp of 25 W was used to light the space of approximately 12 m^2 in surface, simulating a home environment.

---

[1]http://www.gipsa-lab.fr/projet/MOOV3D/

It was performed both in English and French, in function of the preferences of the participants.

The SSQ was filled in by all the participants before and after the visualization session.

2.5. Use of the Simulator Sickness Questionnaire

An adapted form of the SSQ was used in our experiment. From the set of 16 questions originally presented in [6], we kept only the 12 that we considered fit for the context of visualizing static stimuli. They are illustrated in Table 3, along with their weights relative to each symptom cluster assessed: N (nausea), O (oculomotor), and D (disorientation). This modification did not reduce the correctness of the interpretation of our results, since the reasoning on the data we collected was done mostly as a comparative analysis of results obtained in the same conditions.

Table 3. The adapted SSQ used in our study (the questions used and the weights for computing the N, O, and D symptom scores).

	Weight		
SSQ Symptom	N	O	D
General discomfort	1	1	0
Fatigue	0	1	0
Headache	0	1	0
Eyestrain	0	1	0
Difficulty focusing	0	1	1
Nausea	1	0	1
Difficulty concentrating	1	1	0
Fullness of head	0	0	1
Blurred vision	0	1	1
Dizzy (eyes open)	0	0	1
Dizzy (eyes closed)	0	0	1
Vertigo	0	0	1

The original purpose of the use of the SSQ in [6] was to compare a large number of simulators from the point of view of the symptoms they generate, thus the authors of the questionnaire recommended to administer it only after the visual immersion period. In our case, the purpose was to evaluate how the human factors evolve from before to after the visualization of images on a given stereoscopic display, therefore we complemented the post-exposure data collection with a pre-exposure application of the questionnaire, adapting its utilization to our experimental context.

The scale used for each question was the same as in the original questionnaire: a 4-point adjectival scale using the keywords: *none* (score 0), *slight* (score 1), *moderate* (score 2), and *severe* (score 3).

2.6. Participants

For this experiment, we recruited naïve participants. In order to avoid influences among tasks or among display types, different persons participated to each sub-experiment and a minimum of 10 and a maximum of 15 persons were accepted per task and per display. All the candidates were screened for visual acuity (Snellen chart - 20/25), color vision (Ishihara test) and stereoscopic acuity (TNO Stereo test - 60 arcsec) and a total of 102 participants were kept. They included 37 females and 65 males, with a mean age of 25 and median age of 23.

The results of the users that had simulator sickness symptoms that were too strong before the test (scores larger than 1 for any of

the questions) were not considered during the analysis. The final number of valid participants after this screening is given in Table 4, with the original number of participants accepted after the vision tests in brackets.

Table 4. The number of valid participants for each sub-experiment; in brackets, the original number of participants accepted after the vision tests.

	task (1)	task (2)	task (3)	task (4)
3DTV	15 (15)	11 (15)	12 (15)	10 (11)
3D glasses	12 (13)	10 (12)	11 (11)	10 (10)

3. RESULTS AND OBSERVATIONS

The concentrated results of our SSQ study are illustrated in Figure 1, where the differences between the pre-exposure and post-exposure symptoms for both displays (3DTV, 3D glasses) and for each visualization task can be observed. The figure also shows the statistical level of significance of these differences, computed using the Wilcoxon rank sum test.
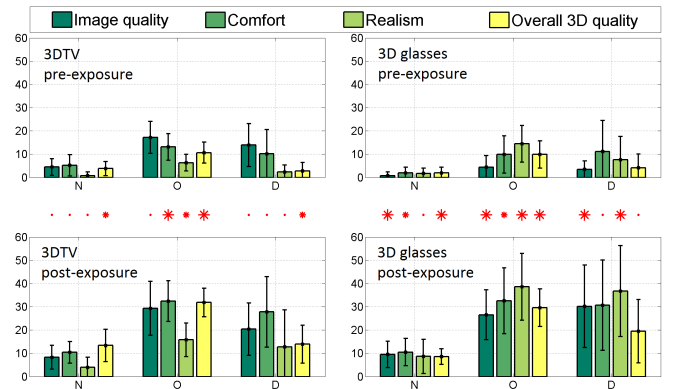
Fig. 1. The N, O, and D cluster symptom scores for the 3DTV screen (left) and the 3D glasses (right), as recorded before (top) and after (bottom) the visualization experiments for each of the four tasks (in different colors). The * symbols indicate the significance levels of the Wilcoxon rank sum test on the differences between corresponding categories before and after the test (1%, 5%, and >5% significance level from the largest to the smallest * symbol).

After the visualization experiments on the 3DTV, no symptom differences were observed for the image quality task. This can be easily justified by the high resolution and large size of the screen, on which subtle differences of blur can be distinguished with no considerable effort. For the comfort task, however, an important augmentation of the oculomotor symptoms was recorded ($p = 0.0034$). This can be explained by the fact that the database contained images with crossed disparities close to the comfort limit of the human visual system. Therefore, when the participants focused on rating this task, they were particularly attentive and sensitive to the fact that the visual exploration was demanding in terms of convergence and accommodation. The realism task led as well to higher post-exposure oculomotor scores ($p = 0.0411$), but the largest human factor changes for the 3DTV experiment were recorded for the overall 3D quality task: statistical differences have been determined between the pre- and post-exposure scores for the nausea ($p = 0.0483$), oculomotor ($p \ll 0.01$), and disorientation ($p = 0.0244$) symptom clusters. This happened because rating the overall 3D quality task meant taking into account at once several factors related to image quality, comfort, and realism. Therefore the visualization effort involved by this

multi-dimensional task can be seen as an implicit cumulative effort of all the other three tasks.

In the case of the tests on the 3D glasses, a larger number of changes than for the 3DTV were observed between pre- and post-immersion. The image quality task generated an increase in all the three symptom clusters: nausea ($p = 0.0096$), oculomotor ($p \ll 0.01$), and disorientation ($p = 0.0024$). In contrast to the results on the 3DTV screen, this shows that on the micro size screens of smaller resolution of the prototype 3D glasses, distinguishing subtleties of image quality can be more demanding. The comfort task on the 3D glasses augmented the oculomotor symptoms ($p = 0.0210$), but less than when it was performed on the 3DTV screen. The result is in accordance with the geometry of the 3D glasses, for which the representation of crossed disparities is less aggressive for the human visual system than the 3DTV configuration considered for our test. The comfort task on the prototype 3D glasses has increased however the nausea symptoms as well, which gives a hint of the fact that an exploration with focus on the different levels of disparities, given the immersion of the 3D glasses, can have consequences on the general state of well being. After the realism task, a considerable deterioration was observed in the scores on the oculomotor ($p = 0.0074$) and disorientation ($p = 0.0094$) symptoms. For explaining this result, we mention that the ratings the participants gave for the realism attribute during the test on the prototype 3D glasses were lower than those during the test on the 3DTV. Therefore, we can conclude that an exploration in order to decide whether a 3D image is similar to what can be seen in reality is more strenuous when this resemblance can be found more difficultly. The overall 3D quality task lead to changes in the nausea ($p = 0.0093$) and oculomotor ($p = 0.0019$) symptom clusters, following the same logic as for the 3DTV. No important changes of disorientation were observed after this task ($p = 0.0579$) on the 3D glasses.

In order to better understand what led to this results, we continued our analysis at the question level. We could observe that there were a series of questions for which only the experiments on the prototype 3D glasses brought changes in the post-exposure symptoms: the image quality task generated more general discomfort, more fatigue, and more difficulty focusing; the comfort task led to augmented scores for the headache; the realism deteriorated the general comfort and the capacity to focus, and augmented the blurred vision and the dizziness with the eyes open. However, on the 3DTV the comfort task increased the fatigue and the overall 3D quality task augmented the eyestrain.

4. CONCLUSION

In the presented study, we considered a consumer situation for which we tested the acceptance of two different stereoscopic displays (from which one was a prototype) for tasks of 3D image visualization from the human factors perspective. We were also interested in better understanding what aspects influence the symptom evolution during the stereoscopic visualization.

The results allowed us to conclude that the task performed during the visualization has an influence on the evolution of the simulator sickness symptoms. Also, we found that different stereoscopic displays generated different symptoms, in function of the content displayed.

The augmentation of certain symptoms in the case of an HMD used as a personal screen is in line with the findings in [7] and this could be an indicator of the fact that the HMDs still leave place for improvements in visualization comfort. In the mean time, it is expected that watching 3D content in a completely immersive en-

vironment like an HMD, but also *e.g.* in a movie theater, would cause more simulator sickness symptoms compared to watching the same content on a TV screen, where the natural ambient vision is still present. For this reason, for typical commercial 3D content like 3D movies, there is a post-processing phase where the 3D depth is adapted to the scene and in general reduced most of the time, while quality artifacts are avoided, in order to prevent phenomena related to visual fatigue. In our particular test case, we did not use such content, but specific still images containing artifacts, which limits the results to a particular case of watching static 3D photographs, that is to just one type of possible consumer utilization.

Moreover, a physiological adaptation to this type of immersive visualization was found in [11], while an adaptation between successive sessions of testing was noted in [6], therefore the lack of familiarity of the consumers with a personal display type of system might be another explanation for our results.

5. REFERENCES

[1] R. Zone, *Stereoscopic Cinema and the Origins of 3-D Film, 1838-1952*, University Press of Kentucky, 2007.

[2] L. Lipton, *Foundations of Stereoscopic Cinema*, Van Nostrand Reinhold Company, 1982.

[3] M. Lambooij, W. IJsselsteijn, and I. Heynderickx, "Visual discomfort in stereoscopic displays: a review," *SPIE, Stereoscopic Displays and Virtual Reality Systems XIV*, vol. 6490, pp. 64900I.1–64900I.11, 2007.

[4] B. Mendiburu, *3D Movie Making Stereoscopic Digital Cinema from Script to Screen*, Focal Press, 2009.

[5] L. Meesters, W. IJsselsteijn, and P. Seuntiens, "A Survey of Perceptual Evaluations and Requirements of Three-Dimensional TV," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 3, pp. 381–391, 2004.

[6] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness," *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993.

[7] P. A. Howarth and P. J. Costello, "The occurrence of virtual simulation sickness symptoms when an HMD was used as a personal viewing system," *Displays*, vol. 18, no. 2, pp. 107–116, Dec. 1997.

[8] J. Häkkinen, T. Vuori, M. Puhakka, B. Postural, and A. Participants, "Postural stability and sickness symptoms after HMD use," in *IEEE International Conference on Systems, Man and Cybernetics*, 2002, pp. 147–152.

[9] J. Häkkinen, J. Takatalo, J. Komulainen, H. Särkelä, J. Havukumpu, and G. Nyman, "Simulator sickness symptoms in virtual display gaming," in *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services*, 2006, pp. 227–230.

[10] J. Häkkinen, M. Polonen, J. Takatalo, and G. Nyman, "Simulator Sickness in Virtual Display Gaming : A Comparison of Stereoscopic and Non-stereoscopic," in *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services*, 2006, pp. 227–229.

[11] S. Jumisko-Pyykko, T. Utriainen, D. Strohmeier, A. Boev, and K. Kunze, "Simulator sickness – Five experiments using autostereoscopic mid-sized or small mobile screens," in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2010, vol. 3, pp. 1–4.

[12] R. Vlad, P. Ladret, and A. Guerin, "Three factors that influence the overall quality of the stereoscopic 3D content: image quality , comfort and realism," in *SPIE, Image Quality and System Performance X*, 2013.

[13] S. Pastoor, "Human Factors of 3D Imaging : Results of Recent Research at Heinrich-Hertz-Institut Berlin," *Proceedings of the International Display Workshop 95*, vol. 3, pp. 66–72, 1995.

[14] "Recommendation, ITU BT.500-13 - Methodology for the subjective assessment of the quality of television pictures," Tech. Rep., ITU, 2012.