

Objective Quality Assessment of Stereoscopic Images with Vertical Disparity using EEG

Forooz Shahbazi Avarvand¹, Sebastian Bosse¹, Klaus-Robert Müller^{2,5}, Ralf Schäfer¹, Guido Nolte³, Thomas Wiegand^{1,6}, Gabriel Curio⁴, Wojciech Samek¹

¹ Department of Video Coding & Analytics, Fraunhofer Heinrich Hertz Institute, Berlin, Germany

² Machine Learning Group, Technische Universität Berlin, Berlin, Germany

³ Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

⁴ Neurophysics Group, Department of Neurology and Clinical Neurophysiology, Charité, Berlin, Germany

⁵ Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea

⁶ Image Communication Group, Technische Universität Berlin, Berlin, Germany

E-mail: gabriel.curio@charite.de, klaus-robert.mueller@tu-berlin.de, thomas.wiegand@hhi.fraunhofer.de and wojciech.samek@hhi.fraunhofer.de

Abstract. Objective: Neurophysiological correlates of vertical disparity in 3D images are studied in an objective approach using EEG technique. These disparities are known to negatively affect the quality of experience and to cause visual discomfort in stereoscopic visualizations.

Approach: We have presented 4 conditions to subjects: one in 2D and three conditions in 3D, one without vertical disparity and two with different vertical disparity levels. Event Related Potentials (ERPs) are measured for each condition and the differences between ERP components are studied. Analysis is also performed on the induced potentials in the time frequency domain.

Main results: Results show that there is a significant increase in the amplitude of P1 components in 3D conditions in comparison to 2D. These results are consistent with previous studies which have shown that P1 amplitude increases due to the depth perception in 3D compared to 2D. However the amplitude is significantly smaller for maximum vertical disparity (3D-3) in comparison to 3D with no vertical disparity. Our results therefore suggest that the vertical disparity in 3D-3 condition decreases the perception of depth compared to other 3D conditions and the amplitude of P1 component can be used as a discriminative feature.

Significance: The results show that the P1 component increases in amplitude due to the depth perception not only in non-stereoscopic images but also in the three dimensional stimuli. On the other hand the vertical disparity in the stereoscopic images is studied here and we suggest that the amplitude of P1 component is modulated with this parameter and decreases due to the decrease in the perception of depth.

1. Introduction

In order to act in their natural environment, it is crucial for humans to perceive and interpret surfaces of objects organized in three-dimensional space. Visual cues that allow for the perception of depth can be categorized in monocular and binocular cues [1]. Examples of binocular cues are convergence and stereopsis and are possible due to the horizontal displacement of the two eyes. Convergence is a binocular oculomotor cue resulting from the inward movement of the eyeball when the eyes focus an object at a certain depth. Stereopsis (or retinal disparity) allows for the most compelling experience of depth; the projection of an object has a slightly different position on each of the retina of the two eyes. This retinal disparity allows the human brain estimate the depth of the object by triangulation [1].

Stereopsis has been used for at least two centuries for evoking the illusion of depth [2, 3] and its technical applications comprise entertainment (e.g. 3D movies), medical engineering (e.g. digital stereo microscopes) and visualizations in virtual or augmented reality [4]. It has been shown that stereoscopic depth leads to a more engaging viewing experience for the observers [4, 5]. In all of these applications, quality of experience (QoE) is an important factor for the acceptance, if not even for the feasibility, of the technology and the perceptual quality is an important aspect of QoE [6]. A factor that can act as a source of quality degradation can be categorized by its position in the line of signal capturing, signal transmission and signal presentation. During the capturing, the camera system setup and the size and intensity of temporal convergence changes affect the perceived quality. Coding and transmission of the 3D visual signal introduce compression artifacts, transmission errors and potentially frame synchronization errors. The viewing environment influences the QoE via the used displays, the viewing distance and angle, crosstalk, and ambient light [7].

Some of the perceptual artifacts arising in this pipeline are also observable in a 2D video transmission system, e.g. reduced visual quality by compression or annoyance due to buffer time given by the throughput of the channel. Other artifacts are 3D exclusive, such as the puppet-theater effect or the cardboard effect [7]. Next to these degradations in perceptual quality affecting psychological states of the observers, there is, particular to 3D visualizations, the notion of visual discomfort that affects the (neuro-)physiological state as well and occurs with visual fatigue, eyestrain or even nausea [8]. Factors leading to visual discomfort include conflicts between vergence and accommodation, exceeding binocular parallax and errors in dichoptic presentation. One source of the latter, potentially occurring due to misaligned cameras, are vertical disparities, as these do not occur in natural scenes observed by horizontally separated eyes. However, the definitions of visual discomfort and visual fatigue are still blurry and thus a precise relation is still an unsatisfactorily answered question [8]. While visual discomfort can be perceived within an instant, the sense of visual fatigue arises after a certain duration of visual stimulation [8, 9].

The assessment of perceived 3D video quality is current subject to research and ITU-

T SG9 in cooperation with the video quality experts group (VQEG) is making progress in the development of psychophysical assessment methods. The standardization efforts recognize the different natures of quality of experience with regard to stereoscopy and thus follow two lines: [10] drafts the recommendation for the subjective assessment methods of 3D video quality. In principle, those methods are closely related to those developed for the assessment of 2D visual quality [11, 12] and address similar types of degradations. The quality factors relevant to stereoscopy, such as visual fatigue or discomfort, are addressed in [13]. Typically, the results of psychophysical quality tests are reported in terms of mean opinion scores (MOS) that are calculated as averaged individual ratings of the test participants presented a certain stimulus. Psychophysical tests have some well known limitations, as they require a sufficient large number of participants to obtain statistical significance, also the judgment of participants might be biased by several factors [14]. Furthermore, in the case of 3D quality and in order to assess visual fatigue, presentation sessions have to be of longer duration as the effect might not be consciously perceivable instantaneously.

This led to a parallel line of research on perceived quality assessment investigating psychophysiological approaches. In these studies, electroencephalography (EEG) is used to assess the perceived quality of experience of several types of multimedia signals, such as speech signals [15, 16, 17], 2D images [18, 19, 20, 21, 22, 23] and 2D videos [24, 25, 26, 27, 28]. In [29] EEG-based assessment of multi-modal quality of experience is evaluated. The correlation between the perceived quality of 3D videos subject to compression quantified in terms of MOS and the power in different neural oscillatory bands is studied in [30]. In [31, 32] the EEG-based assessment of the accommodation-vergence conflict is addressed within a framework using event related potentials (ERP). The effect of different shutter frequencies of shutter glasses on the neural workload is analyzed in [33] by using EEG. For a more detailed overview on neurophysiological approaches to multimedia quality assessment see [34, 35].

In the present study the neurophysiological correlates of vertical disparity in 3D images are studied using EEG. Vertical disparity causes visual discomfort for the subjects which can be crucial specifically in medical applications of 3D technology. The goal is therefore to find discriminative features of brain signals [36, 37] correlated with the vertical disparity. This features can be potentially applied in the prediction of visual discomfort before causing visual fatigue or medical errors caused by visual discomfort. For this purpose we have conducted an experiment in which four different conditions are compared including a 2D image as a ground truth and three 3D figures of the same image. In two 3D images two levels of vertical disparity are introduced, i.e. 3D-2 with minimum vertical disparity and 3D-3 with maximum vertical disparity and in one 3D image, i.e. 3D, there is no vertical disparity. In Section 2, more details of the experimental set up and the analysis methods are described. The results based on the ERP analysis as well as time frequency analysis of the induced potentials are presented in Section 3. Our results show the P1 component which increases in amplitude for 3D compared to 2D condition. However, in 3D-3 this component has significantly smaller

amplitude compared to 3D and 3D-2. This phenomenon is explained by the reduction of depth perception in 3D-3. Large vertical disparity has caused the subjects to have a lower perception of the 3D effect. Another significant difference occurs at N1 component due to an implicit discrimination which is going to be discussed in detail. The summary and discussion follow in Section 4.

2. Methods

The purpose of this study is to develop an objective procedure in the evaluation of quality of experience when the parameter of interest is the vertical disparity. We would like to extract features in brain signals recorded by EEG electrodes which correspond to the change of vertical disparity. For this purpose a simple image of a cube is presented to the subjects in four different conditions. In one condition the image is presented in 2D as a ground truth condition for our comparisons. The second condition is the same image but in 3D without any vertical disparity and the other two conditions are 3D images with two different levels of vertical disparity. The experimental setup and methods of the analysis are described below.

2.1. Experimental Setup

A cube is presented to the subjects in four different conditions and two categories of 2D and 3D. Three images are in 3D and the vertical disparity is the parameter which is varying between them. One image is the reference image with no vertical disparity in 2D (2D) which is shown in Figure 1(a).

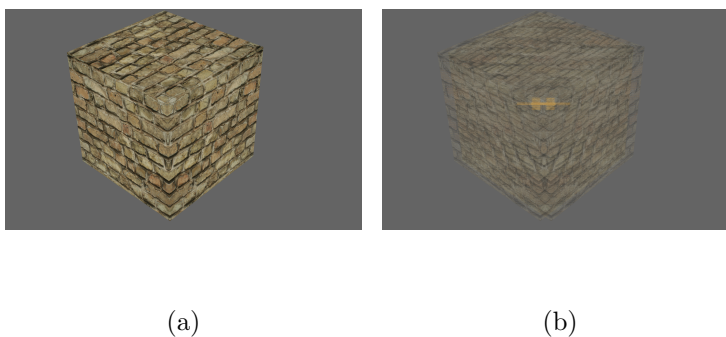


Figure 1. a) The cube is presented in 3D with three vertical disparities (3D, 3D-2 and 3D-3) as well as in 2D. b) A three dimensional cross is presented in between the epochs as a pause interval for subjects to relax their eyes. The cross is projected on the cube in this figure to show its exact location relative to the cube.

The second image is in 3D without a vertical disparity (3D) and the third image has a medium vertical disparity (3D-2). The vertical disparity is introduced in a simulation by moving the right camera upwards. The fourth image has the maximum vertical disparity (3D-3). 3D-2 has 0.4 vertical disparity compared to the maximum amount

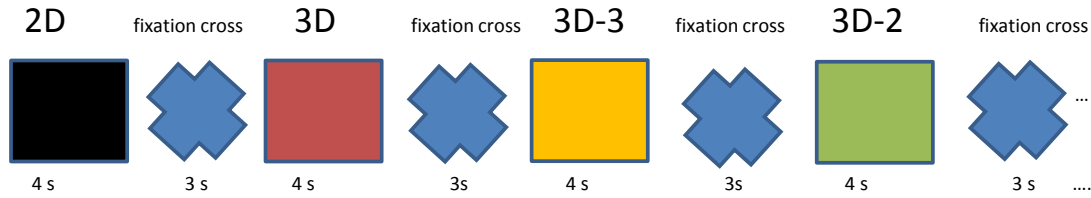


Figure 2. Visualization of each trial in the experimental setup. The order of images are randomized in 120 trials.

in 3D-3. Each image is presented randomly for 120 times (epochs) and each time for 4 seconds. Between the images a cross is presented in 3D for 3 seconds. This time period is considered for subjects to rest their eyes and it helps to reduce the amount of ocular artifacts and therefore during this period subjects are allowed to move or blink.

In Figure 1(b), the cross is projected on the cube for the visualization of the exact position of the cross relative to the cube. A diagram in Figure 2 is a visualization of four epochs including all the images and the pauses in between. In order to keep the subjects attentive they were supposed to press a button when an image of a cat was presented. For that purpose, they are presented by 120 images of a dog (80%) and a cat (20%). Each of these images was presented between two fixation crosses. If the subject hit the target image (cat) by minimum 90% accuracy he/she was rewarded at the end 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. 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 K Ω . 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).

2.2. Pre-processing of EEG Data

During the analysis, data is low-pass filtered at 30 Hz and the amplifier hardware-setup applied a high pass filter at 0.016 Hz to the data. FCz is selected as the original reference electrode during the measurement and the data is re-referenced to the common average. Muscle artifacts are removed from the data by removing the trials with the variance larger than a threshold. The threshold is estimated as the 90th percentile plus the difference between the 10th and 90th percentile multiplied by 10. The baseline in the time interval between -200 ms, i.e. 200 ms before the stimulus onset, and 0 ms is subtracted from each epoch of the data individually. Ocular artifacts are removed by regression. The algorithm projects out part of the data which is correlated with EOG electrodes. For this purpose, a short measurement was conducted in which the subject was supposed to blink upon the appearance of a circle on the screen. An electrode was installed below the right eye of the subject. The horizontal ocular activity of the subjects was also recorded in a 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 EOG electrode under the right eye and Fp2 is calculated as the vertical component and the difference between F7 and F8 as the horizontal component. Part of the EEG data which was correlated with these two components is then projected out from the data as it is described in [38]. In a further step, epochs within which the difference of maximum and minimum amplitude exceeds 70 μ V are rejected.

2.3. Event Related Potential Analysis

When the above mentioned stimuli are presented to the subjects, responses are caused in the brain which are measured by EEG electrodes. Part of these responses which are phased-locked with the stimuli with a very low signal-to-noise ratio 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, i.e. an average over 120 trials of each condition for each subject. The time window of this average is selected between 200 ms before the stimulus onset and 700 ms after that. The BBCI toolbox, which is a Matlab based toolbox [39], is used for the ERP analysis.

2.4. Statistical Test

To test the significance of the results two statistical tests are applied on the data: 1) Student's t-test 2) Bootstrap re-sampling test. The first statistical test is the method which is applied here to test the significance of the amplitude differences in the ERP analysis. In this approach the amplitude differences are measured for each condition within single-participant average. A two sample Student's t-test is applied on two ERPs belonging to each condition averaged over the epochs of each single subject to verify if the means of two populations are different. Alpha value is set to 0.01. Student's t-test

is applied to the data and the time intervals with insignificant differences are set to zero. Bonferroni correction for the problem of multiple comparisons is applied.

In the second approach, we deal with the significance of the time frequency analysis. In this approach we have verified the significance of the differences between the absolute values of the wavelet transformations for each two conditions. In this approach we have 17 subjects and two conditions for each subject which makes 34 data sets in total. In a permutation test we mix up these data sets and randomly choose two sets of data each containing 17 data sets and label these two sets with two conditions. Note that the data sets in each group do not necessarily have the right label of that condition. In the next step we estimate the grand average of the absolute values of wavelet transformations corresponding to each group by averaging over all 17 data sets and estimate the differences between the newly estimated grand averages. This new estimation is expected to be smaller than the original differences. The same procedure is repeated for N times and finally the p-values are estimated.

2.5. Time Frequency Analysis

A wavelet transformation is applied to measure the time-frequency behavior of signals in four conditions. The wavelet coefficients are estimated by the convolution of ERPs by a Morlet wavelet.

$$X(a, b) = \frac{1}{\sqrt{a}} \int \Phi\left(\frac{t-b}{a}\right) x(t) dt \quad (1)$$

where b is the shifting factor and a is a scaling factor which is a function of frequency and Φ is the Morlet function.

3. Results

3.1. Vertical and Horizontal Ocular Activity

Before the main experiment started, a short test was conducted in order to record the eye blinks and horizontal ocular activities of the subjects. The method mentioned in [38] is applied to the data to project out the components corresponding to the ocular activities. These components are estimated from each epoch and each condition and the mean and the standard deviation of both components averaged over all epochs and all subjects are plotted in Figure 3. In the standard deviation of vertical component as well as the mean there is a rise in the amplitude which appears gradually at around 270 ms after the stimulus onset. The peak has the largest amplitude for the standard deviation for 3D-3 while other conditions have similar amplitude levels. 3D-2 has the second largest amplitude and finally 3D and 2D are the smallest ones respectively.

The peak in the horizontal component is not as sharp as the vertical component and seems to happen again at around 400 ms. Variance between conditions seems to be smaller for the horizontal activity compared to the vertical component. We would like

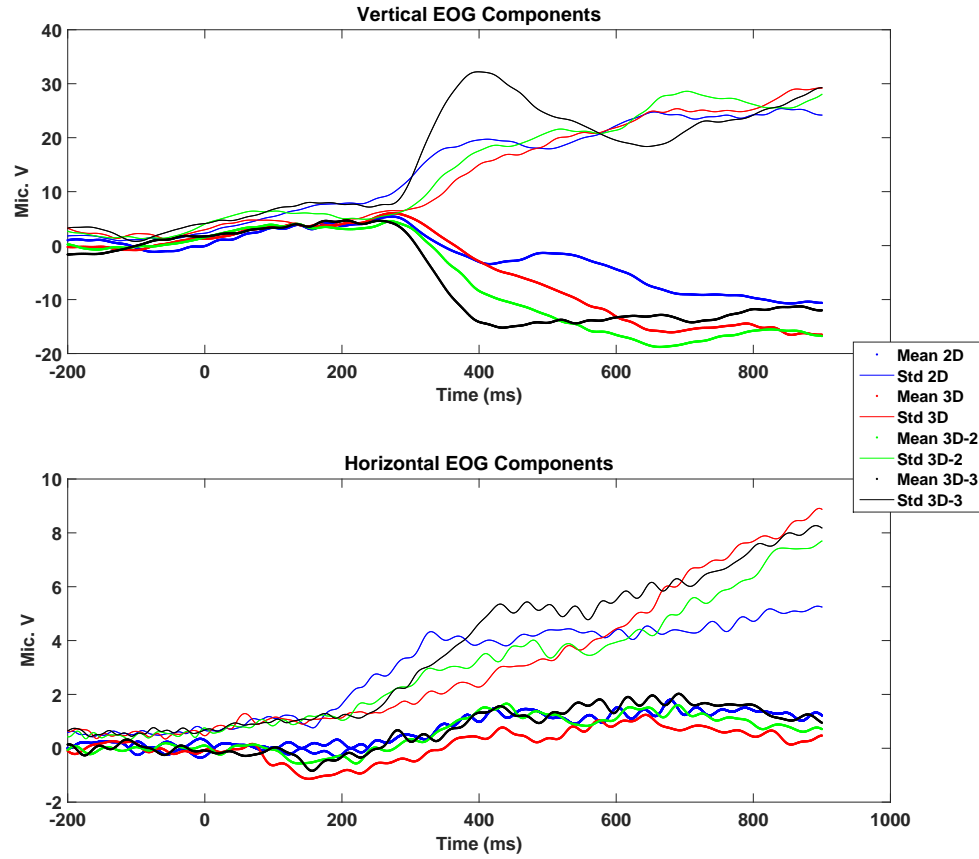


Figure 3. Grand average of the vertical EOG (on the top) and horizontal EOG (on the bottom) are plotted in this figure for all conditions. Zero point in time represents the stimulus onset. The vertical activity is calculated from the difference between the EOG electrode underneath the right eye and Fp2. The horizontal component is calculated as the difference between F7 and F8.

to point out that both of these components have been removed from the data using the regression algorithm.

These two plots however convey an important message for the further analysis, i.e. the brain activities before 200 ms are less probable to be affected by ocular activities as these artifacts start to increase at around 270 ms for the vertical activity and 200 ms for the horizontal activity considering that this component is not large compared to the vertical one. Therefore, we can assure that the ERP components below 200 ms are not caused or strongly affected by the phase locked blinking of the subjects as a reaction to the stimulus e.g., due to the strong vertical disparity effect or appearance of 3D images.

3.2. Event Related Potentials (ERPs)

ERP components are estimated by averaging EEG signals over all epochs belonging to a single condition over all subjects. The average over all subjects can be meaningful only

if we make sure that the subjects have consistent topographies. We have evaluated this by visualizing the topographies of single subjects and single conditions in time windows of 50 ms and found it reasonable to study the over all subjects average.

Figure 4(a) shows four curves corresponding to the ERPs of all four conditions averaged over all subjects in channel O2. All epochs are averaged in the interval between -200 ms and 900 ms. The time courses in time points larger than 900 ms converged for all conditions. The baseline in all conditions in time interval of -200 ms and 0 ms are in the same level as the result of baseline correction and therefore a comparison between ERP components is reasonable. Figure 4(b) shows the same results for Oz. In Figure 4(c), the standard errors estimated for ERPs of different conditions across all subjects are shown. The standard errors appear to be in a reasonable range which shows consistency between the ERPs of single subjects.

In Figure 4(a) there are two very large and positive peaks, one at around 130 ms and the other one at around 300 ms and a smaller negative-going peak at around 150 ms. The same holds for Oz however the negative-going peak has a negative and much larger amplitude in Oz. In the first positive peak corresponding to the P1 component, which appears before the rising point of the ocular activity at 200 ms, the 2D curve corresponds to the condition with the lowest amplitude compared to all 3D conditions as well as shortest latency. 3D-3 has a larger amplitude than 2D followed by 3D and 3D-2 has the largest amplitude. The next peak is the negative-going peak at around 150 ms. This peak which corresponds to N1 component is more obvious in 2D and 3D-3 for O2. 2D has the shortest latency as well as the smallest amplitude. Then 3D-3 appears followed by 3D and 3D-2.

The next largest peak occurs at around 300 ms after the stimuli. In this ERP component 2D condition has significantly larger amplitude compared to other 3D conditions. This significant difference is probably caused by the implicit oddball characteristic of the experiment. 2D is a condition with 25% probability of appearance and 3D conditions (all three conditions together) with 75% probability. It is known that P300 has a larger amplitude when the stimulus is less probable [40] which explains why P3 has significantly larger amplitude in 2D compared to all other conditions. Therefore this difference does not reflect any direct physiological differences caused by the conditions.

The reason we chose these channels becomes clearer in Figure 5 when we look at the scalp topographies. In this image two curves corresponding to the grand average of 2D and 3D in channel O2 as well as their differences are depicted. The intervals in which the signed squared biserial correlation coefficient have positive peak are heat mapped in red and the negative peaks in blue in the horizontal color bar beneath the ERP curves. Three time intervals with large squared biserial correlation coefficients have been chosen to be mapped on the scalp. As we see the brain activity in the O2 region is large in all three time intervals which was the motivation for choosing this channel for the following analysis. The Comparison between 3D and 2D conditions shows an increase in the occipital activity for the 3D condition in the first interval. This interval corresponds to

the interval in which the difference curve in blue is maximum. This amplitude increase has proved to be statistically significant in Figure 5(b) using Student's t-test. In this figure, all the time points which correspond to insignificant differences are set to zero and the ERP difference is plotted in significant intervals. The first significant interval starts already at around P1 peak but reaches its maximum at N1.

Further comparisons are performed between 2D and 3D-2 conditions in Figure 6 which confirm the significant increase in P1 for 3D-2 and a larger negativity N1 amplitude for 2D. In Figure 7, 3D-3 is compared to 2D. There we see a significant decrease in the negativity of N1 however the negativity decrease is detected to be smaller than other 3D conditions. In this figure, the interval corresponding to significant differences is much smaller and does not cover the P1 component anymore unlike previous comparisons. 3D conditions are also compared to each other. The comparison between 3D and 3D-2 did not show any significant differences in the amplitudes however the comparison between 3D-3 and 3D in Figure 8 shows a significant decrease in the amplitude of P1 and increase in the negativity of N1 in 3D-3.

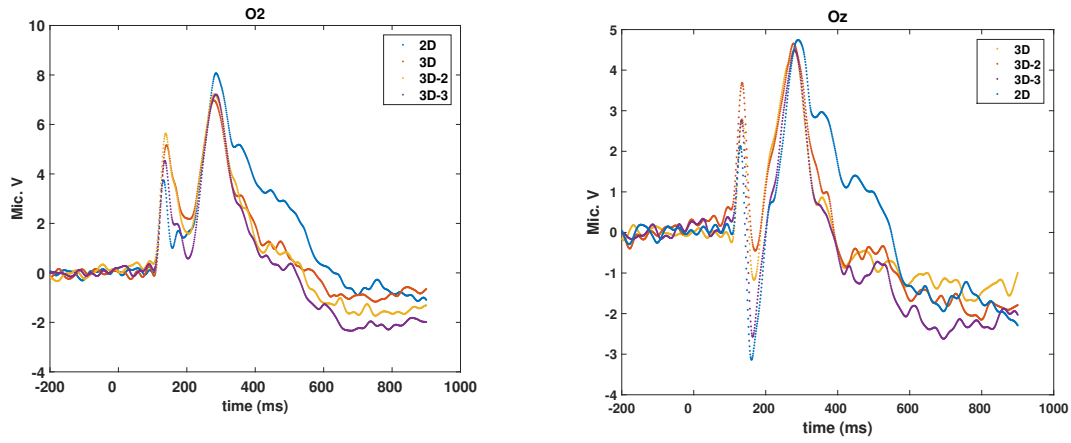
In a nutshell, the analysis of the ERP amplitudes in all four conditions reveals three large ERP components. The first one which corresponds to P1 component shows an increase for all 3D conditions compared to the 2D condition. This increase has been significant in all conditions except for the difference between 3D-3 and 2D. The increase in P1 amplitude is in line with previous studies showing that the perception of depth increases the amplitude in this component. However, the comparison within the 3D conditions reveals an interesting effect, i.e., the P1 amplitude has decreased in 3D-3 compared to 3D. This change could be explained by a decrease in the perception of depth by the subjects due to the large amount of vertical disparity. That also explains why the amplitude difference is not significantly different between 3D-3 and 2D in this interval.

Another significant difference appears between the conditions in the N1 component. In this component larger negativity in Oz (and negative-going component in O2) appears for 2D compared to all other conditions. Previous studies [41, 42, 43, 44, 45] show that spatial attention is one factor which causes an increase in the negativity of N1 component. These studies have addressed several paradigms including the covert and overt attention [46] or increased N1 in the discrimination of attended locations versus unattended-locations. However Ritter and his colleagues in [47, 48] addressed the hypothesis that N1 is correlated to the operation of a limited-capacity discrimination mechanism outside the context of spatial attention experiments. They found out that the amplitude of N1 component was larger in the discrimination tasks than the detection tasks. Vogel and Luck in [43, 44] showed that this increase appeared in the stimuli with color and form discrimination tasks compared to a simple detection task. Authors also conclude that the N1 attention effect and the N1 discrimination effect are caused by a general-purpose visual discrimination mechanism. Indeed discrimination effect could be a very reasonable explanation for the negativity increase in the amplitude of N1 component in our results. As it was described in Section 2.1 the the fixation cross

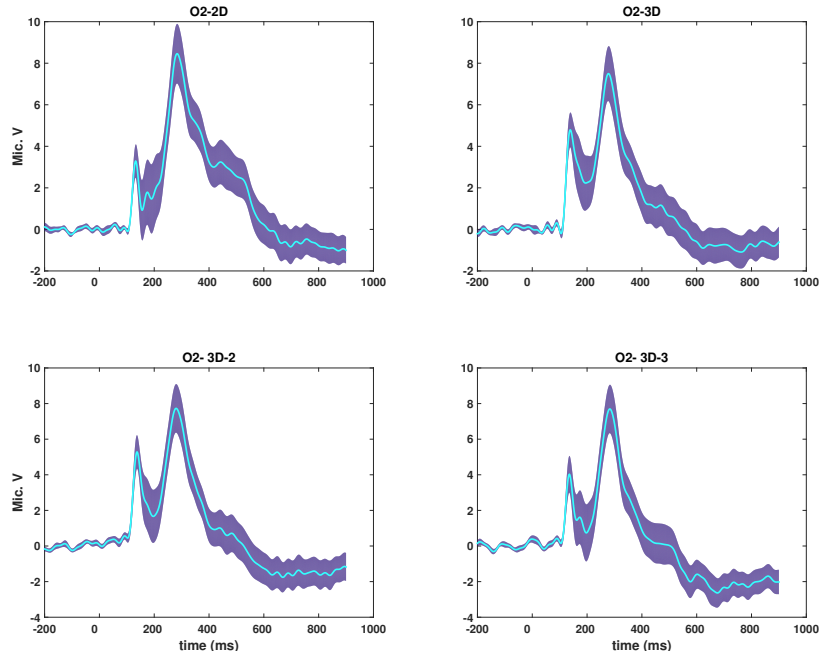
which is presented between the conditions is a three dimensional object. In the 2D condition this object is followed by a two dimensional cube. In other words, here the subjects experience a strong dimensionality switch which might explain the increase in N1 due to the discrimination process between the cross and the 2D cube. The same argument holds for the 3D-3 which has the largest negativity after 2D. In that case also the level of vertical disparity is large enough to give the subject a strong contrast to the three dimensional presentation of the fixation cross without vertical disparity. However for 3D condition the discrimination effect between the three dimensional cross and the 3D cube is not very strong and therefore the smaller negativity appears in N1.

3.3. Time Frequency Analysis of Induced Potentials

Wavelet transformation is applied to each epoch and each subject in the time interval between -1000 ms and 1000 ms as described in Section 2.5. These time frequency signals include the activities from the evoked potentials as well as the induced potentials. However, we are particularly interested in the induced potentials since we have already studied the evoked potentials in the ERP analysis. For this reason we have subtracted the ERPs corresponding to the average of all epochs of each single subject from single epoch activities of that subject in each specific condition before applying wavelet transformation. The absolute values of wavelet transformations is then averaged over all epochs and all subjects. A large amount of the activity disappears after the subtraction of ERPs. In Figure 9 the differences between the absolute values of wavelet transformations for different conditions are plotted. The main difference happens between 25 Hz and 30 Hz at around 400 ms after the stimulus onset. Comparisons show a decrease in the power for 3D conditions compared to 2D (therefore blue in the heat-map figures). 3D-2 shows an increase compared to 3D however this increase does not hold in the comparison between 3D-3 and 3D. To verify the significance of the comparisons a permutation test is applied to the differences according to the description in Section 2.5. The wavelet transformed values from 17 subjects are mixed into one single group of data sets including 34 sets (17 subjects and two conditions). Two data sets are pulled out of this data set randomly each set containing 17 data sets. This selection is repeated in a bootstrap of 1000 iterations. The null hypothesis is that the original difference is a coincidence and the real difference is equal to zero, i.e. the new data set is a random selection of the conditions and therefore the differences between two data sets should be smaller than the original differences. To verify that p-values are estimated as one minus the number of iterations in which the bootstrap set shows a larger difference between absolute values compared to the original absolute value of difference divided by 1000. The subtraction of one is to avoid a p-value equal to zero in an infinite number of iterations, i.e. even if all the bootstrap values are smaller than the original values still the p-value will not be exactly zero but it will be 0.001. The p-values for a one sided test are estimated and it becomes obvious that there is no significant difference between conditions caused by induced potentials.

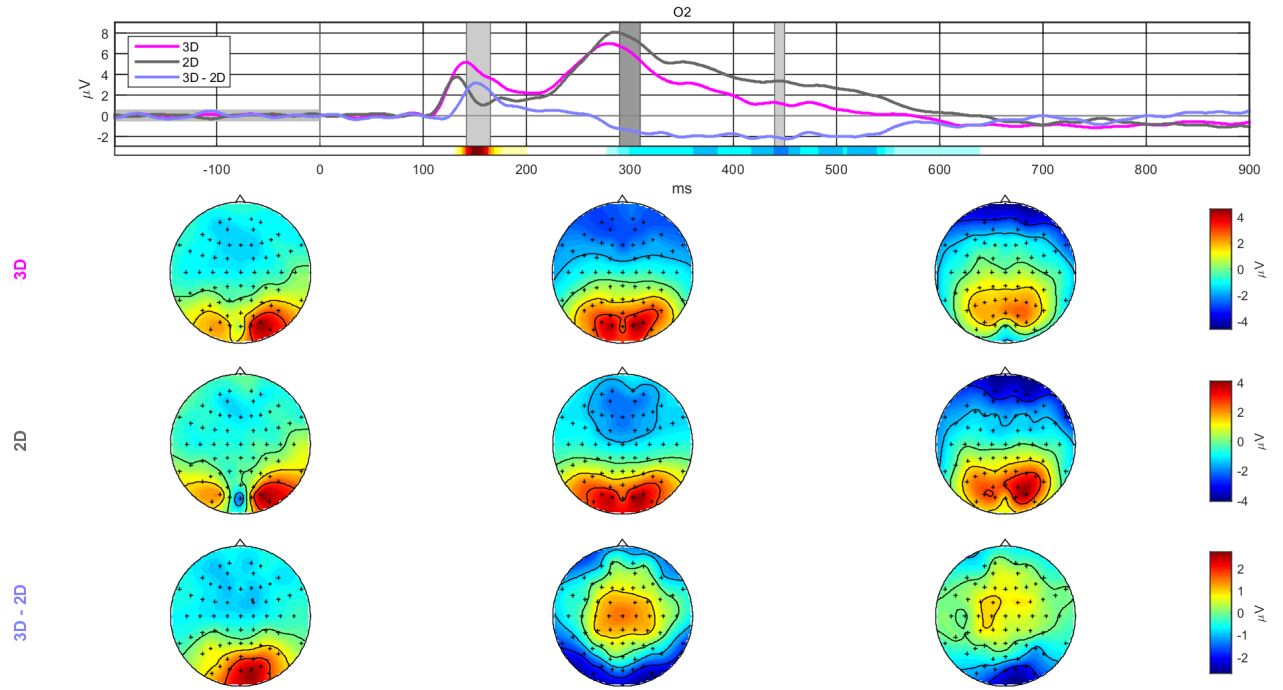


(a) ERPs of grand averaged signals in O2 for (b) ERPs of grand averaged signals in Oz for all conditions are plotted in a single figure to all conditions are plotted in a single figure to have a direct comparison between the amplitude differences as well the latencies.

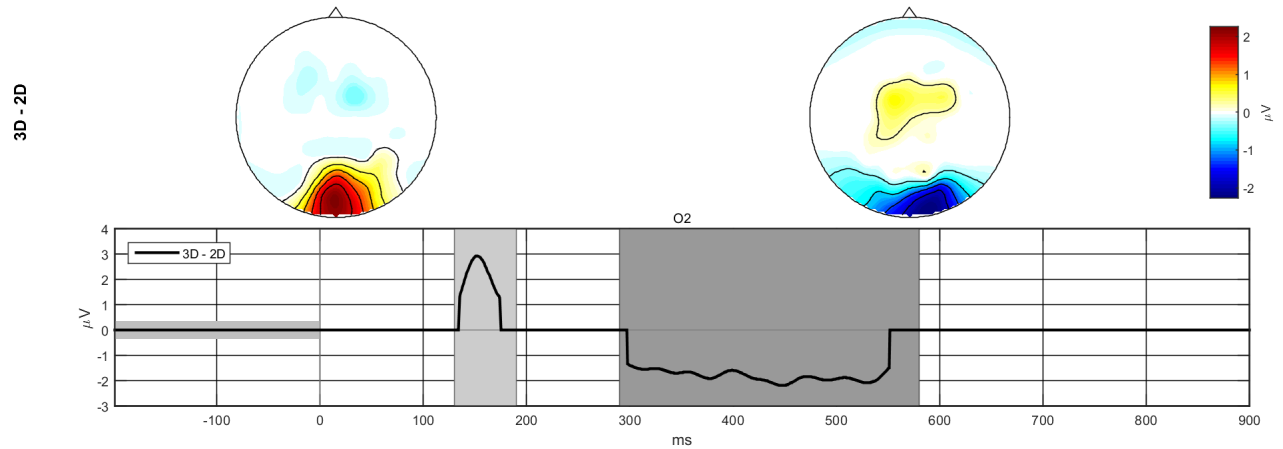


(c) Standard error across all subjects estimated for each condition on O2.

Figure 4. ERPs corresponding to different conditions averaged over 17 subjects for four conditions.



(a) Comparison of ERP between 2D and 3D. The intervals corresponding to significant differences are shown in a horizontal heat map underneath the time courses and topographies corresponding to selected areas with significant differences are plotted beneath.

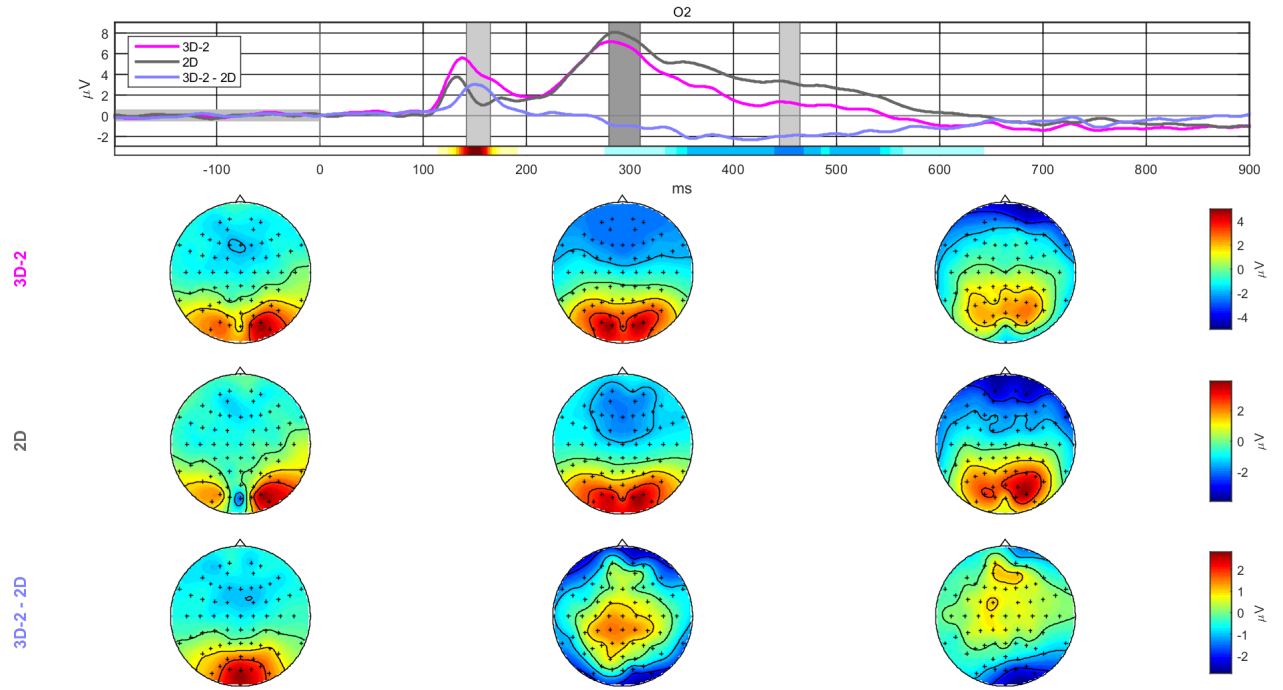


(b) Statistical test on the significance of the differences between ERPs of 2D and 3D.

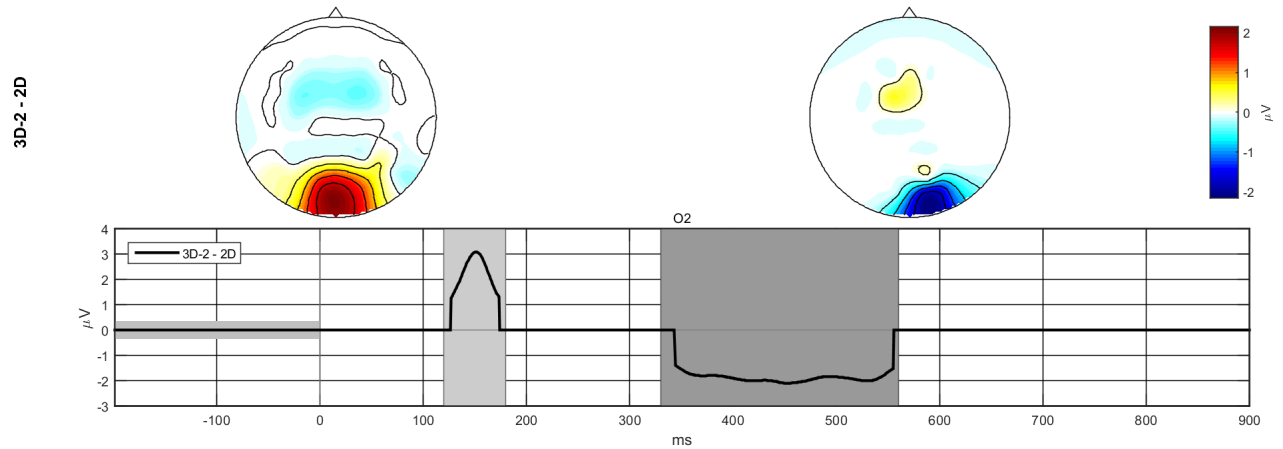
Figure 5. Comparison between the grand averaged ERP over 17 subjects for 2D and 3D.

4. Summary and Conclusion

We have studied the neurophysiological correlates of vertical disparity in an objective approach using EEG signals. In this study subjects were presented by a simple cube in four different conditions randomly: 2D, 3D without vertical disparity, 3D with a medium level of vertical disparity, i.e., 3D-2 and 3D with the maximum vertical disparity, i.e.,



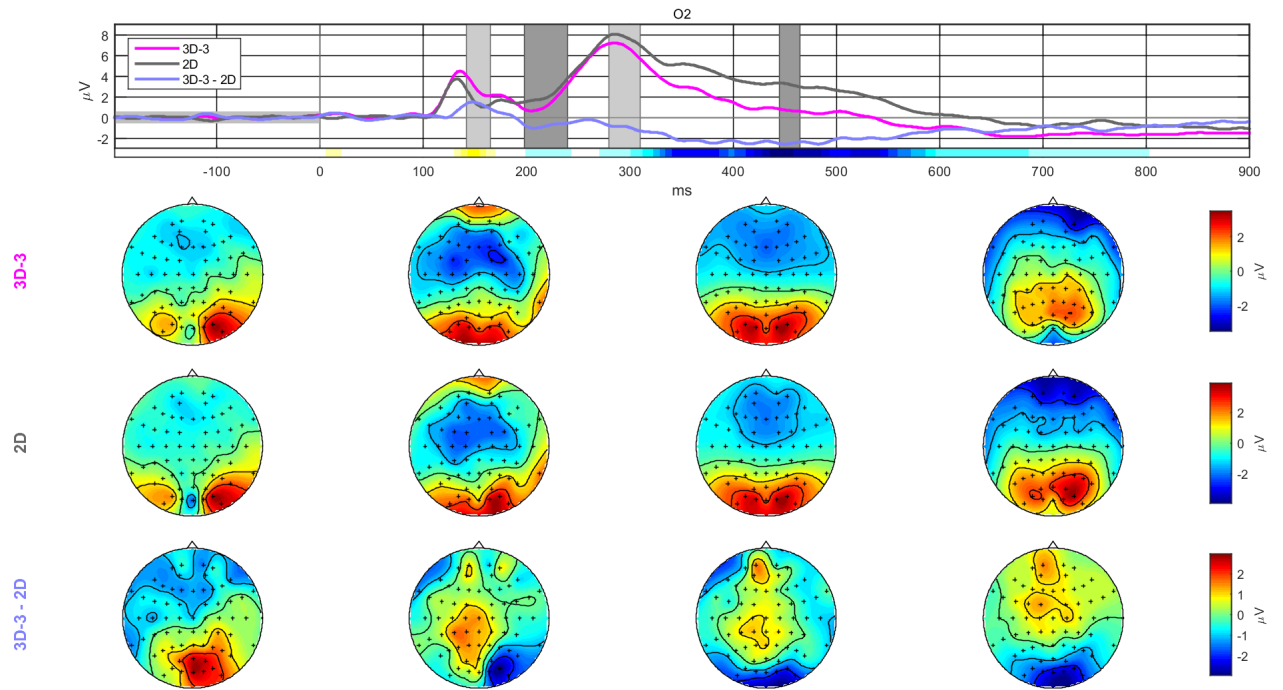
(a) Comparison of ERP between 2D and 3D-2. The intervals corresponding to significant differences are shown in a horizontal heat map underneath the time courses and topographies corresponding to selected areas with significant differences are plotted beneath.



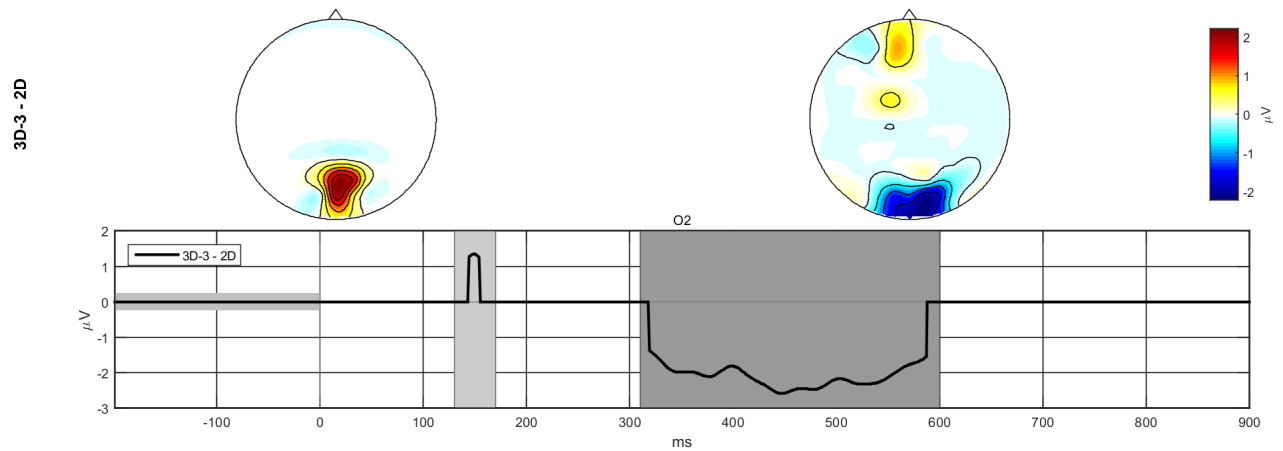
(b) Statistical test on the significance of the differences between ERPs of 2D and 3D-2.

Figure 6. Comparison between the grand averaged ERP over 17 subjects for 2D and 3D-2.

3D-3. Event related potentials have been compared between these conditions in order to extract a feature which correlates with the level of vertical disparity in 3D images. The comparisons between 2D and two 3D conditions, i.e. 3D and 3D-2 show a significant increase in the amplitude of P1 component in the 3D conditions. These results are in-line with previous studies [49, 50, 51, 52, 53] which confirmed that P1 increases due to the depth perception. These studies however consider non-stereoscopic depth perception



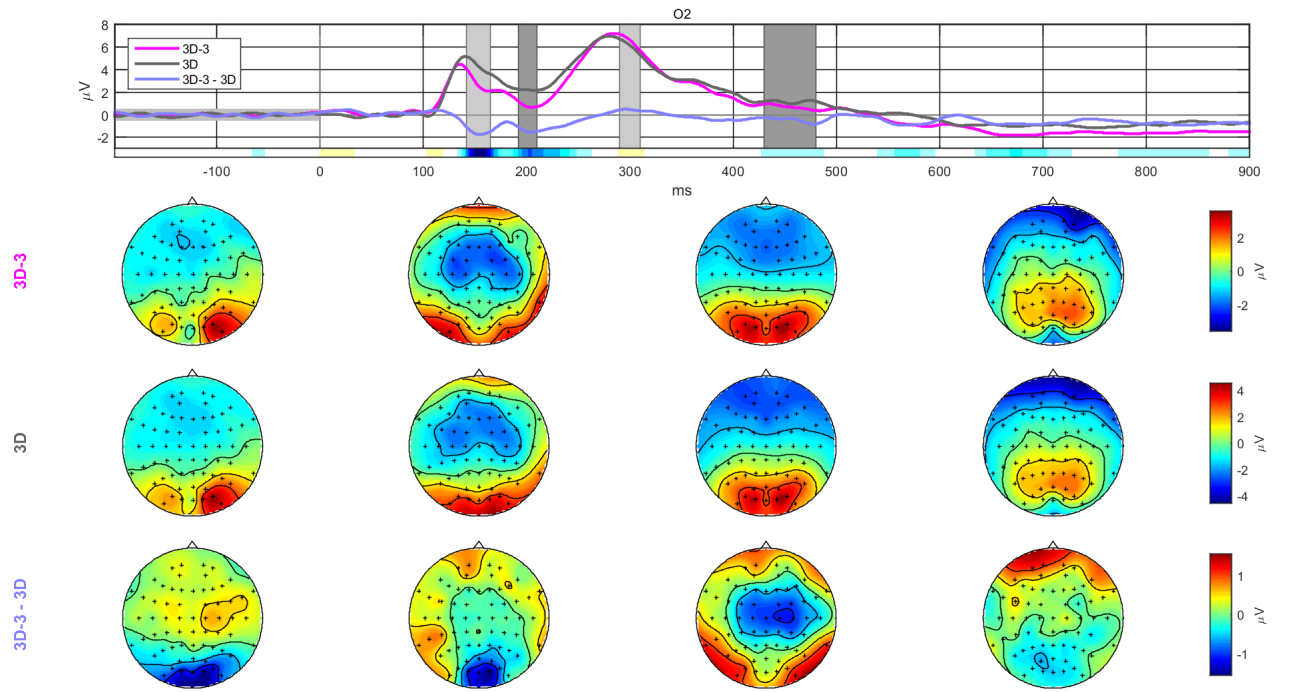
(a) Comparison of ERP between 2D and 3D-3. The intervals corresponding to significant differences are shown in a horizontal heat map underneath the time courses and topographies corresponding to selected areas with significant differences are plotted beneath.



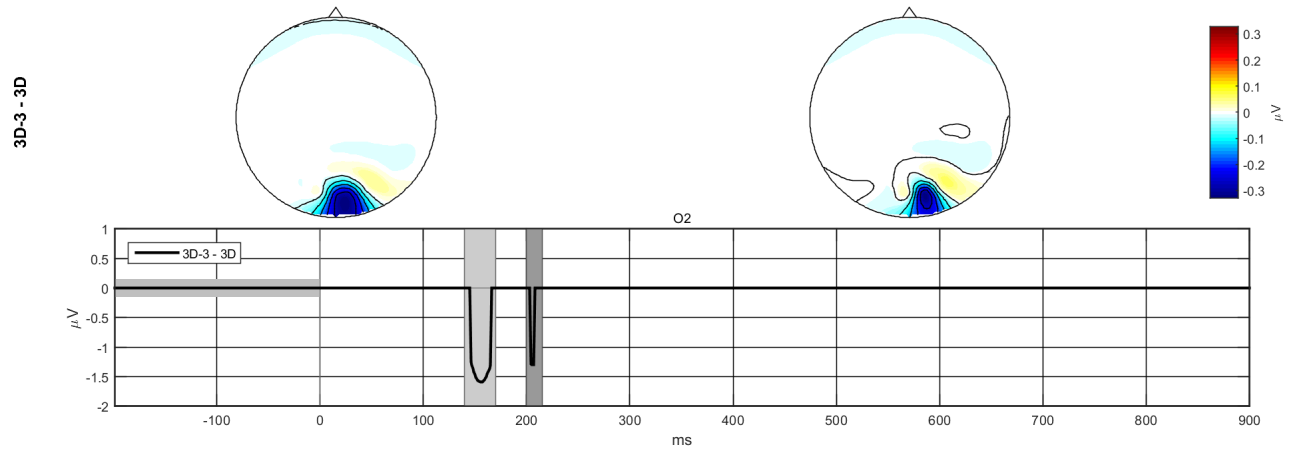
(b) Statistical test on the significance of the differences between ERPs of 2D and 3D-3.

Figure 7. Comparison between the grand averaged ERP over 17 subjects for 2D and 3D-3.

Objective Quality Assessment of Stereoscopic Images with Vertical Disparity using EEG 16



(a) Comparison of ERP between 3D and 3D-3. The intervals corresponding to significant differences are shown in a horizontal heat map underneath the time courses and topographies corresponding to selected areas with significant differences are plotted beneath.



(b) Statistical test on the significance of the differences between ERPs of 3D and 3D-3.

Figure 8. Comparison between the grand averaged ERP over 17 subjects for 3D and 3D-3.

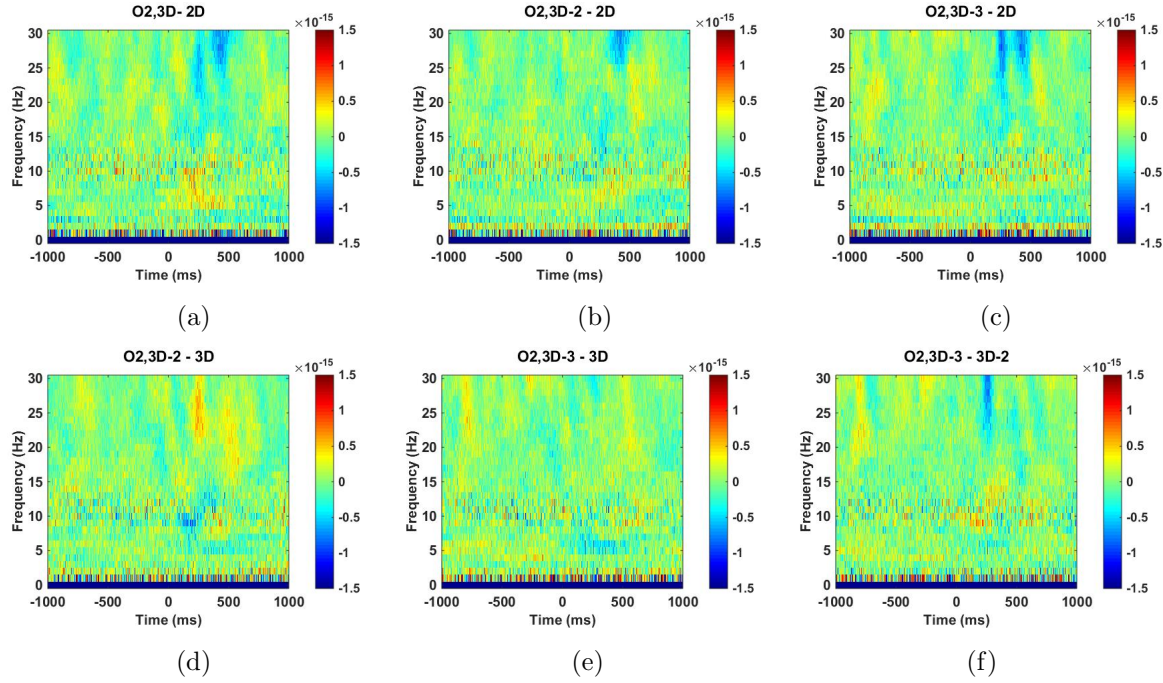


Figure 9. Difference between absolute values of wavelet transformed data after the removal of ERP components between all couples of conditions is plotted.

with concave/convex stimuli although, we here confirm the same results in the case of stereoscopic images. Further comparisons between 3D and 3D-3 however shows a significant decrease in this component for 3D-3. We interpreted this phenomenon as a decrease in the depth perception due to the large amount of vertical disparity in 3D-3. In other words, we have successfully extracted a neurophysiological feature from the EEG data which is correlated with the increase in vertical disparity. This feature can be used in crucial applications of 3D technology in which the visual discomfort caused by vertical disparity are desired to be predicted and avoided in advance such as in the medical applications of 3D imaging.

Significant differences are also observed in the amplitude of N1 component in the occipital electrodes specially in the Oz channel. In several studies N1 amplitude modulations have been linked to the engagement of attention either in the context of spatial attention or non-spatial attention [41, 42, 43, 44, 45]. In these reports different paradigms have been studied including (c)overt attention and discrimination between subjects of different color and form compared to simple detection tasks. Several studies show that the N1 negativity has increased in the case of spatial attention when the targets have been attended compared to the ignored target. However Ritter et. al in [47, 48] have studied the attention effect in the context of non-spatial attention which is more relevant for our paradigm. They showed that discrimination between conditions increased the N1 component compared to the simple detection cases. Vogel and Luck in [43] also showed that this discrimination appears to be larger when the difference appears in color as well as form compared to the simple detection task. This phenomenon can

explain large N1 component in 2D due to a switch happening in the dimensionality of the presentation from the fixation cross (three dimensional) to the two dimensional cube in 2D condition. This switch introduces an implicit discrimination effect causing the larger increase in the N1 component compared to other conditions. The next largest N1 negativity occurs in 3D-3 condition. This is also explainable by the same theory because the vertical disparity in 3D-3 which is missing in the fixation cross causes a strong discrimination effect. However in two other conditions, i.e., 3D and 3D-2 lack of vertical disparity or small amount of it causes smaller discrimination effects and therefore smaller N1 amplitudes. Although the argument seems reasonable but we would like to point out that this result needs to be studied more closely in the future work in which the discrimination effect of the dimensionality switch is measured precisely through a behavioral subjective feedback.

We have also observed that the P1 and N1 components tended to have predominantly larger power in the right hemisphere than the left one which has been reported in the previous study in [49] as well. This results are consistent with neurophysiological studies using fMRI, PET and MEG that showed right hemisphere is associated with visual attentional activity pertaining to letters and figures [54, 55, 56] as well as in the recognition of faces [57, 58]. We also compared the induced potentials between all conditions but no significant differences were observed. In future work we will investigate these effects with more advanced data fusion schemes [59] and more powerful methods such as deep neural networks [60]. Furthermore we will estimate and analyze cognitive properties in neurophysiological data from quality assessment experiments [61].

Another aspect which is of interest in the future work is the analysis of underlying sources generating the ongoing brain activities either at a single frequency or in an inter-frequency approach [62, 63] as well as the event related potentials using inverse methods [64, 65]. It is interesting to know if different brain regions are active for different conditions and if these regions have different connectivity patterns in their interactions [66].

5. Acknowledgment

The authors would like to acknowledge Dr. Matthias S. Treder for his useful comments during this study. This research was partially funded by the German Research Foundation (DFG, SFB938/Z3 and TRR-169/B4). Klaus-Robert Müller thanks DFG (DFG SPP 1527, MU 987/14-1) and the Federal Ministry for Education and Research (BMBF) as well as by the Brain Korea 21 Plus Program funded by the Korean government. Correspondence to WS, GC, TW and KRM.

Reference

- [1] S. E. Palmer, *Vision science: Photons to phenomenology*, vol. 1, MIT press Cambridge, MA, 1999.

- [2] R. A. Crone, "The history of stereoscopy," *Doc. Ophthalmol.*, vol. 81, no. 1, pp. 1–16, 1992.
- [3] C. Wheatstone, "Contributions to the Physiology of Vision. Part II. On Some Remarkable, and Hitherto Unobserved, Phaenomena of Binocular Vision,(Continued).," *Proc. R. Soc. London*, vol. 6, no. 0, pp. 138–141, 1850.
- [4] M. V. Sanchez-Vives and M. Slater, "From presence to consciousness through virtual reality.," *Nat. Rev. Neurosci.*, vol. 6, no. 4, pp. 332–339, 2005.
- [5] M. Gaebler, F. Biessmann, J.-P. Lamke, K.-R. Müller, H. Walter, and S. Hetzer, "Stereoscopic depth increases intersubject correlations of brain networks.," *Neuroimage*, vol. 100, pp. 427–34, oct 2014.
- [6] A. Raake and S. Egger, "Quality and quality of experience," in *Qual. Exp.*, pp. 11–33. Springer, 2014.
- [7] L. M. J. Meesters, W. A. IJsselsteijn, and P. J. H. Seuntjens, "A Survey of Perceptual Evaluations and Requirements of Three-Dimensional TV," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 3, pp. 381–391, 2004.
- [8] M. Urvoy, M. Barkowsky, and P. Le Callet, "How visual fatigue and discomfort impact 3D-TV quality of experience: A comprehensive review of technological, psychophysical, and psychological factors," *Ann. des Telecommun. Telecommun.*, vol. 68, no. 11-12, pp. 641–655, 2013.
- [9] M. Lambooi, M. Fortuin, W. IJsselsteijn, and I. Heynderickx, "Measuring visual discomfort associated with 3D displays," *Spie*, vol. 7237, no. 1, pp. 72370K–72370K–12, 2009.
- [10] ITU-T SG 9 (Study Period 2013) Temporary Document 643-GEN, "Draft New Recommendation J.3D-sam: "Subjective assessment methods for 3D video quality"," Tech. Rep., 2015.
- [11] 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.
- [12] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," Tech. Rep., International Telecommunication Union, Geneva, Switzerland, apr 2008.
- [13] ITU-T SG 9 (Study Period 2013) Temporary Document 642-GEN, "Draft New Recommendation-P.3D-fatigue "Information and guidelines for assessing and minimizing visual discomfort and visual fatigue from 3D video"," Tech. Rep., 2015.
- [14] D. R. Risky, "Use and Abuses of Category Scales in Sensory Measurement," *J. Sens. Stud.*, vol. 1, no. 3-4, pp. 217–236, dec 1986.
- [15] A. K. Porbadnigk, M. S. Treder, B. Blankertz, J.-N. Antons, R. Schleicher, S. Möller, G. Curio, and K.-R. Müller, "Single-trial analysis of the neural correlates of speech quality perception.," *J. Neural Eng.*, vol. 10, no. 5, pp. 056003, 2013.
- [16] A. K. Porbadnigk, J.-N. Antons, B. Blankertz, M. S. Treder, R. Schleicher, S. Möller, and G. Curio, "Using ERPs for assessing the (sub) conscious perception of noise," in *Eng. Med. Biol. Soc. (EMBC), 2010 Annu. Int. Conf. IEEE. IEEE*, 2010, pp. 2690–2693.
- [17] J.-N. Antons, R. Schleicher, S. Arndt, S. Moller, A. K. Porbadnigk, and G. Curio, "Analyzing Speech Quality Perception Using Electroencephalography," *IEEE J. Sel. Top. Signal Process.*, vol. 6, no. 6, pp. 721–731, oct 2012.
- [18] H. Hayashi, H. Shirai, M. Kameda, S. Kunifuji, and M. Miyahara, "Assessment of extra high quality images using both EEG and assessment words on high order sensations," *SMC 2000 Conf. Proceedings. 2000 IEEE Int. Conf. Syst. Man Cybern. 'Cybernetics Evol. to Syst. Humans, Organ. their Complex Interact. (Cat. No.00CH37166)*, vol. 2, pp. 1289–1294, 2000.
- [19] L. Lindemann and M. Magnor, "Assessing the quality of compressed images using EEG," in *2011 18th IEEE Int. Conf. Image Process.*, 2011, pp. 3109–3112.
- [20] J. Perez and E. Delechelle, "On the measurement of image quality perception using frontal EEG analysis," *2013 Int. Conf. Smart Commun. Netw. Technol.*, pp. 1–5, 2013.
- [21] S. Bosse, L. Acqualagna, A. K. Porbadnigk, G. Curio, K.-R. Müller, B. Blankertz, and T. Wiegand,

- “Neurophysiological assessment of perceived image quality using steady-state visual evoked potentials,” *Proceedings of SPIE*, vol. 9599, pp. 959914, 2015.
- [22] S. Bosse, L. Acqualagna, A. K. Porbadnigk, B. Blankertz, G. Curio, K.-R. Müller, and T. Wiegand, “Neurally informed assessment of perceived natural texture image quality,” in *Image Process. (ICIP), 2014 IEEE Int. Conf.* oct 2014, pp. 1987–1991, IEEE.
- [23] L. Acqualagna, S. Bosse, A. K. Porbadnigk, G. Curio, K.-R. Müller, T. Wiegand, and B. Blankertz, “EEG-based classification of video quality perception using steady state visual evoked potentials (SSVEPs),” *J. Neural Eng.*, vol. 12, no. 2, pp. 026012, 2015.
- [24] M. Mustafa, S. Guthe, and M. Magnor, “Single-trial EEG classification of artifacts in videos,” *ACM Trans. Appl. Percept.*, vol. 9, no. 3, pp. 1–15, 2012.
- [25] S. Scholler, S. Bosse, M. S. Treder, B. Blankertz, G. Curio, K.-R. Müller, and T. Wiegand, “Toward a direct measure of video quality perception using EEG,” *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2619–29, 2012.
- [26] L. Lindemann, S. Wenger, and M. Magnor, “Evaluation of video artifact perception using event-related potentials,” *Proc. ACM SIGGRAPH Symp. Appl. Percept. Graph. Vis. - APGV '11*, vol. 0, no. 0, pp. 53, 2011.
- [27] S. Arndt, J.-N. Antons, R. Schleicher, S. Moller, and G. Curio, “Using electroencephalography to measure perceived video quality,” *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 3, pp. 366–376, 2014.
- [28] S. Bosse, L. Acqualagna, W. Samek, A. K. Porbadnigk, G. Curio, B. Blankertz, K.-R. Müller, and T. Wiegand, “Assessing perceived image quality using steady-state visual evoked potentials and spatio-spectral decomposition,” *IEEE Transactions on Circuits and Systems for Video Technology*, 2017.
- [29] A. N. Moldovan, I. Ghergulescu, S. Weibelzahl, and C. H. Muntean, “User-centered EEG-based multimedia quality assessment,” in *IEEE Int. Symp. Broadband Multimed. Syst. Broadcast. BMSB*, 2013.
- [30] E. Kroupi, P. Hanhart, J.-S. Lee, M. Rerabek, and T. Ebrahimi, “EEG correlates during video quality perception,” in *Signal Process. Conf. (EUSIPCO), 2014 Proc. 22nd Eur.* IEEE, 2014, pp. 1–4.
- [31] J. Frey, A. Appriou, F. Lotte, and M. Hachet, “Classifying EEG Signals during Stereoscopic Visualization to Estimate Visual Comfort,” *Comput. Intell. Neurosci.*, vol. 2016, 2015.
- [32] J. Frey, L. Pommereau, F. Lotte, and M. Hachet, “Assessing the zone of comfort in stereoscopic displays using EEG,” *Proc. Ext. Abstr. 32nd Annu. ACM Conf. Hum. factors Comput. Syst. - CHI EA '14*, pp. 2041–2046, 2014.
- [33] M. A. Wenzel, R. Schultze-Kraft, F. C. Meinecke, F. Cardinaux, T. Kemp, K.-R. Müller, G. Curio, and B. Blankertz, “EEG-based usability assessment of 3D shutter glasses,” *J. Neural Eng.*, vol. 13, no. 1, pp. 016003, 2016.
- [34] S. Bosse, K.-R. Müller, T. Wiegand, and W. Samek, “Brain-computer interfacing for multimedia quality assessment,” pp. 002834–002839, 2016.
- [35] U. Engelke, D. P. Darcy, G. H. Mulliken, S. Bosse, M. 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*, 2016.
- [36] K.-R. Müller, C. W. Anderson, and G. E. Birch, “Linear and nonlinear methods for brain-computer interfaces,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 165–169, 2003.
- [37] B. Blankertz, L. Acqualagna, S. Daehne, S. Haufe, M. Schultze-Kraft, I. Sturm, M. Uumluc, M. Wenzel, G. Curio, and K.R. Müller, “The berlin brain-computer interface: Progress beyond communication and control,” *Frontiers in Neuroscience*, vol. 10, pp. 530–550, 2016.
- [38] 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.
- [39] B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. E.

- Ramsey, I. Sturm, G. Curio, and K. R. Mueller, "The berlin brain-computer interface: Non-medical uses of bci technology," *Frontiers in Neuroscience*, vol. 4, no. 198, 2010.
- [40] T. W. Picton, "The p300 wave of the human event-related potential.," *Journal of clinical neurophysiology*, vol. 9, no. 4, pp. 456–479, 1992.
- [41] S. Hillyard and A.-V. Lourdes, "Event-related brain potentials in the study of visual selectiveattention," *Proceedings of the National Academy of Sciences*, vol. 95, no. 3, pp. 781–787, 1998.
- [42] M.D. Rugg, A.D. Milner, C.R. Lines, and R. Phalp, "Modulation of visual event-related potentials by spatial and non-spatial visual selective attention," *Neuropsychologia*, vol. 25, no. 1, Part 1, pp. 85 – 96, 1987.
- [43] E. K. Vogel and S. J Luck, "The visual n1 component as an index of a discrimination process," *Psychophysiology*, vol. 37, no. 2, pp. 190–203, 2000.
- [44] J. Theeuwes A. Wilschut and C. N. L. Olivers, "The time course of attention: Selection is transient," *PLOS ONE*, vol. 6, no. 11, pp. 1–10, 11 2011.
- [45] S. J. Luck, G. F. Woodman, and E. K. Vogel, "Event-related potential studies of attention," *Trends in Cognitive Sciences*, vol. 4, no. 11, pp. 432 – 440, 2000.
- [46] M. S. Treder and B. Blankertz, "(c)overt attention and visual speller design in an erp-based brain-computer interface," *Behavioral and Brain Functions*, vol. 6, no. 1, pp. 28, 2010.
- [47] W. Ritter, R. Simson, and H. G Vaughan, "Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks," *Psychophysiology*, vol. 20, no. 2, pp. 168–179, 1983.
- [48] W. Ritter, R. Simson, H.G Vaughan, and M. Macht, "Manipulation of event-related potential manifestations of information processing stages," *Science*, vol. 218, no. 4575, pp. 909–911, 1982.
- [49] 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.
- [50] 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.
- [51] N. A. Parks and P. M. Corballis, "Attending to depth: electrophysiological evidence for a viewer-centered asymmetry," *Neuroreport*, vol. 17, no. 6, pp. 643–647, 2006.
- [52] A. Séverac Cauquil, Yves Trotter, and Margot J. Taylor, "At what stage of neural processing do perspective depth cues make a difference?," *Experimental Brain Research*, vol. 170, no. 4, pp. 457–463, 2005.
- [53] P. Mamassian, I. Jentzsch, B. A. Bacon, and S. R. Schweinberger, "Neural correlates of shape from shading," *NeuroReport*, vol. 14, no. 7, pp. 971–975, 2003.
- [54] M. Corbetta, J. M. Kincade, J. M. Ollinger, M. P. McAvoy, and G. L. Shulman, "Voluntary orienting is dissociated from target detection in human posterior parietal cortex," *Nature neuroscience*, vol. 3, no. 3, pp. 292–297, 2000.
- [55] W. J. M. Levelt, P. Praamstra, A. S. Meyer, P. Helenius, and R. Salmelin, "An meg study of picture naming," *Journal of cognitive neuroscience*, vol. 10, no. 5, pp. 553–567, 1998.
- [56] G. R. Fink, P. W. Halligan, J. C. Marshall, C. D. Frith, R. S. Frackowiak, and R. J. Dolan, "Neural mechanisms involved in the processing of global and local aspects of hierarchically organized visual stimuli.," *Brain*, vol. 120, no. 10, pp. 1779–1791, 1997.
- [57] K. Linkenkaer-Hansen, J. M. Palva, M. Sams, J. K. Hietanen, H. J. Aronen, and R. J Ilmoniemi, "Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto-and electroencephalography," *Neuroscience letters*, vol. 253, no. 3, pp. 147–150, 1998.
- [58] E. I. Olivares, J. Iglesias, and M. A. Bobes, "Searching for face-specific long latency erps: a

- topographic study of effects associated with mismatching features,” *Cognitive Brain Research*, vol. 7, no. 3, pp. 343–356, 1999.
- [59] S. Fazli, S. Dähne, W. Samek, F. Bießmann, and K.-R. Müller, “Learning from more than one data source: data fusion techniques for sensorimotor rhythm-based brain-computer interfaces,” *Proceedings of the IEEE*, vol. 103, no. 6, pp. 891–906, 2015.
- [60] 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.
- [61] W. Samek, D. Blythe, G. Curio, K.-R. Müller, B. Blankertz, and V. V. Nikulin, “Multiscale temporal neural dynamics predict performance in a complex sensorimotor task,” *NeuroImage*, vol. 141, pp. 291–303, 2016.
- [62] F. Chella, V. Pizzella, F. Zappasodi, G. Nolte, and L. Marzetti, “Bispectral pairwise interacting source analysis for identifying systems of cross-frequency interacting brain sources from electroencephalographic or magnetoencephalographic signals,” *Phys. Rev. E*, vol. 93, pp. 052420, May 2016.
- [63] F. Shahbazi, A. Ewald, and G. Nolte, “Univariate normalization of bispectrum using hlder’s inequality,” *Journal of Neuroscience Methods*, vol. 233, pp. 177 – 186, 2014.
- [64] 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.
- [65] 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.
- [66] F. Shahbazi Avarvand, A. Ewald, and G. Nolte, “Localizing true brain interactions from eeg and meg data with subspace methods and modified beamformers,” *Computational and mathematical methods in medicine*, vol. 2012, 2012.