MID-TERM PROJECT REPORT

PROJECT: P300 BASED REAL-TIME EEG SPELLER WITH CUSTOMIZABLE MACHINE LEARNING MODELS

NAME: SHRIYA SURAVARAPU

ROLL NUMBER: 241003

PROJECT SUMMARY:

This project aims to develop a P300 based EEG speller system, a type of Brain Computer Interface (BCI) that enables character selection using neural responses instead of physical input. The system relies on detecting the P300 Event-Related Potential (ERP), a positive voltage deflection occurring approximately 300 ms after a target stimulus is perceived, which is widely used in attention-based BCI paradigms.

The proposed workflow involves EEG signal preprocessing, epoch extraction around stimulus events, feature extraction from ERP responses, and supervised machine learning-based classification to distinguish target P300 and non-target responses. In the long term, the project aims to integrate customizable machine learning models into a real-time P300 speller framework.

WORK COMPLETED:

1) UNDERSTANDING EEG, ERP AND P300 PARADIGM (from lecture 0)

- Studied fundamentals of EEG signal acquisition, frequency bands, and electrode placement.
- Understood ERP extraction through epoching and averaging techniques
- Learned the physiological basis and significance of the P300 response in attention-based tasks.
- Studied the row-column flashing paradigm used in classical P300 spellers.

2) MACHINE LEARNING FOUNDATIONS (from lectures 1 and 2)

- Covered basics of supervised learning, classification vs regression and evaluation concepts.
- Studied commonly used classifiers such as Logistic Regression, SVMs, Decision Trees, and Random Forests.
- Understood model training, loss functions, gradient descent and evaluation pipelines.

3) NEURAL NETWORKS AND SEQUENTIAL MODELS (from lecture 3)

- Studied feedforward neural networks, CNNs and RNNs.
- Understood limitations of RNNs with long-term memory, such as vanishing/exploding gradients.
- Learned how LSTMs address long-term dependency issues.
- Explored example notebooks demonstrating RNN-based learning pipelines.

4) EEG PREPROCESSING PIPELINE (from week 4)

- Studied and implemented the standard EEG preprocessing pipeline using MNE-Python.
- Performed preprocessing steps on a given dataset, including:
  - Band pass filtering and resampling of EEG signals
  - Event definition and epoch extraction aligned to stimulus events
  - Artifact correction using Independent Component Analysis (ICA)
  - Visualization of EEG signals and inspection of evoked responses.
- Understood the role of EEG signals and inspection of evoked responses.

5) MACHINE LEARNING FOUNDATIONS FOR EEG (from week 5)

- Progressed to Phase 3 of the project flowchart, focusing on feature preparation and classification readiness
- Studied supervised learning pipelines relevant to EEG classification tasks
- Worked on assignment 3, which builds preprocessed EEG data for downstream analysis
- Understood how preprocessed EEG data feeds into machine learning models for target vs non-target classification in P300 paradigms

CODE LINKS:

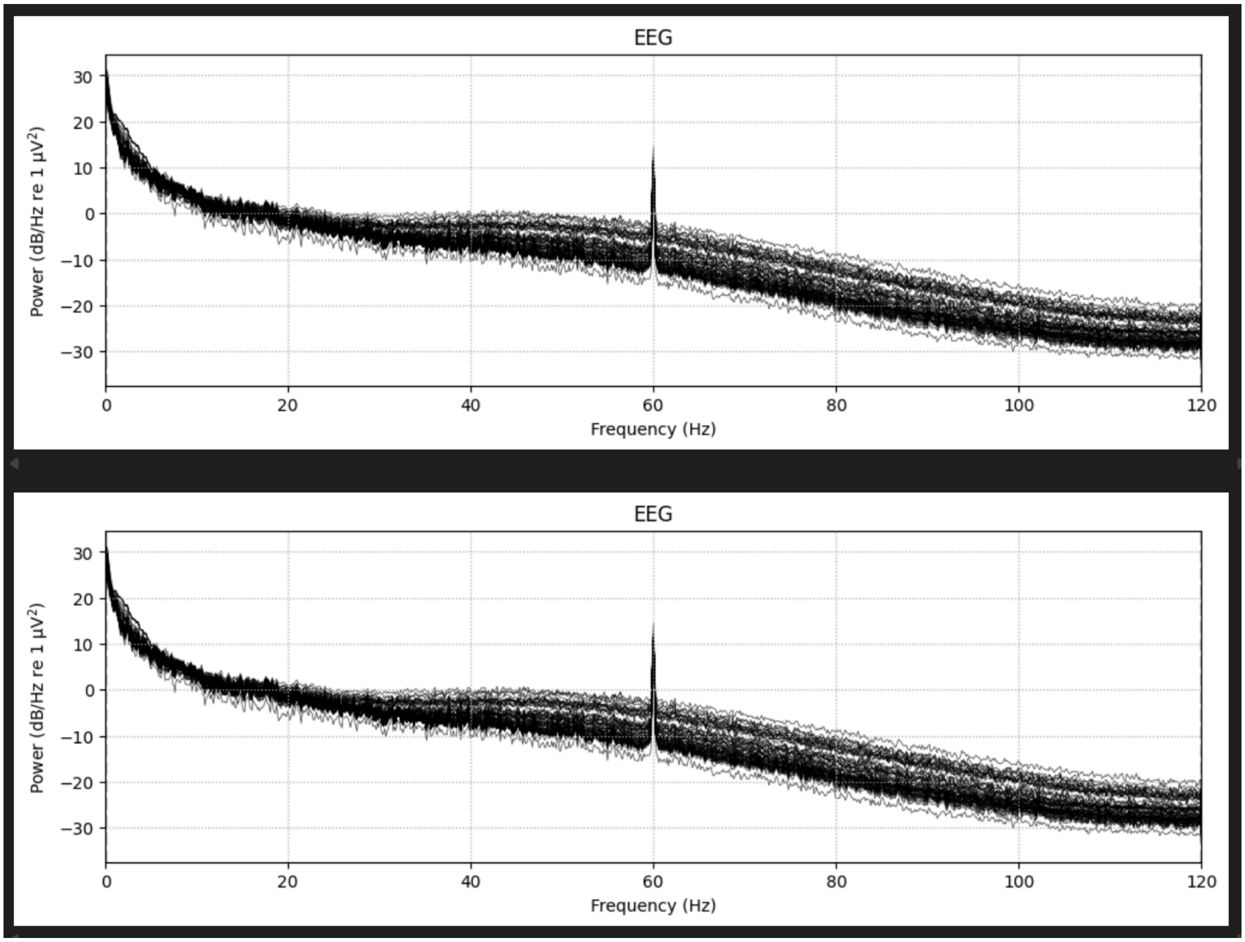
ASSIGNMENT 0:
https://github.com/shriya456/Winter-projects-25-26/blob/main/EEG-Based%20P300%20Speller/assignments/assignment_0/241003_Shriya_Suravarapu.ipynb
241003_Shriya_Suravarapu.ipynb

ASSIGNMENT 1:
https://github.com/shriya456/Winter-projects-25-26/blob/main/EEG-Based%20P300%20Speller/assignments/assignment_1/241003_Shriya_Suravarapu.pdf

ASSIGNMENT 2:
https://github.com/shriya456/Winter-projects-25-26/blob/main/EEG-Based%20P300%20Speller/assignments/assignment_2/241003_Shriya_Suravarapu_A2.ipynb
241003_Shriya_Suravarapu_A2.ipynb

ASSIGNMENT 3:
https://github.com/shriya456/Winter-projects-25-26/blob/main/EEG-Based%20P300%20Speller/assignments/assignment_3/EEG_Assignment3_Shriya_Suravarapu_241003.ipynb
EEG_Assignment3_Shriya_Suravarapu_241003.ipynb

RESULTS AND OBSERVATIONS:



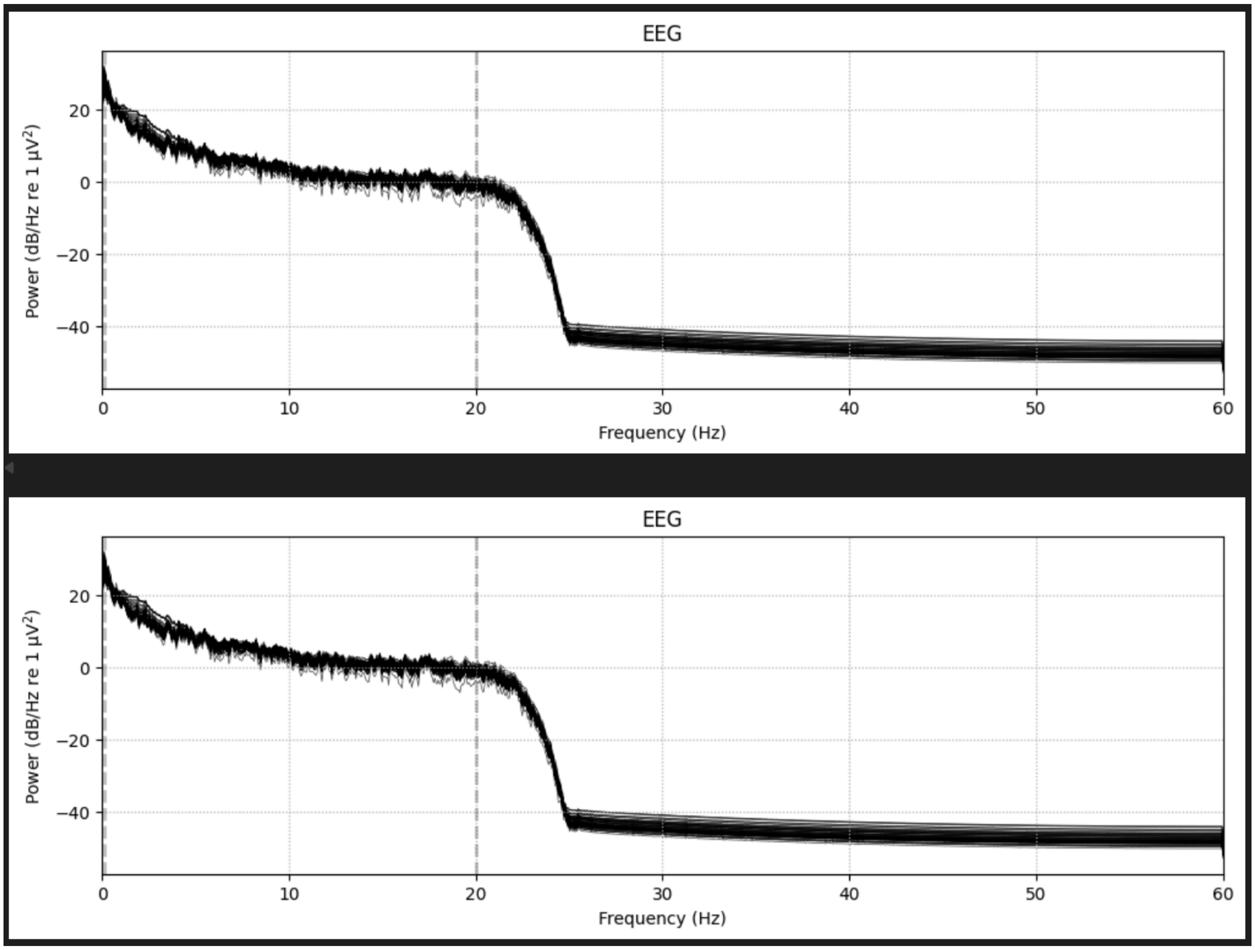
Power spectral density of raw data from SubjectA_Train dataset

```python
#2. applying band pass filter
raw.filter(
    l_freq=0.1,
    h_freq=20,
    fir_design='firwin'
)
```

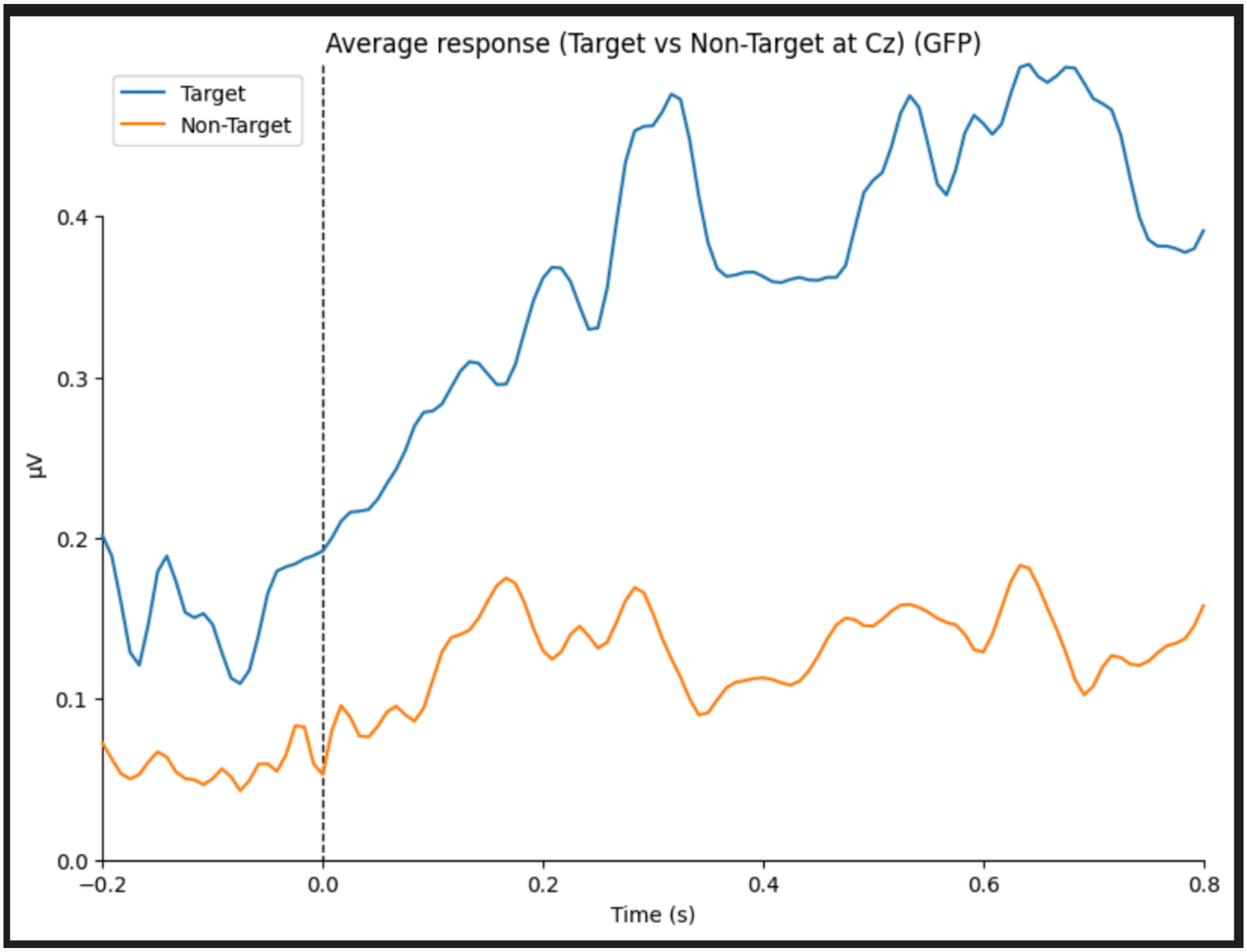
```
Filtering raw data in 1 contiguous segment
Setting up band-pass filter from 0.1 - 20 Hz

FIR filter parameters
---------------------
Designing a one-pass, zero-phase, non-causal bandpass filter:
- Windowed time-domain design (firwin) method
- Hamming window with 0.0194 passband ripple and 53 dB stopband attenuation
- Lower passband edge: 0.10
- Lower transition bandwidth: 0.10 Hz (-6 dB cutoff frequency: 0.05 Hz)
- Upper passband edge: 20.00 Hz
- Upper transition bandwidth: 5.00 Hz (-6 dB cutoff frequency: 22.50 Hz)
- Filter length: 7921 samples (33.004 s)
```

General	
MNE object type	RawArray
Measurement date	Unknown
Participant	Unknown
Experimenter	Unknown
Acquisition	
Duration	00:46:01 (HH:MM:SS)
Sampling frequency	240.00 Hz
Time points	662,490
Channels	
EEG	64
Stimulus	1
Head & sensor digitization	Not available
Filters	
Highpass	0.10 Hz
Lowpass	20.00 Hz



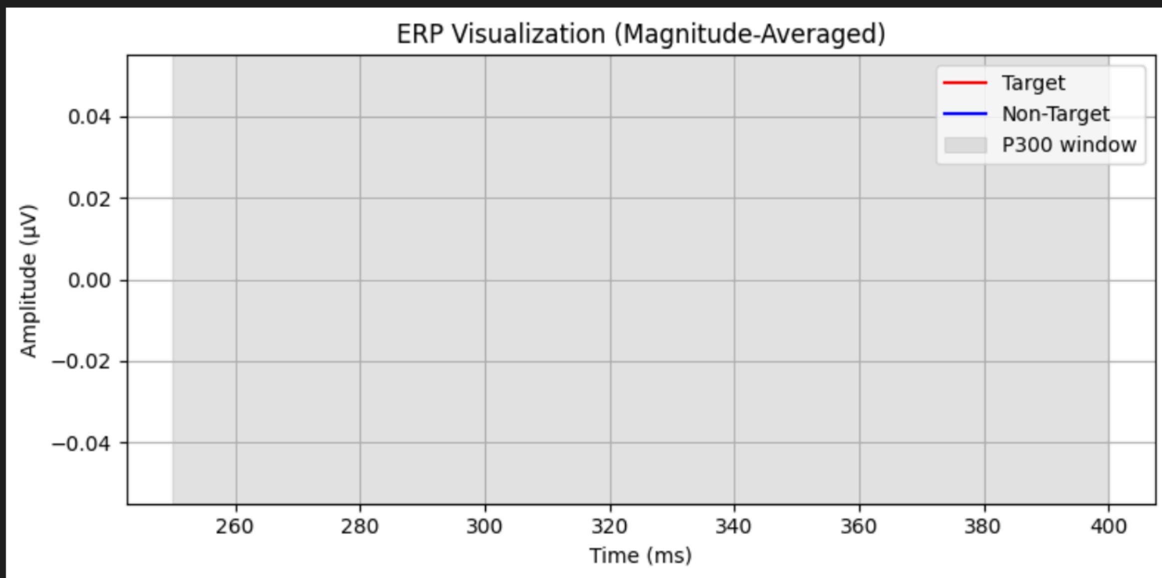
Plot after applying band-pass filter and downsampling the data to 120Hz



Visualizing P300 responses after creating epochs

```
======================================================================
STEP 4: VISUALIZING ERP RESPONSES
======================================================================

--- Subject A ---
Epochs shape: (15299, 60, 64)
Targets: 2550 Non-targets: 12749
```



The ERP visualization did not show due to heavy downsampling, despite all the values coming through correctly, despite several approaches

```
======================================================================
STEP 6: BASELINE CLASSIFIERS (Subject A)
======================================================================

=== Logistic Regression Evaluation ===
Accuracy: 0.8333333333333334
F1 Score: 0.0
Confusion Matrix:
 [[2550    0]
 [ 510    0]]
Classification Report:
              precision    recall  f1-score   support

         0.0       0.83      1.00      0.91      2550
         1.0       0.00      0.00      0.00       510

    accuracy                           0.83      3060
   macro avg       0.42      0.50      0.45      3060
weighted avg       0.69      0.83      0.76      3060
```

Shows an accuracy of 83.33% of the model

CHALLENGES FACED:

- Mapping theoretical ML concepts to EEG specific applications required effort
- EEG preprocessing involves many interdependent steps, making debugging time-consuming.
- Understanding MNE's data structures (Raw, Epochs, Evoked) required a learning curve.