Analysis of Neural Data Final Report

Natali Kostadinovic

May 2025

Contents

1 Introduction

There have been studies in the past examining corresponding brain rhythms from induced emotions from music, and findings from EEGs that suggest relationships between certain frequency bands and emotional states induced by music; the data we used was part of a study that aimed to investigate neural states from emotions induced by music for carefully selected musical stimuli with a wide variety of styles. The music was selected such that a range of emotions would be induced (Daly et al., 2014) and such that the music is not well known to reduce the likelihood that participants would be familiar with the music

(Eerola, 2019), thus minimizing pre-selection bias (Daly et al., 2014). Past literature findings have suggested that brain rhythms will enter into the alpha frequency range when listening to happy music (Daly et al., 2014), and as the study we examined questioned participants if the music made them feel happy such that they could agree or disagree, we had a way of examining the relation between the alpha, beta, gamma, theta, and delta frequency bands and happiness induced by music for the data collected by Daly et al.

Our first hypothesis was that if people agreed that the music made them feel happy, their brain rhythms would fall into the alpha frequency range; for this, we conducted an independent t-test to compare the alpha band z-scores of the songs, grouped by either agreement for the emotion of happiness or disagreement by ratings. Our second hypothesis was that we can use the alpha, beta, gamma, theta, and delta frequency bands as predictors of whether the rating is in the category of low, medium, or high, corresponding to disagreement, neutrality, and agreement with the statement that the song made the participant feel happy. We tested this by performing multinomial regression with the bands of alpha, beta, gamma, theta, delta as predictors for the low, medium, or high happy ratings of the songs. We aimed to answer the following question: Does happiness induced by music correspond with any of the alpha, beta, gamma, theta, and delta frequency bands, and if so, can we use the corresponding band power to predict whether the song induced happiness?

2 Methods

2.1 Dataset Participants and Procedures

We are using a dataset from a research project called BCMI-MIdAS (Brain-Computer Music Interface for Monitoring and Inducing Affective States), which was collected between 2012 and 2017 and funded by the UK's Engineering and Physical Sciences Research Council (Daly et al., 2020). The dataset includes EEG recordings from 31 healthy adult participants (ages 18–66, 18 of whom were women). In the experiment, each participant listened to 40 short music clips, each lasting 12 seconds. These clips came from movie soundtracks and were picked because they were known to make people feel certain emotions. Each participant completed six sessions of EEG recording. The first and last sessions were resting sessions where they sat still for 5 minutes. The other four sessions were the main part of the experiment. In these, each person listened to 10 music clips in random order. A fixation cross appeared on the screen for 3 seconds to help the participant focus. Then, a 12-second music clip played. After a short pause (0.5 seconds), participants were shown 8 questions in random order, asking how they felt. These questions included words like happy, sad, tense, energetic, etc., and participants rated how strongly they agreed or disagreed using a Likert scale, rating from 1 (strongly disagree) to 9 (strongly agree). During all of this, EEG data was recorded using 19 sensors placed on the scalp. (Daly et al., 2014)

2.2 Data Cleaning

Participant 9 was removed because of missing data (ratings). For the other 30 participants, each listened to 40 songs, but brain activity for certain songs were excluded based on emotions for statistical analysis such as t-tests and multinomial regression since they were missing ratings, i.e. if there was missing ratings for the happy question for a music clip for some participant, brain activity for that clip was removed. In addition, the spectrum for each music clip of each participant was computed and the electrodes averaged together, then after averaging frequencies into five bands for each song, z-score normalization was performed. The Data Preparation section describes this normalization process in more detail.

2.3 Data Preparation

2.3.1 Extracting The Score of Happy

To extract the happy ratings from the dataset, we processed EEG event files for 30 participants across four experimental runs each. For every participant and run, we read the corresponding .tsv event file and parsed the sequence of trial codes. We identified trials marked with the "song" event code (788) and then searched forward in time for the corresponding "happy rating question" event (code 805). After locating the happy question, we searched for the participant's response code, which fell between 901 and 909. These numeric codes were converted into a 1-to-9 rating scale by subtracting 900. For each valid response, we recorded the participant ID, run number, song onset time, question onset time, and answer onset time, along with the final extracted rating. These were stored as dictionaries and appended to a list. After processing all available data, we combined the list into a single pandas dataframe called *happy_df*, containing 1,107 complete happy trials. This dataframe was used for subsequent statistical analysis, clustering.

```
Skipping participant sub-09 due to missing files.
Combined 'Happy' trials across all participants:
     participant  run  song_index  song_time  question_index  question_time  \
0         sub-01    2           4      5.689               8         22.290
1         sub-01    2          79     65.482             124        106.086
2         sub-01    2         150    119.310             173        151.002
3         sub-01    2         224    172.721             277        212.187
4         sub-01    2         295    224.824             332        266.202
...          ...  ...         ...        ...             ...            ...
1102      sub-30    5         287    247.304             333        285.802
1103      sub-30    5         364    303.030             401        340.073
1104      sub-30    5         434    356.567             486        405.021
1105      sub-30    5         503    417.895             548        466.002
1106      sub-30    5         572    480.298             594        512.202

      answer_index  answer_time  answer_code  rating  score
0                9       22.290          903       3      3
1              129      107.187          909       9      9
2              177      151.603          907       7      7
3              281      212.788          901       1      1
4              335      266.402          901       1      1
...            ...          ...          ...     ...    ...
1102           335      285.802          904       4      4
1103           403      340.073          904       4      4
1104           487      405.021          904       4      4
1105           549      466.002          903       3      3
1106           595      512.202          902       2      2

[1107 rows x 11 columns]
```

Figure 1: Dataframe of happy trials (see code chunk 8)

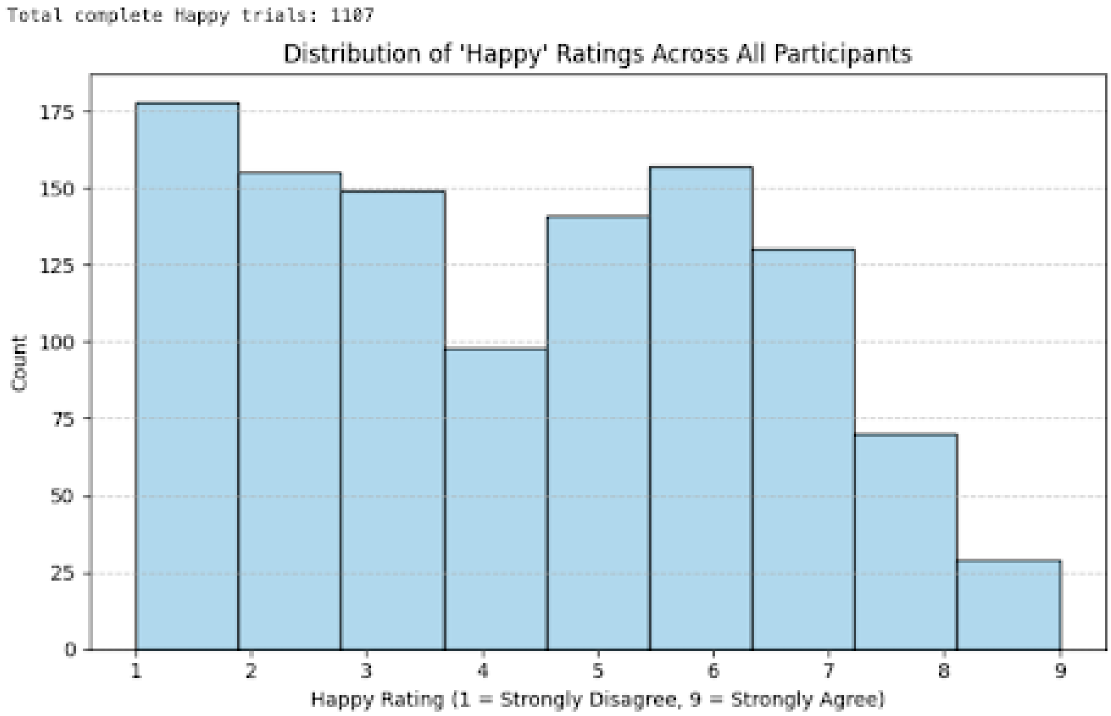

Figure 2: Distribution of happy ratings for all participants (see code chunk 8)

2.3.2 Extracting The Score of Angry

To extract the angry ratings from the dataset, we processed EEG event files for 30 participants across four experimental runs each. For every participant and run, we read the corresponding .tsv event file and parsed the sequence of trial codes. We identified trials marked with the "song" event code (788) and then searched forward in time for the corresponding "angry rating question" event (code 803). After locating the angry question, we searched for the participant's response code, which ranged from 901 to 909. These numeric codes were converted into a 1-to-9 rating scale by subtracting 900. For each valid response, we recorded the participant ID, run number, song onset time, question onset time, and answer onset time, along with the final extracted rating. These were stored as dictionaries and appended to a list. After processing all available data, we combined the list into a single pandas dataframe called *angry_df*, which contained all complete angry trials. This dataframe was used for subsequent statistical analysis and comparison to other emotion-specific responses.

```
Skipping participant sub-09 due to missing files.
Combined 'Angry' trials across all participants:
     participant  run  song_index  song_time  question_index  question_time  \
0         sub-01    2           4      5.689              45         45.202
1         sub-01    2          79     65.482             133        109.402
2         sub-01    2         150    119.310             206        164.202
3         sub-01    2         224    172.721             254        202.802
4         sub-01    2         295    224.824             341        268.717
...          ...  ...         ...        ...             ...            ...
1165      sub-31    5         351    284.506             355        301.007
1166      sub-31    5         420    330.058             443        358.602
1167      sub-31    5         494    381.289             538        423.402
1168      sub-31    5         572    443.430             595        469.618
1169      sub-31    5         641    497.511             645        514.012

      answer_index  answer_time  answer_code  rating  score
0               46       45.202          904       4      4
1              139      110.603          901       1      1
2              211      165.303          904       4      4
3              259      203.903          901       1      1
4              345      269.318          901       1      1
...            ...          ...          ...     ...    ...
1165           361      303.521          901       1      1
1166           445      358.602          903       3      3
1167           539      423.402          903       3      3
1168           603      476.032          904       4      4
1169           647      514.112          901       1      1

[1170 rows x 11 columns]
```

Figure 3: Dataframe of angry trials (see code chunk 9)

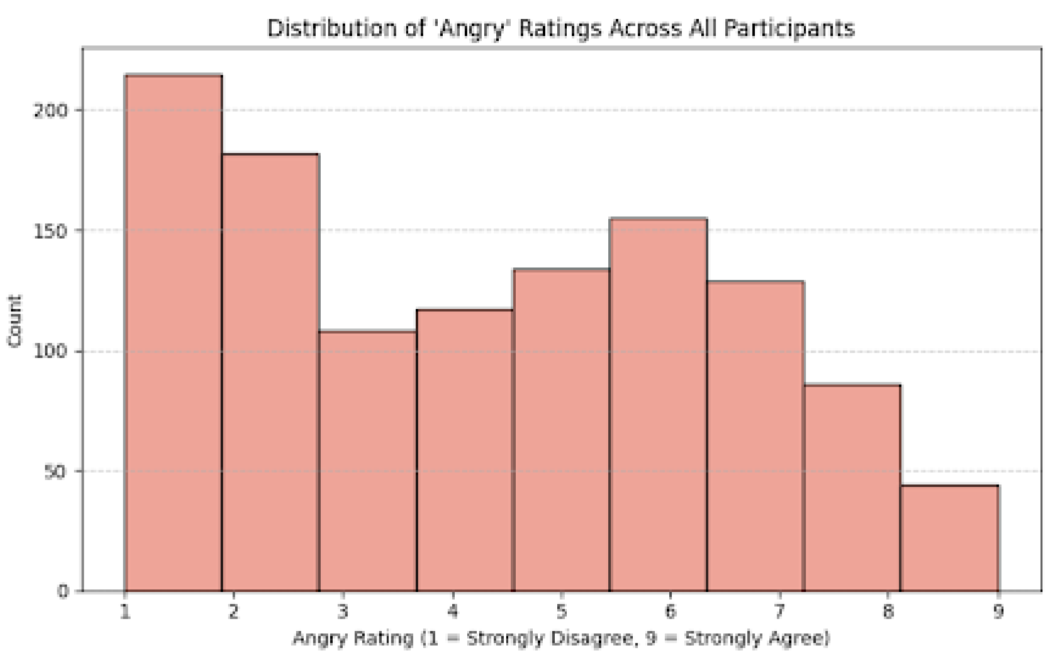

Figure 4: Distribution of angry ratings for all participants (see code chunk 9)

2.3.3 Extracting Power Spectral Density

For each participant, EEG data were segmented into 12-second music-listening epochs based on onset timings from the associated events files. We used the multitaper method to compute the power spectral density (PSD) for each trial, focusing on frequencies between 0–60 Hz. The frequency range of 0–60 Hz covers EEG frequency bands, including delta (0–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (13–30 Hz), and gamma (30–60 Hz). PSDs were computed using all 19 EEG channels, and for each trial, we averaged across channels to obtain a single representative PSD curve. This resulted in one PSD vector per song per participant. All PSD vectors were stacked across trials and subjects to form the final dataset for analysis. The final PSD matrix has a shape (1240, 721), where each row corresponds to one single song trial and 721 denotes the number of frequency bins between 0 and 60 Hz.

To mitigate individual differences in spectral power, we applied within-subject z-score normalization to use for bootstrapping later on. Specifically, for each participant, PSDs across the 40 songs were standardized by subtracting the participant's own mean PSD and dividing by their standard deviation. This normalization emphasized relative fluctuations in power across songs within each subject while controlling for between-subject variability.

2.3.4 Band-Level Features

After computing full-spectrum power spectral densities (PSDs) for each trial, we segmented the spectrum into five EEG bands: delta (0–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (13–30 Hz), and gamma (30–60 Hz). For each band, we computed the mean power by averaging across all frequency bins that fell within the corresponding range. This resulted in a five-dimensional feature vector per trial, where each value represents the average spectral power within a defined frequency band. Therefore, each participant contributed 40 trials, resulting in a matrix with 40 rows (one per song) and 5 columns (corresponding to delta, theta, alpha, beta, and gamma band power). All averaged band power vectors were stacked across trials and subjects to form the final dataset for analysis.

Since the band-level features were derived from unstandardized PSDs, we applied within-subject z-score normalization to the resulting power values. Specifically, we subtracted each participant's mean power (averaged across all 40 songs) from each trial's band power and divided it by the standard deviation across their 40 trials. This transformation was applied separately for each frequency band, ensuring that each subject's data was scaled based on their own distribution.

To reduce the influence of extreme values, we identified and removed outliers based on z-scored band power values. Specifically, after visual inspection of the distribution using histograms, we defined any trial with an absolute alpha band z-score greater than 3 as an outlier. These trials were excluded from subsequent analyses using conditional filtering in code.

2.3.5 Linking Band-Level Features with Happy Rating Score

To enable further statistical modeling, the band-level features were linked with corresponding behavioral ratings (happy rating score). This was achieved by merging the rating score with standardized z-scored band power matrix. Each row represents one song from one participant.

Before merging, we organized the rating dataset by sorting trials within each participant and run (i.e., runs 2–5) according to song presentation order. A trial index was then assigned to each song within a run using a cumulative counter, ensuring consistent alignment between behavioral ratings and EEG-derived features. For the band power features, we also added trial-level indexing to enable accurate merging with the rating data. Specifically, we generated a new column called "trial_index" for each song within each run (runs 2 to 5), assuming 10 songs per run and 4 runs per participant. This indexing approach ensured that each trial could be matched to its corresponding behavioral rating based on participant ID, run number, and song order.

Finally, we merged the behavioral ratings and band power features using participant ID, run number, and trial index as matching keys. Since the rating dataset only contained 1,146 entries due to the exclusion of participant 9 and missing ratings of some songs in other participants, we used a left merge to retain only trials for which both a valid behavioral rating and corresponding band power features were available. This ensured that the final dataset included only matched trials with complete data, suitable for subsequent statistical analyses.

To enable frequency analysis, we also merged the full-spectrum PSD data (721 frequency bins) with the rating dataframe. First, each PSD row was associated with its corresponding participant ID, run, and trial index to form a full PSD dataframe. This was then merged with the rating dataframe on "participant", "run", and "trial_index", resulting in a final dataset where each row represented a single song trial, with both spectral EEG features and behavioral ratings aligned for subsequent analysis.

2.4 Statistical Analysis

We began our statistical analysis by using a multinomial logistic regression model to predict which of the three clusters each data point belonged to, based on EEG frequency band values. To prepare the data, we first extracted the happy rating from participants' responses to the music clips. This rating was used as the target variable because it represents a clear emotional dimension and could reflect differences in brain activity across the three frequency bands. In this analysis, we included only three EEG frequency bands: theta (4 to 8 Hz), alpha (8 to 12 Hz), and beta (13 to 30 Hz).

Before building the model, we divided the dataset into training and test sets. The model was trained on the training data and evaluated on the test data. After fitting the model, we looked at the confusion matrix and the classification report. The results showed that the model did not perform well and had trouble correctly identifying which cluster each data point belonged to.

```
X shape: (1107, 3), y shape: (1107,)
Training Accuracy: 35.03%
Cross-Validated Accuracy: 35.05% ± 0.18%

Classification Report:
              precision    recall  f1-score   support

           0       0.35      1.00      0.52        78
           1       0.00      0.00      0.00        77
           2       0.00      0.00      0.00        67

    accuracy                           0.35       222
   macro avg       0.12      0.33      0.17       222
weighted avg       0.12      0.35      0.18       222
```

Figure 5: Classification report of the model (see code chunk 14)

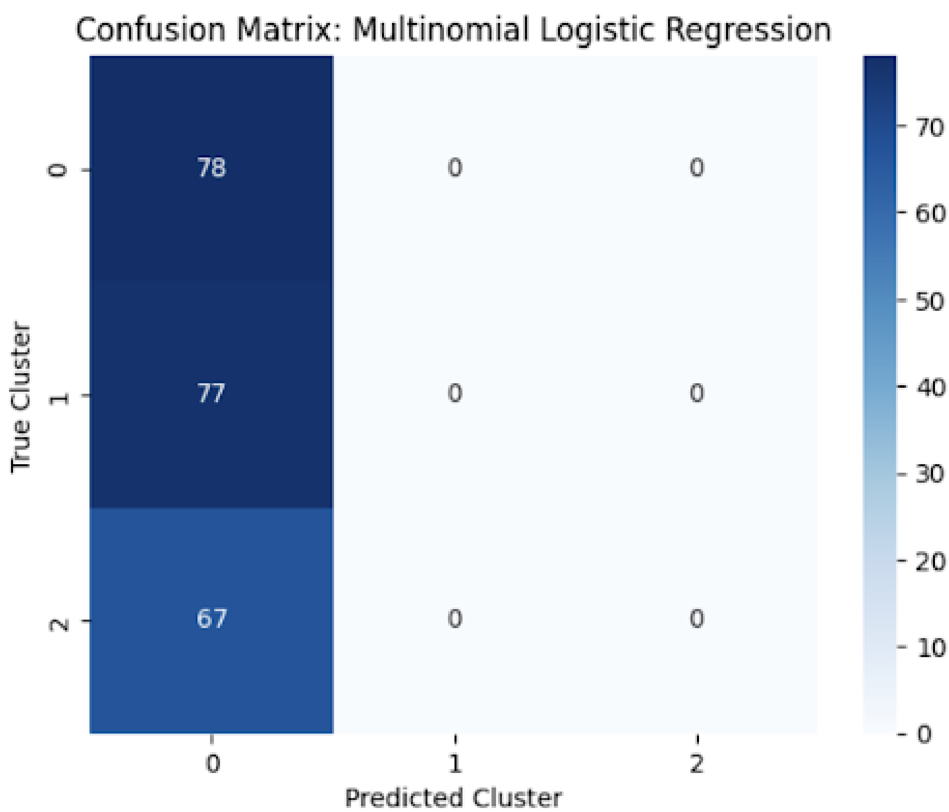

Figure 6: Confusion matrix of the model (see code chunk 14)

Because the logistic regression model performed poorly, we tried using a random forest classifier. We used five-fold cross-validation to test how well this model could generalize to new data. However, the accuracy was still low, around 32%, which is worse than what we would expect by guessing at random.

```
Cross-Validation Results (5-fold):
Accuracy: 0.3207 ± 0.0159
Precision (macro): 0.3211 ± 0.0166
Recall (macro): 0.3190 ± 0.0169
F1-score (macro): 0.3183 ± 0.0167
```

Figure 7: Cross validation results of the random forest classifier (see code chunk 15)

To better understand why the models performed so poorly, we checked whether the clusters had similar numbers of data points. We also did t-tests to compare the average happy ratings between the clusters. The results showed that all the differences were statistically significant. Cluster 0 was different from both Cluster 1 and Cluster 2, and Cluster 1 was different from Cluster 2. The p-values were all much smaller than 0.05, which means the differences were very unlikely to be due to chance. This confirmed that the clusters were clearly separated based on the ratings and the low accuracy was not caused by class imbalance.

```
Cluster Statistics:
   rating_cluster  cluster_size  mean_rating
0               0           388     3.979381
1               1           386     6.924870
2               2           333     1.465465

Inertia: 713.51
Silhouette Score: 0.633
```

Figure 8: Cluster statistics and t-test between clusters results (see code chunk 10)

```
Class distribution:
   Class  Count
0      0    388
1      1    386
2      2    333
```

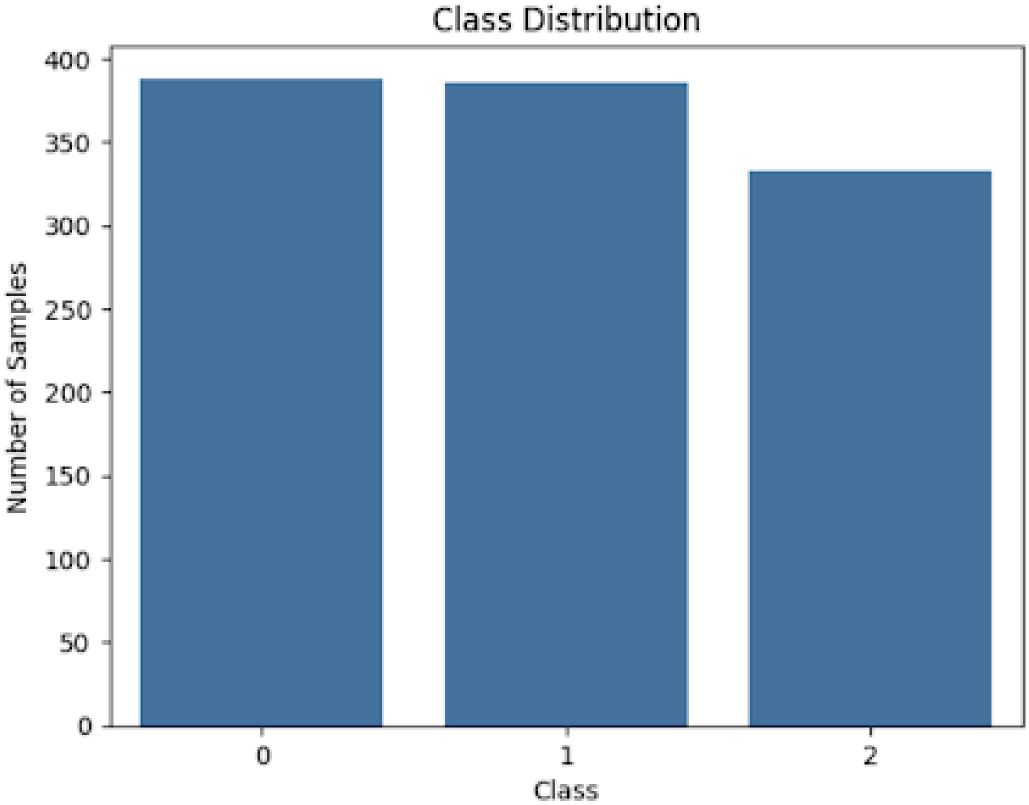

Figure 9: Class distribution (see code chunk 16)

We then looked at the EEG features and realized they might not be informative enough. One possible reason is that we did not normalize the data and used the frequency bands in their raw form. Another issue is that we only used three frequency bands, which may not have captured enough detail in brain activity to allow for accurate classification. We made pairwise distribution plots to see how the features were spread across the clusters. These plots showed that there was a lot of overlap between the clusters, meaning there were no clear boundaries that could help separate the groups based on the EEG data.

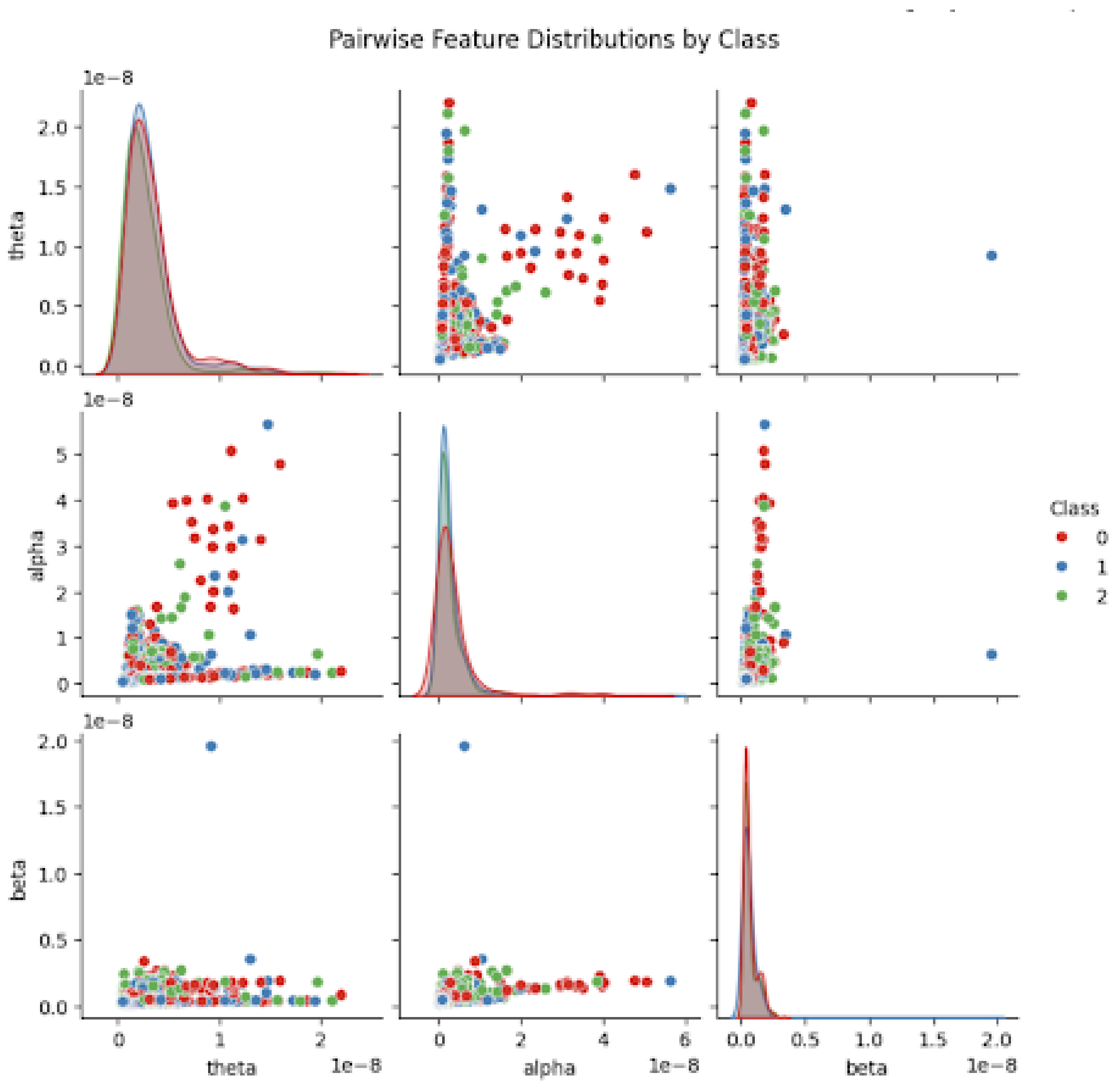

Figure 10: Pairwise feature distributions grouped by class (see code chunk 16)

To make sure the clustering itself was valid, we used the elbow method to find the best number of clusters. This method showed that three clusters were a good choice. This confirmed that the problem was not with the clustering, but rather with the features used in the classification models.

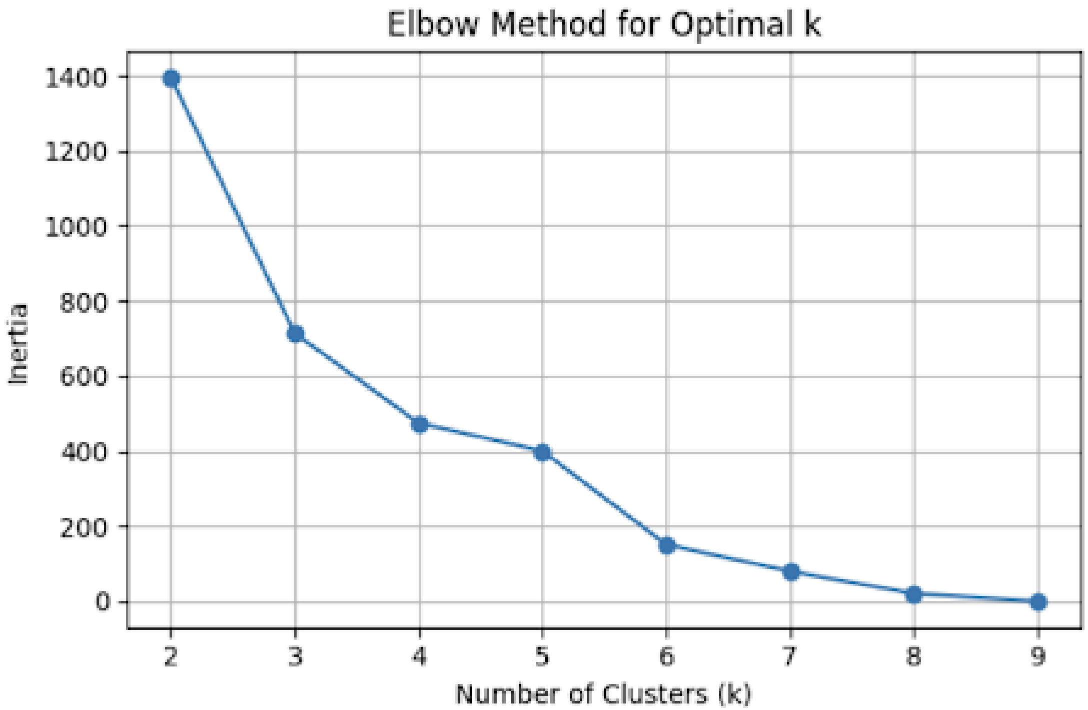

Figure 11: Graph of elbow method (see code chunk 10)

After checking all of these results, we concluded that the EEG features we used did not contain enough information to predict the emotional rating clusters. For future analysis, we should include five frequency bands instead of three by adding delta and gamma bands. We should also apply z-score normalization to the data so that all features are on the same scale and differences between participants are reduced. These changes may help the models perform better in identifying the emotional states based on EEG signals.

3 Results

3.1 Alpha Comparison of Listening to Happy vs Unhappy Songs

To investigate whether EEG band power differed between songs rated as happy (scores 6-9) and unhappy (scores 1-4), an independent sample Welch's t-test was conducted to compare alpha band z-scores between songs rated as happy and songs rated as not happy. The result was not statistically significant, $t(818.27) = -0.356$, $p = .7221$. This suggests that, across participants, alpha power did not reliably differentiate between emotionally positive and negative musical experiences.

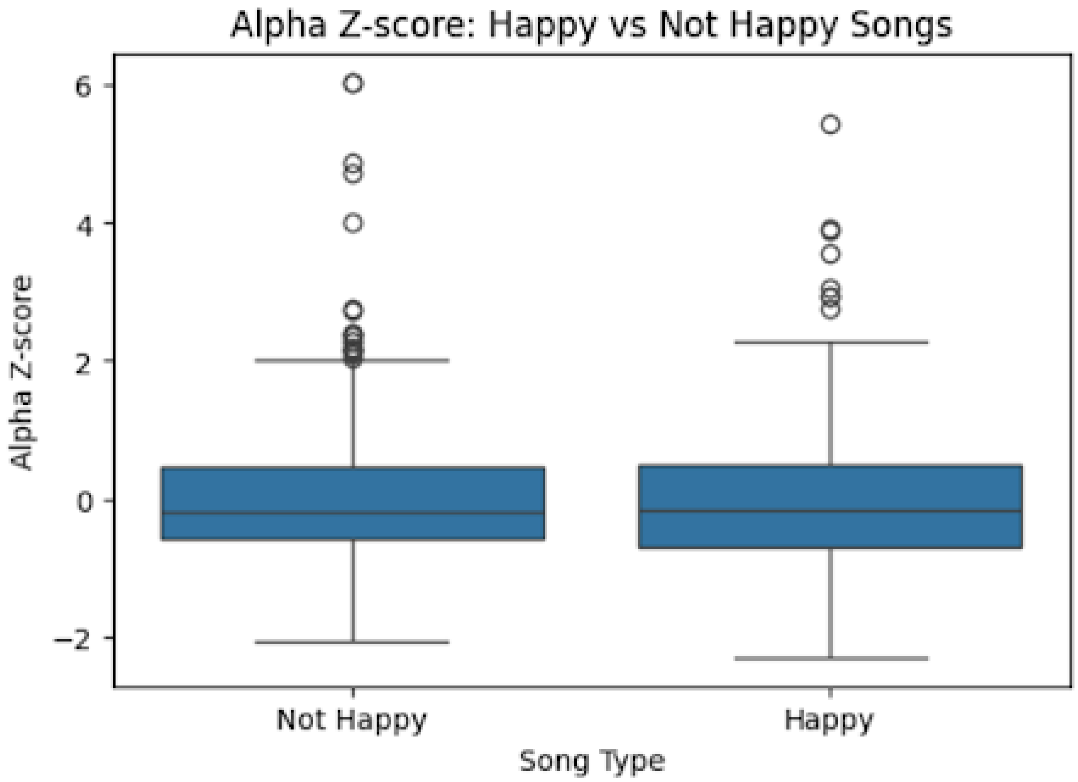

Figure 12: Box plots of the alpha z-scores of songs grouped by happy and not happy (see code chunk 28)

3.2 Alpha Comparison Across Groups (ANOVA)

A one-way analysis of variance (ANOVA) was conducted to examine the effect of subjective happiness ratings (scores 1–9) on alpha band z-scores. The analysis revealed that the effect of score on alpha power was not statistically significant, $F(8, 1107) = 1.182$, $p = .3065$. This suggests that average alpha activity did not vary significantly across different subjective happiness ratings.

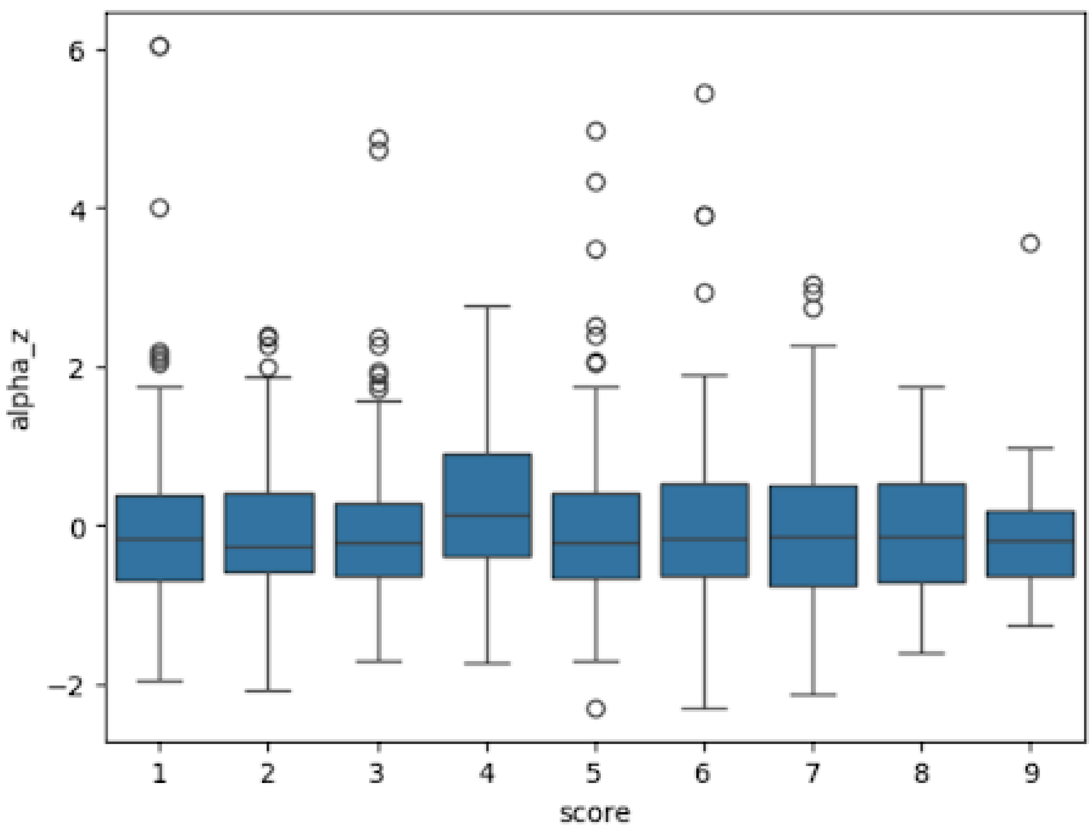

Figure 13: Box plots of the alpha z-scores of songs by rating of happiness (see code chunk 29)

3.3 Logistic Regression Predicting Song Valence from EEG Band Activity

To examine whether band power features predicted the binary classification of song valence (happy vs. not happy), a logistic regression was conducted using standardized z-scores of alpha, beta, gamma, theta, and delta bands as predictors. The overall model was not statistically significant, $\chi^2(5) = 5.01$, $p = .671$, indicating that the model did not reliably distinguish between happy and not happy songs based on EEG band activity. The pseudo R^2 was very small ($R^2 = .0024$), suggesting negligible explanatory power. None of the individual EEG frequency bands significantly predicted happiness status. For example, the alpha band was not a significant predictor, $B = -0.085$, $SE = 0.080$, $z = -1.06$, $p = .288$, 95% CI $[-0.243, 0.072]$.

```
Optimization terminated successfully.
         Current function value: 0.671196
         Iterations 4
                           Logit Regression Results
==============================================================================
Dep. Variable:               is_happy   No. Observations:                  966
Model:                          Logit   Df Residuals:                      960
Method:                           MLE   Df Model:                            5
Date:                Sun, 11 May 2025   Pseudo R-squ.:                0.002448
Time:                        16:46:40   Log-Likelihood:                -648.38
converged:                       True   LL-Null:                       -649.97
Covariance Type:            nonrobust   LLR p-value:                    0.6719
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.4106      0.066     -6.233      0.000      -0.540      -0.281
alpha_z       -0.0854      0.080     -1.062      0.288      -0.243       0.072
beta_z         0.1262      0.101      1.245      0.213      -0.072       0.325
gamma_z       -0.1249      0.096     -1.303      0.192      -0.313       0.063
theta_z        0.0366      0.076      0.484      0.628      -0.112       0.185
delta_z        0.0522      0.071      0.733      0.464      -0.087       0.192
==============================================================================
```

Figure 14: Logistic regression results (see code chunk 38)

3.4 Multinomial Regression Predicting Song Valence from EEG Band Activity

A multinomial logistic regression was conducted to predict song rating categories (Low, Medium, High) based on z-scored EEG band power features (alpha, beta, gamma, theta, delta). The model was statistically non-significant overall, $\chi^2(10) = 12.89$, $p = .240$, indicating that the predictors did not reliably distinguish among rating groups.

However, alpha band power was a significant predictor for both low-rated and medium-rated songs compared to high-rated songs. Specifically, for low-rated songs, α_z was negatively associated with the likelihood of a low rating, $B = -0.2005$, $SE = 0.086$, $z = -2.33$, $p = .020$. Similarly, α_z was also a significant negative predictor of medium ratings compared to high, $B = -0.2034$, $SE = 0.090$, $z = -2.26$, $p = .024$. No other frequency band reached statistical significance at the $p < .05$ level.

```
Optimization terminated successfully.
         Current function value: 1.090445
         Iterations 4
                          MNLogit Regression Results
==============================================================================
Dep. Variable:           rating_label   No. Observations:                 1107
Model:                        MNLogit   Df Residuals:                     1095
Method:                           MLE   Df Model:                           10
Date:                Sun, 11 May 2025   Pseudo R-squ.:                0.005240
Time:                        16:49:04   Log-Likelihood:                -1207.1
converged:                       True   LL-Null:                       -1213.5
Covariance Type:            nonrobust   LLR p-value:                    0.2400
=======================================================================================
   rating_label=Low       coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------------
const                   0.0008      0.072      0.011      0.991      -0.141       0.143
alpha_z                -0.2005      0.086     -2.327      0.020      -0.369      -0.032
beta_z                  0.2046      0.111      1.845      0.065      -0.013       0.422
gamma_z                -0.1378      0.103     -1.334      0.182      -0.340       0.065
theta_z                 0.1232      0.084      1.460      0.144      -0.042       0.289
delta_z                 0.0862      0.079      1.093      0.275      -0.068       0.241
---------------------------------------------------------------------------------------
rating_label=Medium       coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------------
const                  -0.1463      0.075     -1.945      0.052      -0.294       0.001
alpha_z                -0.2034      0.090     -2.260      0.024      -0.380      -0.027
beta_z                  0.0986      0.115      0.855      0.393      -0.128       0.325
gamma_z                -0.0354      0.106     -0.333      0.739      -0.244       0.173
theta_z                 0.1487      0.087      1.700      0.089      -0.023       0.320
delta_z                 0.0592      0.083      0.717      0.474      -0.103       0.221
=======================================================================================
```

Figure 15: Multinomial regression results (see code chunk 39)

Model Equations for Multinomial Regression

$$\log \frac{\mathbb{P}(\text{Low})}{\mathbb{P}(\text{High})} = 0.0008 - 0.2005 * \alpha_z + 0.2046 * \beta_z - 0.1378 * \gamma_z + 0.1232 * \theta_z + 0.0862 * \delta_z$$

$$\log \frac{\mathbb{P}(\text{Medium})}{\mathbb{P}(\text{High})} = -0.1463 - 0.2034 * \alpha_z + 0.0986 * \beta_z - 0.0354 * \gamma_z + 0.1487 * \theta_z + 0.0592 * \delta_z$$

3.5 Alpha Comparison of Listening to Strongly Happy vs Strongly Unhappy Songs

To examine whether alpha power differed between extremely happy and unhappy songs, we conducted independent-samples of Welch's t-tests comparing alpha band z-scores between songs rated 9 (very happy) and songs rated 1 (very unhappy). Prior to analysis, outliers were removed based on visual inspection, and only the most extreme ratings (scores = 9 and 1) were retained to maximize contrast between groups.

No statistically significant difference was found in alpha band activity between happy and unhappy songs, $t(45.19) = -0.912$, $p = .3665$. Similar comparisons for other frequency bands (beta, gamma, theta, delta) also yielded non-significant results: beta: $t = 0.50$, $p = .6214$; $gamma : t = -1.47$, p = .1494; theta: $t = -0.81$, $p = .4254$; $delta : t = 1.17$, $p = .2519$.

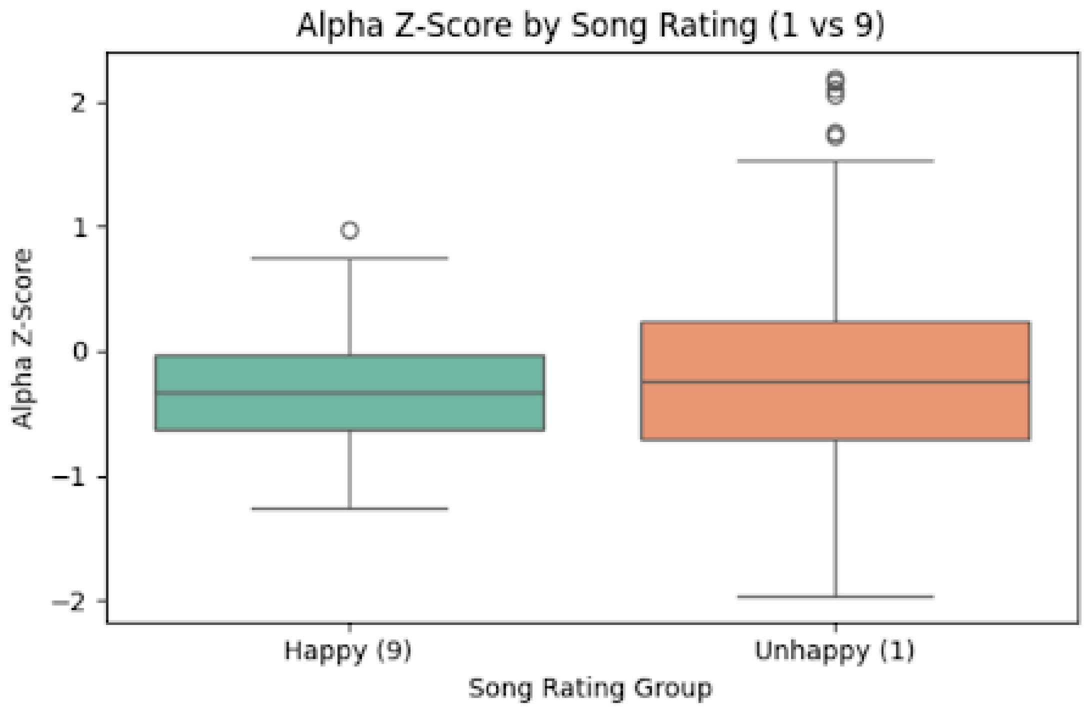

Figure 16: Box plots of the alpha z-scores of songs grouped by strongly happy and strongly unhappy (see code chunk 34)

```
alpha_z: T = -0.91, p = 0.3665
beta_z: T = 0.50, p = 0.6214
gamma_z: T = -1.47, p = 0.1494
theta_z: T = -0.81, p = 0.4254
delta_z: T = 1.17, p = 0.2519
```

Figure 17: Results of strongly happy and strongly unhappy comparison for all five frequency bands (see code chunk 36)

3.6 Alpha Power Differences by Rating Quartiles

To enhance the sensitivity of our analysis to neural differences associated with emotional appraisal, songs were divided into two groups based on participant ratings: the top 25% (high-rated) and the bottom 25% (low-rated). This quartile-based grouping allowed us to maximize contrast between clearly preferred and clearly disliked songs, while also avoiding the interpretive ambiguity of mid-range ratings and maintaining balanced group sizes. A Welch's t-test was conducted to compare alpha band power (z-scored) between the two groups. The analysis revealed no statistically significant difference in alpha power between high-rated and low-rated songs, $t(700.29) = .360$, $p = .7186$.

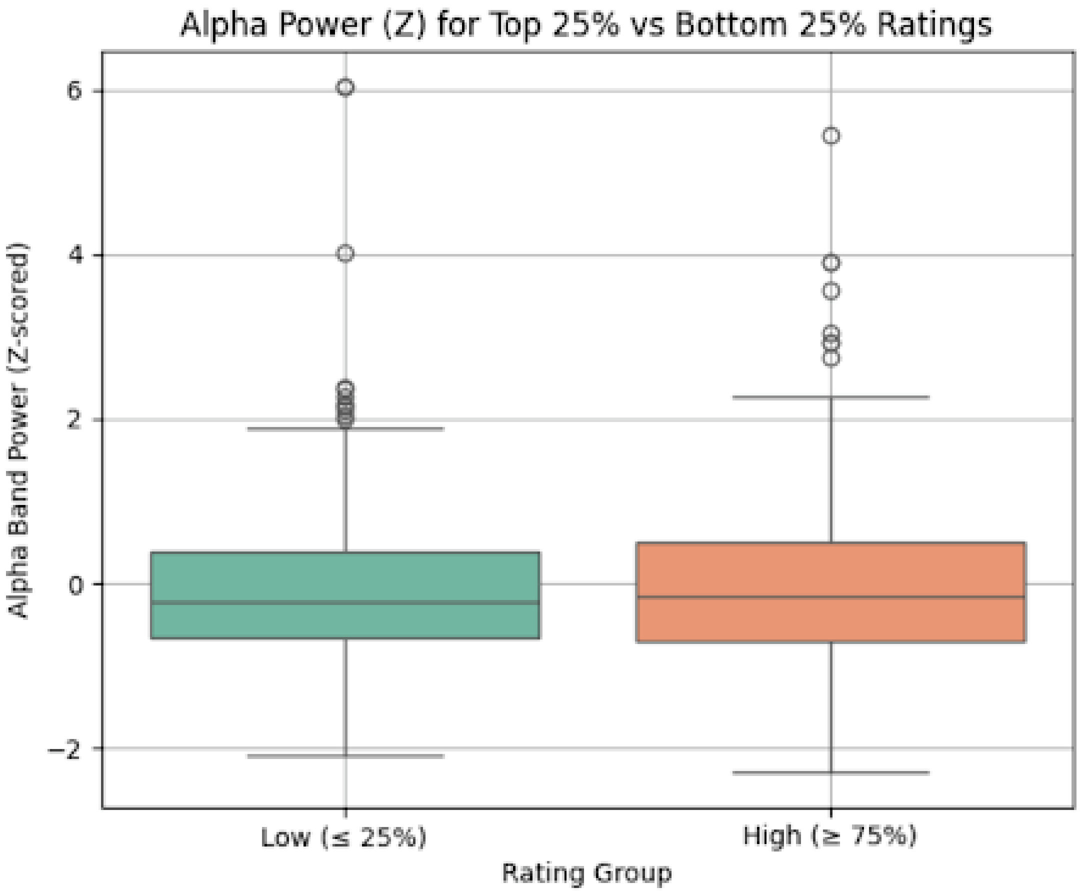

Figure 18: Box plots of the alpha z-scores of songs for the top 25% and bottom 25% ratings (see code chunk 42)

3.7 Bootstrap Analysis of Differences Between High and Low-Rated Songs

To evaluate whether the difference in brain activity between high and low-rated songs was statistically significant, a bootstrap analysis (1,000 iterations) was conducted on the mean absolute difference of normalized power spectral density (PSD). The observed average absolute difference was 238.87, resulting in a p-value less than .0004. This suggests that the magnitude of PSD differences between song ratings is unlikely to have occurred by chance, even though the direction of differences varied across individuals. The results indicate consistent within-subject neural differentiation between happy and unhappy songs, despite the absence of between-subject effects.

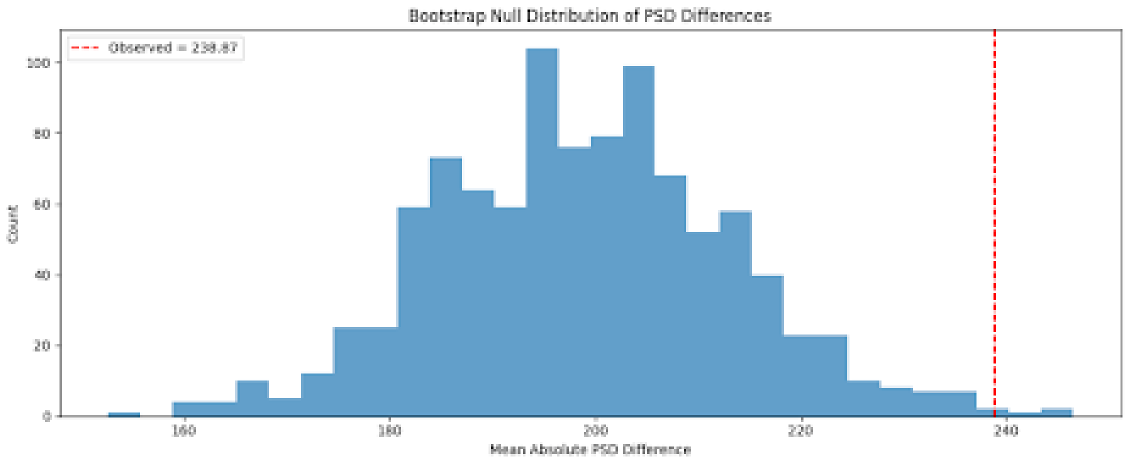

Figure 19: Distribution of Mean Absolute PSD Differences from Bootstrapping (see code chunk 52)

4 Discussion

As mentioned earlier, we first hypothesized that when participants listened to happy songs, the brain rhythms would fall in the alpha frequency range of 8-12 Hz. This was based on past literature involving studies conducted on emotional analysis from music listening through brain rhythms. Secondly, we thought that the different band frequencies would be suitable predictors of emotion based on the participants' ratings. After many attempts, we decided to hone in on Welch's independent t-test, multinomial regression, ANOVA, logistic regression, and k-means clustering. Our results do not seem to support our first hypothesis, as the t-test and ANOVA both showed that the alpha band was not a dependable source for differentiating positive and negative happiness ratings. Similar results were also seen with the logistic regression using the standardised z-scores. However, while the general multinomial regression model was non-significant, the alpha band individually was shown as a significant predictor of higher rated happy songs. Showing that an increased alpha frequency is linked to a more favorable happiness rating. This can coincide with our second hypothesis, as some prediction accuracy was observed. Finally, the bootstrap conducted had significant results revealing that high and low rating songs were not due to chance within participants. The bootstrap allowed us to test the absolute difference in power spectral density across all participants and see if the ratings assigned were at random. Noting this, we know that the difference spectra at absolute values were different from randomness, but the signed difference was not significant as the t-test found. We can denote this to the participants appearing to have vastly different brain rhythms, causing the analyses to not align. It is also important to remember that we had to remove a participant and any remaining song ratings missing across all participants. This could have played a role in the results of our statistical analyses. Overall, we can see that while one hypothesis did not align, the other has more promising indicators that could be explored further.

Limitations we encountered during our data analysis included the lack of availability of the music sources used on the participants during the experiment and an absence of consistent ratings. The soundtracks played for the twelve-second intervals were not included in the description file where the questions and answers were placed. There was no event code correlated to the music stimuli used either. From this, we had to solely base our hypotheses on past literature, making it harder to determine if the likelihood of our chosen dataset would correlate with past studies. This also made it difficult to determine the participants' emotions based on ratings, as we could not confirm the song they listened to. Furthermore, not all participants had forty song ratings that could be extracted for data analysis, resulting in varying amounts of data. Using this information for further directions, it would first be to either create or use data that includes the music sources played, so they can be referred back to. Also, utilizing the music to then compare different emotions based on band frequencies or the full spectrum frequency. There are a lot of research questions that can be explored through emotional analysis, and it can be a beneficial research area that can help to pinpoint important brain rhythms.

5 Sources

References

Daly, I., Malik, A., Hwang, F., Roesch, E., Weaver, J., Kirke, A., Williams, D., Miranda, E., & Nasuto, S. J. (2014). Neural correlates of emotional responses to music: An eeg study. *Neuroscience Letters*, *573*, 52–57. https://doi.org/https://doi.org/10.1016/j.neulet.2014.05.003

Daly, I., Nicolaou, N., Williams, D., Hwang, F., Kirke, A., Miranda, E., & Nasuto, S. J. (2020). *"an eeg dataset recorded during affective music listening"*. OpenNeuro. https://doi.org/10.18112/openneuro.ds002721.v1.0.1

Eerola, T. (2019). *Music and emotion stimulus sets consisting of film soundtracks*. https://osf.io/p6vkg/wiki/home/