Signals and Systems Project

Student Name: Iman Alizadeh Fakouri

Student Number: 401102134

Instructor: Hamid Karbalaei Aghajan

June 6, 2024

Contents

1 Preliminaries

1.1 Project Explanations

In this phase of the project, the main goal is to preprocess the EEG data to train machine learning models for detecting potential epileptic seizures. The raw EEG signals contain significant noise from various sources such as electrical interference and physiological activities. Therefore, it is imperative to apply various filtering techniques to isolate the cerebral signals. The subsequent sections will detail these preprocessing steps.

1.2 Answer to Questions

1.2.1 Question 1: Electrode Naming in the 10-20 EEG System

The 10-20 EEG system's electrode naming convention reflects the brain region each electrode monitors. The letters Fp, F, T, P, O, and C denote the pre-frontal, frontal, temporal, parietal, occipital, and central regions, respectively. Even-numbered electrodes are positioned on the right hemisphere, while odd numbers are on the left.

1.2.2 Question 2: Frequency Band Activities

Each frequency band corresponds to different brain activities:

- Delta (0.5-4 Hz): Associated with deep sleep and slow-wave activity.
- Theta (4-8 Hz): Linked to drowsiness, meditation, and early sleep stages.
- Alpha (8-13 Hz): Indicates a relaxed state, often with closed eyes.
- Beta (13-30 Hz): Reflects active thinking, problem-solving, and concentration.
- Gamma (30-100 Hz): Pertains to higher cognitive functions and information processing.

1.2.3 Question 3: Sampling Frequencies for EEG Signals

Following the Nyquist criterion, the sampling frequency should be at least twice the highest frequency in the signal. For EEGs, with gamma waves up to 100 Hz, a minimum of 200 Hz is necessary to prevent aliasing. However, frequencies of 256 Hz or more are preferable for capturing finer details.

1.3 EEG Data Preprocessing Steps

1. **Load data**

   We can import data by choosing .mat format and setting frequency rate to 256Hz and adjusting number of channels according to data.

   After that we should edit chnnel locations and load electrodes location.I chose spherical model for channel locations.Here is channel locations plot for both of datasets:

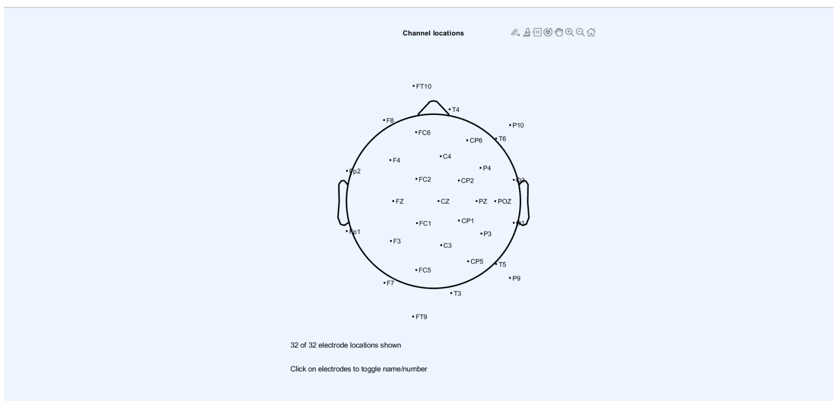

Figure 1: First dataset channel locations

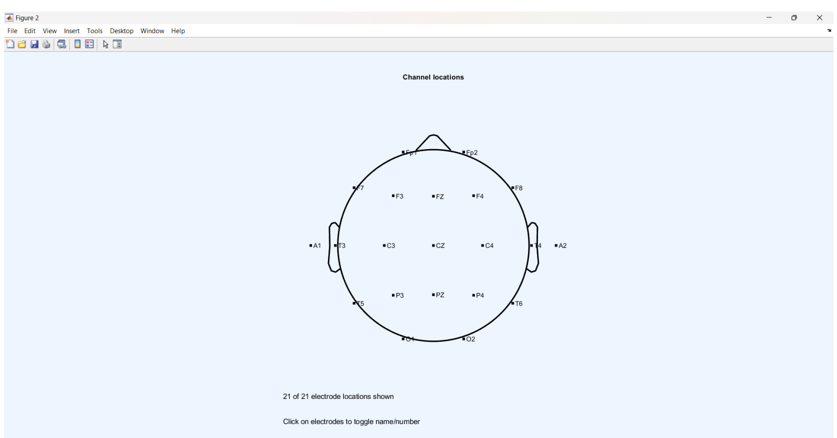

Figure 2: Second dataset channel locations

We can also plot signal recorded by each channel:

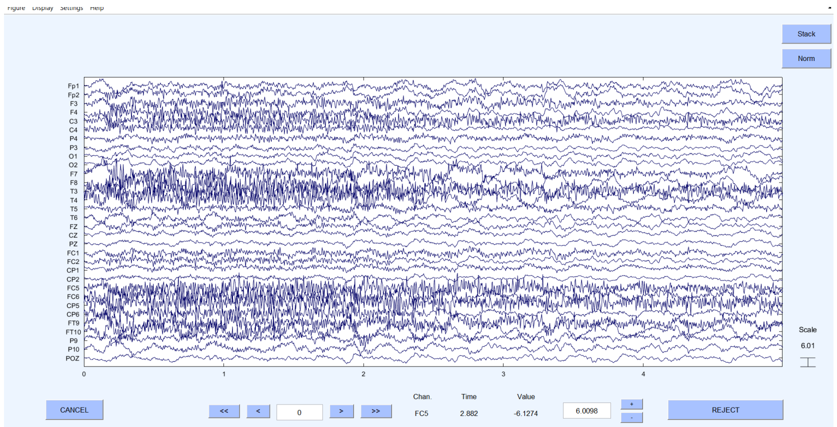

Figure 3: Recorded signals of first dataset channels

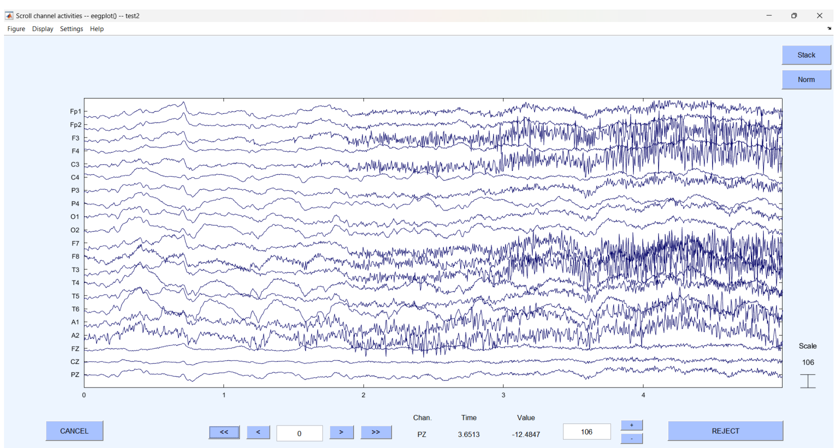

Figure 4: Recorded signals of second dataset channels

2. **Apply a 1 Hz high pass filter to remove baseline drifts**
   Baseline drifts are low-frequency fluctuations that can occur due to various factors, such as head movements, breathing, or even heartbeats. These drifts can obscure the true EEG signals, which are typically in the higher frequency range.

The reason for using a 1 Hz high pass filter specifically is that it allows the removal of these slow fluctuations without significantly affecting the integrity of the EEG data. This frequency cutoff is chosen because it is low enough to preserve the important brainwave frequencies that are typically above 1 Hz, while still eliminating the unwanted drifts that can interfere with data analysis.

Moreover, when preparing data for Independent Component Analysis (ICA), which is often used to identify and remove artifacts from EEG data, it's crucial to minimize the influence of low-frequency drifts. ICA assumes that the source signals are stationary, but baseline drifts can violate this assumption. By filtering out these drifts, the data becomes more suitable for ICA, leading to more accurate artifact detection and removal. Here is the bode diagram of magnitude and phase of the high-pass filter applied on channels signals:

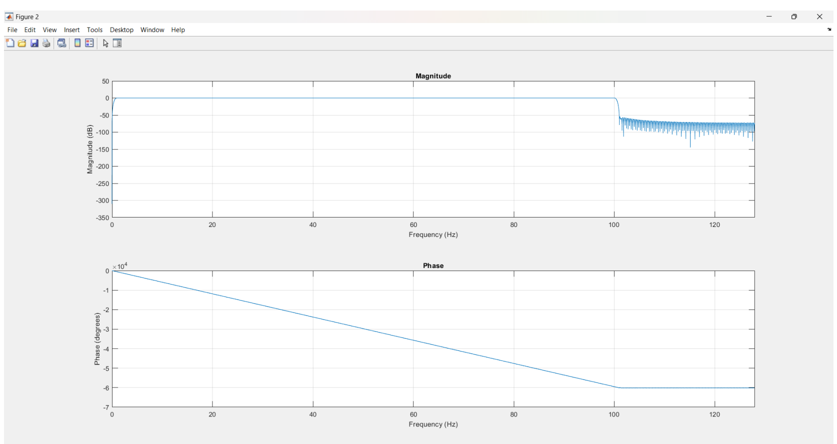

Figure 5: Magnitude and Phase Bode plot

I also removed frequencies higher than 100Hz because it is the highest frequency that can be generated by brain.

3. **Apply a relevant notch filter to remove the 50 Hz line noise**
The 50 Hz line noise is a common type of electrical interference found in many countries where the power supply operates at 50 Hz. This interference can be introduced into EEG recordings from various sources, such as power lines, electrical equipment, and even the building's wiring. The notch filter is designed to target and remove this 50 Hz frequency and its harmonics without affecting the rest of the EEG signal. Here is the bode diagram of magnitude and phase of the notch filter applied on channels signals:

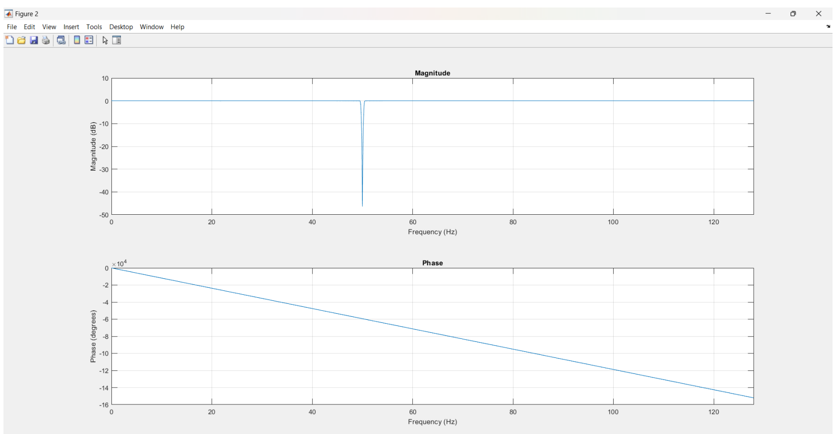

Figure 6: Magnitude and Phase Bode plot

4. **Initial Re-Referencing**
This step is crucial for several reasons:
Consistency: It ensures that all EEG channels are measured relative to the same baseline, which is necessary for comparing signals across different channels.
Reduction of Bias: Without a common reference, each channel could be biased by different reference potentials, leading to misleading interpretations of the EEG data.
Improved Signal Quality: Re-referencing can help to reduce the influence of reference electrode noise and improve the overall quality of the EEG signal.

5. **Reject bad channels as a critical step before average referencing using the `clean rawdata()` EEGLAB plugin**
   Bad channels are those that do not record electrical brain activity accurately. They may be due to broken electrodes, poor scalp contact, or external noise. These channels can introduce significant artifacts into the data, which can distort the analysis.By removing these channels, we reduce the noise in dataset, which enhances the signal-to-noise ratio. This makes it easier to detect the true brain signals from the EEG recordings.
   Average referencing involves re-referencing the EEG data to the average of all channels. If bad channels are included in this calculation, they can skew the average, leading to a distorted reference signal. This can affect all subsequent analyses. Here is the affect of clean rawdata on channels of both datasets:

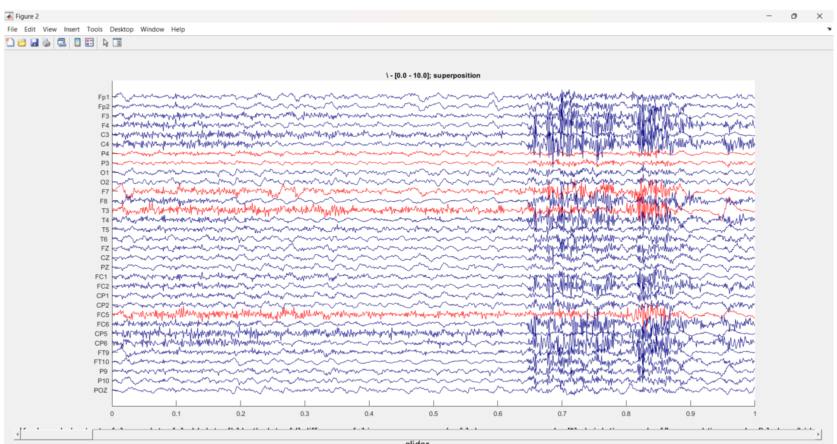

Figure 7: Removing bad channels of first dataset

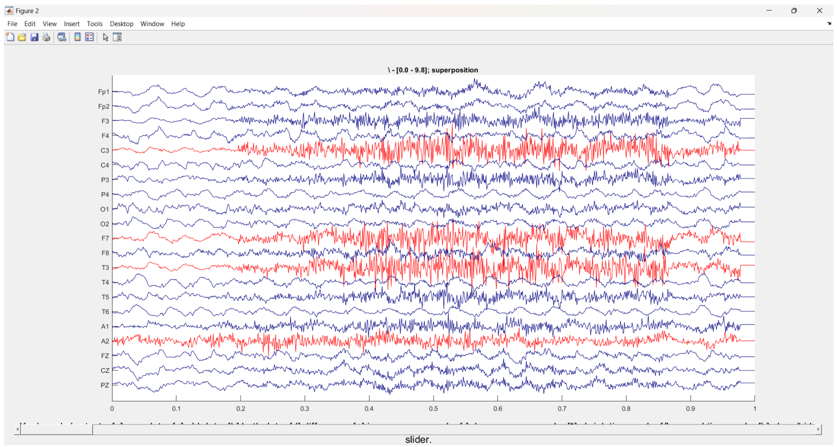

Figure 8: Removing bad channels of second dataset

6. **Interpolate the removed channels**

   Interpolating removed channels in EEG signal processing is an important for several reasons:

   Maintain Data Completeness:

   When bad channels are identified and removed, it leaves gaps in the data. Interpolation is used to estimate the missing data based on the information from the surrounding channels.

   Preserve Topographical Information:

   EEG data is often analyzed in terms of its spatial distribution across the scalp. Interpolating removed channels helps to preserve the topographical integrity of the data, which is important for analyses like topographic mapping and source localization.

   Improve ICA Results: Interpolation is recommended before running Independent Component Analysis (ICA). ICA assumes that the data is from a full rank matrix. If there are missing channels, the rank of the data matrix is reduced, which can negatively impact the ICA results.

7. **Re-reference the data to the average of all channels to obtain a good estimate of reference-independent potentials**

8. **Apply `clean rawdata()` for cleaning the data by running artifact subspace reconstruction (ASR)**
   ASR is a sophisticated algorithm designed to identify and remove artifacts from EEG data. Artifacts can be caused by eye blinks, muscle movements, or external electrical noise, and they can significantly distort the EEG signals.
   Data Quality: By removing these artifacts, ASR helps to improve the overall quality of the EEG data, making it more reliable for further analysis.Here is the result of cleaning the data of channels of first dataset:

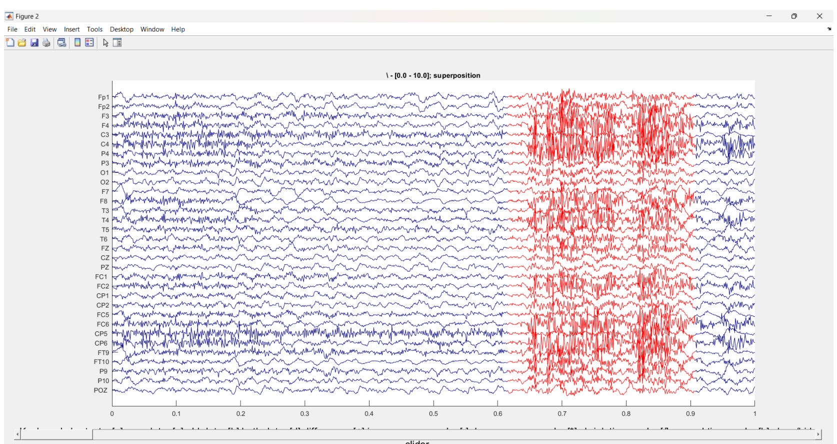

Figure 9: Cleaning effect on first dataset channels

9. **Re-reference the data to the average again to compensate for any potential changes in the data caused by the previous step**

10. **Run independent component analysis (ICA) to identify EEG sources as well as the sources associated with noise and artifacts**

ICA is a computational method that separates a multivariate signal into additive independent subcomponents. This is particularly useful in EEG, where the recorded signals are mixtures of various brain and non-brain sources.
Artifact Identification: ICA can identify components associated with eye blinks, muscle activity, heartbeats, and other sources of noise, which can then be selectively removed from the EEG data.
The theory behind ICA is based on the assumption that the observed EEG signals are linear mixtures of statistically independent sources. The goal of ICA is to find a linear transformation that maximizes the statistical independence of the components. Here's a simplified explanation of the theory:
Generative Model:
ICA assumes that the observed signals ((X)) are generated by a mixing process from several independent sources ((S)). Mathematically, this can be expressed as (X = AS), where (A) is the mixing matrix.
Statistical Independence:
The core principle of ICA is that the source signals are statistically independent of each other. This means that the occurrence of one source signal does not provide information about the occurrence of another.
The process of ICA involves estimating the unmixing matrix ((W)), which is the inverse of the mixing matrix ((A)), to recover the independent sources ((S)) from the observed mixed signals ((X)). Mathematically, this is expressed as (S = WX).
In eeglab this process is done using ICA decompose which first gives rank of matrix A which is number of independent sources(for first dataset rank was 26 and for second dataset rank was 16) and then gives source(component) signals which can be plotted.Here is the plot of component signals of first dataset:

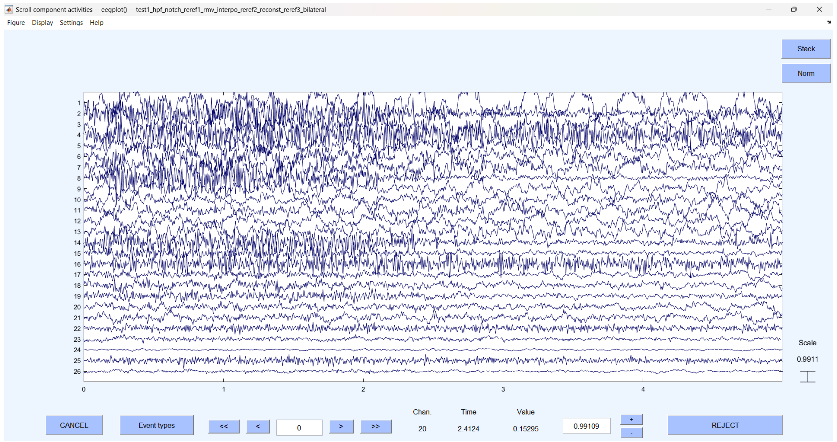

Figure 10: First dataset components

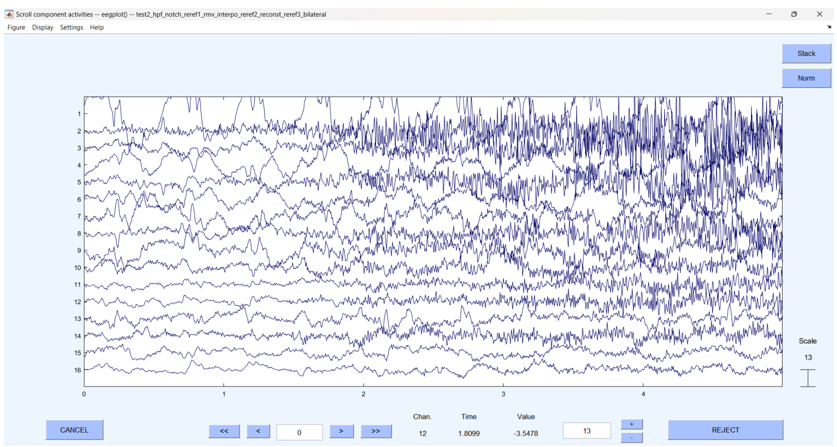

Figure 11: Second dataset components

Before explaining the next step we should select our head model in order to use autofit in eeglab.The best and most realistic option would be MNI head model which is designed to be like a real head and therefore it will be the best option for us.But since we chose spherical model for channel locations at the beginning we should transform locations and coordinates of electrodes from spherical model into MNI model which is done using warp montage in eeglab.Here is the result of this transformation:

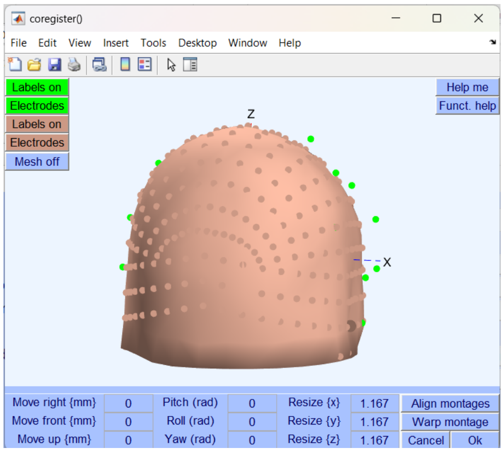

Figure 12: Initial coordinates of electrodes

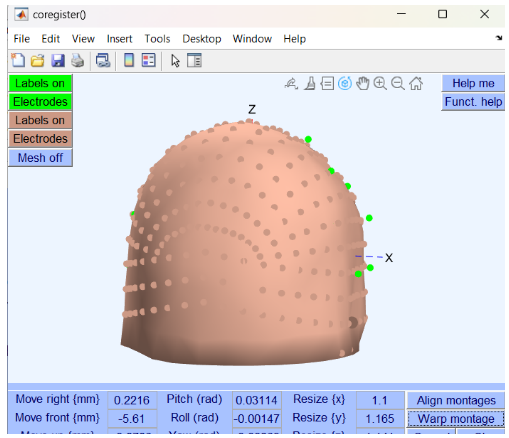

Figure 13: Fitting Coordinates on MNI head model

11. **Fit single and bilateral (if available) current dipoles**

    In this part i used autofit and bilateral fit option in eeglab.I set RV parameter to 100.Here is a short explanation of what they do:

    Autofit:

    This is a feature that automates the process of finding the best-fitting dipole for each independent component. It usually involves a two-step process:

    Coarse Fit:

    A broad search across the whole brain to find an approximate location of the dipole.

    Fine Fit:

    A more detailed search around the initial estimate to refine the dipole parameters.

    Bilateral Fit:

    This refers to fitting two dipoles simultaneously, one in each hemisphere of the brain, for a single independent component. It's used when the scalp map suggests that the activity is generated by sources in both hemispheres. Results of this are shown in next parts.

12. **Further clean the data by source (dipole) selection using `IClabel()` plugin in EEGLAB**

    By removing noise sources from components and regenerating channel signals we can significantly decrease the noise of channels.Here is the remaining and brain components of first dataset:

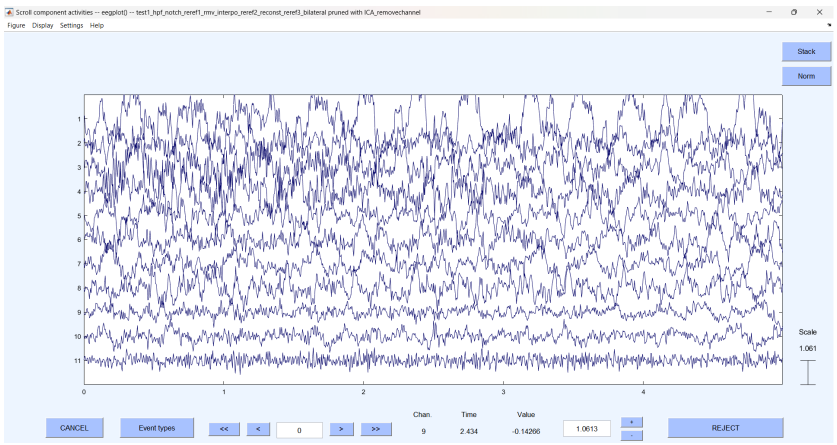

Figure 14: First dataset brain components

13. **Epoch Data**

    Epoching in EEG data analysis is a fundamental step that involves dividing the continuous EEG signal into smaller segments, called epochs, which are time-locked to specific events of interest.

    Why Epoch Data?
    Focus on Relevant Data:
    Epoching allows researchers to focus on the time periods that are most relevant to the study, such as the moments before, during, and after a stimulus is presented.
    Event-Related Potentials (ERPs):
    By isolating these segments, it's possible to average them to observe event-related potentials, which are brain responses that are directly the result of a

specific sensory, cognitive, or motor event.
Comparative Analysis:
Epochs make it possible to compare brain activity across different conditions or groups within the same time frame.

How to Epoch Data in EEGLAB?
Defining Time Window:
WE define the time window for each epoch. This typically includes a pre-stimulus baseline period and a post-stimulus period to capture the brain's response.
Extracting Epochs:
Using EEGLAB's interface, select Tools → Extract Epochs. This will segment the continuous data into epochs based on the events.

14. **Reshape Data**

    EEGLAB provides functions to reshape data. For example, after epoching, the data is typically in a 3D matrix (electrodes x time points x epochs). If we need to perform time-frequency analysis, we might need to reshape this into a 2D matrix (electrodes x [time points * epochs]).

15. **Save Processed Data**

    This process is done for all steps of preprocess for both of dataset 1 and dataset 2 and has been delivered alongside project report.

16. **Select Relevant Channels**

    I did this part by using pop-select command in eeglab.Here is the channels related to epilepsy for first dataset.

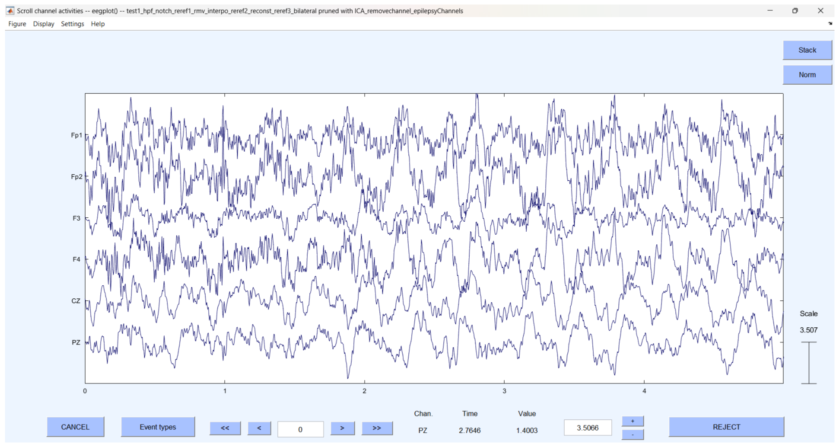

Figure 15: First dataset epilepsy channels

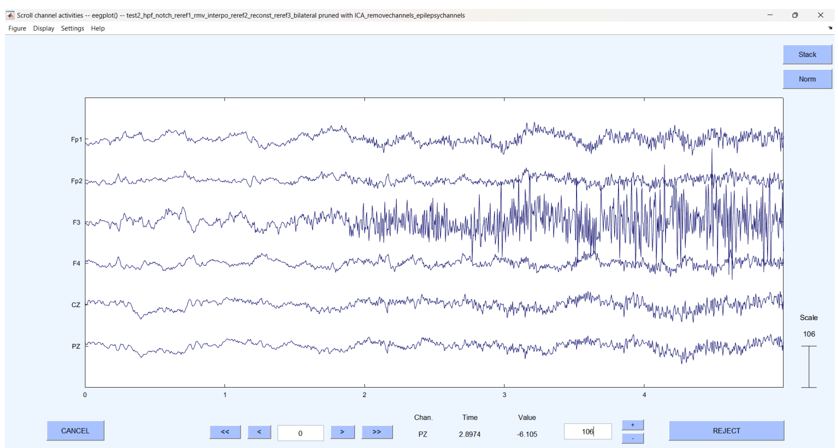

Figure 16: Second dataset epilepsy channels

17. **Subsample Data**

    I did this part by using pop-resample command in eeglab.I set sampling frequency to be 256Hz.

2 Pre-processing First Dataset and Deliverables

2.1 Frequency Spectrum

When comparing the Fz channel frequency spectra between noisy and pre-processed (clean) EEG datasets, several differences may become apparent.

1.Noise Reduction:

In the noisy dataset, we'll likely observe higher power in various frequency bands due to artifacts and interference. These artifacts can include muscle activity, eye blinks, and other non-neural sources. In the pre-processed (clean) dataset, the noise components should be significantly reduced. Artifacts related to muscle movements and eye blinks should be minimized or removed.

2.Frequency Peaks and Bands:

In the noisy dataset, we might find irregular or exaggerated peaks due to artifacts. In the clean dataset, these peaks should be more consistent and representative of neural activity.

3.Artifact Removal:

The pre-processed dataset should exhibit fewer sharp spikes or sudden fluctuations caused by artifacts. Artifacts, such as eye blinks or muscle contractions, can distort the frequency spectra.

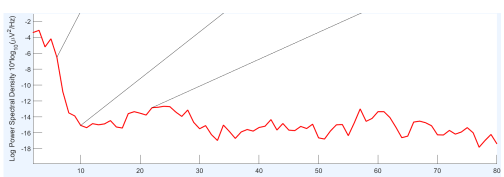

Figure 17: Frequency spectrum of the Fz channel of initial data

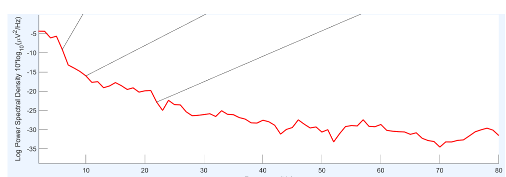

Figure 18: Frequency spectrum of the Fz channel of filtered data

2.2 ICA Components

Picture below shows 26 independent sources that generate signals recorded by electrodes.We have two main sources.First of them is noise sources which has different types like eye,heart(ECG),muscle(EMG),line noise and others.Second is the brain sources which are really important for us.By only keeping brain sources we will regenerate electrode signals in order to have EEG signals recorded by electrodes without any noise.

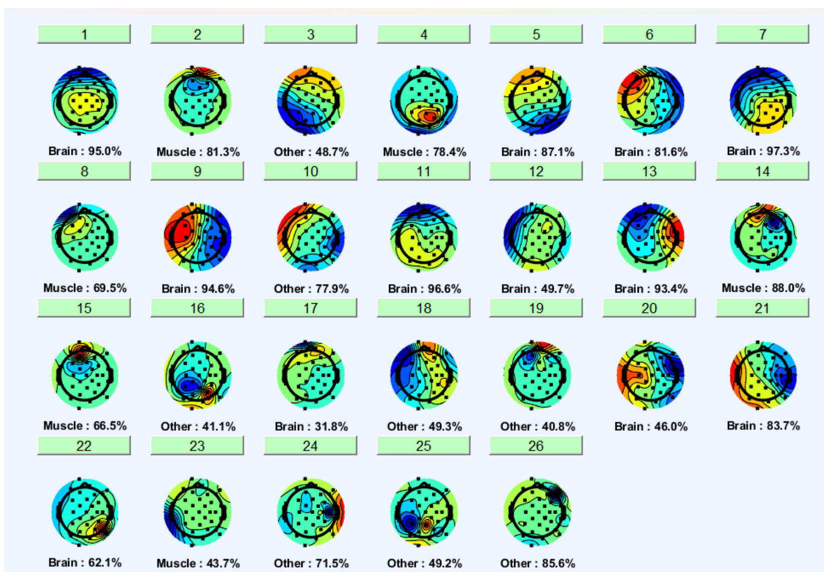

Figure 19: ICA Components of first dataset

Here we have two types of noise sources.The first type is noise caused by muscles and second label of noise is other.
In the context of EEG signal processing, “other noises” in ICA labeled components can refer to a variety of signal sources that are not of cerebral origin. These can include Channel Noise which is often due to a single channel that has been disturbed during recording, such as by being bumped or having poor contact.
Muscle Activity (EMG Artifacts):
Muscle activity or EMG artifacts are generally higher in frequency and can be

quite variable in amplitude. In the time domain, they may appear as irregular, high-frequency bursts. .

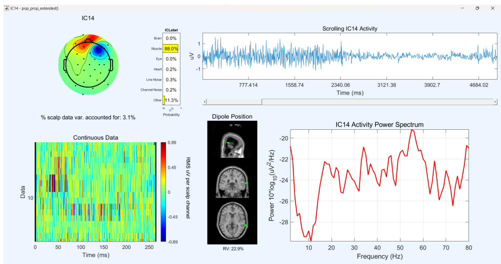

Figure 20: IC14 component(EMG source)

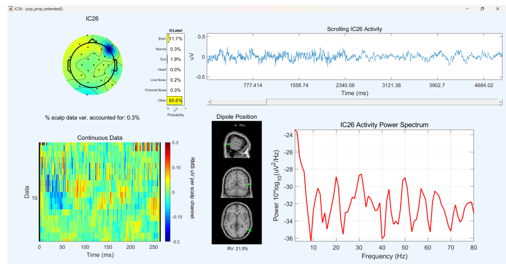

Figure 21: IC26 component(other source)

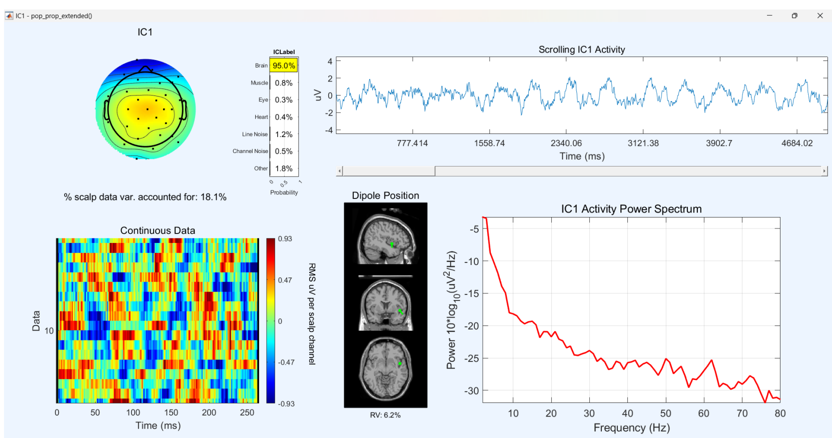

Figure 22: IC1 component(brain source)

Brain Sources:

Brain sources typically show distinct patterns in the frequency domain, often related to different types of brain waves (e.g., alpha, beta, gamma). These can be identified by their characteristic frequency ranges and are usually associated with different states of brain activity.

An important property of brain sources is that in their frequency spectrum,domain of frequency falls by 1/f as the frequency increases.

2.3 Processed Data

Below are the signals related to the first data electrodes before and after the filtering. It is clear that after applying different filters, the noise caused by city electricity, blinking, heartbeat and muscle contraction, etc are removed from EEG signals. So signals are ready to be processed and applied to machine learning models.

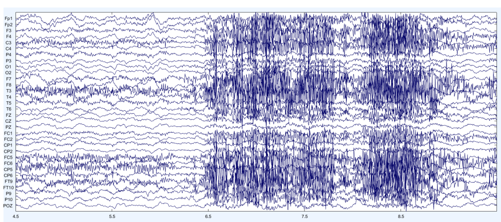

Figure 23: Plot of channels of first dataset before filtering(4.5s-9s)

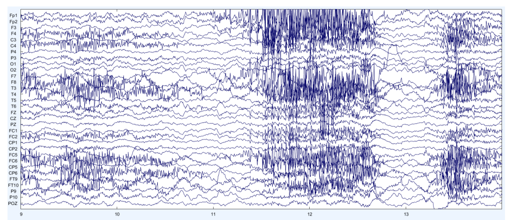

Figure 24: Plot of channels of first dataset before filtering(9s-14s)

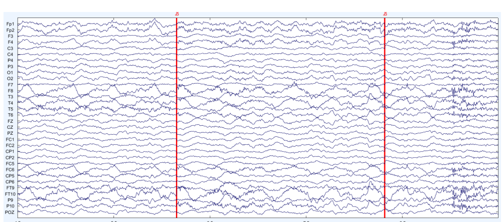

Figure 25: Plot of channels of first dataset after filtering(4.5s-9s)

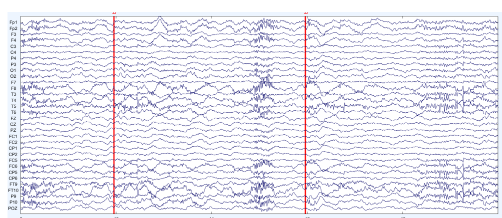

Figure 26: Plot of channels of first dataset after filtering(9s-14s)

3 Pre-processing Second Dataset and Deliverables

3.1 Frequency Spectrum

For dataset 1 we talked about differences between the initial dataset and final dataset's frequency spectrum.In this part we will get deeper into differences between normal person's and a person with a probability of having epileptic seizures frequency spectrum.

When comparing the Fz channel frequency spectra between normal individuals and those with a probability of epilepsy seizures, there are several key differences we might observe:

1.Epileptiform Discharges:

Epileptiform discharges are transient waveforms that occur during seizures. These abnormal electrical patterns can be detected in EEG recordings. In the dataset for the person with a probability of epilepsy seizures, we may find specific patterns associated with epileptiform discharges. These could include spikes, sharp waves, or other abnormal brain waveforms1. In contrast, the normal EEG dataset should not exhibit these epileptiform discharges.

2.Frequency Peaks and Patterns:

Individuals with epilepsy may show specific frequency patterns associated with seizure activity. These could include increased power in certain frequency bands (e.g., theta, alpha, beta, or gamma) during seizures. Normal individuals typically exhibit consistent frequency patterns without significant deviations.

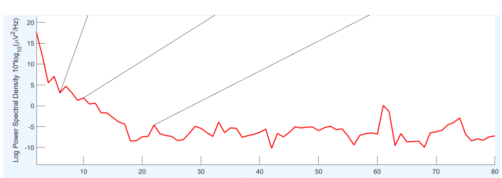

Figure 27: Frequency spectrum of the Fz channel of initial data

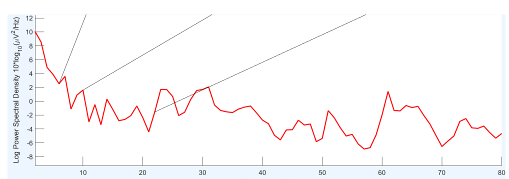

Figure 28: Frequency spectrum of the Fz channel of filtered data

3.2 ICA Components

Picture below shows 16 independent sources that generate signals recorded by electrodes.We have two main sources.First of them is noise sources which has different types like eye,heart(ECG),muscle(EMG),line noise and others.Second is the brain sources which are really important for us.By only keeping brain sources we will regenerate electrode signals in order to have EEG signals recorded by electrodes without any noise.

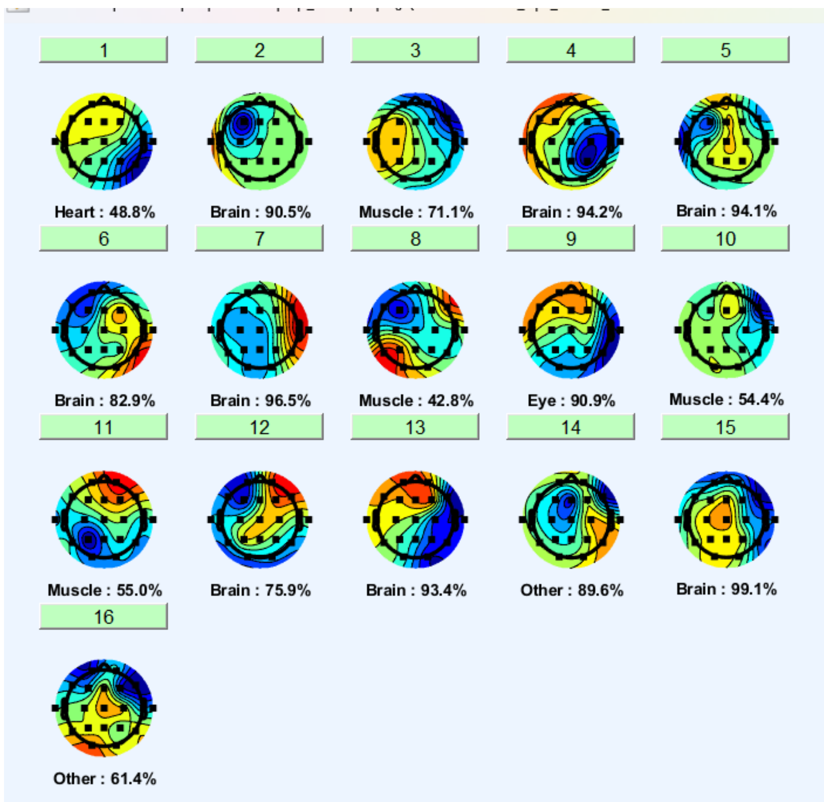

Figure 29: ICA Components of second dataset

Here we have two types of noise sources.The first type is noise caused by eyes and the second is the noise caused by heart.

Eye Movements (EOG Artifacts):
Eye movements and blinks produce sharp spikes in the time domain and are typically seen in the lower frequency range in the frequency domain. They have large amplitudes and can be easily identified due to their distinct shapes.
Heart Signals (ECG Artifacts):
Heartbeat-related signals or ECG artifacts appear as regular, rhythmic activity in the time domain and are usually found in the lower frequency range. They can be identified by their repeating pattern corresponding to the heart rate. .

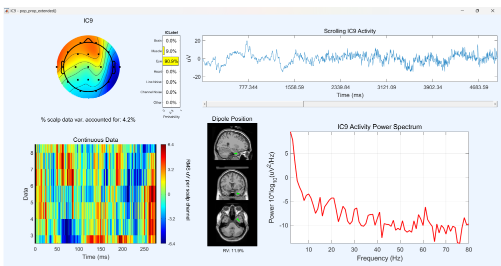
Figure 30: IC9 component(eye source)

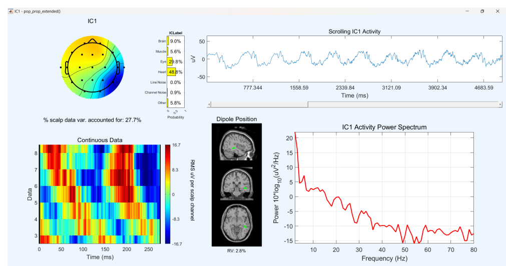
Figure 31: IC1 component(heart source)

For first dataset we explained what are properties of a normal persons brain source signal and frequency spectrum.In this part we will do the same for a person with probability of having epileptic seizures.
Frequency Spectrum:
In individuals with epilepsy, the frequency spectrum of ICA components may show spikes or sharp waves that are characteristic of epileptic activity. These can appear as abnormal rhythmic patterns or bursts at certain frequencies.
Time Signal:
The time-domain signal of ICA components in epilepsy may exhibit sudden, brief episodes of synchronized neuronal activity, which manifest as spikes or sharp waves. These patterns are distinct from the normal background EEG activity and are indicative of seizure activity. .

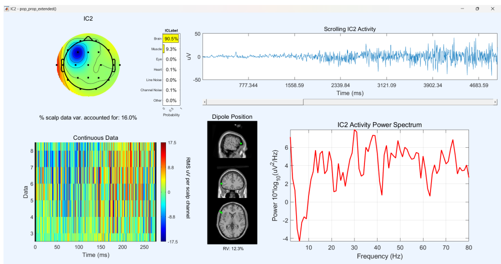

Figure 32: IC2 component(brain source)

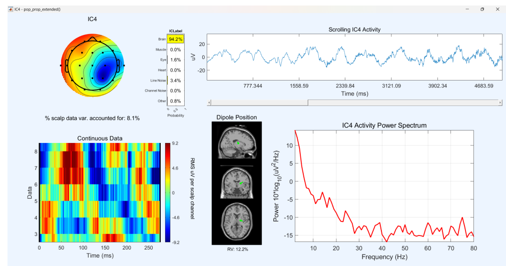

Figure 33: IC4 component(brain source)

One of the important techniques to detect the seizures caused by epilepsy is to examine two components related to the brain. In case of seizures, spikes appear asymmetrically in the brain and its effect on the frequency spectrum of the components is such that one of the components will have a normal frequency spectrum, but the other will have different peaks at different frequencies in its frequency spectrum.

3.3 Processed Data

In dataset one we explained how filtering decreased the noise in the data significantly.In this part we will discuss differences that can be seen in EEG signals of normal human and a person with probability of having epileptic seizures.

When comparing EEG signals from normal individuals to those from a person with a probability of epilepsy seizures, we might observe the following differences:

1. Epileptiform Discharges:

In the EEG of a person with a probability of epilepsy, we may see epileptiform discharges, which are unusual patterns like spikes or sharp waves. These discharges are typically not present in the EEG of a normal person.

2. Background Activity:

The background EEG activity in a normal person is usually regular and rhythmic, with clear alpha waves when the person is relaxed and with eyes closed.In contrast, the background activity in an individual with epilepsy may be disrupted or show slower (delta or theta) waves even outside of seizures.

3. Interictal Activity:

Even when seizures are not occurring, the EEG of a person with epilepsy may show interictal spikes or sharp waves that indicate a tendency for seizures. Normal EEGs do not show these interictal changes. .

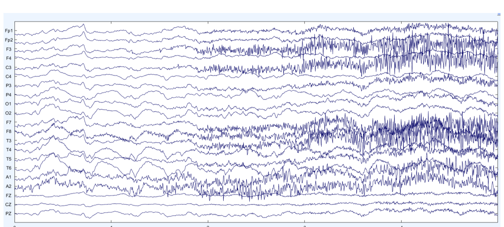

Figure 34: Plot of channels of second dataset before filtering(0s-5s)

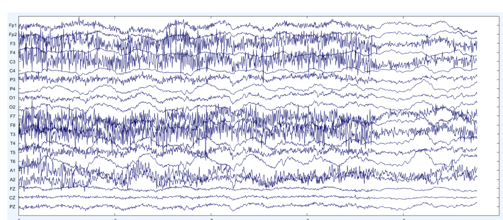

Figure 35: Plot of channels of second dataset before filtering(5s-9s)

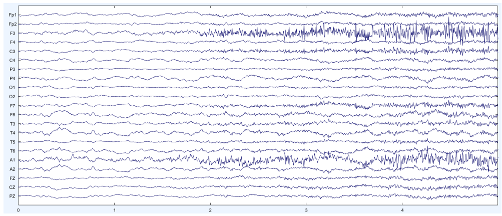

Figure 36: Plot of channels of second dataset after filtering(0s-5s)

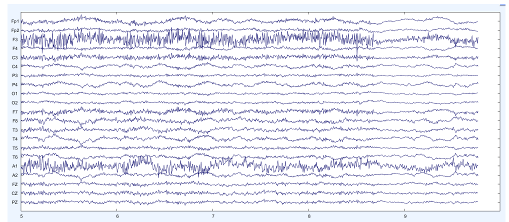

Figure 37: Plot of channels of second dataset after filtering(5s-9s)