P300-Based Real-Time EEG Speller

Name: Shubham Joshi
Roll Number: 240500
Project: EEG Speller (Prof. Nikunj Bhagat)
Department of Electrical Engineering

January 9, 2026

1 Project Summary

Brain–Computer Interfaces (BCIs) enable direct communication between the human brain and external systems by interpreting neural signals. Among various BCI paradigms, the **P300 speller** is one of the most reliable and widely researched systems, especially for assistive communication. It is based on the P300 Event-Related Potential (ERP), a positive voltage deflection occurring approximately 300 ms after a subject perceives a rare or meaningful stimulus.

This project focuses on building a **real-time P300-based EEG speller system** using OpenViBE, while extending the traditional pipeline with customizable machine learning models. The system covers the complete workflow—from EEG signal acquisition and preprocessing to model training, evaluation, and real-time deployment. Classical machine learning models such as LDA, SVM, and Random Forest are implemented alongside deep learning models like EEGNet to study performance trade-offs in accuracy, latency, and robustness.

The final outcome is a modular and extensible EEG speller system capable of detecting user-intended characters in real time. The project emphasizes practical BCI challenges such as noisy signals, low signal-to-noise ratio, and real-time constraints, making it suitable for real-world deployment and future research extensions.

2 Dataset Description

The experiments in this project are conducted using the **BCI Competition III – P300 Speller Dataset (Wadsworth 2004)**, using data from **Subject A**. The dataset follows a row–column visual speller paradigm designed to elicit the P300 event-related potential.

The EEG data consists of recordings from 64 EEG channels sampled originally at 240 Hz. After preprocessing, the data is downsampled to 60 Hz to reduce computational complexity while preserving temporal resolution relevant to ERP analysis.

Each stimulus flash is labeled as either *Target* or *Non-Target* based on whether it corresponds to the attended character. The dataset exhibits strong class imbalance, with Non-Target trials significantly outnumbering Target trials.

A total of **3938 epochs** were extracted, consisting of **655 Target epochs** and **3283 Non-Target epochs**. This imbalance motivates the use of class-weighted learning and evaluation metrics beyond raw accuracy.

3 Overall System Workflow

The complete system follows a multi-phase pipeline inspired by standard EEG-based BCI design principles. The workflow begins with understanding the P300 paradigm and ends with a fully functional real-time speller system.

The pipeline is divided into six major phases:

1. Background study and paradigm understanding
2. EEG data acquisition and preprocessing
3. Feature extraction and model development
4. Model evaluation and selection
5. Real-time system integration
6. Testing, optimization, and finalization

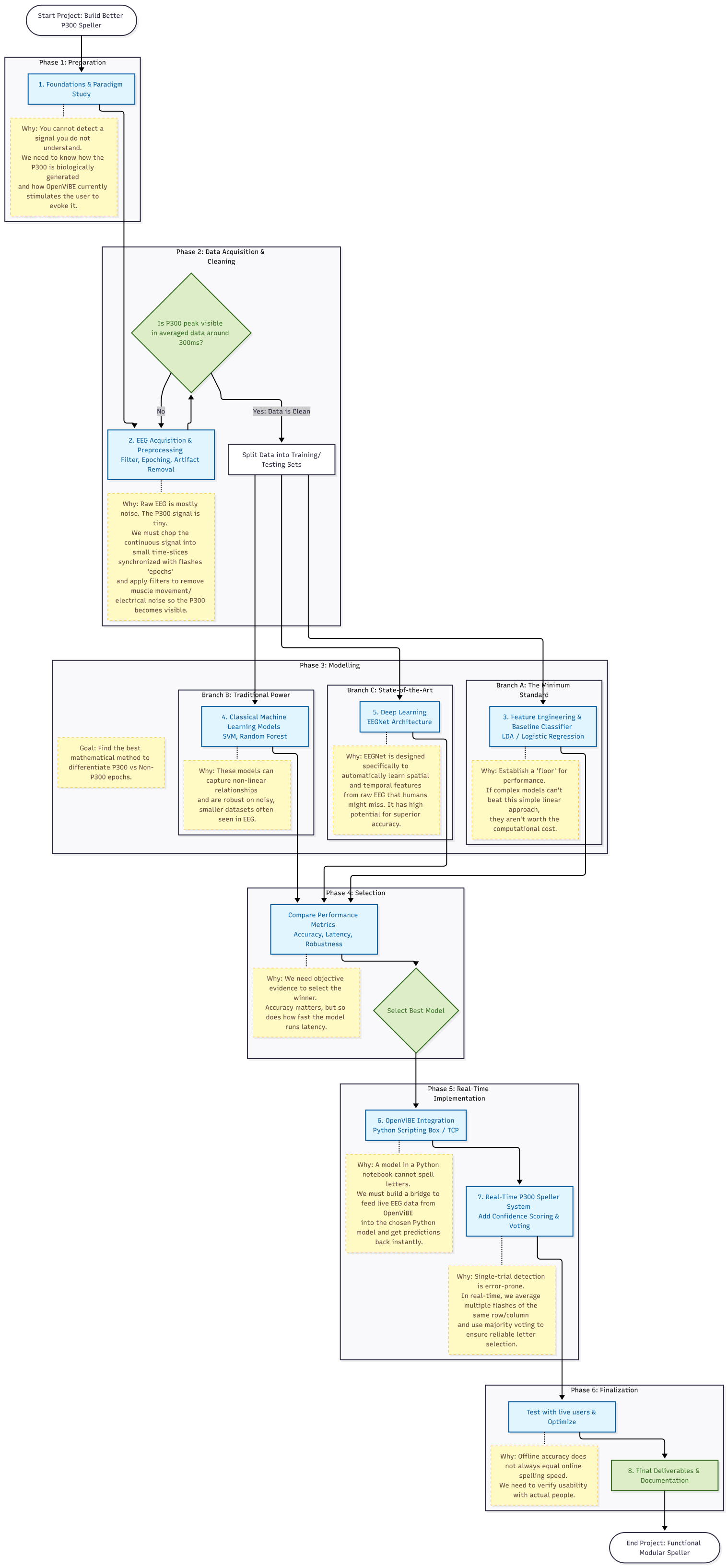

Figure 1: End-to-end workflow of the P300-based EEG speller system

4 Detailed Working of the Model and System

4.1 Epoch Extraction and Baseline Correction

Continuous EEG data is segmented into stimulus-locked epochs of length **0–800 ms post-stimulus**. Each epoch begins at the stimulus onset and captures the temporal dynamics of the P300 response.

Baseline correction is applied by subtracting the mean of the first **50 ms** of each epoch from all subsequent samples. This step reduces slow drifts and inter-trial variability while preserving task-related neural activity.

5 Feature Engineering

To transform preprocessed EEG epochs into machine learning–compatible representations, multiple feature extraction techniques are implemented and evaluated.

5.1 Time-Domain Features

Each EEG epoch is represented by concatenating the raw time-domain samples across all channels. For an epoch of shape (samples $\times$ channels), this results in a single high-dimensional feature vector. This approach preserves the complete temporal structure of the ERP response.

5.2 Principal Component Analysis (PCA)

To reduce dimensionality and suppress noise, Principal Component Analysis (PCA) is applied to the time-domain feature vectors. PCA is performed with whitening enabled, and 20 principal components are retained, capturing the most discriminative variance in the data.

5.3 Common Spatial Patterns (CSP)

For binary classification, Common Spatial Patterns (CSP) are employed to learn spatial filters that maximize variance differences between Target and Non-Target classes. Log-variance features extracted from spatially filtered signals are used for classification.

5.4 Train–Test Split

Extracted features are split into training and testing sets using a stratified 75–25 split to preserve class distribution. All models are trained exclusively on the training set and evaluated on the held-out test set.

6 Classical Machine Learning Models

Multiple classical machine learning classifiers are trained and evaluated on the extracted features.

6.1 Linear Discriminant Analysis (LDA)

LDA is used as a baseline classifier with equal class priors, providing a simple and interpretable reference model.

6.2 Logistic Regression

A class-weighted logistic regression model is trained using the liblinear solver to compensate for class imbalance.

6.3 Support Vector Machine (SVM)

An SVM with a radial basis function (RBF) kernel is implemented with balanced class weights to improve sensitivity to Target trials.

6.4 Random Forest

A Random Forest classifier with 100 trees is trained using balanced class weights. While highly accurate, it exhibits bias toward the majority class.

6.5 Gradient Boosting

Gradient Boosting is trained using manually computed sample weights derived from class frequencies, improving recall for P300 detection.

7 Results and Observations

Model	Accuracy (%)	F1-score
LDA	60.15	0.35
Logistic Regression	58.38	0.33
SVM (RBF)	54.57	0.32
Random Forest	83.25	0.04
Gradient Boosting	60.53	0.33

Table 1: Performance of baseline and classical machine learning models on Subject A

Due to the strong class imbalance in the dataset, accuracy alone is a misleading evaluation metric. The F1-score provides a more reliable measure of P300 detection performance by balancing precision and recall for the minority Target class.

Key observations:

- Random Forest achieves the highest accuracy but performs poorly on Target detection, as reflected by a very low F1-score.
- SVM and Gradient Boosting provide better sensitivity to Target trials despite lower overall accuracy.
- Baseline linear models offer moderate performance and serve as effective reference points.

8 Conclusion

This project demonstrates a complete P300-based EEG speller system integrating signal preprocessing, feature engineering, and classical machine learning models. The results highlight the importance of proper evaluation metrics and class imbalance handling in EEG-based BCIs. Future work can explore adaptive learning, cross-subject generalization, and fully online deployment.