**EEG Preprocessing Protocol for P300-Based Classification**

This document outlines the standardized preprocessing steps for EEG data collected using the Unicorn headset and the P300 GUI system (Phase 1 and Phase 2). The goal is to generate consistent, high-quality training data for machine learning models that classify directional intent based on P300 signals.

**📋 Overview**

* **Headset**: Unicorn EEG
* **Channels**: 8
* **Sampling Rate**: 250 Hz
* **Stimuli**: Single image flashes on 1 of 4 directions
* **Flash Rate**: Every 0.2 seconds (FLASH\_DURATION = 0.2, INTER\_FLASH\_DELAY = 0.2)
* **Trial Duration**: 10 seconds of flashing per trial (plus 10s baseline and 2s cue)

**✅ Preprocessing Pipeline**

**1. Load Raw EEG Data**

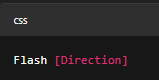
* Import the raw .csv file output by the Unicorn system
* Ensure the data has correct headers and 8 EEG channels

**2. Discard Initial Calibration Period**

* **Remove the first 30 seconds** of data (used for settling electrode impedance)
* At 250 Hz, discard the first 30 × 250 = 7500 rows

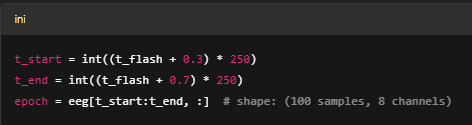
**3. Parse Event Log (from GUI)**

* Load the corresponding GUI-generated trial log .csv for the same session
* Extract timestamps of all events of type:
* Each flash time corresponds to a stimulus onset



**4. Epoching Around Flashes**

* For each flash timestamp t\_flash, extract a **400 ms EEG segment starting 300 ms after the flash** (i.e., from t\_flash + 0.3s to t\_flash + 0.7s)
* Convert this time window to sample indices:



**5. Label Each Epoch**

* If the **flashed direction** matches the **true direction the participant was focusing on** (provided by the trial log context), label the epoch as:
  + 1 → Target (P300 expected)
* Otherwise:
  + 0 → Non-target (no P300 expected)

**6. (Optional) Bandpass Filter**

* Apply a bandpass filter: **1–15 Hz**
* Helps isolate slow ERP components like the P300 and remove muscle noise / drift

**7. Baseline Correction (Optional but Recommended)**

* Subtract the mean of a pre-flash baseline (e.g., 200 ms before flash) **if available** — if not, you may skip this if you filter the data

**8. Artifact Rejection (Optional)**

* Reject any epochs with:
  + Voltage swing > ±100 µV (eye blinks, motion)
  + Channels with flat lines

**9. Flatten and Store**

* Each epoch (100 samples × 8 channels) → Flatten into 1D feature vector (shape: 800)
* Store the matrix:
  + X: Features, shape (n\_epochs, 800)
  + y: Labels (0 or 1), shape (n\_epochs,)

**🔗 Summary Table of Parameters**

| **Step** | **Value** |
| --- | --- |
| Sampling Rate | 250 Hz |
| Epoch Duration | 400 ms |
| Epoch Start Offset | 300 ms after flash |
| Samples per Epoch | 100 |
| Filter Range | 1–15 Hz (optional) |
| Baseline Period | 200 ms before flash (optional) |
| Labeling Method | Target vs. Non-Target Flash |

**❗ Final Notes**

* **Consistency is key**: Use the exact same parameters and file naming each session.
* **GUI logs and EEG files must be matched** for accurate flash-aligned segmentation.
* **Don’t include data from cue (2s) or baseline (10s)** when training the model.
* For trial-level prediction, you may **average model scores per direction** and pick the one with the highest mean confidence.