**Constructing an economic Lie Detector based on EEG Data**

*Group Members:* Soumyadipta Banerjee 2018H1030044G

Sarah Biju 2018H1030188G

Madhusmita Oke 2018H1030194G

***Phase I Project Report***

Problem Statement:

The correct assessment of whether some suspect in a criminal incident is telling the truth or not when interrogated is of great importance when it comes to pinpointing the actual perpetrator of the act. Existing methods such as the usage of polygraphs for monitoring vital signs, skin conductivity, etc. have provided us a plausible way to tell apart the lies from the truth, although such methods require complex set-up procedures and can possible be tricked. But recent studies in Neurosciences have shown that the act of lying also results in certain detectable changes in a person’s brain activity, which are registered as variations in electrical impulses generated in the cortical region of the brain. By employing electroencephalography, we can capture these variations as visual EEG data which can then be fed to appropriate machine learning algorithms to learn the features in that dataset.

With the advancement of Brain Computer Interfaces in the past decade, several companies have made available easy to use commercial EEG detection headsets that can acquire such data with ease and has expanded the reach of these technologies to a wider user community. The aim of this project is to use one such headset, the single electrode dry Neurosky Mindwave Mobile, to acquire EEG data and construct a lie detection mechanism employing appropriate machine learning methods to learn the features in the constituent frequency bands that make up the EEG data. It is expected that the electrical variations that are reflected in a subject’s EEG on lying and on telling the truth will be picked up by this specific headset and can serve as usable data to train an ML model to later classify lies from truths.

Problem Definition:

The problem to be addressed can be defined as follows:

*Given a subject to be interrogated, we have to devise an easy to use, economic Brain Computer Interface to utilize the EEG data generated during the interrogation via appropriate learning algorithms to determine if the subject is lying or not.*

Literature Review:

There have been a number of attempts in the recent past to deal with this problem of lie detection using EEG data, reporting varying accuracy. Although conducted by different groups of researchers, there are two features that stand out in all these works:

* Most of these works seem to use very expensive multi-channel electrode headsets that are very time-consuming and difficult to set up, often involving special conductive solutions that need to be applied to the contact pads and other additional materials that make it a cumbersome process altogether.
* Most of these attempts focus on identifying the P300 Event Related Potential, a specific electrical impulse that shows up in the EEG data when a person is subject to certain “oddball” stimuli (e.g. an image of the murder weapon). The researchers aim to train the ML model to identify such ERP’s so that they can be later detected when a guilty person is interrogated regarding the crime committed. This impulse can often be difficult to detect in the presence of noise and artifacts and, again, require complex, expensive equipments that may not always be affordable.

Some of the notable contributions in this direction are listed below:

* Haider et al. [1] devised a way to detect lies based on the afore-mentioned idea of detecting the P300 ERP where the patterns in the data were learned by using Linear Discriminant Analysis. For acquiring data, the Emotiv EPOC 14 channel wet headset was used and the entire software process was implemented on Matlab and FPGA tools. An accuracy of around 85% was achieved after testing it on 15 to 20 subjects.
* The work done by Simbolon et al. [2] again made use of the P300 detection principle and employed the Support Vector Machine approach to learn the relevant features in the data. The 19-channel ElectroCap headset was used, and a Matlab based program was constructed to implement the learning mechanism. The entire set-up was tested on 11 male subjects in their twenties and an accuracy of around 70% was reported.
* Relationships between the EEG generated in the frontal lobe of the brain and the act of lying were reported in the work of Cakmak et al. [3]. Feature extraction from the dataset was performed with the use of Multiplayer Neural Networks and Short Term Fourier Transform. Again, the Emotiv EPOC 14 channel headset was used in the data acquisition phase and an accuracy of around 90% was achieved.
* Usage of F-Score and Extreme Learning Machines formed the basis of the work by Gao J. et al. [4] to detect the P300 ERP’s in guilty subjects, apart from optimizing the number of hidden layer nodes in the neural networks.
* Multimodal data acquisition was leveraged in the work of Gupta et al.[5]. Video, audio, EEG and eye movements were collectively analyzed. The combined effect of these modalities and their permutations is major contribution of this paper.
* Different types of deception account for varied neural brain activity. fMRI study monitors different dimensions in this respect. Effects of spontaneous and memorized lying were studied as well.

Action Plan:

The EEG data that represent a brain’s electrical activity can be categorized into the following frequency bands, each band having a direct relationship to the thought process of the person under consideration:

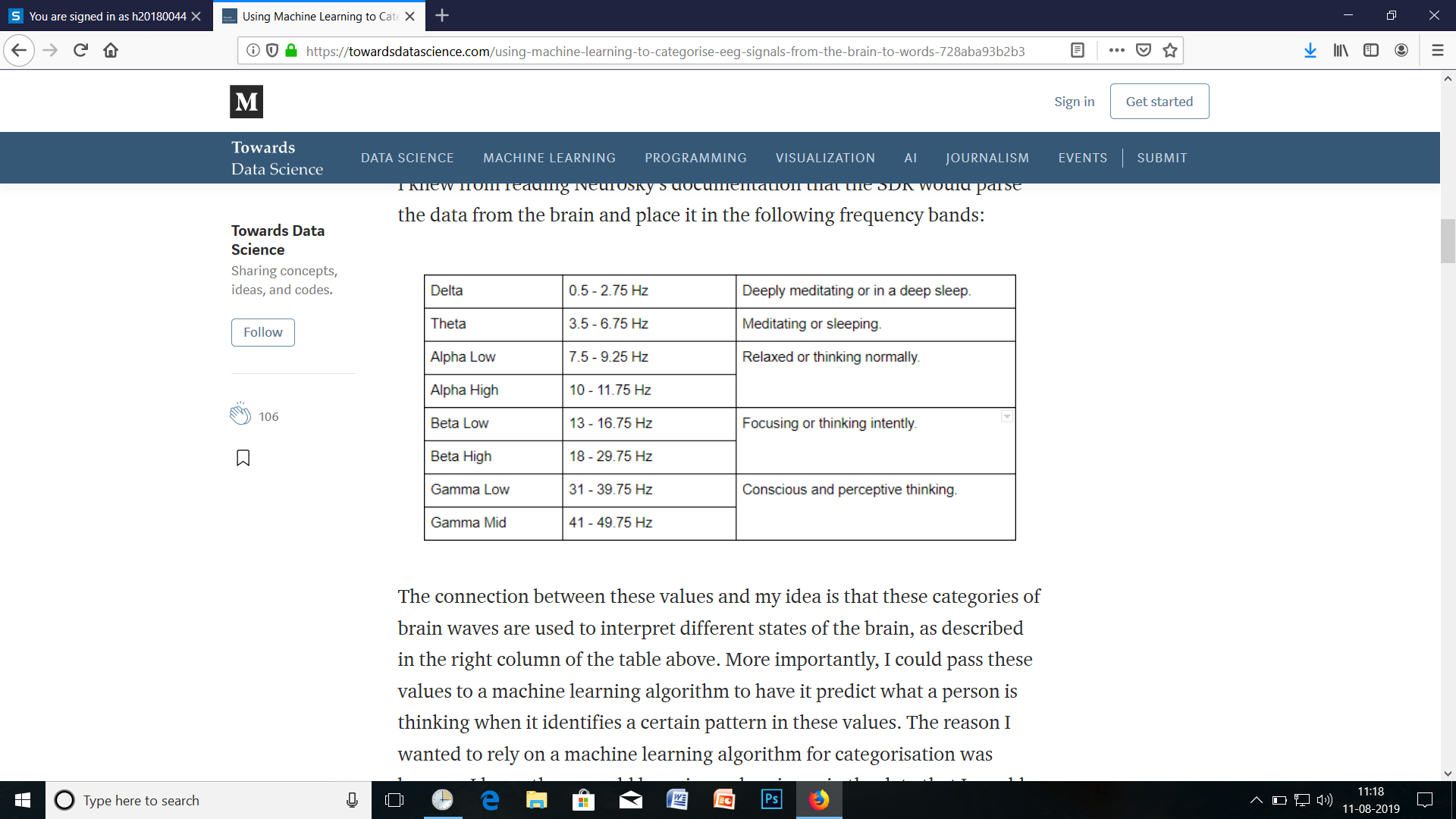


Fig.1: Frequency bands of Brain waves and their correlation to brain activity

The basic idea behind this project is that an innocent subject is without any fear of apprehension will have a normal and relaxed thought process. Thus his/her brain activity will be dominated by the Alpha band of frequencies. On the other hand, a person guilty of some crime will always have some fear of being caught and therefore his/her brain activity will be more agitated and tense, thus skewing more towards the Beta and probably the Gamma band of frequencies. Such aspects of the EEG data should be captured by the Neurosky Maindwave Mobile headset, which is a single electrode dry headset, considerably cheaper and easier to use than other similar equipments. The headset does indeed capture the band separated EEG data and thus the above-mentioned features can be fed to an autoencoder that would primarily reduce the data features to just those important ones that can tell apart a truth from a lie.

If, for instance, we can train the autoencoder on EEG data generated during some baseline questions, answers to which are generally known to be true, the dimensionality reduction and reconstruction by the autoencoder will be skewed towards EEG data related to situations where the subject doesn’t lie. Now during the testing phase, if we feed the autoencoder with EEG data generated when the subject is possibly lying, the autoencoder would not be able to successfully reconstruct the data as it would differ in nature from the data with which it was trained. So anomalies in the test EEG data might be a possible indication that the subject is lying.

This, for now, is the plan of action with which the project will proceed.

***Phase II Project Report***

Data Acquisition

Our initial attempts to capture data using the Neurosky Mindwave Mobile headset, as mentioned in the previous phase plans, was not entirely successful as the data obtained was not sufficient and accurate enough. This was due to the fact that the headset had just a single sensor and it’s precision was not up to the requirements of our experiment. This prompted us to use a more advanced headset.

We used the Emotiv EPOC+, a 14-channel EEG wet headset to acquire data relevant to our project. A python-coded open-source software utility CyKIT ran in the background which connected to the headset via a wireless Bluetooth connection. The utility captured streaming signal data from the device and sent it to OpenVibe, an EEG signal analysis tool, which enabled us to store the captured data stream as 14-channel wide CSV file. The sampling rate of the signals was 128 Hz. To test our hypothesis as outlined in the previous phase, we conducted the following experiment:

One of us was selected as the subject whose EEG signals will be captured. For each session of the experiment, the subject was asked to first describe an image (obtained arbitrarily off the Internet) as it is, truthfully. The corresponding EEG signals were captured for a period of 1 minute. The subject was then asked to describe another arbitrary image deceptively and, in doing so, completely deviate from the actual content of the image shown. This, as we expect, would lead the subject to generate *spontaneous improvised lies*, which requires and more intense thought process than describing an image as it is. Again, the EEG signals were recorded for a period of 1 minute.

Signal Processing

The EEG signals obtained from our experiments were simply raw voltage samples in the range of micro volts. To obtain a better understanding of the nature of the signals obtained and how it fits with our working hypothesis, we performed Fourier analysis using a Python-coded software utility. This yielded a band-wise power spectrum of the EEG data, indicating the power of each frequency band. The EEG data was broken into separate channels, (AF3, F3, F7, etc.) and the spectral analysis was performed on data from a channel of interest.

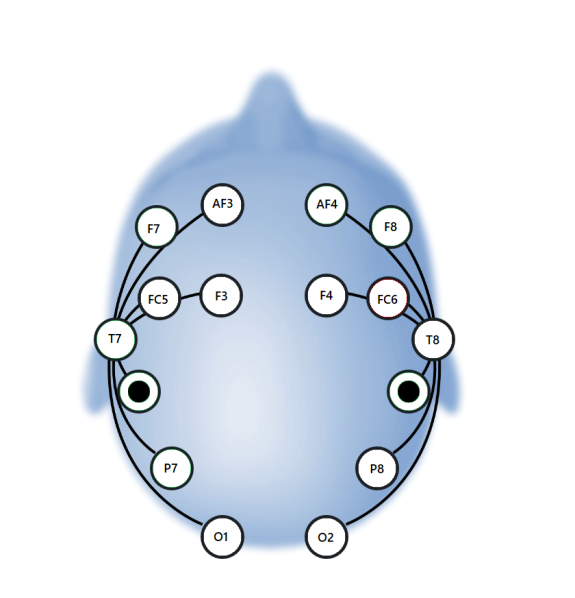


Fig. 2: Position of the electrode sensors and the corresponding channels they represent for Emotiv EPOC+ headset

We focused our attention on the frontal lobe, which is primarily concerned with higher brain functions related to intense thought processes, judgment, and problem solving. We show here the resultant power spectrum obtained for channel F3 in the frontal lobe of the left hemisphere.

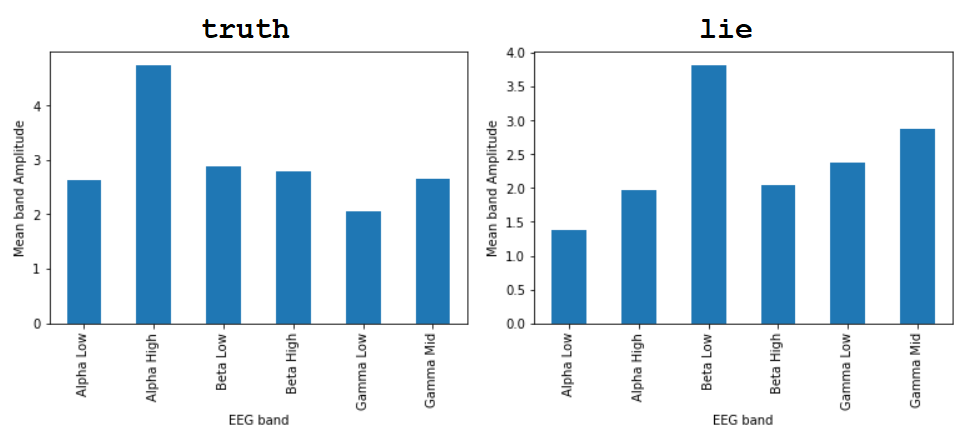


Fig. 3: Power Spectrum corresponding to truth and lie scenarios for F3 channel

From the above power spectrum figures, it is apparent that there is an *increase in beta activity* when the subject lied or falsely described the images. This signifies that the subject’s mind was more excited and her thought processes were more intense considering she had to come up with improvised, spontaneous descriptions that were completely imaginary and contradicted the subject matter of the image presented. This was in support of our hypothesis, which stated that the act of composing a lie and “making up” a false, fictitious scenario, increases brain activity and thus generates more brainwaves in the higher frequency bands.

The above spectral analysis thus enables us to understand which part we would focus on in order to look for the data that can segregate truth-telling scenarios from those where the subject is lying. We expect that the visual demarcations between the two scenarios will be reflected in the machine learning model once we train it on our processed dataset.

*Delay in completion of the basic machine learning model’s training*: The Emotiv EPOC+ headset is a complicated piece of research-grade hardware that requires knowledge of several third party tools and utilities (as Emotiv does not allow for an open source option to extract data) and good understanding in the basics of EEG signal processing. Since we have now used this device for the first time to acquire data, and are new to the field of Brain Computer Interfaces and EEG signal processing, a considerable amount of time was consumed in correctly understanding and implementing the technical requirements that can effectively collect and store data from the device, and then properly process the raw data according to our project’s requirements. Owing to this, we were unable to complete the third and final part of this phase, i.e. training a basic machine learning model on our processed dataset. However, we are positive that this will be completed in the early stages of the third phase, and we can display the variations discussed above in the trained model when we consider more complex machine learning models.

***Phase III Project Report***

*Note:-* For this phase we required more data and hence longer sessions of the experiment we carried out for phase II. So, we began with the task of acquiring this data, as discussed below.

Further Data Acquisition

This time, we carried out 18 minute sessions for each of the two scenarios, i.e., one where the subject truthfully describes images shown to her and the other where the subject lies about the contents of the images shown. Nine random images were shown in each session, with the subject having to describe each of them for 2 minutes. The corresponding EEG time varying signal samples were recorded as in phase II. The acquired data was stored in two CSV files, one for each scenario. As before, the CSV files contained signal samples across all the 14 channels, the data being recorded at a sampling frequency of 128 Hz. The same subject form phase II participated in these sessions.

Data Processing

Since the EEG signal obtained had noise arising from the skin and from other factors, the data was filtered using a Butterworth bandpass filter, with the pass band lying between 3 Hz to 30 Hz, the range where the desired alpha and beta bands could be located. This was achieved via OpenVibe, where we used the *Temporal Filter* component with the desired passband and filter type parameters, taking in data from the unfiltered CSV file via the *CSV Reader* component and writing the filtered values in a separate file using the *CSV Writer* component. This is shown in the screenshot below.

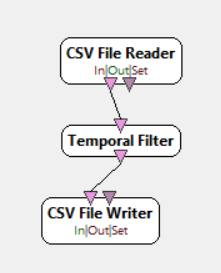


Fig. 4: Using an OpenVibe “scenario” to bandpass filter the CSV files

Each filtered CSV file was then segregated into 30 second, or 3840 tuple-wide, windows. As before, we focused on the data from the F3 channel, located above the left frontal lobe of the brain. Fast Fourier Transform was performed individually on each of these windows to obtain the power spectra values across all the bands for a window. The following figure shows two window samples, one each from the truth and lying scenarios. Higher alpha activity was mostly noted from the windows taken from the truth scenario, whereas those from the lying scenario had mostly higher levels of beta activity. However, there was some degree of deviation from the expected trend due to the varying attention levels of the subject.

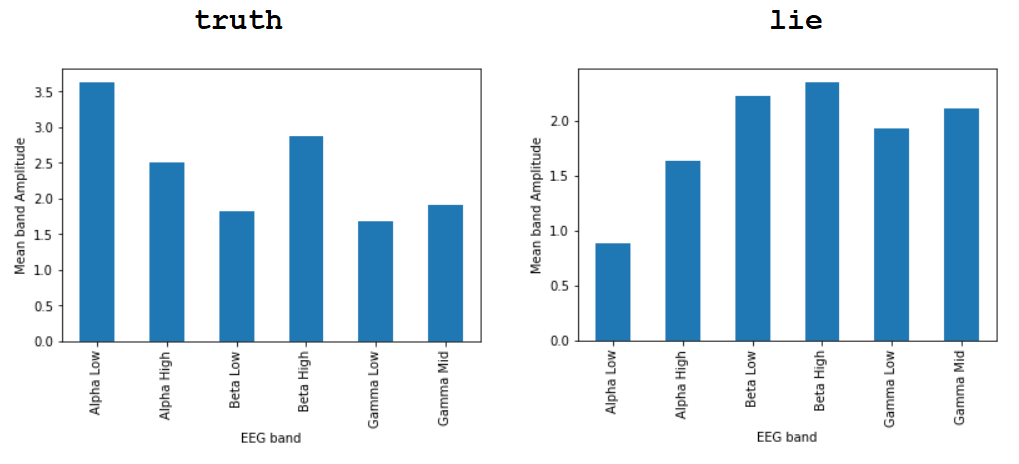


Fig. 5: Power Spectrum corresponding to 30-second windows from the truth and lie scenarios for F3 channel

We now had all the mean band spectral amplitude values for each of the 30 second windows obtained from the two scenarios. For each window performed an average on the Alpha Low, Alpha High and Beta Low, Beta High amplitude values. As there were varying ranges of fluctuation for the averaged Alpha and Beta band values, we normalized them using simple “min-max” normalization to lie within the range of 0 to 1. The obtained normalized values were then plotted onto a two-dimensional “Alpha-Beta” space, as shown below.

Fig. 6: Normalized band spectra variations of averaged band amplitudes for truth and lying sessions

Clustering

Now we proceeded to define two clusters for the points representing the truth and lie data points on the Alpha-Beta space. For this we applied the *k-means* clustering algorithm with k=2. The output of the clustering after 3 iterations when the variation in cluster center positions converged is shown below.

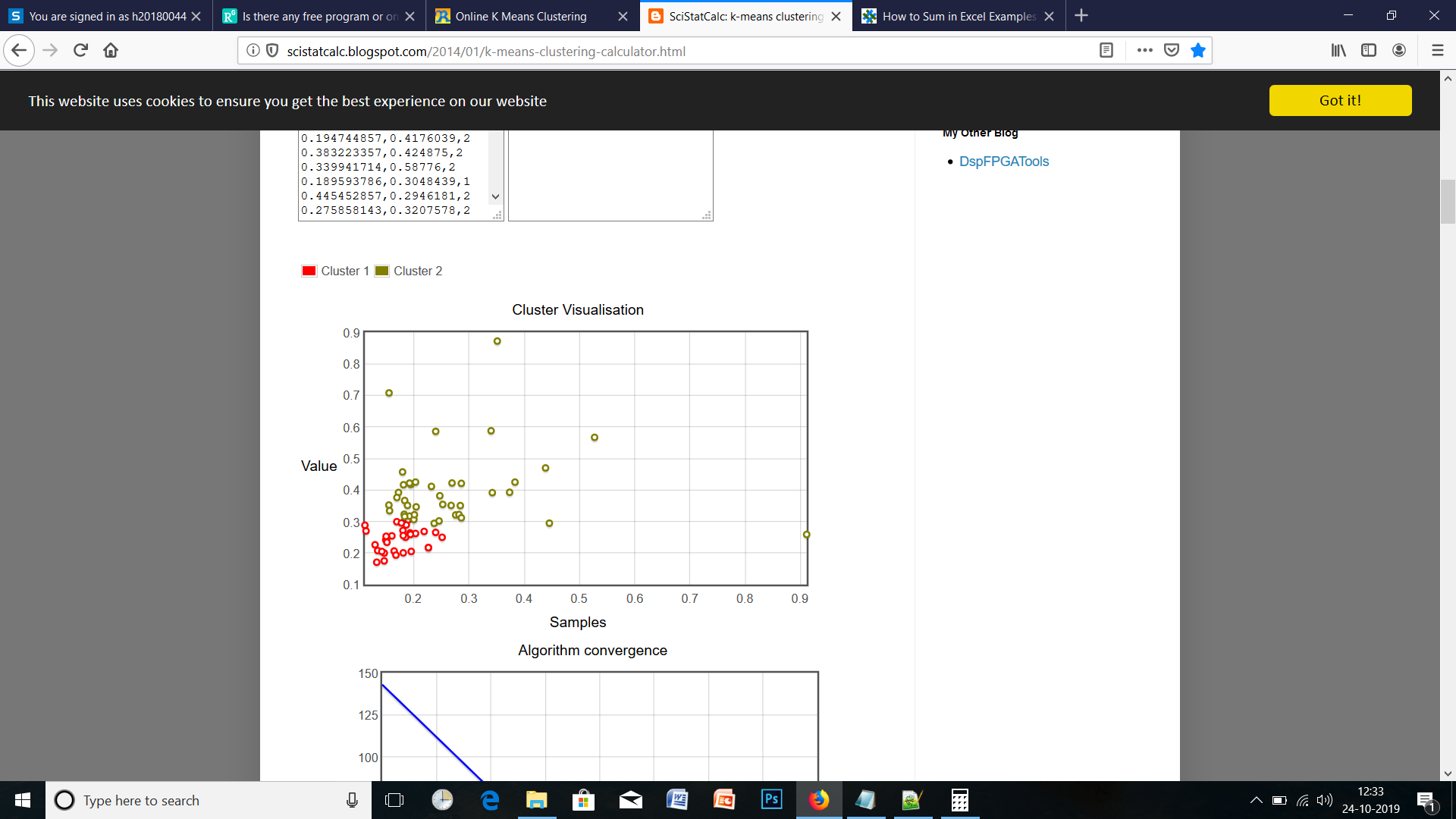


Fig. 7: Result of the clustering using k-means algorithm for 3 iterations

Comparing with the actual classification of the data points, a clustering accuracy of 75.00% was observed. As our hypothesis suggests, the truth set does appear to have higher levels of Alpha band values, as most data points in this set are scattered towards the higher end of the alpha axis. However, a point that contradicts our initial understanding is that the false set registers lower levels of Beta activity than the truth set. We are currently looking into this contradiction, attempting to understand why this deviation occurred and if it is related to (a) the way the experiment was conducted, including the mental disposition of the subject at the time of the sessions, (b) the channel in question or (c) processing of the data. This would constitute a part of our work for the next phase, including an extension of this approach to other users for whom we shall test how accurately we can separately cluster the band values corresponding to truths and lies.

***Phase IV Project Report***

We continued our work into phase IV, focusing on the following objectives:

* To check whether other channels on the left hemisphere of the brain yield better results in separating the two scenarios.
* To test our lie detection procedure on another random subject.

To achieve the above objectives, we decided upon carrying our experiment on another subject. The following section elaborates on this.

Testing the Lie Detection Procedure

One of our batchmates, *Shantanu Kulkarni*, volunteered to be the subject in our experiment. The experiment was scheduled according to his availability to participate, and there was no conflict of interest on the either our or the subject’s side. We carried out a similar experiment as in phase 2 and 3, where the subject would describe images shown to him either truthfully or deceptively for a pre-specified period. However, this time, we had not set any restrictions on which image to describe truthfully and which to describe falsely. The subject was free to make this decision prior to the displaying of the image. Although, we did request the subject to keep the number of truthfully and falsely described images the same, just to keep an equal distribution of data points for processing.



*Fig. 10: The test subject, Shantanu Kulkarni wearing the Emotiv EPOC+ headset*

Eight images were shown to the subject sequentially. Each image was displayed to the subject for a period of 2 minutes, during which the subject described it according to his choice, and his EEG signals were recorded in this period of time. Between each image display, the subject was shown a blank white screen with a “+” sign at the centre and asked to relax by looking at it. None of the images were shown to the subject prior to the experiment’s commencement as it might have influenced the subject’s decision-making beforehand.

The EEG recordings were carried out as described in the previous phases. After acquiring the data, we attempted to process it to obtain any observable patterns.

To evaluate the accuracy of our experimental findings, we asked the subject to note down which images he had lied for. This “confession chit” would be referenced after we completed the clustering of our data points.

Data Processing

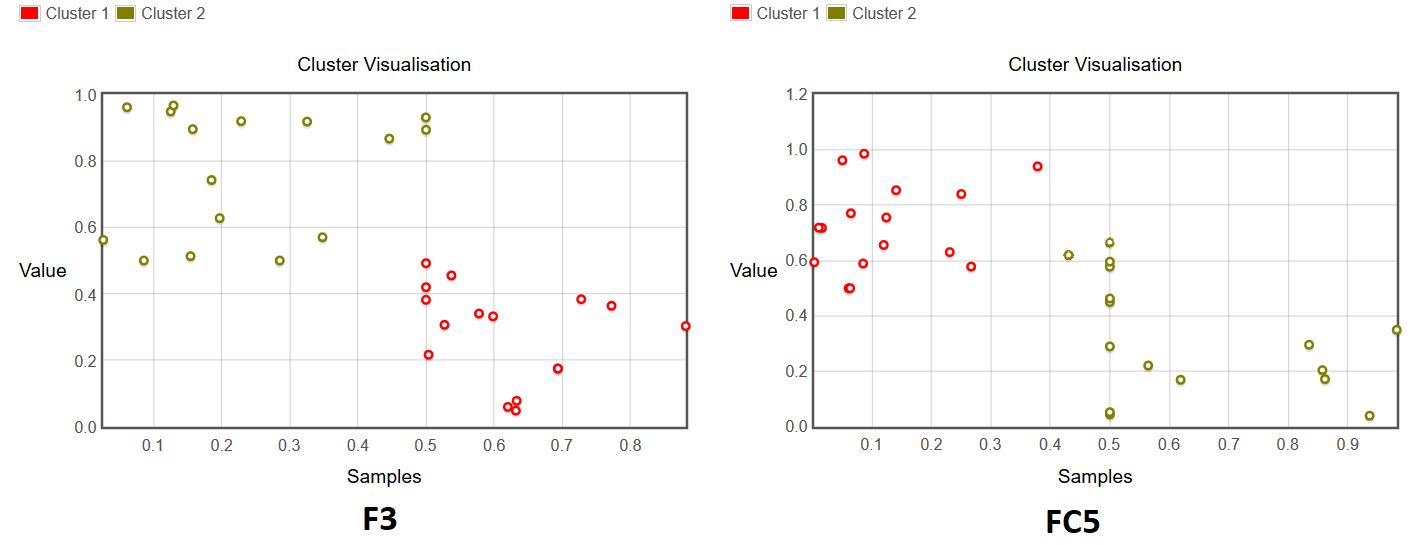
As before, we fed the acquired temporally varying voltage data through a bandpass filter with the pass band lying between 3 Hz to 30 Hz, the range where the desired alpha and beta bands could be located. This was achieved via OpenVibe, where we used the *Temporal Filter* component with the desired passband and filter type parameters.

Each filtered CSV file was then segregated into 30 second, or 3840 tuple-wide, windows. We focused on the data from multiple channels this time, particularly the F3 and FC5 channels, located above the left hemisphere of the brain as shown in Figure 2. Fast Fourier Transform was performed individually on each of these windows to obtain the power spectra values across all the bands for a window.

We now had all the mean band spectral amplitude values for each of the 30 second windows obtained from the two scenarios. For each window, we performed an average on the Alpha Low, Alpha High and Beta Low, Beta High amplitude values. As there were varying ranges of fluctuation for the averaged Alpha and Beta band values, we normalized them using simple “min-max” normalization to lie within the range of 0 to 1.

Clustering

As in phase 3, we carried out the clustering on the data set using k-means algorithm (k=2). The data points were plotted on the 2-dimensional “Alpha-Beta” space as before, and the clustering converged after 3 iterations, giving the following data distribution across clusters, as shown in Figure 11.



*Fig. 11: Output of the k-means clustering on processed data from channels F3 (left) and FC5 (right)*

Now, to evaluate the accuracy of our clustering we referred to our subject’s “confession chit”. Figure 12 lists out the chit entries showing the subject’s behavior with respect to describing the images.

|  |  |
| --- | --- |
| Picture ID | Described Truthfully? |
| 1 | ✓ |
| 2 | ✓ |
| 3 | 🗶 |
| 4 | 🗶 |
| 5 | ✓ |
| 6 | ✓ |
| 7 | 🗶 |
| 8 | 🗶 |

*Fig. 12: Subject’s response to describing each image*

Each image, as we know, corresponds to 4 3-second windows, or 4 data points on the alpha beta space. Taking cluster 1 as the “truth cluster” (✓) and cluster 2 as the “false cluster”, we obtained the following lie detection results as given in the following tables. For an image to detected as “truthfully described” or otherwise we took the majority count of the clustering outcome over all the four 30-second windows that the image comprised of.

|  |  |  |  |
| --- | --- | --- | --- |
| **Channel F3** | | | |
| Picture ID | Clustering Outcome | Detected As | Subject’s Behavior |
| 1 | 🗶 ✓ ✓ ✓ | ✓ | ✓ |
| 2 | 🗶 🗶 ✓ ✓ | ? | ✓ |
| 3 | 🗶 ✓ 🗶 ✓ | ? | 🗶 |
| 4 | ✓ 🗶 🗶 🗶 | 🗶 | 🗶 |
| 5 | ✓ ✓ 🗶 ✓ | ✓ | ✓ |
| 6 | ✓ 🗶 🗶 ✓ | ? | ✓ |
| 7 | 🗶 ✓ ✓ 🗶 | ? | 🗶 |
| 8 | ✓ 🗶 🗶 🗶 | 🗶 | 🗶 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Channel FC5** | | | |
| Picture ID | Clustering Outcome | Detected As | Subject’s Behavior |
| 1 | ✓ ✓ 🗶 ✓ | ✓ | ✓ |
| 2 | ✓ ✓ 🗶 ✓ | ✓ | ✓ |
| 3 | ✓ 🗶 🗶 🗶 | 🗶 | 🗶 |
| 4 | ✓ 🗶 🗶 🗶 | 🗶 | 🗶 |
| 5 | 🗶 ✓ ✓ ✓ | ✓ | ✓ |
| 6 | ✓ 🗶 🗶 🗶 | 🗶 | ✓ |
| 7 | 🗶 🗶 ✓ 🗶 | 🗶 | 🗶 |
| 8 | ✓ ✓ ✓ 🗶 | ✓ | 🗶 |

From the above observations, data from the F3 channel gives a detection accuracy of 50%, as half of the images are left undecided due to equal counts of points clustered as ✓ or 🗶. For FC5, however, the detection accuracy improves to 75%, as 6 out of 8 image descriptions are correctly detected. Combining the detection results from both channels therefore gives us a detection accuracy of 87.5%, as image 8’s description, which was wrongly detected as being the truth using data from FC5, is correctly detected as a lie using data from F3, increasing the number of correct detections to 7 out of 8. As a result, we could correctly detect all the false descriptions, which was the primary aim of this experiment.

Discussion

We now discuss some of our inferences regarding the data collection in phases 3 and 4, and some issues regarding the technical difficulties in the experimental setup. Firstly, we return to our contradiction from phase 3 findings, which show the “false” set having lower levels of Beta activity than the “truth” set. This contradiction arises also in the data in phase 4, where data for the “false” set from the F3 channel again registers lower Beta activity (and also higher Alpha activity) compared to the “truth set”. However, this is not the case for data from the FC5 channel, which follows the trend predicted by our hypothesis. As seen in Figure 11, the FC5 data has higher Alpha and lower Beta activities for the “truth” set and vice versa.

We believe this might arise as due to the characteristics of the portion of the brain that the individual electrodes monitor. As we do not know much further about the inner workings of the human brain, being beyond the scope of our course of study, we can only say that, based on the behavior observed in our acquired data over the last two phases, EEG data corresponding to telling the truth can considerably vary from that corresponding to the subject lying. This difference is what leads to an effective clustering of the processed data and enables the detection of deceptive answers to questions, especially those lies which are formed spontaneously without any prior preparation.

We also state here some issues faced during our experimental sessions which might have impacted the accuracy of our results. Firstly the device used is a wet-EEG headset, which requires the use of special conductive solutions to moisten the sensor pads and ensure proper contact. The degree of moisture on the pads and their quality of contact can degrade, as the pads come in contact with hair and can dry out with time. Even so, we ensured that the contact quality was always above or close to 90% throughout the course of our experiments. This, coupled with surface electrical phenomena at the skin can sometimes diminish the accuracy of the readings, and introduce discrepancies. However, given all these obstacles, we consider our experiments fairly successful as our results, for the most part, match the actual behavior of the subjects and deliver a decent accuracy of detection.

References:

[1] Syed Kamran Haider, Malik Imran Daud, Aimin Jiang, Zubair Khan, *“Evaluation of P300 based Lie Detection Algorithm”*, Electrical and Electronic Engineering 2017, 7(3): 69-76

[2] Simbolon, Artha & Turnip, Arjon & Hutahaean, Jeperson & Siagian, Yessica. (2015). *“An experiment of lie detection based EEG-P300 classified by SVM algorithm.”* 10.1109/ICACOMIT.2015.7440177

[3] R. Cakmak and A. M. Zeki, *"Neuro signal based lie detection"* *2015 IEEE International Symposium on Robotics and Intelligent Sensors (IRIS)*, Langkawi, 2015, pp. 170-174

[4] Gao J, Wang Z, Yang Y, Zhang W, Tao C, et al. (2013) “*A Novel Approach for Lie Detection Based on F-Score and Extreme Learning Machine*.” PLOS ONE 8(6): e64704.

[5] Viresh Gupta, Mohit Agarwal , et al. “*Bag-of-Lies : A Multimodal Dataset for Deception Detection.*”

[6] G. Ganis, S. M. Kossyln, et al. “*Neural Correlates of Different Types of Deception: An fMRI Investigation*.” Cerebral Cortex Aug 2003;13:830–836.