MID-TERM PROJECT REPORT

EEG-Based Speller System

Name: Harsh Chandwani

Roll Number: 240428

Project Summary

The objective of this project is to design and develop an EEG-based speller system capable of interpreting human brain signals to select intended characters. Using brain-computer interface (BCI), the project aims to build a direct communication channel between the human brain and an external computing system.

The workflow involves collecting EEG signals, followed by preprocessing to remove unwanted noise and artifacts. Important features are extracted from the signals and analyzed using machine learning and signal processing methods to determine the intended character. The selected character is then displayed on a graphical interface to form words and sentences.

Overall, the project aims to implement a complete EEG-based speller system and evaluate its performance in detecting user-intended characters from EEG responses. The focus is on building a reliable processing pipeline and analyzing the effectiveness of different signal processing and classification approaches in achieving accurate character selection as output.

Progress Till Mid-Term

The progress of the project up to the mid-term evaluation has been carried out in a structured, assignment-wise manner. The completed work is summarized below:

Assignment 0: Python Fundamentals

In Assignment 0, we revised fundamental Python programming concepts, including array manipulation using NumPy, data handling and cleaning using Pandas, and data visualization through Matplotlib. These exercises helped develop a practical understanding of data structures, numerical computations, dataset preprocessing, and effective graphical representation of data..

Assignment 1: Machine Learning Concepts

Assignment 1 was oriented toward strengthening our conceptual understanding of fundamental machine learning principles through theoretical reasoning. The assignment focused on learning about how and why different learning algorithms work, covering topics such as ensemble methods (bagging, boosting, and random forests), loss functions, and regularization techniques used in models like SVM, LASSO, and Ridge regression. It also emphasized understanding the bias–variance trade-off, overfitting and underfitting behavior, and the role of splitting criteria and tree depth in decision tree learning. Overall, this assignment helped clarify the assumptions, limitations, and performance-related factors that influence machine learning models.

Assignment 2: EEG Signal Processing

In Assignment 2, we performed signal processing on EEG data using a real-world P300 speller dataset from the BCI Competition III, with the objective of understanding the complete preprocessing and analysis pipeline. The steps carried out in the assignment is as follows:

- Loaded and reshaped raw EEG data from a P300 speller dataset, and constructed event markers using stimulus and flashing information for proper event alignment.

- Developed an MNE-based preprocessing pipeline including band-pass filtering, resampling, channel mapping to the standard 10–20 montage, and artifact removal using Independent Component Analysis (ICA). The spatial patterns of the extracted independent components are shown in Figure **??**.
- Extracted target and non-target epochs, applied baseline correction, and visualized averaged evoked responses at key electrode sites to observe the P300 component.

Overall, this assignment provided substantial hands-on experience with EEG data handling, preprocessing, artifact removal, and event-related potential analysis using real BCI data. Experimenting with different methods and parameter values helped reinforce theoretical concepts through practical implementation and demonstrated how these choices affect the results.

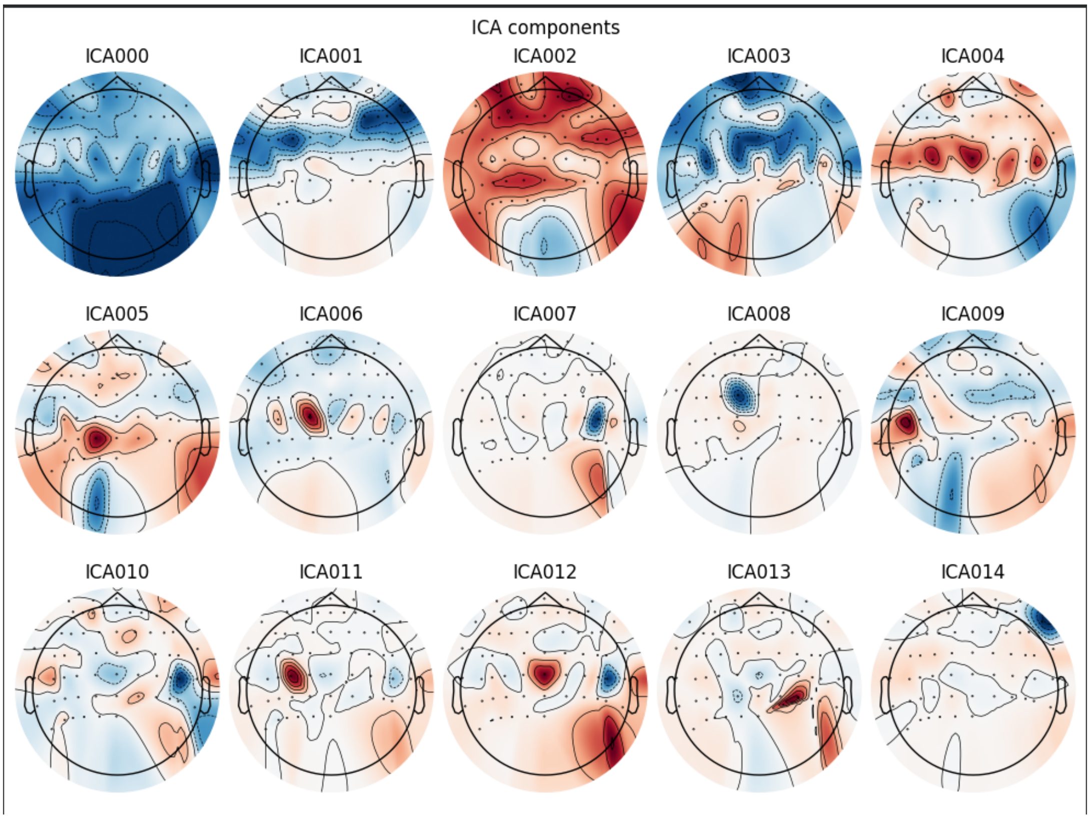

Figure 1: "Independent Component Analysis" (ICA) components of EEG signals showing distinct source patterns used for artifact identification and removal.

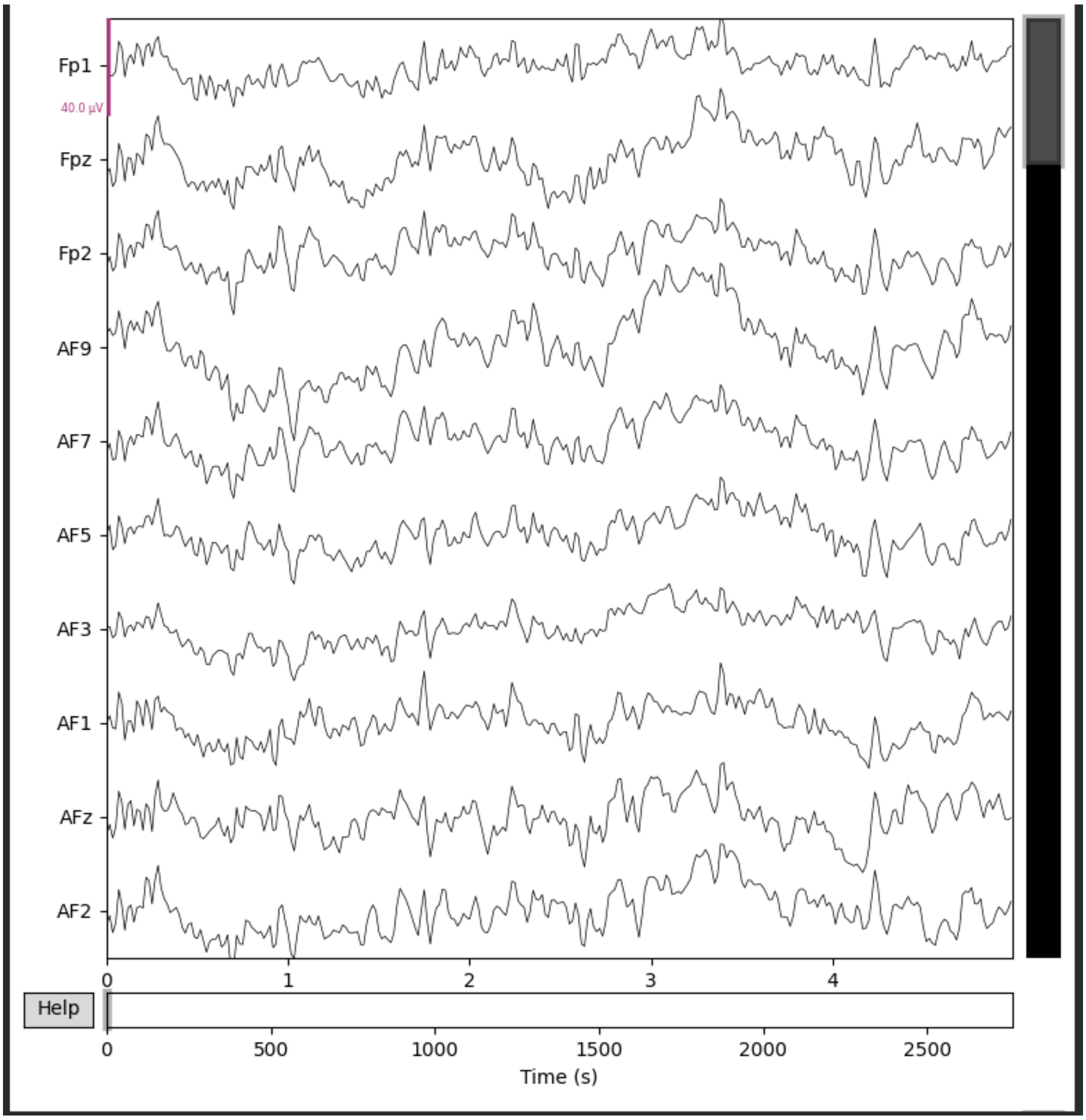

Figure 2: Time-domain EEG signals after preprocessing and artifact removal using ICA.

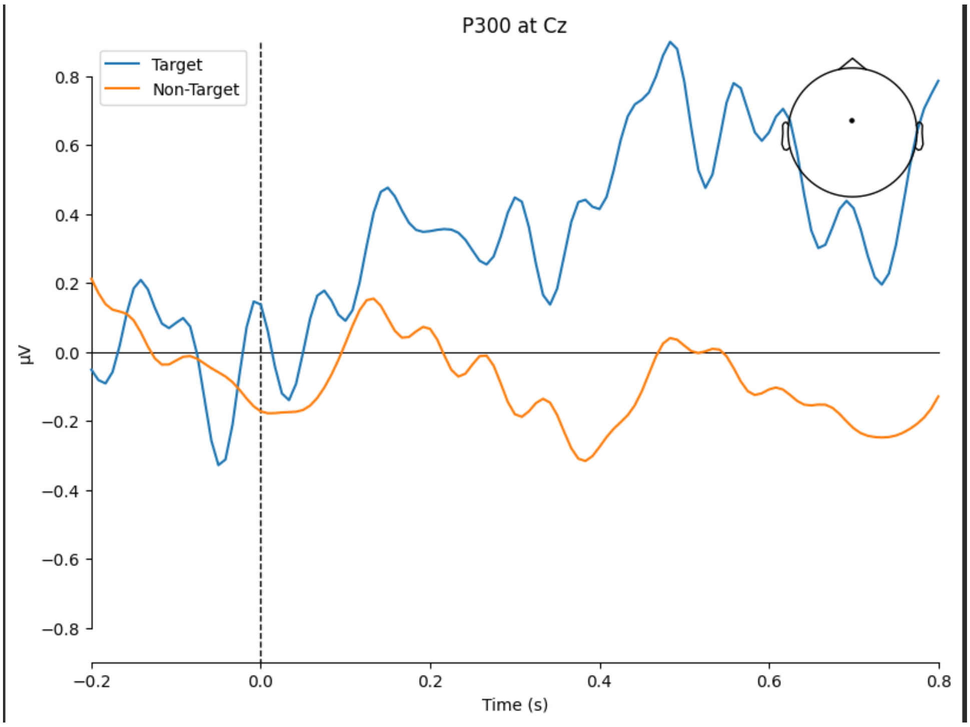

Figure 3: Comparison of target and non-target ERP responses at the Cz electrode.

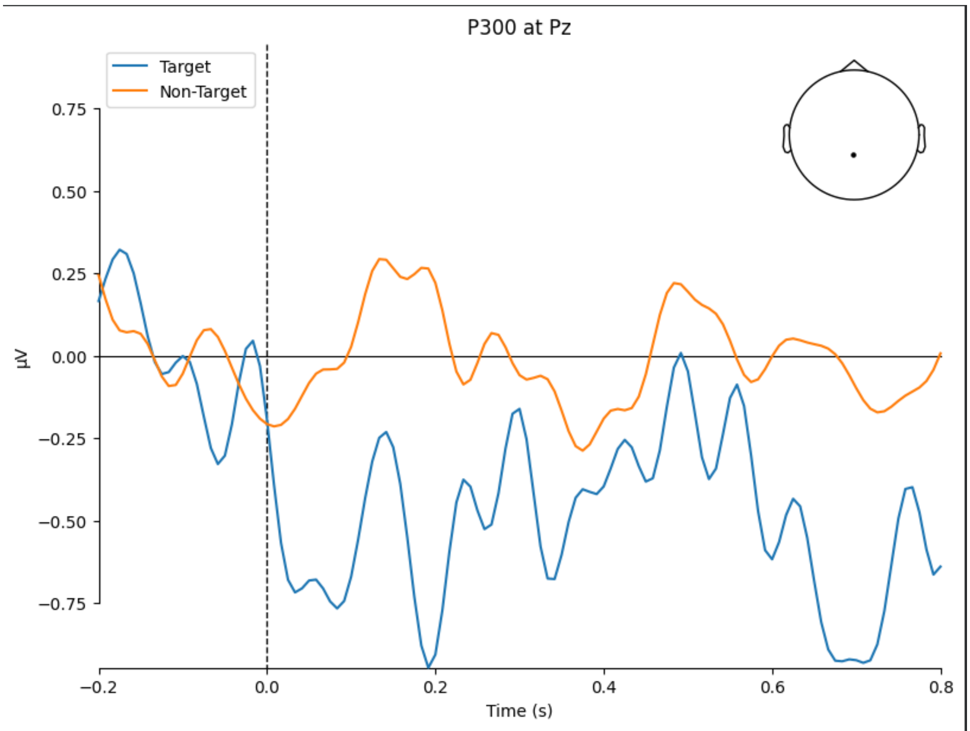

Figure 4: Comparison of target and non-target ERP responses at the Pz electrode.

Assignment 3: EEG Classification Pipeline

Assignment 3 involved implementing a structured EEG processing and analysis pipeline. The primary objective was to integrate multiple stages of EEG analysis into a single workflow and understand how each stage contributes to distinguishing target and non-target brain responses using real EEG data.

- Loaded EEG data and applied preprocessing, including band-pass filtering, resampling, and artifact removal.
- Performed epoch extraction based on stimulus events, applied baseline correction, and computed averaged evoked responses for target and non-target classes to analyze event-related potentials(ERP's).
- Extracted relevant EEG features and applied classical machine learning classification techniques to evaluate the ability to differentiate between target and non-target EEG responses.

Overall, in this assignment we integrated EEG preprocessing, ERP analysis, and classification into a unified pipeline. By trying and comparing different methods, parameter values, and processing choices, we gained a deeper understanding of how each stage influences signal quality and classification outcomes. This iterative process strengthened practical insight into EEG-based analysis and established a strong foundation for further work on feature optimization, model refinement, and the development of real-time EEG-based speller systems.

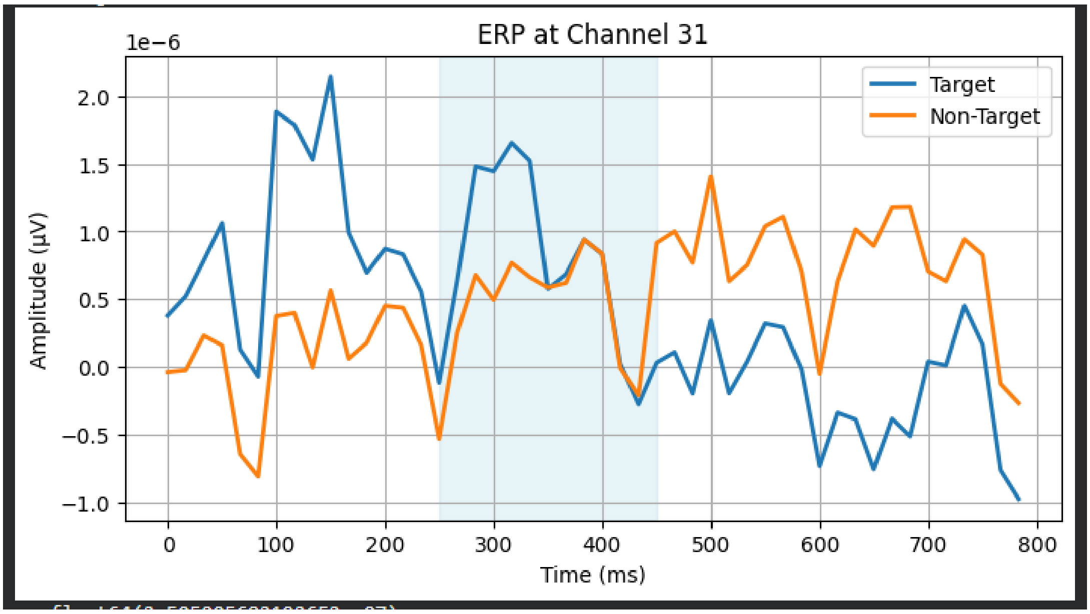

Figure 5: Comparison of target and non-target ERP responses at EEG Channel 31, showing temporal differences within the P300 time window.

Challenges Faced

The primary challenges encountered during this project were concentrated in the implementation of the complete EEG speller pipeline in Assignment 3. A major difficulty involved understanding the structure of the EEG dataset, including what does each channel in different data fields represents. Gaining clarity on how stimulus information, channel data, and labels were organized required careful inspection of the dataset and iterative experimentation.

Another significant challenge was selecting appropriate preprocessing parameters, particularly during signal filtering. Determining suitable frequency bands proved non-trivial, as under-filtering resulted in noisy signals with prominent artifacts, while over-filtering led to loss of important EEG information, including components relevant to the P300 response. Achieving a balance between noise suppression and signal preservation required repeated adjustments using visual inspection.

Combining all steps of the pipeline: preprocessing, epoch extraction, feature extraction, and classification, was challenging. Mistakes or poor choices in the early stages often affected later results, so the system had to be debugged and improved multiple times.

Overall, these challenges emphasized the complexity of working with real-world EEG data and highlighted the importance of a deep understanding of data structure, careful preprocessing choices, and systematic pipeline design in EEG based speller systems.

Work done During the Project

- **Assignment 1 (Machine Learning Concepts):**
  `https://github.com/electricalengineersiitk/Winter-projects-25-26/blob/main/EEG-Based%20P300%20Speller/assignments/assignment_1/240428_Harsh_Chandwani.pdf`
- **Assignment 2 (EEG Signal Processing):**
  `https://github.com/electricalengineersiitk/Winter-projects-25-26/blob/main/EEG-Based%20P300%20Speller/assignments/assignment_2/Assignment_2_240428.ipynb`
- **Assignment 3 (EEG Classification Pipeline):**
  `https://github.com/electricalengineersiitk/Winter-projects-25-26/blob/main/EEG-Based%20P300%20Speller/assignments/assignment_3/Harsh_Chandwani_240428_Assignment3_EEG.ipynb`