

Analysis and Preprocessing of Pure and Contaminated EEG Signals Using Time- and Frequency-Domain Methods

Rith Rajak
24116085

1 Introduction

Electroencephalography (EEG) is a widely used technique for recording neural activity with high temporal resolution. However, EEG recordings are highly susceptible to various sources of contamination, including power line interference, baseline drift, ocular artifacts, and motion-related noise. These artifacts often overlap spectrally with genuine neural activity, making preprocessing a critical and non-trivial step in EEG analysis.

The objective of this report is to analyze an EEG codebase with a focus on understanding signal characteristics, identifying contamination, and evaluating preprocessing strategies through systematic observation rather than descriptive summarization. The analysis progresses from clean EEG data to contaminated recordings and employs both time-domain and frequency-domain techniques to diagnose artifacts and assess the effectiveness and limitations of commonly used filters.

Rather than assuming preprocessing choices a priori, this report adopts a data-driven approach: contamination is first observed visually, then isolated analytically, localized spectrally, and finally addressed using targeted filtering methods. Emphasis is placed on understanding trade-offs between artifact suppression and neural signal preservation.

2 Dataset Description and Environment Setup

The analysis is conducted using EEG data stored in MATLAB `.mat` format and processed within a Python-based scientific computing environment. Core libraries include NumPy for numerical operations, SciPy for signal processing and Fourier analysis, Matplotlib for visualization, and MNE for EEG-oriented abstractions.

Two datasets are examined:

- **Pure EEG dataset:** A cleaned or reference EEG recording used as a baseline for comparison.
- **Contaminated EEG dataset:** A collection of multiple EEG simulations containing structured artifacts and noise.

Both datasets share a consistent structure of 19 channels sampled at 200 Hz, enabling direct comparison without dimensional or alignment adjustments.

3 Time-Domain Analysis of Pure EEG

Initial inspection of the pure EEG data is performed in the time domain. Multi-channel plots of the first 700 samples reveal oscillatory behavior centered around zero with amplitudes typically within $\pm 50 \mu V$, consistent with physiologically plausible EEG signals. Plotting all channels simultaneously allows for detection of inter-channel inconsistencies that would be difficult to identify in isolation.

A channel-specific inspection of the FP1 electrode demonstrates stable oscillations with no sustained drift or extreme transients, indicating suitability for further analysis. This step establishes a baseline expectation for both amplitude and temporal structure against which contaminated signals can be evaluated.

4 Frequency-Domain Characterization of Pure EEG

To complement time-domain observations, the frequency content of the pure EEG signal is examined using the Fourier Transform. A manual implementation of the Discrete Fourier Transform (DFT) is first employed to reinforce conceptual understanding of spectral decomposition, symmetry between positive and negative frequencies, and the relationship between sampling rate and frequency resolution.

Subsequent use of optimized FFT routines confirms that the spectral energy of the pure EEG signal is concentrated in low-frequency bands below approximately 30 Hz, with smooth decay at higher frequencies. This distribution aligns with known EEG rhythms and validates the internal consistency between time-domain and frequency-domain representations.

5 Time-Domain Analysis of Contaminated EEG

The contaminated EEG dataset contains multiple simulated recordings, each representing a different contamination scenario. Examination of a representative simulation reveals substantial deviations from the pure EEG baseline. Multi-channel time-domain plots show increased amplitude variability, abrupt transients, and reduced baseline stability across multiple channels.

Notably, when individual contaminated channels are viewed in isolation, the signals can appear superficially plausible. However, direct overlay of pure and contaminated FP1 signals reveals pronounced low-frequency drift and sustained transients present only in the contaminated recording. This demonstrates that contamination is often relative rather than absolute and may remain undetected without reference comparisons.

6 Isolation of Artifact Contribution

To explicitly characterize contamination, the pure FP1 signal is subtracted from the contaminated FP1 signal. The resulting difference signal provides a direct time-domain estimate of the artifact contribution under the assumption of additive contamination.

The difference signal exhibits large, smooth, low-frequency deviations localized in time, confirming that the dominant contamination is structured and non-neural rather than random noise. This result explains the sustained baseline shifts observed in the contaminated recordings and motivates further frequency-domain investigation.

7 Spectral Analysis of Contamination

Frequency-domain differencing is performed by subtracting the FFT magnitude spectrum of the pure signal from that of the contaminated signal. The resulting artifact spectrum reveals overwhelming dominance of very low-frequency components, with negligible energy above approximately 10–15 Hz.

This finding establishes that baseline drift and slow artifacts, rather than power line interference or broadband noise, are the primary sources of contamination in the dataset. Importantly, this spectral localization provides a principled basis for selecting appropriate preprocessing filters.

8 Notch Filtering for Power Line Interference

A 50 Hz notch filter is applied to the contaminated EEG signal to address power line interference. Zero-phase forward–backward filtering is used to preserve temporal alignment.

Frequency-domain analysis confirms effective attenuation of the 50 Hz component, while time-domain inspection shows minimal visible change. This result demonstrates that notch filtering improves spectral purity but does not address dominant low-frequency artifacts, highlighting the need for additional preprocessing steps.

9 High-Pass Filtering for Baseline Drift Removal

Guided by the artifact spectrum, high-pass filtering is applied to the contaminated signal to suppress low-frequency components. FFT analysis after filtering shows substantial reduction of near-zero frequency energy while preserving mid-frequency EEG activity.

Direct spectral comparisons before and after filtering confirm that high-pass filtering effectively targets the dominant contamination identified earlier. This step illustrates the importance of data-driven filter selection and demonstrates that high-pass filtering is better suited than notch filtering for baseline-related artifacts in this dataset.

10 Band-Stop Filtering and Trade-Off Analysis

As an alternative approach, a band-stop filter targeting the 0.5–4 Hz range is applied to suppress ocular artifacts. Time-domain results show effective reduction of large slow transients; however, frequency-domain analysis reveals substantial attenuation across the entire delta band.

Since the delta band contains legitimate neural activity, especially in frontal electrodes, this approach removes both artifact-related and physiologically meaningful components. This trade-off is confirmed quantitatively through FFT comparison and highlights a fundamental limitation of frequency-based artifact removal when neural and artifact spectra overlap.

11 Discussion

Across all analyses, a consistent theme emerges: effective EEG preprocessing requires diagnosis before treatment. Time-domain inspection alone is insufficient to identify subtle contamination, while frequency-domain diagnostics provide critical insight into artifact structure and guide appropriate filtering strategies.

The report demonstrates that while notch filtering is effective for narrowband interference, it cannot address broadband or low-frequency artifacts. High-pass filtering provides a more suitable compromise for baseline drift, whereas aggressive band-stop filtering risks excessive signal loss. These observations reinforce the need for cautious filter design and motivate the use of more advanced artifact-removal methods when preservation of low-frequency neural activity is essential.

12 Conclusion

This report presented a comprehensive analysis of pure and contaminated EEG signals using a combination of time-domain visualization, frequency-domain diagnostics, and targeted filtering techniques. By progressing systematically from observation to isolation and validation, the study demonstrated how artifacts can be identified, characterized, and mitigated in a principled manner.

Rather than treating preprocessing as a fixed pipeline, the analysis emphasized understanding the spectral footprint of contamination and selecting filters accordingly. The results show that no single filtering method is universally optimal; instead, preprocessing decisions must balance artifact suppression against preservation of neural information.

Overall, the report illustrates that rigorous EEG analysis depends not on applying standard techniques blindly, but on interpreting data-driven evidence across domains. This approach leads to more reliable preprocessing choices and deeper insight into both signal quality and neural dynamics.