Proceedings of the
7th Graz Brain-Computer Interface Conference 2017 DOI: 10.3217/978-3-85125-533-1-81

MEASURING THE QUALITY OF 3D VISUALIZATIONS USING EEG: A TIME-FREQUENCY APPROACH

F. Shahbazi Avarvand[1], S. Bosse[1], G. Nolte[2],
T. Wiegand[1,3], W. Samek [1]

[1]Department of Video Coding & Analytics, Fraunhofer Heinrich Hertz Institute, Berlin, Germany
[2]Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[3]Image Communication Group, Technische Universität Berlin, Berlin, Germany

E-mail: forooz.shahbazi@hhi.fraunhofer.de, wojciech.samek@hhi.fraunhofer.de

ABSTRACT: In this study we have assessed the quality of experience objectively when vertical disparity is introduced in the stereoscopic presentation of images. Four different conditions including a cube in 2D and the same cube in 3D with and without vertical disparities are compared based on the EEG signals recorded from 17 subjects. Two different vertical disparity levels are studied. Event-related potentials (ERPs) corresponding to four conditions are compared in time-frequency domain. The results show an increase in beta power of the occipital region for the images with vertical disparity compared to the 3D image with no vertical disparity. An increase in beta power in this region correlates with an increase in the level of attention which in our case is caused by the vertical disparity component in the stereoscopic images.

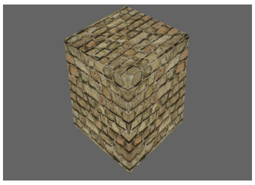
INTRODUCTION

Stereoscopic imaging technologies are used since almost two centuries for inducing the illusion of visual depth by providing images from horizontally slightly shifted focal points to the two eyes of the observer [1]. In recent years, this technique became popular again in various application domains and is used to create or increase the impression of greater realism and immersion e.g. for entertainment, as in movies or computer games, or in data visualization. New application scenarios in virtual and augmented reality environments are in reach with the advent of market-ready head mounted displays such as the Oculus Rift. The imaging systems' quality experienced by the user is crucial for the success of these applications. One aspect driving the quality of experience for systems based on stereoscopy is visual comfort that can be severely impaired e.g. by accommodation/vergence conflicts, excessive binocular parallax, dichoptic errors and other factors [2]. As for most multimedia signals and their impairments, unfortunately, no reliable objective computational model is currently available. Therefore, assessing the perceived quality of imaging systems relies on behavioral tests. These tests are typically carried out as psychophysical judgment experiments, during which a subject has to rank the quality of a set of test stimuli by giving an overt response. Thus, the outcome of these tests is the result of a cognitive process of the subject. This leads to different drawbacks of this method: The ratings are highly variable across subjects and prone to subjective factors as bias, expectations and strategies. As such tests are exhaustive for the subject, participants' ratings may become unreliable over time and the duration of psychophysical quality assessment test should not take longer than 30 minutes [3]. Moreover, it is very difficult, if not impossible, to integrate this kind of behavioral procedure in real-time assessment of perceived quality.
To overcome these limitations, brain activity recordings for directly monitoring the users cognitive state have shown promising results recently. Electroencephalography (EEG) in particular is one of the easiest and most mobile devices to record brain signals and has been used to study neural correlates of perceived quality of multimedia, such as audio [4] and 2D [5, 6, 7, 8] and 3D video [9]. [10] and [11] provide a comprehensive overview on approaches to assessing perceived quality physiologically.
If the two focal points in a stereoscopic imaging system are not aligned vertically, vertical disparities occur alongside horizontal disparities. While the latter serve as depth cue, vertical disparities commonly cause visual discomfort experienced by the observer that can ultimately even lead to physical pain. In this paper we address the assessment of visual discomfort caused by vertical disparities using EEG. For that the brain signals recorded for two amounts of vertical disparity are compared to no vertical disparity and a 2D image on channel level in time-frequency domain. More detailed analysis of this data set with a focus on the ERP components is presented in [12].
The next section describes the experimental setup and explains the event-related potentials (ERPs) which are studied in time-frequency domain. The results show significant differences in time-frequency domain, more precisely larger beta power for occipital region in vertical disparity conditions. The results and discussion are presented in more details in the last two sections.

 DOI: 10.3217/978-3-85125-533-1-81

METHODS

Experimental Setup: In an objective approach we have studied EEG features caused by the vertical disparity in stereoscopic images. The vertical disparity is simulated for a 3D image of a cube in which the right camera is shifted upwards relative to the left one. 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 in two categories of 2D and 3D, i.e. three images in 3D including two 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 a).


a)

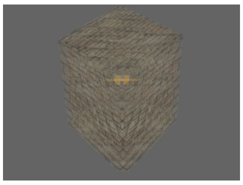

b)

Figure 1: a) The cube is presented in 2D and in three different 3D conditions (3D, 3D-2 and 3D-3). b) A cross in 3D is presented in between the epochs as a pause interval. The cross is projected on the cube in this figure to show its exact location relative to the center of the cube.

Each image is presented randomly for 120 trials (epochs) for 4 seconds and between the images a cross is presented in 3D for an interval of 3 seconds. This interval is implemented for subjects to be able to rest their eyes and it helps to reduce the amount of ocular artifacts.
In Figure 1 b), the cross is projected on the cube to visualize the exact position of it relative to the cube. To keep the subjects attentive they were asked to do a task: they were supposed to press a button when an image of a cat was shown. 120 images of a dog (80%) and a cat (20%) appeared randomly among other stimuli between two fixation crosses. If the subject successfully 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 horizontal angle of view is 20.76 degrees while the distance between the subject and the screen is 280 cm. The subjects wear 3D polarized glasses during the experiment. According to a previous study [13] polarized passive glasses tend to be more comfortable than the active shutter glasses. The experiment was conducted in a dimly lit and silent room. The 3D screen is JVC 3D LCD Monitor (model number: GD-463D10E). EEG signals are recorded by an actiCap from Brain Products GmbH with 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) and the impedance of electrodes was kept below 10 KΩ. 21 subjects have been recorded out of which 4 data sets have been excluded due to the low signal-to-noise values and high number of rejected trials. The data from 17 subjects, 6 male and 11 female, with the average age of 25.83 is analyzed. All subjects have been checked to 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).

Pre-processing of EEG data: EEG data is low-pass filtered below 30 Hz during the analysis besides the filter applied by the EEG amplifier hardware at 0.016 Hz. FCz was selected as the original reference electrode during the experiment and the data is re-referenced to the common average for 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 the time course of each epoch. Ocular artifacts are removed by regression through projecting out part of the data which is correlated with EOG electrodes. For this purpose a short measurement was conducted before the experiment in which the subject was supposed to blink when a circle appeared on the screen. Meanwhile the vertical ocular activity was measured by an electrode underneath the right eye (VEOG) and Fp2. The horizontal ocular activity of the subjects was also recorded in a similar approach 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 two electrodes, i.e. F7 and F8 is estimated as the horizontal component while the difference between Fp2 and VEOG is estimated as the vertical component. Part of the EEG data which is correlated with these two components is then projected out from the data. For more details on the method please refer to [14]. In an additional step epochs in which the difference between maximum and minimum amplitude exceeds 70 μV are rejected.

Event-related potentials: Part of the responses in EEG signals are phased-locked with the stimulus however with a very low signal to noise ratio due to the non-phase locked activities which are considered as noise when the focus is on the event-related potentials (ERP). EEG data is averaged over all trials to cancel out all non-phase-locked activities and therefore to increase the signal-to-noise ratio. The time window of this average is selected between 200 ms before the stimulus onset and 900 ms after that because the brain activity appeared not to be affected by the stimuli after 900 ms anymore. The BBCI toolbox, which is a Matlab-based toolbox [15], is used for the ERP analysis.
The ERP components might vary for different conditions both in amplitudes and latencies of the peak. Different conditions are ideally differentiable from each other based on their different ERP components. We have stud-

 DOI: 10.3217/978-3-85125-533-1-81

ied ERP components in time-frequency domain to extract the features correlated with different conditions.

Statistical tests: To verify the significance of the differences between time-frequency analysis results of different conditions a combination of Jackknife re-sampling method and the student's t-test is applied.
One alternative would be a student's t-test between data sets of ERPs of 17 subjects for two classes. However since the signal-to-noise ratios of single subject ERPs are low, we have taken another approach. As it is described in [16], Jackknife re-sampling test in combination with a one sample Student's t-test is applied. Jackknife is performed as follows: the averaged ERP of each single subject is left out each time and the grand average ERP is estimated by averaging over the rest of subjects (16 subjects in 17 iterations). The differences between the absolute values of wavelet transformed ERPs are then estimated for each two conditions and each iteration and a one sample t-test is applied to test the null hypothesis that the mean of the distribution is zero. Note that in this case the t values i.e., t, should be corrected to t_c

$$t_c = \frac{t}{(n-1)} \tag{1}$$

where n is the number of subjects. The proof of this adjustment is provided in [16].
Another issue to be considered is the correction for the problem of multiple comparison. The correction is applied as it is suggested in [17] for controlling the false discovery rate (FDR). The p-values are sorted and c is estimated as

$$c = \sum_{i=1}^{N} \frac{1}{i} \tag{2}$$

where N is the number of all samples. In the case of time-frequency analysis it will be the number of p-values of all time and frequency points. For the largest i where

$$P_i \leqslant \frac{i \times q}{N \times c} \tag{3}$$

P_i is selected (q is the desired false detection rate which is usually set to 0.05.) and if P_i is larger than the significant threshold of Bonferroni, it will be replaced as the new significance level. Otherwise the level will be set to the Bonferroni level. All p-values are then compared to the new significance level.

RESULTS

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. To increase the signal-to-noise ratio of single subjects ideally we average over all subjects given that subjects are consistent enough. To verify the consistency the ERPs on channel O2 for single subjects averaged over all epochs for condition 2D in Figure 2 are compared to each other and to the ERP of the grand averaged data, i.e., the ERP average over all subjects shown in the thick red curve in the same Figure. It is shown that the peaks in single subject curves have slight differences in amplitude and peak latencies which correspond to normal differences between subjects and therefore we found them consistent enough to be averaged. We have also checked the scalp topographies of single subjects for intervals of 50 ms starting from -200 ms up to 900 ms and observed strong similarities between subjects' topographies. Therefore in the following analysis we only focus on the grand averaged data.
We are interested to figure out how the corresponding differences between conditions shown in Figure 3 are behaving in different frequency ranges. For this reason, we have applied wavelet transformation to the ERP signals averaged over all epochs for single subjects and for each condition. The absolute values are then averaged over all subjects to be studied in the time frequency domain. The wavelet window applied in our analysis is a Morlet window. The baseline correction is applied to the wavelet transformation of averaged ERP of each subject by subtracting the mean of the absolute value of wavelet transformed data in the interval of -200 ms and 0 ms from the absolute value of the rest of the wavelet transformed data. In Figure 4 differences of the absolute values (second row) of wavelet transformations for each two conditions are plotted together with the p-values (first row) corresponding to each time point and frequency point. The statistical test is performed as it is described in Section Statistical Tests using Jackknife and student's t-test. The significance threshold (alpha) is corrected according to the FDR method for the time frequency window to consider the problem of multiple comparisons.

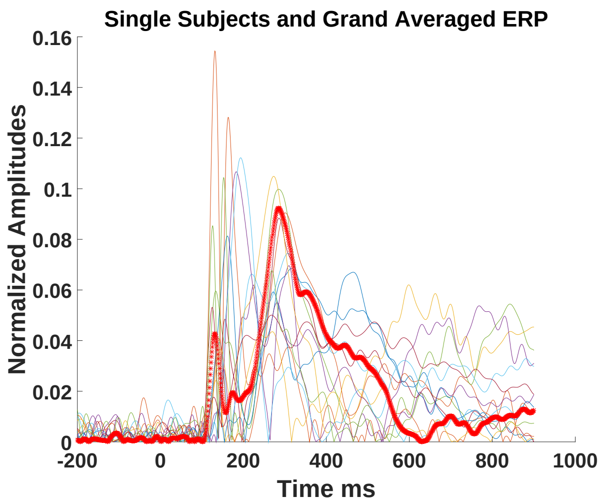


Figure 2: ERPs averaged over all epochs for each subject on channel O2 and condition 2D are plotted in the time interval between -200 ms and 900 ms. The thick line in red shows the grand average ERP, i.e. the average over all subjects. The absolute value of the curves are normalized by their norms. Single ERPs have different amplitudes and differences in latencies but in general they are consistent enough to be averaged over subjects.

Proceedings of the
7th Graz Brain-Computer Interface Conference 2017 DOI: 10.3217/978-3-85125-533-1-81

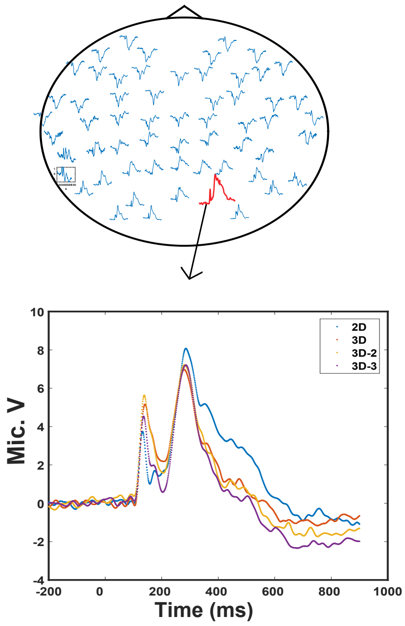


Figure 3: ERPs of grand average corresponding to each channel are plotted on the scalp in the top figure. On the bottom, ERps of channel O2 are compared between conditions.

Except for the condition 2D and 3D-3, the comparisons between 2D and 3D, 3D and 3D-2 as well as 3D and 3D-3 all show significant differences in the frequency band between 14 Hz and 17 Hz starting very early (almost 22 ms) after the stimulus onset up to 150 ms. This effect is shifted towards 200 ms for the comparison between 3D and 3D-2 and 3D and 3D-3 in the same frequency band. The brain activities in the frequency of interest (Beta band) in the occipital region, i.e. O2 channel in our analysis, are reported to be correlated with the level of attention in previous studies [18, 19, 20]. According to these studies, an increase in the attention caused the beta power in the occipital cortex to increase. In Figure 4.a the power in 3D is subtracted from the power in 2D which shows larger power values for 2D compared to 3D. This means the attention level in condition 2D was higher for subjects than in 3D. Figure 4.c shows higher power for 3D-2 than 3D and Figure 4.d shows higher beta power for 3D-3 compared to 3D. Logically the amount of information which is presented to the subjects in 3D-2 and 3D-3 is higher due to the vertical disparity simulated in these two conditions. In 3D-3 most of the subjects reported that the vertical disparity was very strong and noticeable. Since this parameter changes the stimuli to be a less normal stimuli it is expected to observe higher amount of attention in this condition compared to 3D which is confirmed by the results as well. However, what surprises us in the first glance is the higher attention level in 2D compared to 3D. It was expected that the depth information in 3D increases the attention level in the subjects in contrast to our results. Our hypothesis in this case is that this effect is caused by a change in the dimensionality of presentation in the 2D condition. As it was mentioned before, a three dimensional cross is presented before the stimulus onset in all conditions. In the case of 2D stimulus, there is a switch from a three dimensional cross to a two dimensional cube which might be the reason of higher attention level in 2D. However the 3D condition follows a three dimensional cross and therefore no switching is happening in the dimensionality of the cross and the cube and therefore less increase in the attention level is observed.

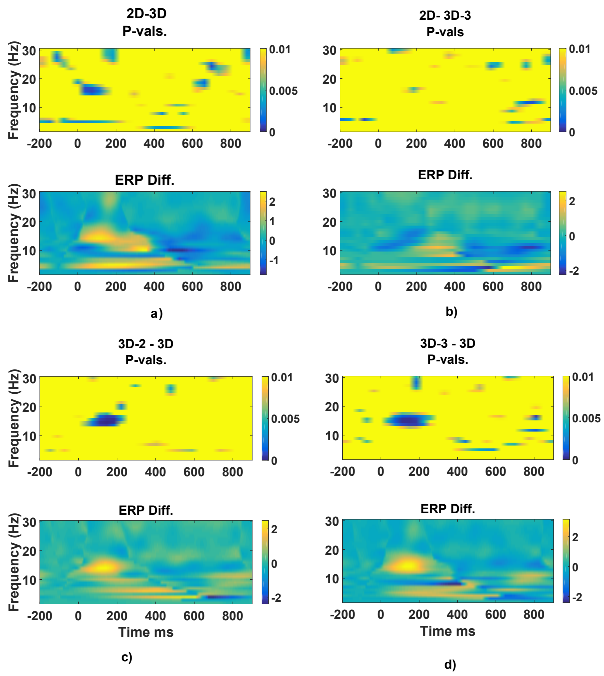


Figure 4: Time frequency analysis on the differences between conditions: Each figure contains two plots. The plot on the second row shows the difference between the absolute values of wavelet transformations for two conditions and the plot on the first row shows the corresponding p-values.

CONCLUSION

In this study the simulated vertical disparity in stereoscopic images was the focus of an objective test for the quality of experience. Time frequency analysis of ERP components has shown a significant increase in the beta band power in 3D-2 and 3D-3 compared to the 3D condition. In previous studies the increased occipital beta power has been linked to the increase in the attention level. An increased attention level in 3D-3 and 3D-2 compared to 3D was observed which sounds reasonable due to extra unexpected amount of information in these two conditions due to the vertical disparity component. We suggest that an increase in the occipital beta power can be extracted from the data as a feature correlated with an increase in the vertical disparity of stereoscopic images even before the subjects show first signs of visual fatigues. However we have also observed an increase in the beta power for 2D compared to 3D. This increase in attention can be explained by the change in the dimensionality

 DOI: 10.3217/978-3-85125-533-1-81

of the cross presented before the cube in 2D. Since the cross is presented in 3D in all pause intervals, the switch happening between the three dimensional cross and the two dimensional cube has been probably the reason of higher attention for 2D condition than the 3D condition in spite of our expectation for a higher attention in 3D than 2D.
In future work we will investigate multimodal [21] and deep [22] approaches for quality assessment of stereoscopic images.

ACKNOWLEDGMENT

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

References

[1] 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, pp. 138–141, 1850.

[2] 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.

[3] 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.

[4] 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.

[5] 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 *IEEE Int. Conf. Image Process. (ICIP)*, 2014, pp. 1987–1991.

[6] 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.

[7] 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.

[8] 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.

[9] J. Frey, A. Appriou, F. Lotte, and M. Hachet, "Classifying EEG Signals during Stereoscopic Visualization to Estimate Visual Comfort," *Comput. Intell. Neurosci.*, vol. 2016, 2016.

[10] 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.

[11] 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.

[12] 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.

[13] A. S. Malik, R. N. H. 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.

[14] 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.

[15] B. Blankertz, L. Acqualagna, S. Dähne, S. Haufe, M. Schultze-Kraft, I. Sturm, M. Ušćumlic, M. A. Wenzel, G. Curio, and K.-R. Müller, "The berlin brain-computer interface: Progress beyond communication and control," *Frontiers in Neuroscience*, vol. 10, pp. 530, 2016.

[16] R. Ulrich and J. Miller, "Using the jackknife-based scoring method for measuring lrp onset effects in factorial designs," *Psychophysiology*, vol. 38, no. 5, pp. 816–827, 2001.

[17] C. R. Genovese, N. A. Lazar, and T. Nichols, "Thresholding of statistical maps in functional neuroimaging using the false discovery rate," *NeuroImage*, vol. 15, no. 4, pp. 870 – 878, 2002.

[18] M. Gola, M. Magnuski, I. Szumska, and A. Wróbel, "EEG beta band activity is related to attention and attentional deficits in the visual performance of elderly subjects," *International Journal of Psychophysiology*, vol. 89, no. 3, pp. 334–341, 2013.

DOI: 10.3217/978-3-85125-533-1-81

[19] A. Wróbel, A. Ghazaryan, M. Bekisz, W. Bogdan, and J Kamiński, "Two streams of attention-dependent beta activity in the striate recipient zone of cat's lateral posterior–pulvinar complex," *Journal of Neuroscience*, vol. 27, no. 9, pp. 2230–2240, 2007.

[20] A. Wróbel, "Beta activity: a carrier for visual attention," *Acta Neurobiol Exp*, vol. 60, no. 2, pp. 247–260, 2000.

[21] 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.

[22] 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.