EEG Preprocessing Pipeline(Project Dataset)

The EEG data was preprocessed using **EEGLAB** (MATLAB toolbox) following below pipeline. Each step is described below with its purpose and methods.

1. Data Loading

- The raw EEG data was stored in **BIDS format** (Brain Imaging Data Structure).
- EEG signals were loaded from `.edf` files using `pop_biosig`.
- Event information was imported from the corresponding `.tsv` files.
- Event markers were mapped into the EEGLAB `EEG.event` structure.

Purpose:
To organize the EEG data and event markers in a format suitable for preprocessing and further analysis.

2. Raw Data Inspection

- A segment of the **first 30 seconds** of raw EEG was plotted.
- Every 8th channel was visualized to check for **spikes, flat lines, or noise**.
- An interactive EEG plot was opened for manual inspection.

Purpose:
To detect potential issues such as excessive noise, artifacts, or flat channels early in the process.

3. Data Quality Statistics

- Standard deviation and mean were calculated for each channel.
- Channels with **very low variability (<1 μV std)** were flagged as **potentially flat channels**.

Purpose:
To identify noisy or non-functional electrodes.

4. Downsampling

- Original sampling rate: **2048 Hz**
- Data resampled to **512 Hz** using `pop_resample`.

Purpose:
To reduce computational load while retaining sufficient temporal resolution for ERP and spectral analyses.

5. High-Pass Filtering

- A **0.1 Hz high-pass filter** was applied using `pop_eegfiltnew`(EEG = pop_eegfiltnew(EEG, 0.1, [], [], false, [], 0))
- **Parameters:**
  - `0.1` → high-pass cutoff.
  - `[ ]` → no low-pass applied here.
- This removed very slow drifts (e.g., sweating, electrode movement).

Purpose:
To stabilize the signal baseline and remove low-frequency noise.

6. Bad Channel Detection and Removal

- Automatic detection performed using `pop_clean_rawdata` with criteria:
  - **Flatline Criterion:** Channels flat for >5s.
  - **Correlation Criterion:** Channels poorly correlated (<0.8) with neighbors.
  - **Line Noise Criterion:** Excessive 50/60 Hz noise.
  - Others set `off` to avoid overly aggressive cleaning.
- Detected bad channels were removed.

Purpose:
To eliminate unreliable electrodes that degrade data quality.

7. Average Referencing

- EEG was re-referenced to the **average of all remaining channels** using `pop_reref`(pop_reref(EEG, [] -> parameters : [] means use average of all channels)

Purpose:
To reduce reference bias and improve spatial consistency across electrodes.

8. Low-Pass Filtering

- A **40 Hz low-pass filter** was applied using pop_eegfiltnew(pop_eegfiltnew(EEG, [], 40, [], false, [], 0) ->means []->no high pass filter, 40->low pass cutoff at 40Hz)
- Bandpass range after filtering: **0.1–40 Hz**.

Purpose:
To remove high-frequency noise (muscle artifacts, line noise) and retain frequencies relevant to cognitive EEG analysis.

9. Epoching

- Data was segmented into **epochs of -1 to +3 seconds** around specific event markers:
  - `startOfRememberedClipFirstWatch`
  - `startOfRecognisedClipFirstWatch`
  - `startOfNotRecognisedClip`
- Each epoch was 4 seconds long (1 sec pre-stimulus, 3 sec post-stimulus(Captures late cognitive or memory responses), done using pop_epoch.

Purpose:
To analyze brain responses time-locked to experimental events.

10. Baseline Correction

- Baseline period: **-200 ms to 0 ms** (pre-stimulus).
- This baseline activity was subtracted from each epoch using pop_rmbase.

Purpose:
To remove pre-stimulus drifts and normalize the signal.

11. Artifact Rejection

- Automatic epoch rejection using `pop_autorej` with:
  - **Amplitude threshold:** 100 μV means rejects epochs exceeding plus or minus 100 microvolts
  - **Maximum rejections per iteration:** 5 means control rejection iterations.
- Bad epochs were removed from the dataset.

Purpose:
To exclude trials contaminated with large artifacts (e.g., eye blinks, muscle movements).

12. ERP and Trial Distribution Analysis

- Number of trials per condition (Remembered, Recognised, Not Recognised) was calculated.
- Distribution was visualized with bar and pie charts.
- **Event-Related Potentials (ERPs):**
  - Grand average ERPs were computed per condition (e.g., at Fz channel).
  - Compared across memory conditions.

Purpose:
To examine condition-wise differences in brain responses.

13. Final Pipeline

- Original sampling rate: **2048 Hz**
- Final sampling rate: **512 Hz**
- Original channels: **64**
- Channels after cleaning: depends on dataset (bad channels removed)
- Epoch length: **4 seconds**
- Trials per condition: reported (Remembered / Recognised / Not Recognised)
- Behavioral performance:
    - **Recognition rate:** (Remembered + Recognised) / Total
    - **Strong memory rate:** Remembered / Total

This preprocessing pipeline:

1. Loaded BIDS EEG data
2. Inspected raw quality
3. Downsampled and filtered signals
4. Removed bad channels and re-referenced
5. Applied bandpass filtering
6. Segmented into epochs
7. Corrected baseline
8. Rejected artifact epochs
9. Produced trial distributions and ERPs
10. Saved the final clean dataset