Quality Assessment of 3D Visualizations with Vertical Disparity: An ERP Approach

Forooz Shahbazi[1], Sebastian Bosse[1], Guido Nolte[2],
Thomas Wiegand[1,3], *Fellow, IEEE*, and Wojciech Samek[1], *Member, IEEE*

***Abstract*— In an objective approach for the assessment of quality of experience the neural correlates of EEG data are studied when stereoscopic images are presented in three different conditions containing vertical disparity. These conditions are compared to a similar image in 2D both on the channel level by studying the ERP components and on the source level by the localization of the corresponding ERP component. Our findings posit that P1 component in the occipital cortex has significantly increased in amplitude for 3D condition without vertical disparity compared to the 2D condition. According to previous studies, this component increases when depth information are added to the stimulus which is in line with our findings. However the amplitude of this component has significantly decreased for 3D condition with maximum vertical disparity compared to the 3D condition without vertical disparity. We have concluded that the perception of stereoscopic depth by subjects have decreased in this case due to the distortion introduced by vertical disparity. The underlying sources corresponding to P1 component are localized. Except for the power differences, the source locations do not differ for different conditions.**

I. Introduction

In recent years stereoscopic imaging technologies became popular in various fields in order to create or increase the impression of greater realism and immersion: Movie theaters present a remarkable number of movies in 3D today, most modern television sets feature a 3D mode and 3D imaging is used widely for data visualization. With the advent of market-ready head mounted displays such as the Oculus Rift the interest in stereoscopic imaging technologies will further increase as new application scenarios in virtual and augmented reality environments are in reach. However, the success of these applications crucially relies on the systems' quality that users experience. Such quality of experience can be severely impaired by visual discomfort that can be introduced to the viewer, e.g. by accommodation/vergence conflicts, excessive binocular parallax, and other factors [1]. Unfortunately, no reliable objective computational model for visual discomfort or most other aspects of perceived quality is available and, thus, the assessment of the visual quality of imaging systems relies on psychometric tests. The typical procedure in these tests is that a subject has to rank the quality of a set of test stimuli. These kind of tests suffer from different drawbacks, the ratings are highly variable across subjects and, principally, are the results of a conscious process, affected by subjective factors such as bias, expectations and strategies. Practically, these psychometric are difficult do integrate in real-time assessment procedures and, in order to avoid unreliability by tiredness of the subject, should not last longer than 30 minutes [2]. To overcome these limitations of behavioral subjective tests, brain activity recordings allow for directly monitoring the users cognitive state. Electroencephalography (EEG) is one of the cheapest and most mobile devices to record brain signals and was recently shown to be a promising tool for assessing perceived quality of audio and 2D and 3D video signals [3], [4].

If the two focal points in a stereoscopic imaging system are not aligned vertical disparities occur alongside horizontal disparities. Vertical disparities commonly cause visual discomfort. In this paper we address the assessment of visual discomfort caused by vertical disparities using EEG. For that the brain signals recorded for different amounts of vertical disparity are compared on channel level and on source level. We find P1 ERP component to be a neural marker of vertical disparities. We have also localized P1 component in the brain and did not see any difference in source locations for different conditions.

II. Methods

A. Experimental Setup

In this experiment EEG features related to the vertical disparity in stereoscopic images are studied. The vertical disparity is introduced by the simulation of a 3D image of a cube. The right camera is shifted upwards compared to the left one to simulate the vertical disparity. Two conditions are simulated in this way while the amount of vertical disparity in one condition (3D-2) is 40 % less than the other condition (3D-3). The cube is presented to the subjects in four different conditions and two categories of 2D and 3D, i.e. three images in 3D including 2 different 3D images with vertical disparity levels and one 3D image without vertical disparity. The same cube is also presented in 2D shown in Figure 1.

Each image is presented randomly for 120 trials (epochs) and each trial lasts 4 seconds. Between the images a cross is presented in 3D for an interval of 3 seconds. This interval is considered for subjects to rest their eyes and it helps to reduce the amount of ocular artifacts. To keep the subjects attentive they were supposed to press a button when an image of a cat was presented. 120 images of a dog (80%) and

[1] F. Shahbazi Avarvand , S. Bosse and W. Samek are with the Department of Video Coding & Analytics, Fraunhofer Heinrich Hertz Institute, 10587 Berlin, Germany. E-Mail: `forooz.shahbazi@hhi.fraunhofer.de`, `wojciech.samek@hhi.fraunhofer.de`

[2] G. Nolte is with the Department of Neurophysiology and Pathophysiology University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany.

[3] T. Wiegand is with Fraunhofer Heinrich Hertz Institute, 10587 Berlin, Germany and with the Image Communication Laboratory, Berlin Institute of Technology, 10587 Berlin, Germany.

Fig. 1. *Left*: The cube is presented in 2D. *Right*: The cube is presented in three stereoscopic conditions (3D, 3D-2 and 3D-3).

a cat (20%) are presented between two fixation crosses. If the subject hit the target image (cat) by a minimum 90% accuracy he/she was rewarded by 5 € extra. All subjects were rewarded for their participation by 8.5 € per hour. The distance between the subject and the screen is 280 cm and the horizontal angle of view is 20.76 degrees. The subjects wear 3D polarized glasses during the experiment. The reason to prefer the polarized passive glasses over the active shutter glasses is the results in the previous research [5] showing that the passive glasses provide more comfortable visualization for the subjects. EEG signals were recorded in a dimly lit and silent room. The 3D screen is JVC 3D LCD Monitor (model number: GD-463D10E). EEG signals are measured by a cap of 64 active electrodes (Fp1,2, F1 to F8, FC2 to FC6, T7 and T8, C1 to C6, Cz, Tp7, Tp9 and Tp10, Pz, P1 to P8, PO3 and PO4, POz, PO7 to PO10, O1 and O2, Oz, AF3, AF4, AF7, AF8, FT7 to FT10, FC3 and FC4, CP3 and CP4, CPz, VEOG) , i.e. actiCap from Brain Products GmbH and the impedance of electrodes was kept below 10 KΩ. 21 subjects have been recorded out of which 4 data sets had a very low signal-to-noise ratio with high number of rejected trials which were therefore excluded from the analysis. The data from 17 subjects, 6 of which are male and 11 are female with the average age of 25.83 is analyzed. All subjects have normal or corrected to normal vision and are tested for their 3D vision and gave informed consent. We have received a permission for the experiment in accordance with the declaration of Helsinki from the IRB of Technische Universität Berlin (TU Berlin).

B. *Pre-processing of EEG Data*

EEG data is low-pass filtered at 30 Hz besides the filter applied by the amplifier hardware at 0.016 Hz. During the measurement FCz is selected as the original reference electrode and the data is re-referenced to the common average during the analysis. Muscle artifacts are removed from the data by removing the trials with the variance larger than a threshold. The baseline in the time interval between -200 ms, i.e. 200 ms before the stimulus onset and the stimulus onset is subtracted from each epoch of the data. Ocular artifacts are removed by regression. The applied algorithm projects out part of the data which is correlated with EOG electrodes. A short measurement was conducted before the experiment in which the subject was supposed to blink when a circle appeared on the screen. The vertical ocular component was measured by an electrode underneath the right eye and Fp2. The horizontal ocular activity of the subjects was also recorded in a similar measurement in which the subject was supposed to follow a circle on the screen which moved from the right end to the left end of the screen and vice versa. The difference between the F7 and F8 are measured as the horizontal component. Part of the EEG data which is correlated with these two components is then projected out from the data as it is described in [6]. In a further step, epochs within which the difference of maximum and minimum amplitude exceeds 70 μV are rejected.

C. *Event Related Potential Analysis*

Part of the responses in EEG signals caused by the above mentioned stimuli are phased-locked with the stimuli with a very low signal to noise ratio. These responses are called Event Related Potentials (ERPs). To increase the signal-to-noise ratios, EEG data is averaged over all trials to cancel out all non-phase-locked activities. The time window of this average is selected between 200 ms before the stimulus onset and 900 ms after that. The BBCI toolbox, which is a Matlab based toolbox [7], is used for the ERP analysis.

The ERP components might vary between conditions both in their amplitude and latency of the peak. While each component is the result of an underlying neural reaction in the brain, different conditions are ideally differentiable from each other based on their different ERP components. In the following analysis we have studied ERP components in order to extract condition dependent features.

D. *Source Localization*

The ERP components of different conditions are not only of interest on the sensor level but also on the source level. Due to the artifacts of volume conduction i.e., mixing of the underlying brain activities on the channel level and contribution of sources which are far from a channel in the activity of that channel, looking at the topographies on the scalp does not necessarily reveal the exact and accurate locations of the underlying sources. ERP signals corresponding to each condition in the time point of the ERP component's peak are localized by applying an inverse method named eLORETA [8]. However we would like to point out the volume conduction artifacts do not necessarily disappear in the process of inverse modeling [9] and have to be considered even on the source level. Although some inverse methods such as SC-MUSIC [10] are based on the imaginary part of cross-spectrum [11] which makes them robust to the artifacts of volume conduction but since we are interested in the underlying ERPs sources i.e., time domain instead of frequency domain, these methods were not beneficial here. As a head model we used a standardized MNI head consisting of 152 averaged brains [12]. The MNI brain is divided into a continuous grid with 2113 voxels. eLORETA filters are estimated from the forward models and applied to the time point of interest. The power at each voxel is then estimated as the sum of the powers of the voxel in all three dipole directions.

III. Results

A. ERP Analysis on the Channel Level

EEG signals are averaged over all trials for single subjects in the interval between -200 ms and 900 ms after the baseline correction is applied. The ERPs on channel O2 plotted in Figure 2 (Top panel) for single subjects appeared to be consistent enough across subjects to be averaged over all subjects. The grand averaged ERP on channel O2 is shown in Figure 2 in the bottom panel. In Figure 3, ERPs corresponding to three conditions (2D, 3D and 3D-3) are plotted and the differences between 2D and 3D as well as 3D and 3D-3 are calculated. Looking at the topographies of the ERPs corresponding to the large ERP components, it was obvious that the brain activity in channel O2 is larger compared to other channels in almost all the intervals of interest. Therefore in the following analysis we focus on this single channel. The scalp topographies of two intervals corresponding to the larger values of signed squared biserial correlation coefficient are visualized for all three conditions as well as for the differences. For all three conditions there are two large ERP components one at around 130 ms and the second one at around 300 ms after the stimulus onset. However there is another negative peak which is clearer in 2D and less obvious in the other two conditions at around 150 ms. We did not focus on the ERP component at 300 ms because we believe this component is caused by the implicit odd ball paradigm in the stimuli, i.e. all conditions are stereoscopic except for 2D and this explains why 2D has the largest amplitude at 300 ms component.

The topography corresponding to the difference between 3D and 2D shows a large difference in the occipital area with an increase for the 3D condition. This result is in line with the previous studies [13], [14] which show that depth perception in 3D images increases the P100 [15] amplitude compared to 2D images. However the stimuli used in these studies were non-stereoscopic stimuli. Here we have shown that the same results are produced in the case of stereoscopic images for P1 component. However the comparison between 3D-3 and 3D shows a large activity in the occipital area but with a decrease in the amplitude of 3D-3 condition. Student's t-test is applied to the differences which shows all the above mentioned differences are significant. The difference between the components in 3D and 3D-2 was not significant and therefore is not presented here. Given that an increase in the depth perception causes an increase in the amplitudes of P1, we believe that the decrease in the amplitude of this component for 3D-3 compared to 3D is a sign of a decrease in the perception of depth by subjects. This hypothesis explains the similarity between the ERP components considering both the amplitude and latency of 3D-3 and 2D. Further analysis is performed on the same data set for N1 ERP component in [16].

B. Source Localization of P1 Component

As described in Section II-D, the underlying brain activities corresponding to the earliest ERP component is localized

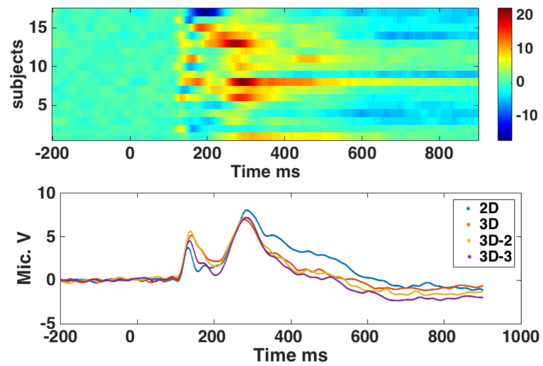

Fig. 2. *Top*: ERPs averaged over all epochs for each subject on channel O2 and condition 2D are plotted in the time interval between -200 ms and 900 ms. The ERP components across subjects are consistent enough to be averaged over all subjects. The same holds for all other 3 conditions. *Bottom*: ERP of all conditions averaged over all subjects in a single channel O2 are plotted in the same time interval as above.

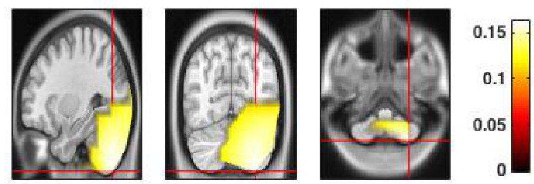

Fig. 4. Underlying sources modeled as dipoles in each grid point localized using eLORETA method for 2D condition. The source locations corresponding to other conditions are identical to 2D with the difference in their amplitudes corresponding to the differences we already saw in the ERP amplitudes.

and compared between conditions. The first ERP component in Figure 3 corresponding to P100 is localized for the time point corresponding to the maximum amplitude in O2 for each condition. Figure 4 shows the corresponding sources for condition 2D localized by eLORETA. Having localized the sources in other conditions, the source activities appear to be at the same locations for all conditions in the occipital cortex. These locations are consistent with previous studies [17] in which the authors refer to the P1 component in this interval as the late P1 and report the source locations in ventral extrastriate cortex of the fusiform gyrus.

IV. Conclusion

In this study the simulated vertical disparity in the 3D image resulted in a decrease in the amplitude of P1 ERP component compared to those components in the 3D condition without vertical disparity. In previous studies it have been shown that the amplitude in P1 increases for the stimulus containing depth information compared to a 2D image. However the 3D stimuli in those studies were all non stereoscopic images while we have shown the same results in the stereoscopic images. On the other hand the amplitude of P1 component has significantly decreased for 3D-3 compared to 3D. The comparison between 3D-2 and 3D did not show significant differences for these components. Considering that P1 amplitude increases by an increase in the perception of depth, we conclude that the significant decrease in 3D-3

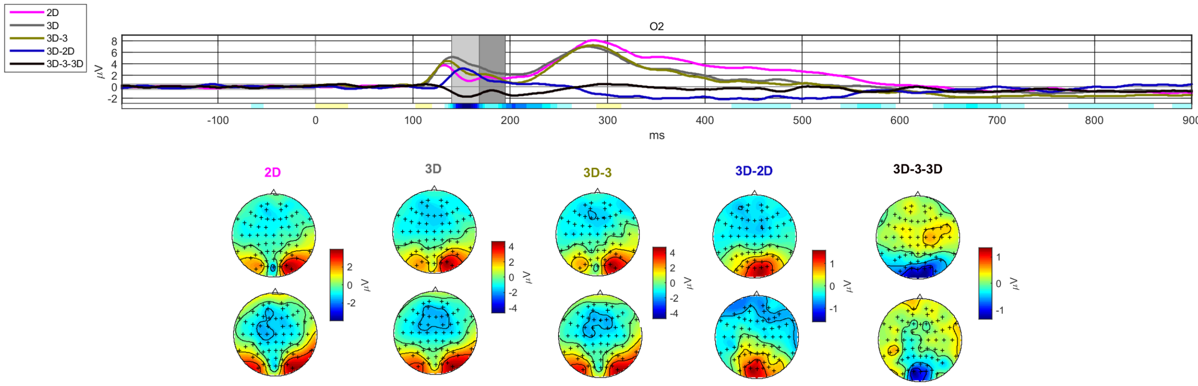

Fig. 3. In the first row the ERPs corresponding to three conditions and the differences between 3D and 3D-3 plus the difference between 2D and 3D in the time interval of -200 ms and 900 ms are plotted. In the next two rows (corresponding to the first and the second selected intervals respectively) the corresponding topographies are shown.

is a sign of a decrease in the perception of depth by subjects. This means that the effect of the vertical disparity in a 3D image reduces the stereoscopic feeling for subjects although the image still includes depth information. Furthermore, we have shown P1 is an important feature which one could extract from EEG data for detection of vertical disparity when it is high enough to affect the depth perception in the subjects. As we mentioned before 3D-2 did not result in significant differences compared to 3D which is potentially because the vertical disparity was not strong enough to affect the depth perception.

Next to the channel level, we also analyzed the ERP components on the source level. We were interested to know if there are any differences between the source locations corresponding to different conditions. We realized the underlying sources corresponding to the P1 component did not differ between conditions. The location of these sources are in ventral extrastriate cortex of the fusiform gyrus which is in line to the previous study [17].

In future work we will investigate multi-modal [18] and deep [19] approaches for 3D quality assessment.

V. Acknowledgment

This research was partially funded by the German Research Foundation (DFG, SFB938/Z3 and TRR-169/B4).

References

[1] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review," *J. Imaging Sci. Technol.*, vol. 53, no. 3, pp. 030201, 2009.

[2] ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," Tech. Rep., International Telecommunication Union, Geneva, Switzerland, 2012.

[3] S. Bosse, K.-R. Müller, T. Wiegand, and W. Samek, "Brain-Computer Interfacing for Multimedia Quality Assessment," in *IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC)*, 2016, pp. 002834–002839.

[4] U. Engelke, D. P. Darcy, G. H. Mulliken, S. Bosse, M. G. Martini, S. Arndt, J.-N. Antons, K. Y. Chan, N. Ramzan, and K. Brunnstrom, "Psychophysiology-Based QoE Assessment: A Survey," *IEEE Journal of Selected Topics in Signal Processing*, pp. 1–1, 2016.

[5] A. S. Malik, R. Khairuddin, H. U Amin, M. L. Smith, N. Kamel, J. M. Abdullah, S. M. Fawzy, and S. Shim, "Eeg based evaluation of stereoscopic 3d displays for viewer discomfort," *BioMedical Engineering OnLine*, vol. 14, no. 1, pp. 21, 2015.

[6] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, "Recipes for the linear analysis of EEG," *NeuroImage*, vol. 28, no. 2, pp. 326 – 341, 2005.

[7] B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. E. Ramsey, I. Sturm, G. Curio, and K.-R. Müller, "The berlin brain-computer interface: Non-medical uses of bci technology," *Frontiers in Neuroscience*, vol. 4, no. 198, 2010.

[8] R. D. Pascual-Marqui, "Discrete, 3D distributed, linear imaging methods of electric neuronal activity. part 1: exact, zero error localization," *arXiv:0710.3341*, 2007.

[9] F. Shahbazi, A. Ewald, A. Ziehe, and G. Nolte, "Constructing surrogate data to control for artifacts of volume conduction for functional connectivity measures," in *17th Int. Conf. on Biomagnetism Advances in Biomagnetism - Biomag 2010*. 2010, pp. 207–210, Springer.

[10] F. Shahbazi, A. Ewald, and G. Nolte, "Self-consistent music: An approach to the localization of true brain interactions from eeg/meg data," *NeuroImage*, vol. 112, pp. 299 – 309, 2015.

[11] G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett, "Identifying true brain interaction from eeg data using the imaginary part of coherency," *Clin. Neurophysiol.*, vol. 115, pp. 2292–2307, 2004.

[12] V. S. Fonov, A. C. Evans, R. C. McKinstry, C. R. Almli, and D. L. Collins, "Unbiased nonlinear average age-appropriate brain templates from birth to adulthood," *NeuroImage*, vol. 47, pp. S39–S41, 2010.

[13] S. Omoto, Y. Kuroiwa, S. Otsuka, Y. Baba, C. W. Wang, M. Li, N. Mizuki, N. Ueda, S. Koyano, and Y. Suzuki, "P1 and P2 components of human visual evoked potentials are modulated by depth perception of 3-dimensional images," *Clinical Neurophysiology*, vol. 121, no. 3, pp. 386 – 391, 2010.

[14] E. Hayashi, Y. Kuroiwa, S. Omoto, T. Kamitani, M. Li, and S. Koyano, "Visual evoked potential changes related to illusory perception in normal human subjects," *Neuroscience Letters*, vol. 359, no. 12, pp. 29 – 32, 2004.

[15] A. Séverac Cauquil, Y. Trotter, and M. J. Taylor, "At what stage of neural processing do perspective depth cues make a difference?," *Experimental Brain Research*, vol. 170, no. 4, pp. 457, 2005.

[16] F. Shahbazi Avarvand, S. Bosse, K-R. Müller, R. Schäfer, G. Nolte, T. Wiegand, and W. Samek, "Objective quality assessment of stereoscopic images with vertical disparity using EEG," *Journal of Neural Engineering*, 2017.

[17] F. Di Russo, A. Martínez, M. I. Sereno, S. Pitzalis, and S. A. Hillyard, "Cortical sources of the early components of the visual evoked potential," *Human brain mapping*, vol. 15, no. 2, pp. 95–111, 2002.

[18] S. Dähne, F. Bießmann, W. Samek, S. Haufe, D. Goltz, C. Gundlach, A. Villringer, S. Fazli, and K. R. Müller, "Multivariate machine learning methods for fusing multimodal functional neuroimaging data," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1507–1530, 2015.

[19] I. Sturm, S. Lapuschkin, W. Samek, and K.-R. Müller, "Interpretable deep neural networks for single-trial eeg classification," *Journal of Neuroscience Methods*, vol. 274, pp. 141–145, 2016.