Using Visual Stress EEG data to create an Auto Focusing system

SAIF ALI KHAN, MUHAMMAD SOHAIB ARSHID, LARS FRANKE, and PAUL SZYMAN-SKI, Saarland University, Saarland

This paper includes a first of its kind auto focusing system, which uses real time visual Stress EEG data to recognize the focus of a screen and readjust it. Although being in its preliminary stages, it promises a system which can be helpful in many industrial and healthcare applications.

Additional Key Words and Phrases: Electroencephalography (EEG), Power Spectrum (PSP), Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA)

ACM Reference Format:
Saif Ali Khan, Muhammad Sohaib Arshid, Lars Franke, and Paul Szymanski. 2020. Using Visual Stress EEG data to create an Auto Focusing system. 1, 1 (September 2020), 8 pages. 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Passive Brain Computer Interfaces (or otherwise referred as BCIs in this paper) are a way of automating (user) interfaces by analyzing brain activity and adapting the user's environment based on the said activity. One of the most frequently used techniques in this case is the measurement of the user's mental state e.g. his/her cognitive workload in order to adjust the complexity of a user interface.Those mental states can be derived from brain activity, mostly by frequency analysis and can be used as an input for a system to automatically react to changes in this state. Another possible measurement is visual effort or stress. This measure is of high interest when it comes to scenarios in which a person has to work a longer period of time in a visually demanding setting, e.g. looking at a screen or through some sort of optics. The parameters of those screens can be adjusted to better fit the current circumstances and the requirements the user has in his current situation. VR settings are a possible field of application, where there are a variety of different adjustments that can be made to best fit the users' needs and those adjustments might even change with a longer lasting VR session. Measuring visual stress while wearing the headset could enable automatic adjustment of the headset parameters in order to adapt to the changing requirements of the user over time. The need for automatic adjustment of settings is even stronger in scenarios which requires a change in focus of a scene, e.g. surgical microscopes. Surgeons need to adjust the focus of their operational microscope frequently depending on changes of the operational scene during a surgery. An automatic adjustment of the focus based on the current visual effort/stress of the surgeon could be a game changer in this field of application. As the research in this field is still in its infancy, we will try to develop and prove basic concepts of visual stress measurement and create a first prototype for automatic parameter adjustment based on this measurement.

2 RELATED WORK

This chapter will briefly go over the state of the art work that has already been done and experimented with in the light of EEG based measurements for visual fatigue. This chapter will go through the research one by one and discuss briefly the approaches.

Authors' address: Saif Ali Khan; Muhammad Sohaib Arshid; Lars Franke; Paul Szymanski, Saarland University, Saarland.

XXXX-XXXX/2020/9-ART $15.00
10.1145/nnnnnnn.nnnnnnn

2.1 Assessment visual fatigue of watching 3D TV using EEG power spectral parameters [1]

Several methods have been proposed to measure 3D TV fatigue. Subjective evaluation done with questionnaires is a reliable method for evaluating an individual's fatigue state. However, it is affected by individual differences and there are non-uniform evaluation criteria. Objective measurements are based on biological signals collected by electrocardiograms (EOG), electrocardiograms (ECG), electroencephalograms (EEG), galvanic skin response (GSR) or photoplethysmograms (PPG). Out of these the EEG signal is perhaps the most promising, reliable and predictive indicator. Prior research propose multiple indicators and methods to quantitatively analyze the EEG signal. However many models need too many parameters and lack accuracy and stability. The authors found in prior research that gravity frequency and power spectral entropy have strong correlations with the fatigue status. This work analyzes these two parameters and compares them with subjective questionnaire evaluations.

From the experiments and user study, the authors find out that the subjective self-reported fatigue increased more significantly in the 3D group than the 2D group. Most subjects reported dizziness, blurred vision and dry eyes after 60-minute watching 3DTV. The power spectral entropy of the electrode P3 was dropped because it did not obey normal distribution using the Shapiro-Wilk test. Booth the gravity frequency and the power spectral entropy changed more significantly in the 3D group than in the 2D group on most electrodes. The experiment indicates that gravity frequency of power spectrum significantly decreased (p less than 0.05) on Fp1, Fp2, F3, F4, C3, C4, P4, O1, F7, F8 and T6 electrodes after watching 3DTV, while power spectral entropy clearly decreased (p less than 0.05) on Fp1, Fp2, F3, F4, C3, F7 and F8 electrodes.

In comprehensive consideration of the authors previous work, the detection result of Fp1 is the most robust and effective in all locations. So in the fatigue evaluation model Fp1 is still chosen as the significant electrode. The coefficients for the fatigue evaluation index are found using linear fitting on EEG data. 80 percent of the data was used to fit the coefficients. The 20 percent remaining data sets was used to assess the fatigue model. The accuracy of the model was 87 percent.

2.2 Estimating Visual Comfort in Stereoscopic Displays Using Electroencephalography: A Proof-of-Concept [2]

Inspired by all the previous work done in this field, the author designed and tested a system that classifies EEG data to measure visual comfort. This type of system is a passive brain-computer interface. Their main contribution is to prove the feasibility of an EEG system that could estimate in near real-time (1s delay) the visual comfort viewers are experiencing as they watch stereoscopic displays. It could be adapted to real-case scenarios by controlling the discrepancy between left and right images depending on the output of the classifier. Then it could be employed in different settings to improve HCI by easing users' comfort, for example when they manipulate 3D contents during prolonged periods of time remote design, video games, and so on or when people are watching 3D movies especially when there are many rapid depth variations, as seen in action sequences.

It was found out that during short exposures to images, participant reported worse vision clarity and less visual comfort in NC condition, thereby validating a clear distinction between the two zones of comfort of our protocol. Participants performed equally well in both conditions during the task, suggesting that even if severe, a VAC does not alter their ability to make rough estimations of virtual depths. In this context, it also highlights the limits of behavioral methods for measuring participants' comfort. A neuroimaging technique, on the other hand, did manage to discriminate two comfort conditions.

The score of 63 percent accuracy for the classifier, while not as high as some other established BCI systems, may be already sufficient to improve users' comfort. Indeed, on-the-fly correction of uncomfortable images can be seen as error correction. In such settings, detection rates from 65 percent are acceptable to improve interactions. These findings depend on the nature of the task; this is why we proposed a mechanism to increase the performance of the classifier by taking into account more than one object appearance.

2.3 Estimating Visual Discomfort in Head-Mounted Displays using Electroencephalography [4]

Being a novel technology, use of VR without proper configurations (Brightness, disparity of objects being displayed) can add mental workload or visual stress on the user. While wearing the Head-Mounted Display (HMD), a user can't simply look away from the display (like in normal TV or mobile screens) to relax his/her eyes. That couples with objects in the display really close to the user's eyes, causes visual stress. When objects appear really close to user's eyes, the disparity between left and right eyes picture increases and vergence-accommodation conflict (VAC) occurs. VAC is the phenomenon when your brain receives mismatching cues between the distance of a virtual 3D object and the focusing distance required for the eyes to focus on that object.

Participants report more visual discomfort in NC condition than in C condition. The symptoms got worse from eye strain, to double vision without the ability to fuse the stereoscopic picture the closer the object appeared in the NC condition. This means our EEG measurements represent the users' actual experience. The EEG analysis gives promising insight on detecting visual stress either through monitoring the parietal or occipital lobes of the brain with 2 electrodes. The high classification rate of 71 percent for C and 83 percent for NC condition when using the O1 and O2 electrode [2] proves the applicability of using a BCI to detect visual stress within a HMD. Furthermore, the finding that only two electrodes are needed make it easy to wear as it might be integrated into the headstraps of an HMD. Also it makes it a relatively low-cost tool that does not interfere with the users experience at all.

3 MATERIAL AND METHODS

3.1 Experiment

Our preliminary goal was to create an auto focus system using EEG data in a VR setting. The scene required a surgical table with some task which would be set up in Unity and then we would collect EEG visual stress and not stressed data for our classification model. But, as we progressed in our setup for the VR, things became rather cumbersome. The task at hand in the VR setting did not require any significant thought process and hence, through our deductive argumentation, we came to the conclusion that since the task is not significantly challenging, it might not be a good idea to use the experimental setup per se for collecting EEG data for our model. Rather, we completely deviated from our VR setting and moved on to 2D screen.

There were two phases for setting up the experimental environment on a 2D screen. Our first approach involved us creating a simple web-app with numbers filled on the whole screen, from left to right, and up to down in a symmetry (like in Figure 1). Over the course of the experiment the numbers would be focused or blurred. A single experiment took 160 seconds, with blurred (out of focus) and non blurred screen taking 20 seconds for each iteration. A number was assigned to the participant, which he had to count during the whole experimental period, during focused and not focused screen visualization. Despite being a promising task of getting wholesome data, there were still some discrepancies in the setup. This particular approach was not reflective to the use case we wanted to implement, that is of a surgical microscope setting. Moreover, when counting numbers, participant complained that they had to move their head while reading, so the artifacts introduced into the data because of that head movement was potentially polluting the data.

440880750857835266441337587256231662
862468693053767802759878393394334745
385489658644780782175672090502281557
527857889744568320839187829497049551
224836212638505317781472101205468275
150713604514421856785366268353478426
985172102630462942542944231246853961

Fig. 1

Finally, after mutual discussion and brain storming, our final experimental setup looked like Figure 2. The patch of numbers only fills a small section of the screen (mimicing a Petri dish), so that the eyes wouldn't have to move much. Moreover, the task assigned to the user now involved counting numbers and once counted, entered into the system, using a keyboard, to check at the end the error rate of the user in counting numbers. This strategy was only used to keep the participant vigilant and focused. This would continue after the 160 second experiment was over.

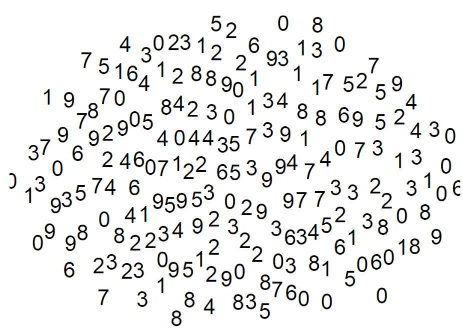


Fig. 2

As already mentioned, each user had to sit through the experiment for 160 seconds. Because of the non availability of participants, we utilized our team members for the task. Each user, counted a number which was assigned to him, and he had to find that number in the cluster of numbers visualized on the screen. Once counted, he entered the number using the keyboard and hit Enter, after which a new Petri dish setting appeared, and the user had to continue counting again until 160 seconds were over. The data from the EEG headset was being recorded and saved into files for each of the study for future processing and EEG data analysis.

For collecting the EEG data we used the g.tec Unicorn headset. It has 8 EEG electrodes with positions as shown in Fig. 3. It uses two reference electrodes behind the ear lobes. The headset can be used with gel to improve the signals. After having a first look at the data we collected we assumed we would be ok with not using gel. However, when we actually tried to train models with that data we couldn't get models that were better than guessing. So we repeated some experiments with gel. That data was much better for training models. However, because of lacking time we couldn't repeat many experiments. In total we conducted 46 experiments without gel and 4 experiments with gel.

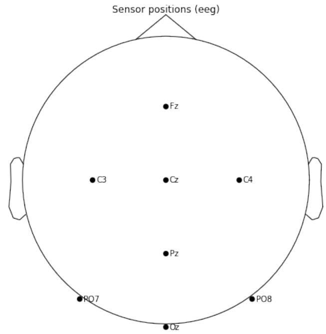


Fig. 3. Sensor Layout

3.2 EEG Data Analysis

3.2.1 Preparation. We first passed the data over a 2 Hz – 40 Hz bandpass filter. We then set the reference to be the common average reference of all channels. We Split the data into one second epochs. We used the Autoreject package [3] to automatically find a threshold on which epochs with high peak-to-peak amplitude are rejected. Figure 4 shows the rate of bad data per channel. At last we normalized booth epoch batches so each batch has the same amount of data to prevent bias later on when training models. This dropped 39 more epochs (for gel data). We were left with 229 epochs per state (focused/ blurred). After seeing that channel Fz was the cause of many rejections we decided to remove Fz completely from our data. That resultet in having 10 more epochs per state (i.e. 239 epochs per state).

3.2.2 Exploring the data. We used the MNE library f We first plotted the power spectral density (PSD) (Fig. 5) and topomaps (Fig. 6) for booth experiment states to get an feeling how significant the states differ in the EEG data. The differences didn't look particularly emphasised.

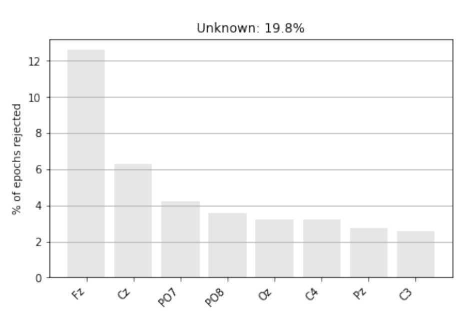
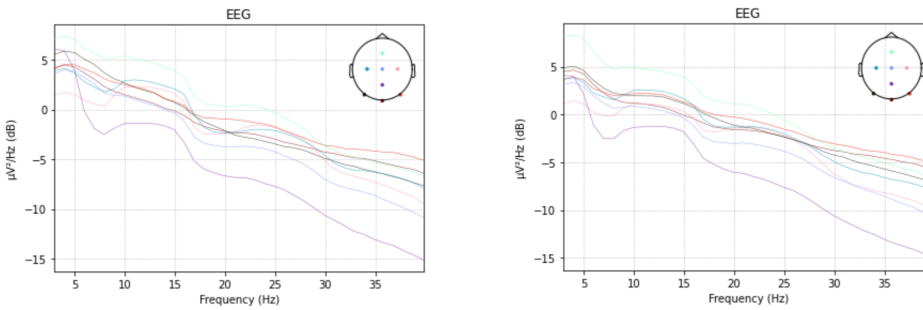


Fig. 4. Rejected epochs due to high peak-to-peak signals

3.3 Machine Learning Models

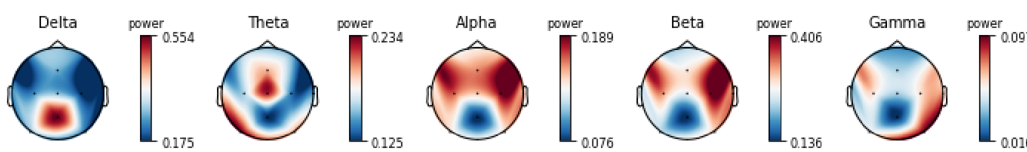
This sub section will highlight the traditional machine learning approaches and shed some light on more complex deep learning approach that were used for the binary classification of our EEG data.

3.3.1 Support Vector Classifier. We started analyzing our data using the basic Support Vector Classifier for our EEG data. Our argument for using a Support Vector Classifier (henceforth

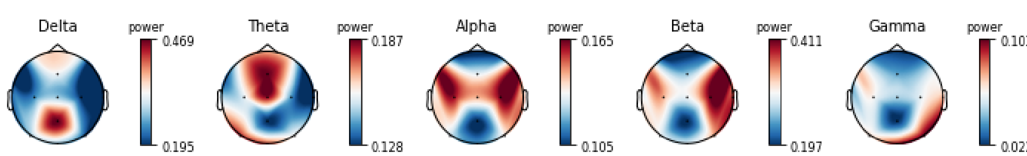


(a) Focused (b) Blurred

Fig. 5. Power spectral density over all Epochs (with gel)



(a) Focused



(b) Blurred

Fig. 6. Frequency band separated topomap for the two experiment states (with gel)

mentioned as SVC) was to first experiment the classification capabilities of state of the art basic machine learning algorithms, and we decided to use SVC for the classification task. We used sklearn for our machine learning pipeline, where the data was split into train and test sets. For the gel data, we had 320 training sample out of total dataset of 478 samples. On the contrary, for the data which was recorded without gel, out of 4302 total samples, training is 2882 samples. Evaluation numbers for both sets of data has been mentioned in the results section. Before moving on, it is necessary to mention that in our SVC pipeline, we actually compute the Power Spectrum Density (PSD) of the data and then feed it into the SVC.

3.3.2 Neural Network Classification. After our basic machine learning algorithm analysis, we moved onto experiment and test with more complex, neural network models, since our literature study suggested neural networks to be really good at classifying raw EEG data. Our model architecture is as follows:

For the hyperparameters, we use binary crossentropy as the loss function, adam as the optimizer and then train the model for 200 epochs. Using neural networks did not significantly increased the accuracy of the model, whether it was gel data or without gel. We argue that it is because we dont

```
model = keras.Sequential(
    [
        keras.Input(shape=(X_train.shape[1],)),
        layers.Dense(1024, activation="relu", name="layer1"),
        layers.Dense(512, activation="relu", name="layer2"),
        layers.Dense(256, activation="relu", name="layer3"),
        layers.Dense(1, activation="sigmoid",name="layer4"),
    ]
)
```

Fig. 7. Model Architecture

have much data to feed the network, and since we know how hungry neural networks can get for data, it was a shortcoming in our data collection because of which it was not able to perform quite as well as expected.

4 RESULTS

The accuracy of the model trained on the data collected by using the EEG-Headset without gel was just about 50% for SVC as well as for the neural network, which means that we could as well choose randomly if the image is blurred or not. When we used SVC trained on the data collected by using the EEG-Headset with gel however, the accuracy was significantly higher which was 68% but the accuracy for the neural network still roamed around 50%, which meant we did not see significant improvement in the neural network. Sadly most of our training data was collected without using gel and therefore the model that used gel had a low amount of training data.

4.1 Application

The Application setup needed to resemble as closely as possible to the training/experiment setup. Before starting the application, the user can select the blurriness level of his/her own preference. Random integers would appear on the screen enclosed in an invisible circle, and the user would start counting the amount of one particular integer visible on the screen. After twenty seconds, all the integers will get blurred and user would have to continue the count as before. After another twenty seconds, new set of random integers will appear without any blurriness, and the cycle goes on.

The Application works on two threads mainly. One thread deals with the twenty second image cycles, while the other thread simultaneously deals with reading the live stream of EEG-signals. The stream of EEG-signals is captured approximately after each second, which gives a total of 250 samples from seven chosen channels. These sample point are processed and filtered, and fed into the already trained Machine Learning model, which outputs the prediction for blurriness of the image.

5 DISCUSSION

One of the main issues faced in getting a good enough accuracy was the lack of data. Because the number of features being used in a sample to be trained was 1757, huge amount of training data was needed to combat the issue of over-fitting.

In the early stages, the experiment was supposed to be based on VR setup, but because of the issue with controlled blurriness and stress of just wearing a heavy VR-headset for a prolonged period of time, the setup was moved to a Monitor / TV screen instead. Another big issue was attaining perfect environment for experiment. The room had to be nearly sound proof, with good temperature conditions. The user was seen exhausted after several experiments, because the room got really warm when the windows and door were closed to make it sound proof. It was understood that such scenarios might also happen in real world, but this can produce noise in training data. Furthermore, experiment setup was redesigned a couple of times after it was noted that user had to move his head to find the integers. The continuous movement of head produced a lot of data that was unusable for training.

In order to keep the data clean and consistent, a lot of it had to be thrown out. By doing that, although we made our training data consistent, but we suffered from a phenomenon called 'Curse of Dimensionality', which happens when we surpass the ratio of number of features with data points. To combat this hurdle, PCA was used for dimensionality reduction and LDA was used as feature selection method, but it reduced the train and test accuracy even more.

6 CONCLUSION AND OUTLOOK

With more training data (especially more gel data) it is probably possible to create some really helpful applications by measuring the visual stress of a person that looks at a blurred image. The first goal of our experiment was to make an application that a doctor could use in a surgery to focus the parts of the body where she is looking at. At the current stage of development however, the model wouldn't do a good job at it and therefore it is not usable in practice. Even with an accuracy of 70% most doctors would probably prefer changing the focus of the microscope manually instead of letting the model decide.

There is not much research done in this field of science yet. With more testing and developing people could very well create applications that are build on the same theory we used for our experiment, which have uses in the daily life of some people.

REFERENCES

[1] Chunxiao Chen, Kun Li, Qiuyi Wu, Haowen Wang, Zhiyu Qian, and Gu Ning. 2014. Assessment visual fatigue of watching 3DTV using EEG power spectral parameters. *Displays* 35 (12 2014). https://doi.org/10.1016/j.displa.2014.10.001

[2] Jérémy Frey, Aurélien Appriou, Fabien Lotte, and Martin Hachet. 2015. Estimating Visual Comfort in Stereoscopic Displays Using Electroencephalography: A Proof-of-Concept. arXiv:cs.HC/1505.07783

[3] Mainak Jas. 2016. Autoreject. http://autoreject.github.io/

[4] Christian Mai, Mariam Hassib, and Rolf Königbauer. 2017. Estimating Visual Discomfort in Head-Mounted Displays Using Electroencephalography. In *16th IFIP TC 13 International Conference on Human-Computer Interaction — INTERACT 2017 - Volume 10516*. Springer-Verlag, Berlin, Heidelberg, 243–252. https://doi.org/10.1007/978-3-319-68059-0_15