Last Two Brain Cells

Souvik Saha (3704112)

Ananya Kulkarni (3647062)

Abstract

In this report, we discuss a framework for preprocessing and analyzing EEG data that builds on the findings of Ian Daly et al. about the neural correlates of emotions induced by music. By adding mne-ICLabel for artifact detection, FastICA for component analysis, and an automated rejection algorithm for epoch selection. Additionally, we suggest changing the filtering procedure such that a 50Hz notch filter is not required, which simplifies the preprocessing step. With more efficiency and data integrity in EEG studies on emotional reactions to music, these improvements should provide deeper insights into the neural basis.

Introduction

The current study undertakes an in-depth study of Electroencephalography (EEG) preprocessing techniques, highlighting the neurological foundations of emotional responses to music. (EEG) is a crucial non-invasive method for recording brain activity, but its data is sometimes affected by artifacts, therefore signal processing and artifact removal are important. A conventional method was carefully developed in the groundbreaking research of Ian Daly et al., which included manual artifact inspection, filtering using a 50Hz notch filter combined with a 0.1-45Hz bandpass filter, and Infomax ICA in EEGLAB, which was accomplished by downsampling to 200Hz. Building on this basis, we propose a novel framework that works with modern computational tools to challenge and possibly even eliminate standard filtering practices. Our goal is to develop a more streamlined and computationally efficient preprocessing pathway by utilizing mne-ICLabel for automatic artifact detection, applying an exclusive bandpass filter, incorporating downsampling early on, and implementing FastICA within mne-python. Our method has the potential to improve the accuracy of EEG data analysis. In this analysis, we highlight our approaches to standard methods, showing the impact on the accuracy and potential of EEG data in the field of music-induced emotional resonance.

Background

According to [Daly et. al., 2014], there are several standard procedures for preprocessing EEG data that are aimed at minimizing noise and artifacts such as electrode drift, electromyographic activity, and power line interference. Bandpass filtering attempts to maintain frequencies of value, while notch filtering at 50 Hz has been used to target power line noise. Though highly computational, Independent Component Analysis (ICA) is a commonly used technique for artifact removal. Considering these factors, we looked at the importance of each step and suggested an updated strategy that puts efficiency ahead of data quality. Our method relies on the idea that a more efficient preprocessing pipeline that maintains the integrity of the EEG can be achieved by properly reorganizing steps and implementing better algorithms.

Approach

The first step in the process is to import raw EEG data that has been structured with the BIDS standard. After the import of the raw data, a PSD analysis is conducted to determine the frequency content of the data. After that, the data is subjected to a bandpass filtering method that MNE-ICLabel supports, which isolates the signal of interest and reduces noise within the frequency range of 1Hz to 100Hz.

After filtering, the data is downsampled to 200Hz to control computational load and file size, maintaining data integrity while making the data easier to analyze. After that, the downsampled data is divided into 1-second interval epochs, resulting in structured data frames that can be processed further.

To identify and eliminate epochs with low-quality signals, the epoched data is subjected to AutoReject. After cleaning, the EEG signal is divided into components using ICA, which breaks down the signal into its parts by applying the Infomax algorithm through MNE-ICLabel. The components related to artifacts can be recognized and eliminated. The ICA components are labeled using MNE-ICLabel, which categorizes them according to the possibility that they originate from the brain. To further purify the signal, non-brain components detected by ICLabel are eliminated.

Finally, by applying the inverse of the ICA and removing the components associated with artifacts, the remaining components (which are thought to be brain-related) are utilized to recreate the signal. This produces a cleaned EEG signal that is suitable for additional research or analysis. The majority of artifacts are eliminated, and the data is reduced to more precisely reflect the neural activity.

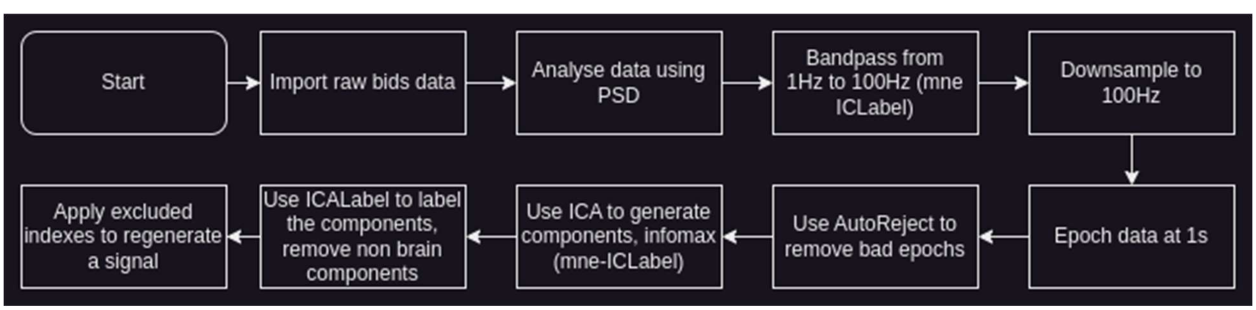

Fig. 1: Proposed Approach (we changed approach – downsample to 200Hz)

Analysis

1.The first step in the preprocessing pipeline - reads the raw data from the BIDS dataset, and adds a new column called value to the events.tsv file, which it gets from the events_keys.json file.
The missing values in the trial_type column are replaced with nan, and the rows with nan values are dropped.

2.The script preprocesses EEG data by importing it, applying bandpass filters, downsampling, and segmenting into epochs. It then uses AutoReject and ICA to remove artifacts, followed by a final filter to clean the signal. The process is visualized at each step, ensuring quality control.

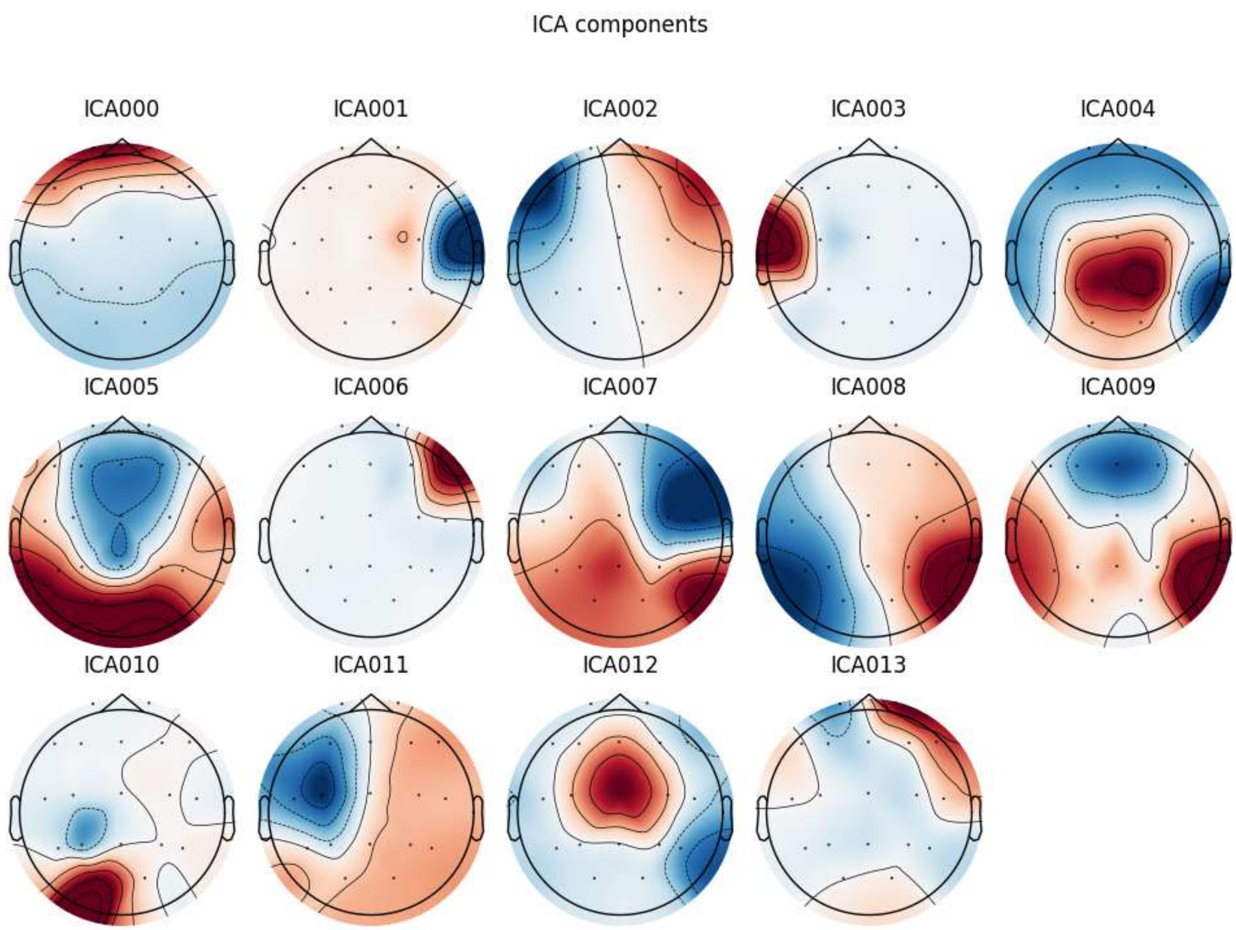

Fig. 2: ICA Components

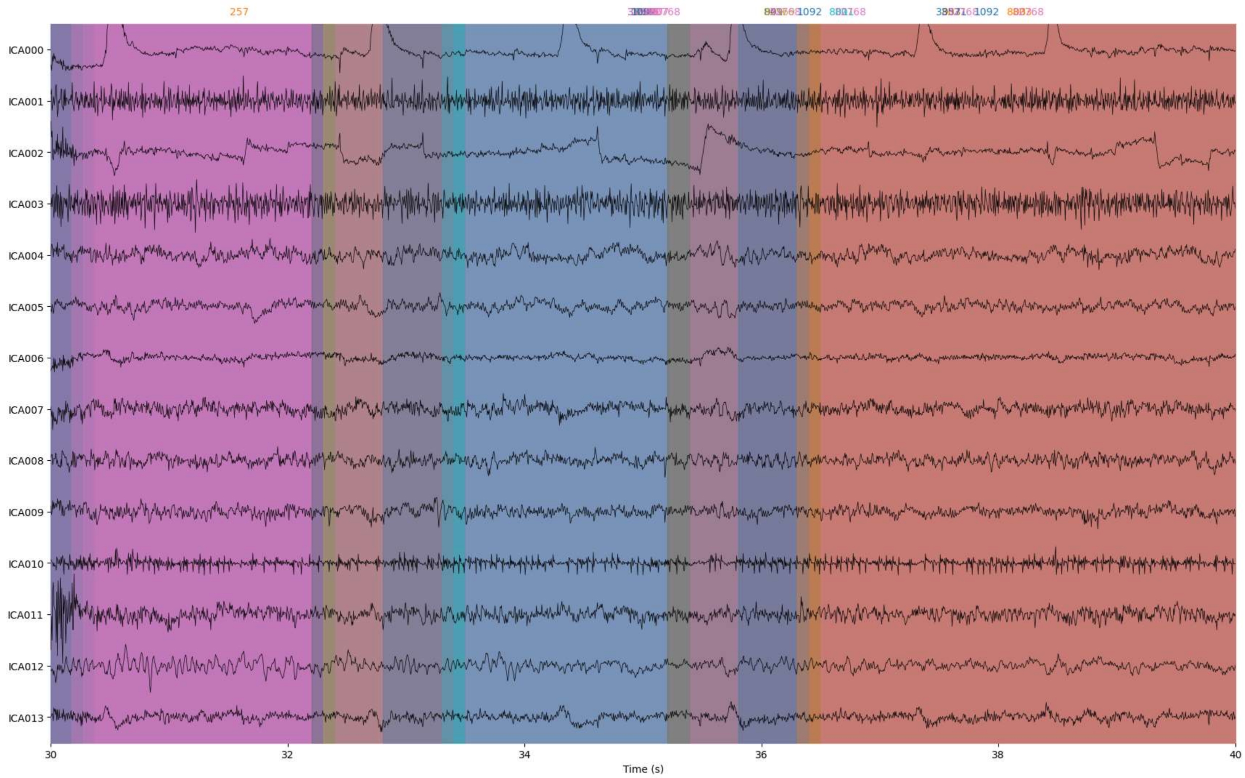

Fig. 3: ICA Components Plotted

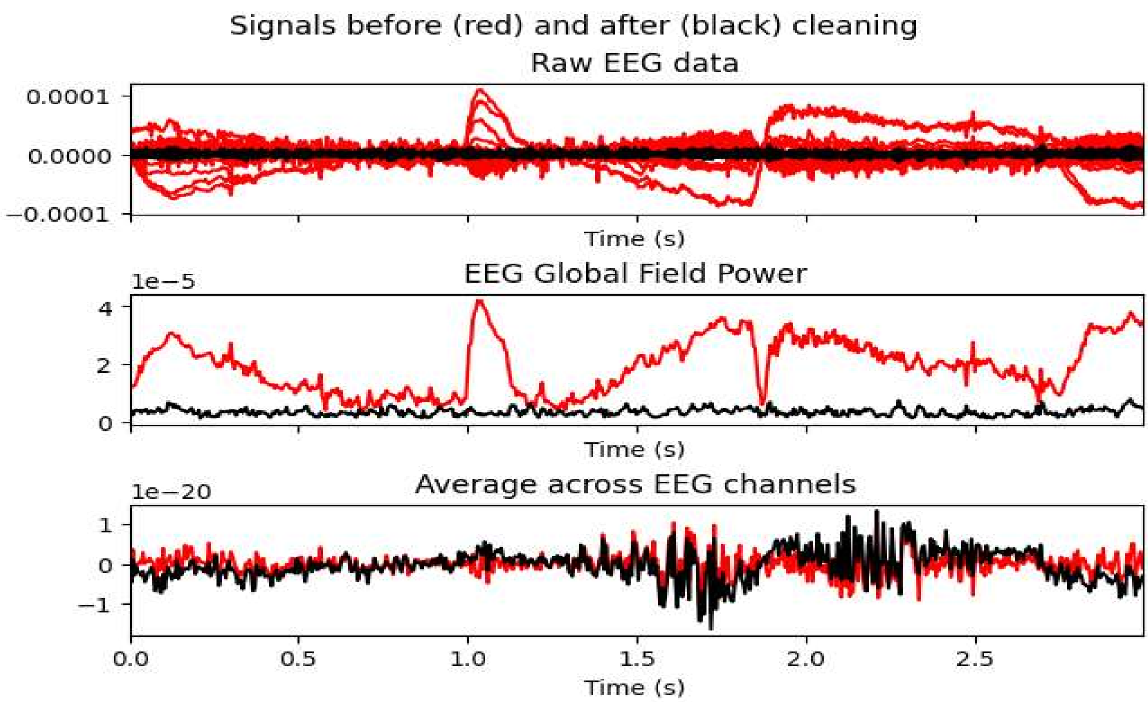

Fig. 4: Signals before and after cleaning raw data

This is how the data looks before and after preprocessing:

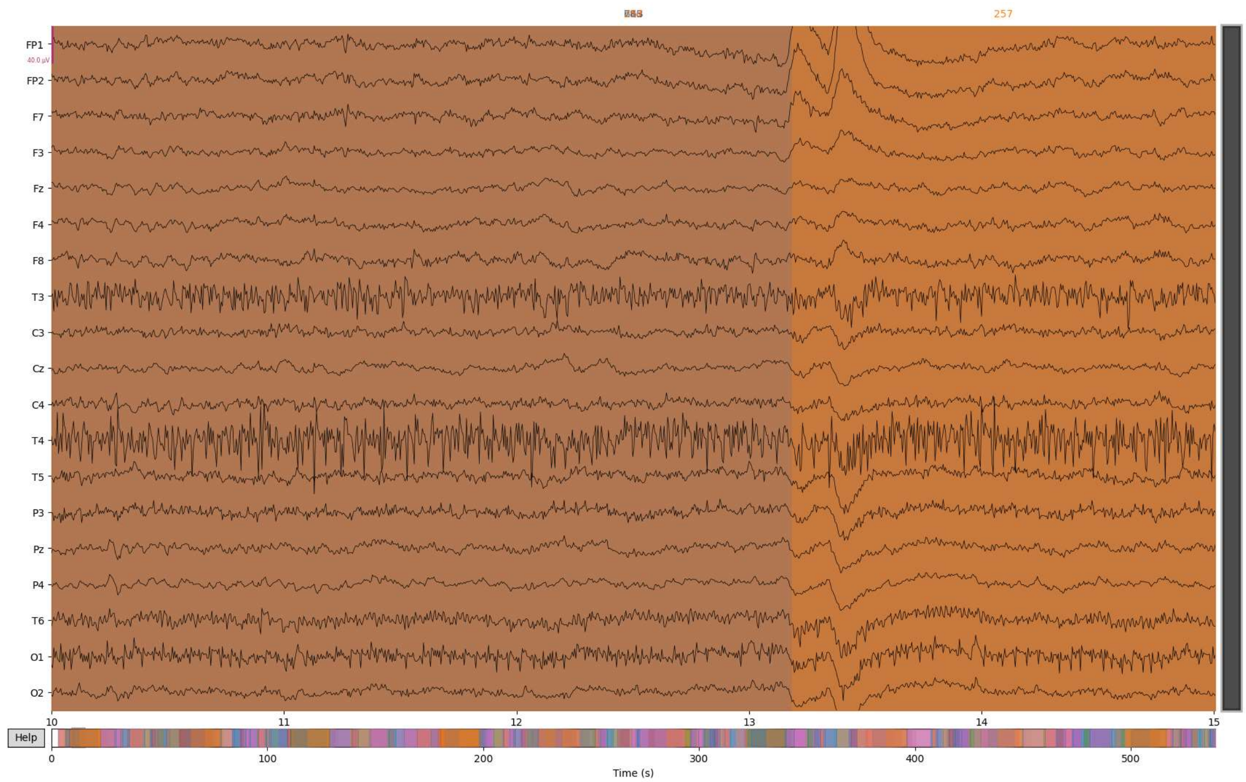

Fig. 5: Data before preprocessing

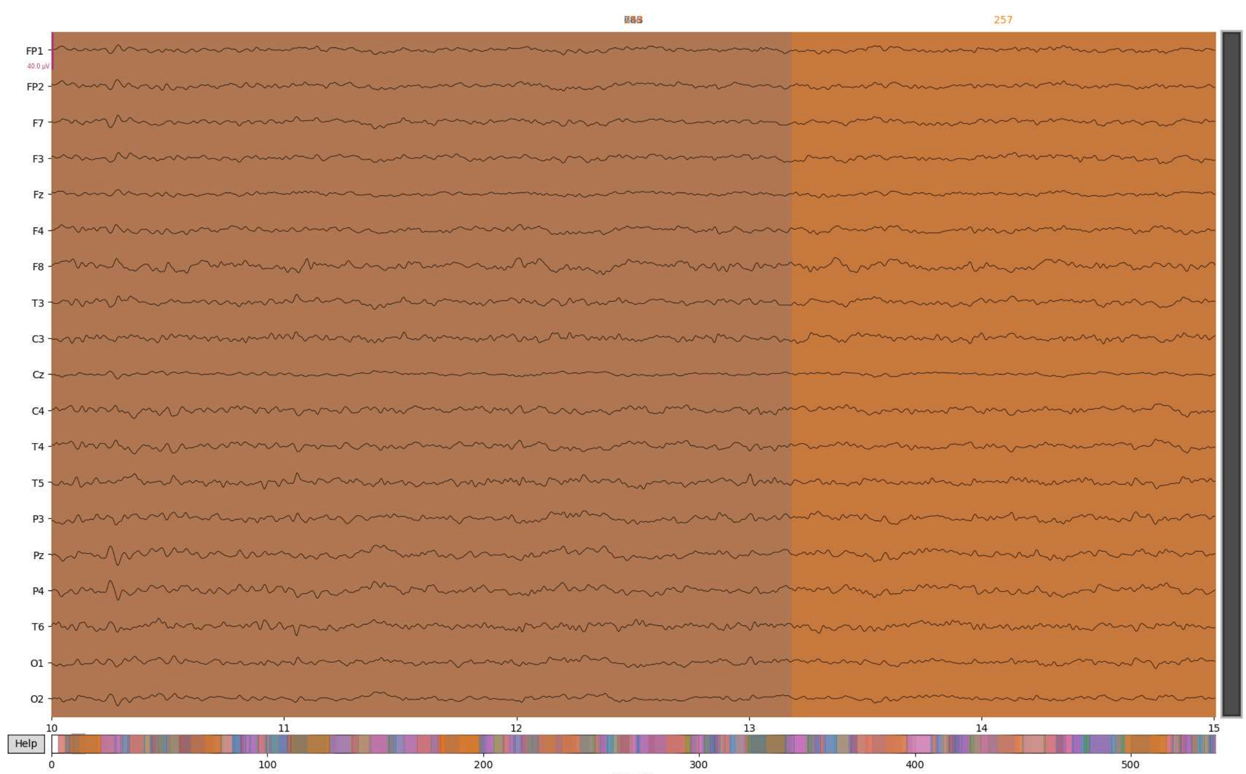

Fig. 6: Data after preprocessing

3. Execute a series of EEG data preprocessing steps on a BIDS dataset: it imports raw data and annotations, sets the EEG reference and montage, filters the frequency range, downsamples for efficiency, segments into one-second epochs, and cleans epochs using AutoReject. It then performs ICA, labels components via ICALabel, removes non-brain components, re-applies ICA to refine the data, filters out line noise, and finally saves the processed data in the EDF format.

4. A sanity check on the data by comparing the no music run psd on the music run run psd. A t-test is performed and the corresponding p value is used to check if the null hypothesis can be rejected.

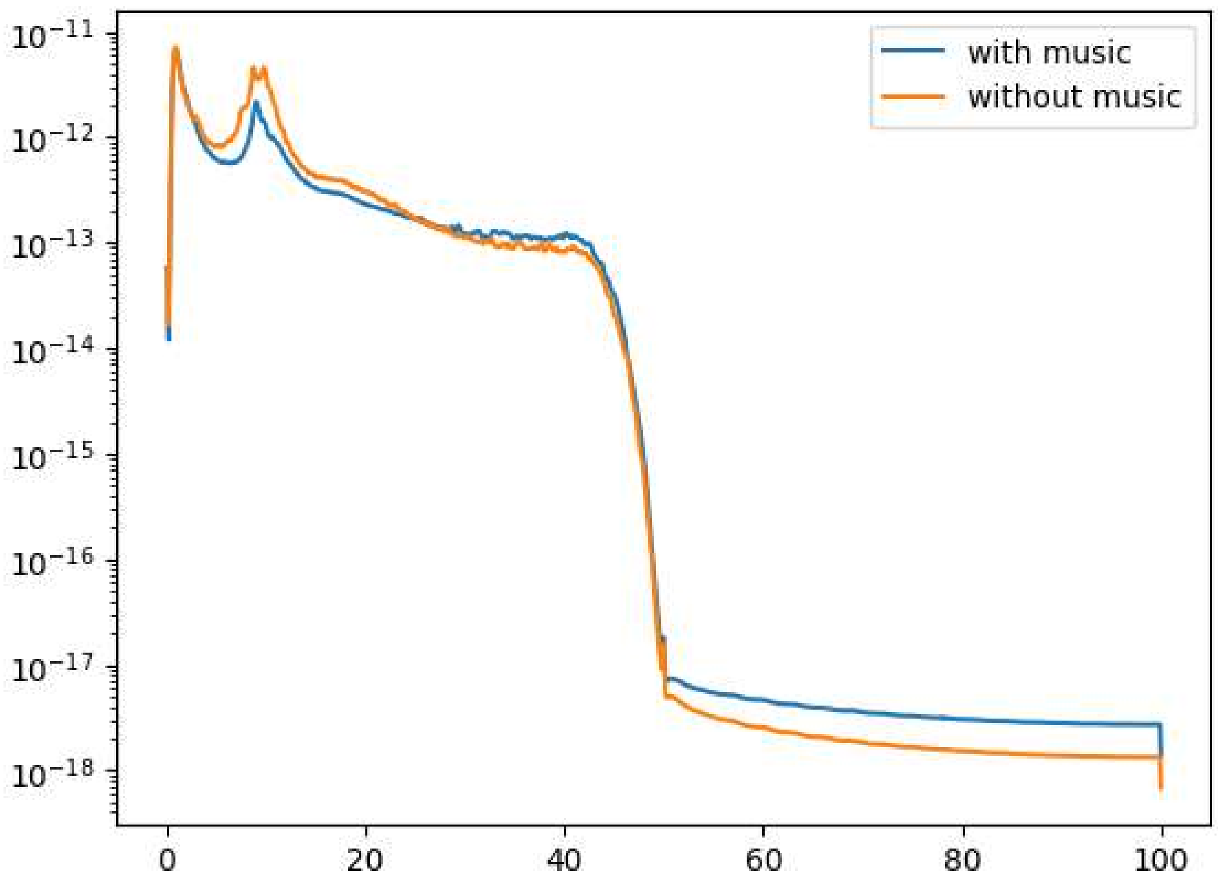

Fig. 7: With music and without music

5. Segmenting EEG epochs to record the brain response a subject has when listening to music. It carefully sorts and saves these segmented epochs into separate folders based on the specific audio files that are played. To provide an initial reference, the code also calculates the average EEG signal during times when no music was played.

6.Extract PCA information from the recordings available. Analyze and compare EEG signals across different music-induced brain states, identify the average EEG signal during silence, and perform a PCA to uncover the main features of the dataset.

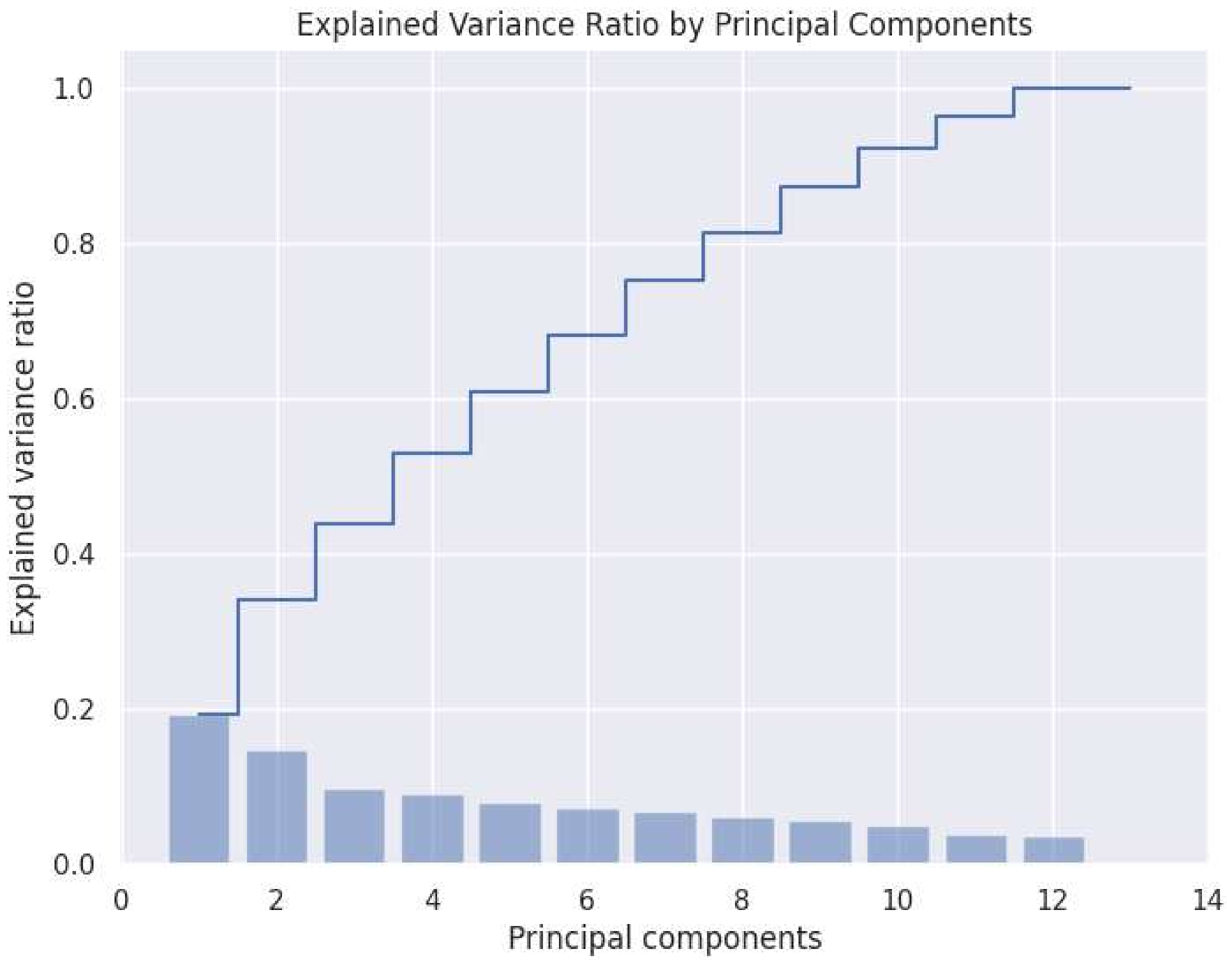

Fig. 8: Explained variance ratio by principal components

The graph, Fig. 8, explains the contribution of each Principal component to the net result. We can observe that first two components explain a large content of the variance.

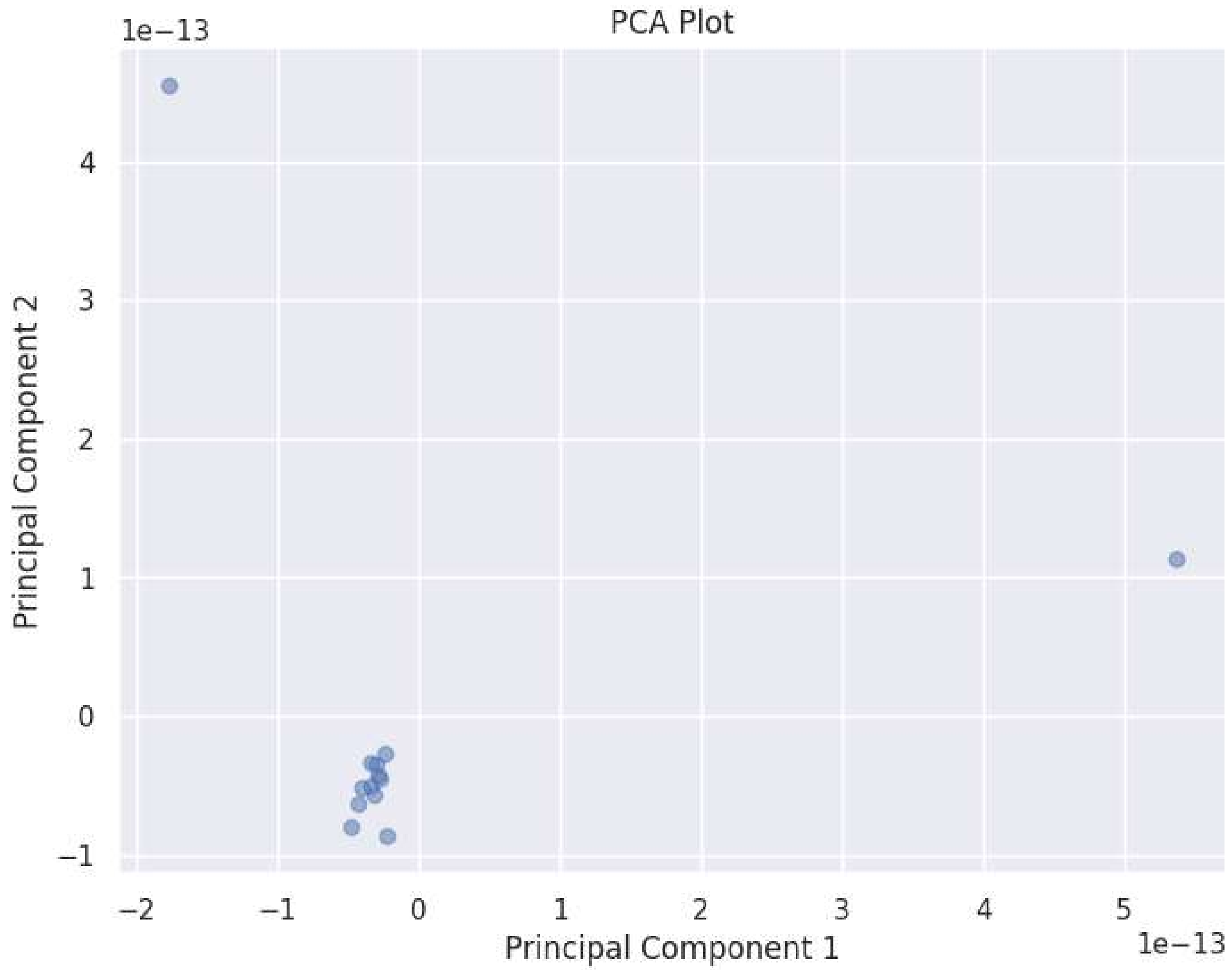

Fig. 9: PCA plot

The graph, Fig. 9, shows us how the first two components explain a large portion. We see 1 groups, and 2 outliers. We cannot come to a better conclusion without more proper data.

Challenges

Discrepancies in the dataset. We were unable to cross-reference the EEG data with the corresponding music because the original link for the dataset was inactive. We were able to identify a secondary group that was responsible for creating the music dataset.

Just 10% of the EEG runs had accurate annotations with the music tracks played, making up an important percentage of the dataset with incomplete annotations. This raised the question of whether to move forward with the available data or look for other ways to improve the dataset to conduct a thorough study.

Uncertainty of the EEG dataset, as it's unclear which set of music was used during the EEG recording sessions. The available documentation—the open neuro readme and the associated paper—offer conflicting information. Obtaining annotations regarding patient responses proved challenging. These annotations are crucial for correlating EEG data with subjective experiences during music listening, and their absence poses an obstacle to interpreting the data accurately.

We noted additional discrepancies. For instance, the event keys specified for music pieces (301-360) were scarcely present, leading to difficulties in identifying the music played. This inconsistency limited the usable data. Furthermore, a trial type labeled '1092-01-01' appeared frequently in the events.tsv files, which we think might represent a significant annotation but currently lack clarity.

Conclusion

Our project aimed to establish a framework for EEG data preprocessing in the context of music-induced emotional response analysis. Building upon the foundational work of Ian Daly et al., we proposed modifications that included utilizing mne-ICLabel for artifact detection and implementing FastICA within mne-python, which aimed to increase efficiency while maintaining data integrity.

Our modified approach streamlined the preprocessing steps, as evidenced by the clarity of signals post-processing and the structure revealed in the PCA analysis. However, our approach was not without its challenges. The inaccessibility of the music dataset, incomplete EEG annotations, and the ambiguity of the music sets used introduced significant challenges. These not only delayed our progress but also limited the scope of our analysis, restricting us to a fraction of the usable data.

Moreover, additional discrepancies within the dataset, such as the presence of essential event keys and the trial type '1092-01-01', only increased the complexity of the task at hand. Such issues highlighted the importance of meticulous dataset documentation and the need for reliable data sources when dealing with such sensitive and nuanced fields of study.

Despite these challenges, our project had insightful preliminary results. The visual and statistical analysis conducted showcases the potential of our proposed methods and highlights the necessity for a complete and accurately annotated data set for more conclusive research. Future studies would benefit from resolving the identified discrepancies and ensuring access to comprehensive music annotations, leading to a better understanding of how our brains react to music and the emotions it evokes.

References

Ian Daly, Asad Malik, Faustina Hwang, Etienne Roesch, James Weaver, Alexis Kirke, Duncan Williams, Eduardo Miranda, Slawomir J. Nasuto, Neural correlates of emotional responses to music: An EEG study, Neuroscience Letters, Volume 573, 2014, Pages 52-57, ISSN 0304-3940, https://doi.org/10.1016/j.neulet.2014.05.003.

Link to the dataset: An EEG dataset recorded during affective music listening - OpenNeuro