University of Reading

Department of Computer Science

Classification of Musical Preference in Generation Z through Machine Learning and Signal Processing

Billy Ward

Supervisor: Dr Varun Ojha

A report submitted in partial fulfilment of the requirements of
the University of Reading for the degree of
Bachelor of Science in *Computer Science*

May 10, 2020

Declaration

I, Billy Ward, of the Department of Computer Science, University of Reading, confirm that all the sentences, figures, tables, equations, code snippets, artworks, and illustrations in this report are original and have not been taken from any other person's work, except where the works of others have been explicitly acknowledged, quoted, and referenced. I understand that if failing to do so will be considered a case of plagiarism. Plagiarism is a form of academic misconduct and will be penalised accordingly.

Billy Ward
May 10, 2020

Abstract

This project proposes a single-dimensional model for classifying emotional responses to audio stimuli, where the dimension used is a measure of like or dislike for a piece of music. An electroencephalogram was taken for 10 participants and their preference rating for each song was recorded. Multiple preprocessing and feature extraction methods are considered with a comparison made on the most appropriate method. Classification using trained machine learning algorithms is attempted, using a multi-layer perceptron and support-vector machine to classify participant rating and recognition of the songs played. Using the single-dimensional classification model produces varying performance, however high accuracies are achieved. Further exploration of the data using k-means clustering is conducted, proving less effective than the supervised classification. Potential patterns are revealed from visual inspection of spectral peaks using independent component analysis.

Keywords: Signal Processing, EEG Signals, Machine Learning, Music, Brain Activity

Report's total word count: 12557

Acknowledgements

I would like to thank my supervisor, Dr Varun Ojha, for helping me throughout the whole process of this project. I would like to thank Dr Yoshikatsu Hayashi, Alec Chetcuti, Ozge Saracbasi and the whole Biomedical Engineering department at the University of Reading for letting me use their equipment and assisting my learning of signal processing and EEG signals. I would like to thank Chandresh Pravin for helping me process my data and find useful features to work on. And finally, the 10 participants who participated in my experiment, I thank you for giving up you free time and making this whole project possible.

Contents

List of Figures

List of Tables

List of Abbreviations

SMPCS	School of Mathematical, Physical and Computational Sciences
EEG	Electroencephalography
MLP	Multi-layer Perceptron
SVM	Support-vector Machine
SNR	Signal-to-noise Ratio
ICA	Independent Component Analysis
PCA	Principal Component Analysis
IIR	Infinite Impulse Response
SWT	Stationary Wavelet Transform

Chapter 1

Introduction

The following sections contain an overview of the report to follow, stating the background of the study and the aims it has. The main areas of focus are demonstrated as research questions and an appropriate approach is outlined as to answer these questions. Finally, the organisation of the remaining report is explained.

1.1 Background

The motivations for this project were to investigate the way in which music affects and is related to human brain activity. Music has become an essential part for most of the human population and plays an integral role in pop / celebrity culture. Music, like many forms of modern media, has the power to elicit strong emotional responses in our brains, so it seems appropriate that an effort into understanding this process be made.

There have been numerous experiments in the past that have looked at the impact music can have on a person's brain activity, some used to treat mental conditions, such as a study by Ramirez et al. (2015) which aimed to help treat depression in elderly people. Other studies explore more commercial approaches and investigate the potential uses of real time brain response functionality (Nathan et al., 2017). The majority look to try and classify emotional states in the participant depending on how their brain reacts to both audio and visual stimuli.

It has been found in past experiments that emotional states can be classified using the electroencephalography (EEG) signals of a given participant. An EEG signal is a recording of brain activity acquired by attaching sensors to a participant's scalp to pick up the electrical signals produced from different parts of the brain. The participant is usually shown or exposed to stimulus to which their neurological responses are recorded. Depending on the emotional classification model being used, varying levels of activity in parts of the brain can directly correlate to a person feeling a certain emotion.

Emotional states are complex and often prove hard to determine, so in most cases a bipolar model is used to do so. The most used dimensions to quantify an emotional state are arousal and valence, the latter of which is defined as a measure of please or displeasure. For this study, a single-dimensional model is implemented to try and identify a participant's response; this being a participant's preference to a given stimuli.

This research presents an experiment to investigate the responses of 10 participating humans to audio stimuli. An original data set was obtained where each participant was given 12 audio snippets from a playlist of popular songs to listen to and rate based on their like or dislike of each song. Recordings were taken for each response to the audio snippets using an EEG device and then classified by song rating using machine learning algorithms.

In many studies conducted, the most appropriate method of classifying emotional states from EEG signals is by using machine learning. The raw EEG signals typically go through two stages; preprocessing to clean the target signals and feature extraction to emphasize these useful signals. These are then used as the input to a machine learning algorithm. The target classes are the emotional states defined by the model used.

1.2 Aims and Objectives

There are two main aims for this study, the first and foremost is to explore the possibility of classification using EEG signals with a single-dimensional model. This is to be explored by applying both supervised and unsupervised machine learning algorithms to the EEG data recorded from the 10 participants. The secondary aim of this study is to investigate which areas of the brain are used when a person likes / dislikes a song. To attempt to meet these aims, the following research questions were created for the study.

1. Can a person's preference to a song be classified using supervised machine learning algorithms solely from their brain activity response to it?
2. Can a person's brain activity be accurately clustered using unsupervised machine learning algorithms?
3. Which parts of the brain are activated when a person listens to a song they like or dislike?

1.3 Solution Approach

To interpret the EEG recordings obtained, a mixture of machine learning algorithms and techniques were applied to the data. Both supervised and unsupervised methods have been used to explore the categorical capacity of the data and underlying patterns in human brain activity. First however, the appropriate procedures to clean and extract important features in the data must be followed.

The most general approach to signal processing consists of the following stages:

1. Data Acquisition
2. Preprocessing
3. Feature Extraction
4. Classification

Different methods of preprocessing the data were investigated and evaluated to try and determine the best method for the data. A similar comparative approach was taken with feature extraction, whereby the classification performance was used to evaluate the most effective method. The data obtained was classified using supervised machine learning methods. In addition to this, the clustering performance and spectral peaks of both positive and negative responses were investigated.

1.4 Organization of the Report

The report has been organised into 7 chapters including this introduction. Section 2 contains a literature review looking at previous studies in this area and considers different methods of approaching emotional classification using EEG signals. Chapter 3 finalises the methods to be used for the study and explains how different algorithms work and will be implemented. It depicts the procedures of preprocessing and feature extraction on raw data to obtain useful features that can be used in machine learning classification. As well as classification, the potential use of spectral peaks and ICA components for pattern recognition are explored. Chapter 4 discusses the results obtained from the filtering, feature extraction and classification using cross-validation performance evaluation. Possible limitations in the study are discussed and how different approaches could produce better results. Chapter 5 is a discussion on the results obtained and what the findings imply. Finally, chapters 6 and 5 contain a conclusion and reflection on the study.

Chapter 2

Literature Review

This chapter contains all research conducted for the study. The sections consist of a literature report of previous studies in relevant areas for this paper. Emotion classification systems are considered, as well as the relevance of signal frequency bands and electrode placement. Following this, the stages of signal processing such as preprocessing and feature extraction are explored, finishing with the many applications of EEG.

2.1 Emotion Classification

There are many ways of classifying emotions that have been explored in previous literature. One work by Ekman (1999) explores the concept of basic emotions. In it he explores the relationship between facial expressions and emotions, claiming that there are 6 emotions associated with facial expressions. These are anger, disgust, fear, happiness, sadness and surprise. He later added amusement, contempt, contentment, embarrassment, excitement, guilt, pride in achievement, relief, satisfaction, sensory pleasure, and shame. Plutchik (2003) built on this and suggested 8 basic emotions, consisting of the same original 6 expressed by Ekman (1999), but adding anticipation and acceptance. Both examples show it is no simple task to classify human emotion, with there being many complex combinations of feelings a person may feel at any given time.

The general approach of trying to represent these emotions tends to be using a dimensional model (Liu et al., 2010, Lin et al., 2010). For a dimensional approach, the two fundamental dimensions needed to measure a participant's emotional state are valence and arousal, valence being a measure of pleasure or displeasure, and arousal representing how stimulated the participant feels. As an example, anger could be categorised as a measure of negative valence and high arousal. By using this two-dimensional model of valence and arousal most basic emotions stated by Ekman (1999) can be accurately mapped from continuous signals to discrete emotions. Liu et al. (2010) claim the most used model for dimensional emotion classification is a bipolar one, however there are cases of more dimensions being added for more accurate classification. Liu and Sourina (2013) agree that the bipolar model is most widely used, however a three-dimensional model is also possible. In this model, a third dimension of "dominance" is added and quantifies how in control the participant is of their current emotional state. The addition of this extra dimension is justified given that emotions such as anger, a state of high arousal and negative valence, could also be interpreted as fear, another emotion of similar dimensional values.

The addition of more dimensions allows researchers to distinguish between difficult emotional states and identify more advanced ones, however it can make classification less accurate as there are more target classes. The study using the three-dimensional model combines all possible combinations of high and low values for each dimension, to classify a total of 8 emotions. In the context of classifying emotions using machine learning, finding the middle ground between catering for all emotional states and achieving high classification accuracy is crucial.

A study by Rozgić et al. (2013) goes further by using "liking" as an additional fourth dimension to the model. However, the classification performed is a binary classification of each individual dimension using the DEAP dataset. The DEAP dataset is a publicly available dataset commonly used to classify emotional states, containing both EEG and peripheral physiological data from 32 participants. It has a total of five dimensions for user ratings to a mixture of audio and visual stimuli, meant to elicit a full range of emotional states.

There are a few studies that have attempted classification using a single dimension, such as the study by Rozgić et al. (2013), which makes use of the fourth dimension "liking". Another study by Li and Lu (2009) classifies a participant's happiness and sadness using only the gamma frequency band. The two emotional states were classified using a support-vector machine and achieved extremely high accuracies. A study by Hadjidimitriou and Hadjileontiadis (2012) investigated classifying musical preference using EEG signals with similar aims to this study, using only two target classes of "like" and "dislike".

Most studies that attempt to perform classification of emotional states use machine learning algorithms with features extracted from EEG signals (Lin et al., 2010, Basterrech and Krömer, 2019, Li et al., 2018). Commonly used algorithms are the multilayer perceptron (MLP), support-vector machine (SVM) and k-nearest neighbour, all explored and compared in a study by Bazgir et al. (2018). Chinmayi et al. (2017) applies k-means clustering using features extracted from the raw EEG data, the aim being to locate centroid positions for each of the 5 emotional responses categorised using topographic plots of the brain. The centroids with the most samples present in their clusters are selected as the dominant centroids. A study by Asif et al. (2019) explores the classification of stress levels in participants with 3 target classes using machine learning algorithms. In this study, one measure that is used to evaluate the efficiency of classification is Cohen's kappa, which represents the agreement of a classification and considers the possibility that correct classifications are made unintentionally. This makes it a good statistic for verifying the reliability of accuracies obtained.

It appears that most emotional classification is done using the bipolar model, however for the intended work depicted in this report a different approach is going to be explored. Whereas most previous studies in emotional classification models use 2 or more dimensions to quantify any given state, the feasibility of reducing this to just one dimension is explored. This single dimension will be preference and is the participants reaction to a given stimulus based on like or dislike.

This approach will bring with it some difficulties distinguishing between emotional states. As shown, emotional classification is not simple, and most feelings are very complex. In the context of music for example, a person may choose to listen to a song that makes them feel sad, but this conversely makes them feel better. This concept is explored further in Chapter 3. However, if it is possible to quantify a participant's preference to a song based solely from brain activity, this could be applied in a wide range of fields. In the music streaming market for example, this would allow for better feedback systems for companies such as music streaming service "Spotify" if widely available commercial EEG devices were to be developed. A study by Nathan et al. (2017) presents the possibility of an emotion-based music player that explores this possible application as well.

2.2 Frequency Bands

An EEG signal consists of different frequency bands that associate signals of a given frequency range to a specific brain state or process. Typically, the signals can be split into 5 frequency bands: delta, theta, alpha, beta and gamma. Table 2.1 shows the frequency bands and their respective ranges in Hertz and associated brain states, however, this is a very simplistic interpretation. Both the frequency ranges and brain states differ slightly depending on the source used. In many of the recordings taken for this study, there are very few frequencies below 8 Hz. This is because frequencies of 7 Hz or below normally only occur in children who are relaxed or sleeping and rarely in awake adults (Chinmayi et al., 2017).

Table 2.1: Frequency Bands

Band	Frequency Range (Hz)	Associated Brain State
Raw EEG	0 - 50	Conscious
Delta	0.5 - 3.5	Deep sleep
Theta	4 - 7.5	Creativity, imaginary
Alpha	8 - 12	Relaxed
Beta	13 - 35	Thinking, alert
Gamma	$>$36	Short term memory

The spectral powers of frequency bands are often used as classification features. In the study performed by Lin et al. (2010), power spectra were examined in these distinctive bands to investigate how they correlate with a person's emotional state. One commonly used indicator of a given emotional state is alpha-power asymmetry which was also used in this study. Comparing symmetrical channel spectral powers is also used as a feature for classification in the study by Rozgić et al. (2013).

A study by Ramirez et al. (2015) made use of the alpha and beta bandpowers to relate to a participant's arousal and valency. The study mapped an increase of participant arousal as an increase in the ratio of activity in the frontal cortex between the beta and alpha bands, and an increase in valence as relative alpha activity in the right lobe compared to the left lobe. It appears that the comparative spectral powers offer the best insight into how a person is responding to a given stimulus.

It is noted that only the alpha and beta bands are used to interpret relative power between bands, however there are some studies choose to specifically target the only the gamma frequency band (Li and Lu, 2009). Although using only these bands have proved to be good features in the literature referenced, there may be interesting data that is being missed contained in the other frequency bands.

2.3 Electrode Placement and Scalp Locations

The 10-20 system is an internationally recognised method to describe the location of scalp electrodes for EEG recording and is used in various studies in this area (Lin et al., 2010, Schultz et al., 2008, Al-Fahoum and Al-Fraihat, 2014). In it, each electrode location is given a label to represent which area of the brain it is recording signals from, these consist of pre-frontal (Fp), frontal (F), temporal (T), parietal (P), occipital (O) and central (C). The "central lobe" does not exist as a lobe in the brain; its location is dependent on the individual. These labels can be combined to represent a point in-between two areas of the brain, such as "PO" being in-between the parietal and occipital lobe. The odd numbered positions account for electrodes on the left side of the brain, whilst even numbers are for the right. There are some nodes that are used as a reference point to the others, which have a "z" placed at the end of their label. The different areas in the brain relate to specific functionality, for example the frontal lobe is commonly associated with emotional processing and the temporal lobe contains the primary auditory cortex so is associated with audio stimuli.

The number of electrodes used can range widely depending on the experiment setup. As seen in the study by Bazgir et al. (2018), a total of 62 channels were used for some tests, in contrast to the study by Basterrech and Krömer (2019) who only used 1 channel. Both studies emphasize the importance of appropriate preprocessing and feature selection of the data, especially when attempting classification using machine learning. When used for clinical or research applications, recording EEG signals can be done using high-density arrays containing up to 256 electrodes.

2.4 Preprocessing

Chinmayi et al. (2017) removed unwanted noise from the signals obtained in the frequency range of 0.2-40 Hz using an 1800 order finite impulse response filter. Unwanted noise could be the participant blinking, external electrical current and muscle movements.

A study by Schultz et al. (2008) uses a combination of filtering methods on the data, first applying a bandpass filter to the data and then Dual-tree complex wavelet transform, a method with optimal performance for the aim of their study. Applying a wavelet transform can also be an effective form of feature extraction (Al-Fahoum and Al-Fraihat, 2014, Ishino and Hagiwara, 2003). The optimal parameter configuration for the bandpass filter is found by calculating the classification accuracy of a machine learning algorithm for every combination of frequency bands. The best classification performance was found when using the alpha, beta and gamma bands, so a bandpass filter of 8-45 Hz is applied.

Since there are many ways of filtering or denoising a signal, a way to evaluate the efficiency of the filtering performed is desirable. A study by Liu and Sourina (2013) takes a base reading without stimuli before performing the rest of the experiment for the participant. Using the base reading obtained, it could be assumed that the signals in it are just pure noise, so when subtracted from a recording with stimuli all that is left is useful response data from the participant. This assumption also allows the calculation of the signal-to-noise ratio (SNR) to evaluate how a filter may improve or worsen signal quality.

The number of filters applied to raw EEG data seems limited from the literature reviewed in this section. As well as the finite impulse response filter used in the study by Chinmayi et al. (2017), there are infinite impulse response (IIR) filters, such as Butterworth and Chebyshev, which could be implemented to try and improve signal quality further and remove unwanted noise.

2.5 Feature Extraction

When recording EEG signals, the signals can become "smeared" with each other, causing meaningful features to become obscured by other less important data. One method to avoid this is Independent Component Analysis (ICA). ICA can be used to try and decompose a multivariate signal into its sub-components or hidden factors, creating less correlated channels of data to be used for further analysis.

In the study by Chinmayi et al. (2017), the data set is decomposed using ICA and the components are then transformed into topographic plots which give both the spatial location and spectral powers of EEG signals. The point of maximum spectral energy is taken as two features, an x and y location value, and the cumulative energy is taken as another, thus providing 3 features for a single ICA component. Out of 14 computed ICA components, only the first 4 are considered as these typically contain the most variance in the data. These features are then used to cluster the data by emotional response, finding the central response location in the brain for each emotion, and to classify the emotional states using the k-nearest neighbour algorithm.

A popular tool used for EEG analysis is EEGLAB, an open source toolbox made for use in MATLAB that specialises in processing EEG signals. A paper by Onton et al. (2006) shows it can be used to produce ICA topographic plots to visualise brain activity, stating it can identify brain activity patterns that directly correlate to a person's behaviour or activity. Another study by Delorme and Makeig (2004) demonstrates the use of the EEGLAB tool for ICA decomposition as well as it is other functions, which include data preprocessing and generating event-related potential image plots.

The use of Short-time Fourier transform is commonly used for feature extraction (Schultz et al., 2008, Lin et al., 2009). Short-time Fourier transform is a form of Fourier transform that evaluates the change of frequency content for a signal over time. In the study by Lin et al. (2009), it is claimed that spectral powers are the most common feature used for EEG signals, where the frequency bands are those as stated in Table 2.1. Schultz et al. (2008) suggests calculating the relative power of each frequency band by taking the target frequency component and dividing it by the sum of all frequency components.

An article by Bos et al. (2006) proposes using principal component analysis (PCA) to reduce the number of features used in classification. In the study the total number of features, 1000, was reduced to 1 - 25. It seems that a common problem when classifying using EEG data is the abundance of redundant information, so as well as feature extraction, a form of data reduction is also desirable.

2.6 Applications of EEG

There are a wide range of fields that can incorporate emotional classification through EEG signals. Typically EEG recordings are performed in medical scenarios for diagnosing conditions and performing music therapy Sourina et al. (2009), Ramirez et al. (2018), Sourina et al. (2012). A study by Ramirez et al. (2015) uses music to try and treat depression in elderly people. It involves the use of real time neurofeedback, in which the volume and tempo of the music being played is increased relative to the participant's EEG signal responses. Another example of real time neurofeedback was used in a study by Liu et al. (2010) which visualised the emotion the participant was feeling whilst listening to music using a 3D avatar.

With recent advances in wireless recording devices, the potential for commercial applications is becoming more realistic. This highlights the importance of being able to interpret brain waves and identify what a person is feeling based on their brain activity. A study investigating musical stimuli specifically and its capabilities within a mobile application called "Emosic", explored the possibility of labelling songs with evaluated valence and arousal values that correspond to an emotion (Nathan et al., 2017). This was done using emotion recognition from an input image of the user, but the same functionality could be produced from EEG recordings as well. As explained in the paper, this would allow for custom made playlists for users based on the emotion they were feeling.

Chapter 3

Methodology

This chapter depicts the methods chosen to conduct the study and what considerations were made when choosing.

3.1 Emotion Classification Model

In most previous studies, a 2 or more-dimensional model is used to classify emotional states such as happiness, sadness or anger for example. However, these emotions are not always so simply defined. Audio and visual stimuli have the potential to make us feel these emotions, and one would assume that feeling a negative emotion would be a negative experience for a person. But this is not always the case.

In some cases, a person may want to feel a negative emotion, such as sadness, as this can be therapeutic and, in some cases, even be a pleasurable experience. A study by Hanich et al. (2014) investigates this strange concept, concluding that there is a highly positive correlation between sadness and enjoyment which comes from a feeling of "being moved" by a stimulus. Reducing the dimensions used to classify an emotional state allows this to be explored. Instead of explicitly defining the emotion felt, the participant's preference rating disregards the underlying emotion as positive or negative, and simply measures the pleasure or displeasure that person is feeling. A person may enjoy the way a song makes them feel, even if that feeling is a negative one. The single-dimensional model used for this study quantifies a participant's liking or disliking of a given song.

3.2 Experiment Setup

To conduct this experiment, a new data set was obtained from 10 participants, 5 men and 5 women, from Generation Z, which is defined by the Merriam-Webster online dictionary as people between the ages of 18 and 24. This seemed an appropriate number of participants as previous literature shows this is enough to obtain well classified data (Liu et al., 2010). Participants were asked to confirm they did not have a history of mental illness, as this has been shown to produce unreliable EEG recordings.

The experiment setup was influenced by previous experiments in this area. For the stimuli, 12 song clips of 30 seconds each were chosen, all of which had received awards in recognition of their quality. To attempt to diversify the music, 3 different awards were chosen: The BRIT Awards for 2019, the Mercury Prize ranging from years between 2010 to 2019 and Rolling Stone's Most Influential Albums of all time. In the case for the Mercury and Rolling Stone awards, a whole album is given the award, so the most played song from that album was chosen according to Spotify statistics. The music chosen is shown in Table 3.1 and the award categories for each song as in Table 3.2.

Table 3.1: Music for Experiment

Index	Song Name	Artist	Award Code
S1	One Kiss	Calvin Harris and Dua Lipa	BRIT
S2	Don't Delete the Kisses	Wolf Alice	MERC
S3	Money	Pink Floyd	ROLS
S4	Shotgun	George Ezra	BRIT
S5	Location	Dave	MERC
S6	Smells Like Teen Spirit	Nirvana	ROLS
S7	God's Plan	Drake	BRIT
S8	Breezeblocks	alt-J	MERC
S9	Lucy In The Sky With Diamonds	The Beatles	ROLS
S10	Thank U, Next	Ariana Grande	BRIT
S11	Shutdown	Skepta	MERC
S12	Billie Jean	Michael Jackson	ROLS

Table 3.2: Award Categories

Award Code	Award Category
BRIT	BRIT Awards (2019)
MERC	Mercury Prize (2010 - 2019)
ROLS	Most Influential Albums (all time) Rolling Stone, The Guardian

Each participant signed a consent form and had the experiment process explained to them. Before playing the selected music to the participant, an EEG recording for when no music is being played was taken whilst the participant closed their eyes. This baseline recording was taken so that a SNR value could be calculated for the preprocessing of the data. Participants were asked to close their eyes during every EEG recording because a lot of noise is produced in EEG signals from eye movement.

The participants were given an answer sheet with a single rating table for each song, ranging from -2, -1, 0, +1 and +2, shown in Table 3.3. They were also asked to indicate whether they recognised the song that had been played to them.

Table 3.3: Rating Table

-2 Hate it	-1 Dislike it	0 Neutral	+1 Like it	+2 Love it

There were two preliminary questions on the answer sheet asking if the participant played a musical instrument and for them to list their favourite / most listened to genres of music. To try and reduce noise in the data, over ear headphones were used to play the music to the participants and they were made to adjust the volume of the music until they could only hear the music and not any background noise.

The recording for the EEG signals was started 5 seconds after the song clip started playing and 5 seconds before it ended, as to ensure that all signals recorded are when the participant is exposed to the audio stimuli. When the clip had finished playing, they could open their eyes and fill in the answer sheet provided. After the participant had completed their answers for that song, the next one was played, and this process was repeated until all 12 songs had been recorded and rated. 12 EEG recordings of 20 seconds long each were obtained for all 10 participants.

For the experiment award winning music was chosen over a range of genres and periods of time in order to investigate how participants would react to different types of music. The genres for the music can be classified into 4 categories using the general opinions from the website "Discogs". The songs and their respective genres are shown in Table 3.4. This information was used later in the study for classification and further analysis.

Table 3.4: Genre Categories

Pop (1)	Rap (2)	Rock (3)	Indie (4)
Shotgun	God's Plan	Money	Breezeblocks
Thank U, Next	Location	Smells Like Teen Spirit	Don't Delete The Kisses
One Kiss	Shutdown	Lucy In The Sky With Diamonds	
Billie Jean			

The Unicorn Hybrid Black device was used to record the participants brain activity. During recording, a bandpass filter ranging from 2-30Hz was applied and a notch filter at 50hz. The notch filter helps to remove any environmental noise, such as electrical current in wiring near to the EEG device or eye movement. The device has the electrodes positioned according to the 10-20 system at Fz, C3, Cz, C4, Pz, PO7, Oz and PO8. These 8 channels were recorded and analysed for all participants.

3.3 Data Preprocessing

Preprocessing the data obtained is an important stage of signal processing, especially with EEG signals as there is a lot of information that may not be beneficial to the aim of the study. The data recorded for this study has already been through one level of preprocessing during recording, where a bandpass filter of 2-30Hz and notch filter of 50Hz were applied. This section discusses the potential of further preprocessing of the data obtained, with the performance of processing evaluated by two different methods. These are calculating the SNR values and classification accuracies before and after filtering. All tests were conducted on data obtained from Participant 2.

3 different methods were chosen to try and filter the data, consisting of 2 analog IIR lowpass filters, Chebyshev Type I and Butterworth, and the Stationary wavelet transform (SWT). The Butterworth filter was of order 6 with a cut off frequency of 35Hz. The Chebyshev 1 filter was also of order 6 and had a peak-to-peak passband rippled 10 decibels and a passband edge frequency of 0.6. The SWT applied was to the "db1" wavelet at 3 levels of decomposition.

To assess which filtering method was best suited to the data set obtained, both evaluation methods were tested and compared to determine whether further preprocessing of the data was appropriate. All the below tests were conducted in MATLAB and made use of the Signal Processing Toolbox extension.

3.3.1 Signal-to-noise Ratio

The base recording taken for each participant can be used to calculate the SNR, which is a measure used to compare the level of a target signal against any background noise in the recording. The SNR ratio is calculated by dividing the power of the useful signal by the power of background noise, as shown in Equation 3.1 where 'P' denotes power. This measure is typically used when a sample of a pure signal and the pure noise added to that signal is present, however for this study it will be assumed that any signals taken in base recordings are not significant and only consist of noise.

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{3.1}$$

A MATLAB script was created to calculate the SNR values before and after filtering. Both IIR lowpass filters were applied using MATLAB built-in functions and applying a SWT was done using the MATLAB "Stationary Wavelet Transform Denoising 1-D" tool. The script was used to identify which filter increases the SNR the most, and therefore removes the most background noise.

3.3.2 Classification Performance

Another way of quantifying how effective filtering has been is to compare classification accuracy values of unfiltered and filtered data. The classification used to test these preprocessing methods was by binary participant rating. All classification tests were performed in the KNIME data analytic platform.

The data obtained for this study provides 1 of 5 possible rating values for any row of data, representing the preference a participant feels towards a song ranging from -2, -1, 0, +1 and +2. The data set was configured as a binary classification problem, where only two classes exist for preference rating: like or dislike. They are made up of like (+1, +2) and dislike / neutral (0, -1, -2).

The performance of the 3 filtering methods were evaluated by their classification of binary participant ratings, where the raw EEG signals recorded were the input features. This was done using a single layered MLP consisting of 10 neurons, with a maximum number of iterations of 100. When performing classification, the accuracy values are subject to change as the samples used to train the algorithms are different each time, meaning there could be data that is never used for training or testing. To avoid this, 10-fold cross-validation was used to evaluate the performance of the classification for each data set. Cross validation is used to ensure all parts, or folds, are used in the algorithm and will provide a more accurate classification evaluation. There is still randomness involved in the selection of the folds meaning there will be differing accuracy results each time, so the algorithms were trained and tested 3 times and a mean average of the classification performance was taken.

The data sets created for this study have many instances of imbalanced class distribution. This means that classification accuracy's obtained could be bias towards the dominant class of the training set. In this case, there is a much higher percentage of positive responses as opposed to negative responses, shown in Table 3.5. To ensure that training was reliable, the data was down sampled to train and test the models. This was done using the "Equal Size Sampling" KNIME node, which randomly removes data rows of the majority class for a dataset until the class distributions are equal. Doing this allowed for equal instances of all classes for machine learning classification and much more reliable results. The 10-fold cross validation implemented used stratified sampling on the rating attribute, which means that an equal number of samples from each class are taken for each fold. This also helps to create balanced training and testing data sets.

Table 3.5: Participant Response Count

Rating	Count	Binary Count
-2	8	-
-1	10	-
0	21	39
1	37	81
2	44	-

The process of this evaluation method was a lot longer and required the filtering of entire databases before testing their performance. Consequently, less variations of parameter classifications were explored. Configurations were chosen based on tests conducted in MATLAB and by following the methodologies of previous studies.

3.4 Feature Extraction

It is common in signal processing to attempt to extract the most significant features from a signal. This is beneficial for both further analysis of the data and providing effective inputs to machine learning algorithms. Feature extraction methods can also result in less data being used for training, which can lead to less irrelevant data being used that may worsen results. The methods applied for this study are calculating the relative band powers of the signals and using PCA to reduce the number of features in the data for classification. As with the preprocessing test, a single layer MLP will be trained using data after applying each feature extraction method and the classification accuracy's will be compared. The same dataset from participant 2 was used as a binary classification problem with down-sampling applied.

The relative power of a frequency band represents how much of that specific frequency is being used in the signal and can reveal insights into how a participant is acting or feeling. For each 20 second song recording, there are 8 channels of signals to be interpreted. Each of these 8 signals will be decomposed into their relative frequency bands, consisting of delta, theta, alpha, beta and gamma, and a relative percentage will be calculated.

PCA aims to reduce the number of dimensions needed to represent given data, therefore retaining useful information but reducing the amount of data required to represent it. By applying PCA to the dataset obtained, the amount of training data should be reduced and only the most important features will remain, therefore resulting a better classification. The number of dimensions PCA reduced the dataset to was 2 and 3.

A comparison between normalised and non-normalised input values is also made during the testing of these feature extraction methods. This is a suggested action when using the MLP node in KNIME and can improve the performance of classification.

3.5 Machine Learning Enabled Data Analysis

There are two main supervised machine learning techniques used for classifying emotional states that appear to be used most frequently, these are the SVM and MLP. Unsupervised methods aim to reveal hidden patterns in the data with minimal interference from humans. For this study, K-means clustering will be applied to the data. All algorithms were executed using the KNIME platform. The datasets used for these classifications consisted of the relative bandpowers calculated during feature extraction.

When classifying the data, a range of different target labels were tested to try and find any significant patterns or results in the data. 4 different labels were tested: Song ID, Song

Genre, Participant Rating and Participant Recognition of Song. For supervised learning, a binary classification was attempted and for unsupervised learning a multi-class clustering was performed.

3.5.1 Support-vector Machine

An SVM can be used for both classification and regression problems and is favoured for its production of reliable accuracies using less computational power. By mapping data points into n-dimensional space the algorithm aims to identify the hyper-plane that differentiates two individual classes the best, ergo the hyper-plane that produces the greatest distance between the data points of each class.

The SVM was originally designed to find an optimal hyper-plane between two non-overlapping classes, although a multi-class classification can also be used (Basterrech and Krömer, 2019). When calculating classification performance for the preprocessing methods, the participant rating was performed as a binary classification problem. For the SVM this method was used again, as well as classifying the data by song recognition. The data for song recognition was obtained for each song played to a participant, where they were asked to write down whether they recognised the song being played. This feature was added to the dataset as binary value of 1 for recognising the song and 0 for not recognising it.

3.5.2 Multi-layer Perceptron

A multi-layer perceptron is a class of feed-forward artificial neural network and is based on the way a human brain works. It represents a connected network of neurons that can take input data, perform a simple operation and feed that result through to other neurons to produce an output. The connections between neurons are associated with a weight value, these are changed when the perceptron is trying to learn an output given a set of inputs. As with the implementation of the SVM, the MLP will be used to classify binary song preference and song recognition.

3.5.3 K-means Clustering

K-means clustering is an unsupervised machine learning algorithm used to group data points that have similar attributes into clusters. It does so by mapping data points in 2-dimensional space and creating a centroid (centre point of the cluster) for each cluster k to assign the points to. This is done by repeatedly measuring the distance between the data point and all centroids in the space, then assigning that point to the closest centroid / cluster. The mean values of the centroids are re-calculated as the sum of all points in that cluster divided by the number of points. Once the centroid means stop changing, the clustering is complete.

Clustering will be performed by 4 different classifications using the relative bandpowers calculated in the feature extraction stage. These consist of clustering by song ID, song genre, preference rating and participant recognition. The EEG data will be normalised to improve the quality of the clustering, and the performance of clustering will be evaluated using the "Entropy Scorer" node in KNIME.

As shown earlier in this chapter in Table 3.4, the songs can be split into 4 categories based on their respective genres. When performing K-means clustering, the song's ID was replaced with a number to represent the genre of that piece.

3.6 Spectral Analysis and ICA Components

The main aim for this method of analysis was to compare the difference in locations of spectral peaks within the brain, between positive and negative responses. The activity present in the different lobes of the brain provide information on the state of the brain or its current function. The 8 channels used for this study consist of Fz, C3, Cz, C4, Pz, PO7, Oz and PO8. The areas of brain these relate to are the frontal, central, parietal, and occipital lobes. The central lobe is dependent on the individual and placement of the electrodes, meaning it can produce signal activity closer to that of the frontal, temporal or parietal-occipital lobes.

To visualise the activity and specific features of the brain for a given song, a topological view can be obtained using ICA. A tool in MATLAB called "EEGLAB" was used to run ICA on a selection of participant's response to songs to extract sub-components depicting which areas of the brain were activated most. In the process, ICA can identify noise elements in the EEG data, such as eye and muscle movements. EEGLAB provides a tool called "ICLabel" attempting to identify these components, which can then be removed. A component is given a percentage accuracy of what the signals most likely represent, being brain, eye or muscle activity. Any component with below 90% accuracy of being predominantly brain signals was removed from analysis.

When the ICA decomposition is performed, the components are reordered by their variance, meaning the first components will contain the most interesting data. Because of this, out of the of 8 ICA components created, only the first 4 components were considered. Songs were investigated that had divided ratings of very high and very low (+2 and -2) from the participant responses. Like the study performed by Chinmayi et al. (2017), the spectral peaks are investigated using the topographic ICA plots created in EEGLAB.

Chapter 4

Results

The following chapter presents the results obtained from the signal processing and classification methods performed. Figures are included to depict the efficiency of clustering on the data, and ICA topographic plots highlighting the areas of the brain activated during song listening are shown.

4.1 Preprocessing

This section discusses the results from the SNR values calculated and the classification accuracies obtained after filtering the EEG data obtained. The data used for both preprocessing tests was the EEG recording of Participant 2 listening to the first song.

The SNR values were calculated for data without filtering and then after each of the 3 filtering methods had been applied to all 8 channels, as shown in Table 4.1. An average value for each filtering method is shown in Table 4.2.

Table 4.1: Channel SNR Values

Filtering	Fz	C3	Cz	C4	Pz	PO7	Oz	PO8
Un-filtered	-3.172	-1.894	-4.004	-1.33	14.595	-0.117	-3.111	-1.796
Butterwoth	-3.205	-1.904	-4.021	-1.362	14.583	-0.181	-3.138	-1.816
Chebyshev 1	-10.796	-10.92	-12.24	-8.717	5.625	-6.483	-10.966	-9.748
SWT	-3.591	-2.017	-4.268	-1.926	14.424	-0.961	-3.529	-2.253

Table 4.2: Average SNR Values

Filtering Method	Average SNR Values
Un-filtered	-0.104
Butterworth	-0.130
Chebyshev 1	-8.031
SWT	-0.515

As shown by the results the SNR values for both filtered and unfiltered data were consistently negative, meaning there is more noise present than signal. This could be due to the selection of noise for the experiment. One common problem found when analysing EEG data is its unclear which signals are useful and which are not. The base reading taken was when the participant was not listening to any music, and therefore it was assumed to be noise. However, the SNR calculations show that this may not be the case, and that signals present in the base reading may have been useful signals as opposed to noise. Another issue is that the filters generally decrease the SNR value, suggesting that useful signals are being lost in the process.

Table 4.3: Classification Accuracy Results

Filtering Method	Test 1	Test 2	Test 3	Average Accuracy
Un-filtered	58.107	58.569	58.146	58.274
Butterworth	58.207	58.007	57.286	57.833
Chebyshev 1	58.625	58.406	58.75	58.594
SWT	58.668	58.188	58.507	58.454

The same filtered and unfiltered datasets used for the SNR calculations are also used for determining classification accuracy. The results in Table 4.3 show there is a slight improvement of classification when the data has the Chebyshev Type 1 filter and the SWT applied to it. However, it appears that the Butterworth filter decreases the accuracy of classification. These results correlate with the SNR calculations in the previous section, showing there was little improvement from filtering the data and in some cases even makes the signal more obscured. These results could vary given a different filtering method with different parameters. It is also possible that the filtering applied during recording from the EEG device was able to remove most of the noise, so filtering the data further could mean losing important features that would assist in training classification algorithms.

As with the SNR calculations, the classification performance did not appear to increase by a significant amount after the data had been filtered. It could be that most of the noise has already been removed from the filtering applied during recording, so filtering further means a loss of signal features. An average classification rate of 58.274% was obtained using raw EEG signals without further filtering. Since a bandpass and notch filter had already been applied to the data during recording, the decision was made that further preprocessing would not be applied, as it could obscure important features that could be used in classification of the data. Feature extraction would be applied to the data before classification, which would be more suitable as there is no risk of removing significant signals.

4.2 Feature Extraction

The two methods of feature extraction chosen, calculating relative bandpowers and applying PCA, were used to create two data sets that would also be trained using a single layer MLP. As well as these two methods, a comparison of their performance when using normalised and non-normalised data is also made. The data used for these tests was from Participant 2 and all song recordings were included in the classification.

Table 4.4: Non-normalised Data

Feature Extraction	Test 1	Test 2	Test 3	Average Accuracy
Bandpowers	51.562	48.438	56.25	52.083
PCA (2-dimensions)	57.562	57.494	57.007	57.354
PCA (3-dimensions)	57.362	56.988	57.889	57.413

Table 4.5: Normalised Data

Feature Extraction (Normalised)	Test 1	Test 2	Test 3	Average Accuracy
Bandpowers	67.188	60.938	64.062	64.063
PCA (2-dimensions)	49.871	50.313	49.384	49.856
PCA (3-dimensions)	50.061	50.159	49.761	49.994

Table 4.4 show that the accuracy decreases after converting the raw signals to their relative band power representations, however this test was performed without normalisation of the data. When normalisation is applied, Table 4.5 shows there is an increase in accuracy by 11.98% against non-normalised band power data and 5.79% against the unfiltered EEG data shown in Table 4.3.

PCA was applied to try and reduce the dimensionality of the dataset, reducing the data features to 2 and 3 dimensions. On both normalised and non-normalised data, the accuracy was less than the unfiltered raw data set. More dimensions were tested; however, the classification rate did not change significantly in all cases. Unexpectedly, normalising the data before applying PCA reduced the classification accuracy further.

Having seen the increase in classification accuracy from normalising the data set, the raw EEG data was normalised and tested under the same conditions to see if this would perform better than the band powers. However, the average classification accuracy was calculated to be 50.94%, so band power representation with normalisation appears to be the most effective form of feature extraction.

The best form of feature extraction appears to be calculating the relative bandpower values for each channel and normalizing that data. This produced an average classification accuracy of 64.063 when trained for binary classification of a participant's song rating. Therefore, the rest of the participants data was transformed to their relative bandpowers and further classifying using these data sets is explored in the following section.

4.3 Machine Learning Classification

The following supervised machine learning algorithms have been applied in KNIME to try and classify the data obtained and have been organised by the data attribute that was used as the classifier. All classification experiments were validated using the 10-fold cross-validation method. The results for the accuracies acquired were run 3 times and an average accuracy was obtained.

It was attempted to find the optimal parameters for both classifiers, however since each participant dataset is unique, the optimal configuration training one model is not necessarily the best for training another. Furthermore, if a validation method such as cross-validation is used, the samples chosen to train the model each time will be different, resulting in greater variation of optimal parameters. The optimal parameters were found through experimentation for one participant and then these were maintained for the rest of the experiment. The data used was for Participant 2.

For the MLP, combinations of values for the number of hidden layers used and the neurons in each layer were explored. From the experiments conducted, it seemed that higher accuracies were consistently obtained using fewer layers and greater neurons per layer. The MLP was configured to have 1 layer with 4 neurons. The polynomial kernel was used for the SVM as this produced the most consistent results. This kernel has 3 parameters to be configured, consisting of Bias, Power and Gamma. As shown in Equation 4.1, for any given pair of vectors (x,y) in the input space, the parameters represent the power (T), the bias added (c) and the multiplier (d). From tests using different configurations, the best results were obtained when all parameters were set to 3.

$$K(x, y) = (x^T y + c)^d \tag{4.1}$$

One problem encountered was the lack of negative ratings in the data set. After performing feature extraction on the data, the number of data rows was significantly lower than with the raw data. Then during analysis, the data was down sampled to equal out the class distributions, leaving little data to be used for training the models. This could also mean some data would never be considered during training, so the 10-fold cross-validation method was used to try and assist this. The 10-fold cross-validation method splits the data into 10 folds and ensures that each fold is used as the testing data at least once, whilst the other folds are used for training.

The values of for the Cohen's kappa coefficient of the classification were also recorded and have been displayed in tables in the following section. Cohen's kappa provides a measure to determine the agreement of a classification, considering correct classification due to chance. Table 4.6 was taken from a paper by Rafieyan (2016) and shows how the measure can be interpreted.

Table 4.6: Cohen's Kappa Interpretation

Values	Interpretation
<0.00	Poor Agreement
0.00 - 0.20	Slight Agreement
0.21 - 0.40	Fair Agreement
0.41 - 0.60	Moderate Agreement
0.61 - 0.80	Substantial Agreement
0.81 - 1.00	Almost Perfect Agreement

The following tables in this section depict 3 separate classification accuracies for each of the 10 participants with the mean values of their respective accuracies in the final column. An overall accuracy is displayed in the bottom right cell of the table, showing the mean of all average accuracies achieved. The Cohen's kappa value shown is also the mean value of the three iterations of the classification performed.

4.3.1 Classification by Song Rating

The data was classified based on the rating values each participant had provided to the 12 songs listened to in the experiment. The data set was treated as a binary classification problem, where a user's preference was converted into either 0 or 1, representing a participant disliking or liking the song, respectively. The tables below shows the classification results for all 10 participants using the MLP in Table 4.7 and SVM in Table 4.8.

Table 4.7: MLP Classification for Participant Rating

Ratings	MLP 1	MLP 2	MLP 3	Cohens Kappa	Average Accuracy
P1	0.641	0.688	0.641	0.313	0.656
P2	0.563	0.578	0.641	0.188	0.594
P3	0.516	0.578	0.563	0.104	0.552
P4	0.672	0.563	0.609	0.229	0.615
P5	0.625	0.578	0.641	0.229	0.615
P6	0.703	0.609	0.563	0.250	0.625
P7	0.688	0.625	0.609	0.281	0.641
P8	0.672	0.516	0.594	0.188	0.594
P9	0.563	0.578	0.609	0.167	0.583
P10	0.578	0.578	0.672	0.219	0.609
Average	-	-	-	0.217	0.608

Table 4.8: SVM Classification for Participant Rating

Ratings	SVM 1	SVM 2	SVM 3	Cohens Kappa	Average Accuracy
P1	0.734	0.703	0.594	0.354	0.677
P2	0.703	0.641	0.672	0.344	0.672
P3	0.656	0.734	0.719	0.406	0.703
P4	0.609	0.766	0.641	0.344	0.672
P5	0.781	0.766	0.750	0.531	0.766
P6	0.734	0.703	0.641	0.385	0.693
P7	0.625	0.625	0.672	0.281	0.641
P8	0.719	0.656	0.766	0.427	0.714
P9	0.656	0.625	0.625	0.271	0.635
P10	0.609	0.672	0.688	0.313	0.656
Average	-	-	-	0.366	0.683

The results show that classification by rating is better than a random guess, with the best performance shown using the SVM as the classifying method. The best average accuracy achieved was for Participant 5, with 76.6% and a moderate agreement from the Cohen's kappa coefficient. The SVM appears to perform better than the MLP in this case, achieving equal or higher accuracies for every participant and considerably higher Cohen's kappa coefficients. As shown from both classifiers, there is high variance in the accuracy values obtained, demonstrating how the selection of training and testing data can heavily affect results.

4.3.2 Classification by Song Recognition

A value of 1 or 0 was appended for each song on the participant dataset to represent whether they recognised the song or not. As with the rating classification, the values were calculated using the labelled song responses for all 12 songs. Table 4.9 shows the classification results for all 10 participants using the MLP and Table 4.10 using the SVM.

Table 4.9: MLP Classification for Participant Recognition

Recognition	MLP 1	MLP 2	MLP 3	Cohens Kappa	Average Accuracy
P1	0.813	0.719	0.781	0.542	0.771
P2	0.844	0.813	0.781	0.625	0.813
P3	0.844	0.813	0.813	0.646	0.823
P4	0.750	0.656	0.813	0.479	0.740
P5	0.719	0.750	0.844	0.542	0.771
P6	0.844	0.688	0.750	0.521	0.760
P7	0.906	0.875	0.875	0.771	0.885
P8	0.719	0.719	0.938	0.583	0.792
P9	0.844	0.844	0.781	0.646	0.823
P10	0.781	0.719	0.719	0.479	0.740
Average	-	-	-	0.583	0.792

Table 4.10: SVM Classification for Participant Recognition

Recognition	SVM 1	SVM 2	SVM 3	Cohens Kappa	Average Accuracy
P1	0.563	0.563	0.406	0.021	0.510
P2	0.563	0.594	0.406	0.042	0.521
P3	0.625	0.563	0.625	0.208	0.604
P4	0.625	0.531	0.563	0.146	0.573
P5	0.563	0.656	0.500	0.146	0.573
P6	0.563	0.625	0.531	0.146	0.573
P7	0.625	0.594	0.563	0.188	0.594
P8	0.500	0.594	0.563	0.104	0.552
P9	0.781	0.500	0.438	0.146	0.573
P10	0.625	0.563	0.563	0.167	0.583
Average	-	-	-	0.131	0.566

The data classifies well for song recognition using an MLP with relative bandpower frequencies as the input features. It obtained the highest average accuracy of 88.5% and substantial agreement from the Cohen's kappa coefficient. In contrast, the performance of the SVM model is not as effective, achieving the lowest accuracies of all tests performed. The Cohen's kappa values only display slight agreement, suggesting that the model's performance could be completely due to chance. As with the classification by song rating, the values for separate executions of the model vary a large amount, this will be explored in Chapter 5.

4.4 K-means Clustering

The clustering performance of the data is revealed in this section by implementing the unsupervised machine learning algorithm k-means clustering. At the end of the section, Table 4.11 shows the quality scores obtained from the "Entropy Scorer" node in KNIME, where quality is defined as 1 minus the entropy of the classification. The performance of the clusters is explored in their respective classifier sections, where 4 different classes were used for target clusters. These were song ID, song genre, participant recognition and participant rating. Scatter plots of the data clustered by both the k-means outputs and by class label is displayed using PCA to reduce the data to 2 dimensions.

4.4.1 Clustering by Song

Participant data was clustered by the song ID, to see if a song would provoke similar brain activity throughout its recording. 12 clusters were used for the 12 songs used in the experiment. The quality of the clusters showed a relatively good average in comparison with the other methods used, but there was clearly a lot of Entropy in the clusters. The best clustering quality achieved was 46.7%. In Figure 4.1 the colours are based on the clusters created using k-means clustering, whereas in Figure 4.2 the colours are assigned based on the song ID.

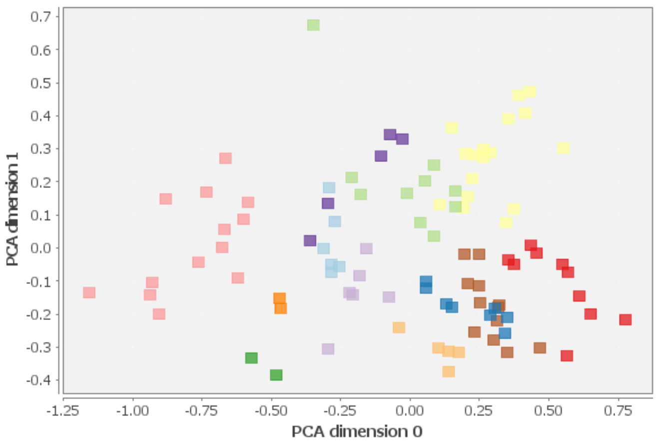


Figure 4.1: K-means Song Clusters

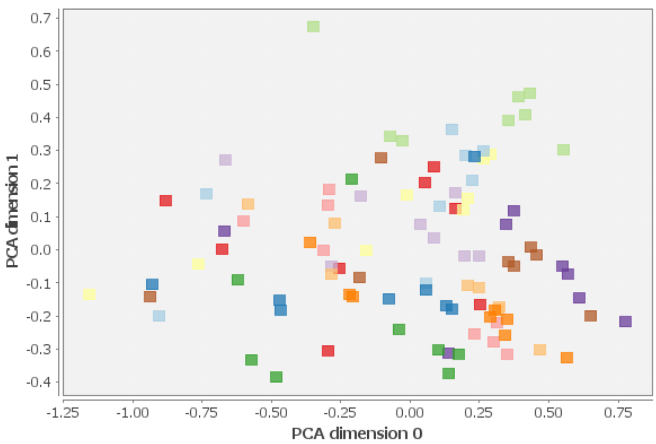


Figure 4.2: Song ID Clusters

The clusters created using k-means are reasonably split, but not enough to easily distinguish from each other. Looking at the distribution of song points shown in Figure 4.2, the clusters created do not accurately classify the data points to their respective song IDs.

4.4.2 Clustering by Genre

The data collected was classified by each song's respective genre, there being 4 in total consisting of: Rock, Pop, Rap and Indie. In KNIME, the song IDs were replaced by their genre ID, labelled 1 to 4, and the data set was classified by this attribute. Figure 4.3 shows the clusters with colours assigned by k-means clustering, and Figure 4.4 shows the colours assigned by song genre.

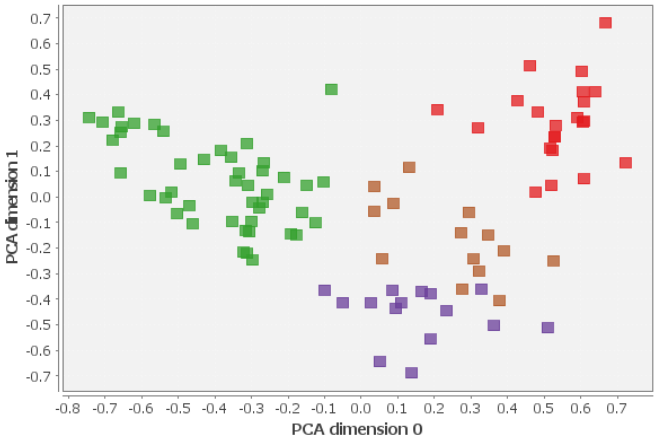


Figure 4.3: K-means Genre Clusters

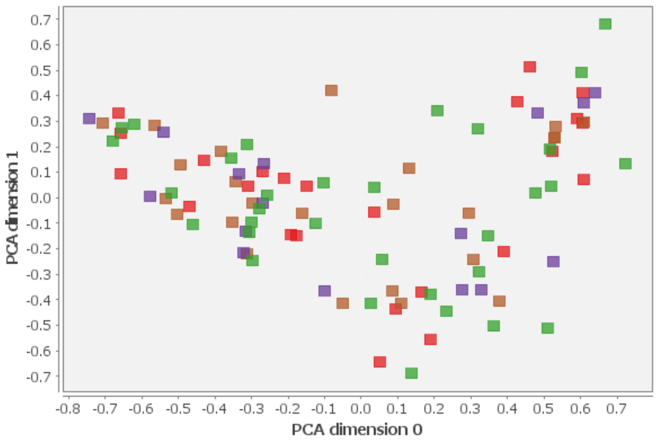


Figure 4.4: Song Genre Clusters

Clustering by genre also seemed to be ineffective and performed considerably worse than the clustering by song ID producing low quality scores for all participants, the highest being 10.2%. Although the clusters visually appear well separated, the genre class is not what separates them. The distribution of points show 3-4 possible clusters, however looking at Figure 4.4, the genre points are split with almost no clear spatial patterns.

4.4.3 Clustering by Song Recognition

Clustering by recognition of the song generated greater accuracy scores than both genre and song ID but were still ineffective, with the highest quality value being 60.2%. This clustering also provided the best overall clustering for the data at 33.9%. Figure 4.5 shows the clusters coloured by the predictions using k-means. Figure 4.6 shows the points coloured by song recognition, with the red colour points representing recognition of a song and green colour points showing an unheard song.

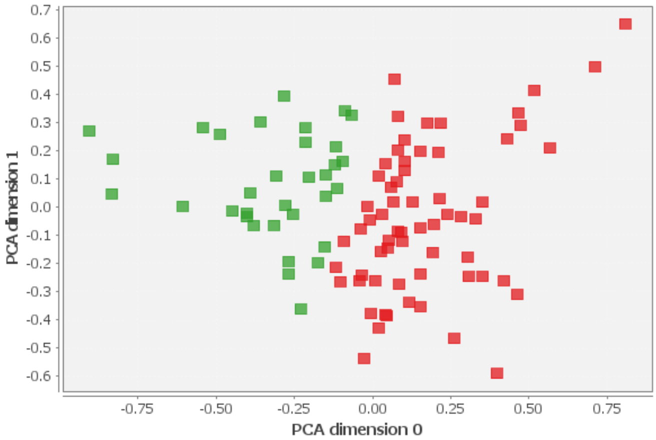


Figure 4.5: K-means Song Recognition Clusters

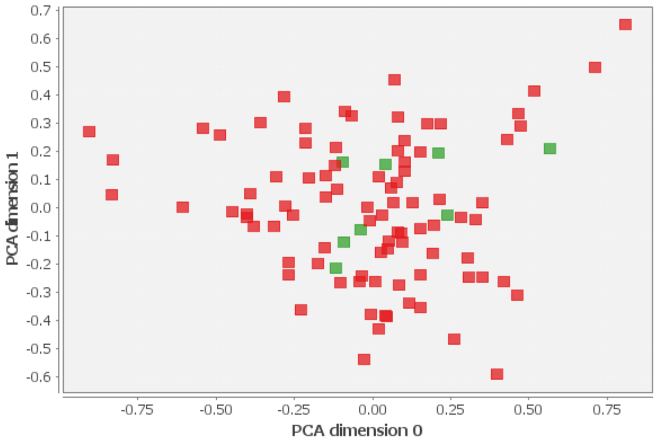


Figure 4.6: Song Recognition Clusters

The accuracy achieved suggests there is some correlation for song recognition, however looking at Figure 4.6, it is made apparent that the high accuracy values are only obtained because the class distribution is heavily weighted towards a participant recognising a song. Because of this accuracy values obtained for this clustering are not considered to be reliable.

4.4.4 Clustering by Song Rating

Clustering by rating was explored with all responses and when converted to a binary classifier. Because the number of rating responses is dependent on the participant, the number of ratings was determined before clustering and used as the number of clusters for the k-means algorithm to produce. For binary classification, 2 clusters were used, and the scatter plots are displayed below.

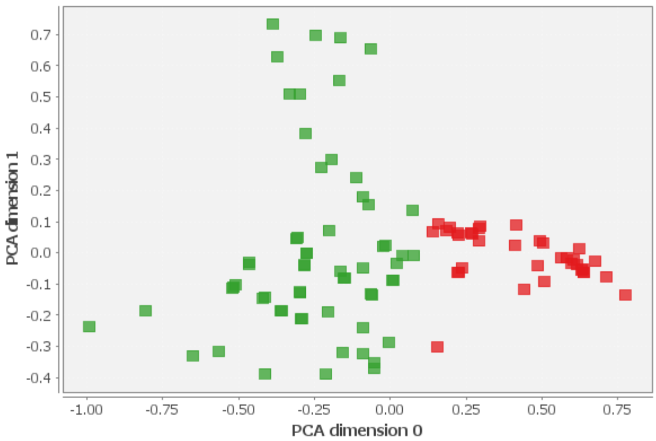


Figure 4.7: K-means Binary Song Rating Clusters

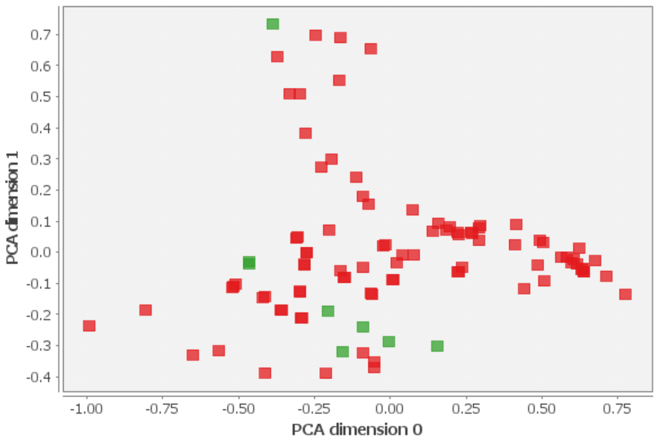


Figure 4.8: Binary Song Rating Clusters

Figure 4.7 shows the clusters produced using k-means, Figure 4.8 shows the coloured points by participant binary rating. As with the clustering by song recognition, there is an imbalanced class distribution where there are a lot more positive ratings (red) than negative ones (green), making the accuracy values obtained unreliable.

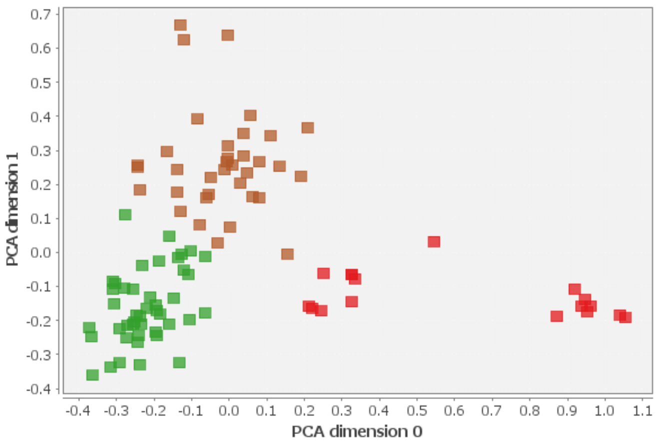


Figure 4.9: K-means Multiple Song Rating Clusters

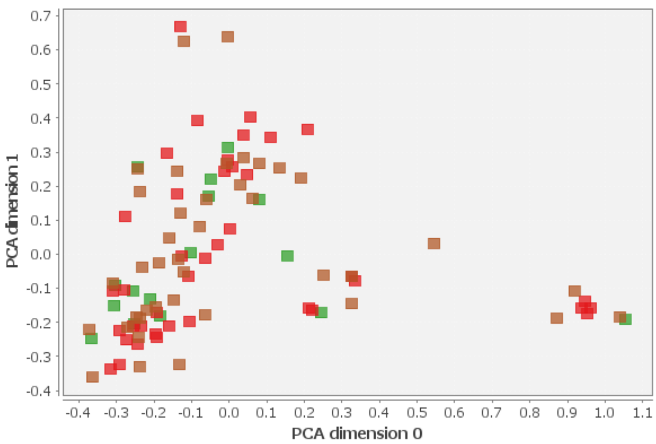


Figure 4.10: Multiple Song Rating Clusters

Figure 4.9 shows the clusters by rating for all values obtained using K-means and Figure 4.10 shows the true distribution of rating values. These figures demonstrate the data does not clearly separate by ratings values. The binary clusters accuracy values were a lot higher, with the highest quality value being 60.5%. When clustering by all rating values the performance was decreased, the highest of the participants being 21.7%. However, the higher accuracy is likely due to the class imbalance present when classifying the data, shown in Figure 4.8.

Table 4.11: Quality of K-means Clusters

Participant	Song	Genre	Recognition	Rating (all)	Rating (binary)
P1	0.205	0.060	0.363	0.093	0.216
P2	0.291	0.069	0.358	0.116	0.086
P3	0.423	0.087	0.361	0.208	0.605
P4	0.155	0.035	0.088	0.066	0.355
P5	0.404	0.102	0.602	0.217	0.157
P6	0.210	0.033	0.350	0.160	0.086
P7	0.258	0.054	0.379	0.160	0.022
P8	0.467	0.052	0.189	0.090	0.082
P9	0.365	0.069	0.350	0.154	0.204
P10	0.177	0.076	0.350	0.148	0.197
Average	0.296	0.064	0.339	0.141	0.201

4.5 Spectral Analysis and ICA Components

Producing ICA components allow for descriptive analysis of the data. As shown in the study by Chinmayi et al. (2017), the spectral peaks can be observed and used as features, however in this instance they will be used to visually interpret which areas of the brain are being used depending on the song being listened to and whether the participant's response is positive or negative.

The topographic plots shown in this section depict rough areas of activation as electric potentials and are measure in volts. The components were generated using the 20 second recordings obtained, a 30 second clip minus 5 seconds at the start and 5 seconds at the end, so depict the activity over the entire duration of the audio clip.

The labels of the channels reference the lobe of the brain where the activity is being recorded. Figure **??** shows the locations and labels for the channels used in this study, and Table 4.12 maps their respective lobes. As discussed in previous literature, the different lobes are associated with different functions, shown in Table 4.13.

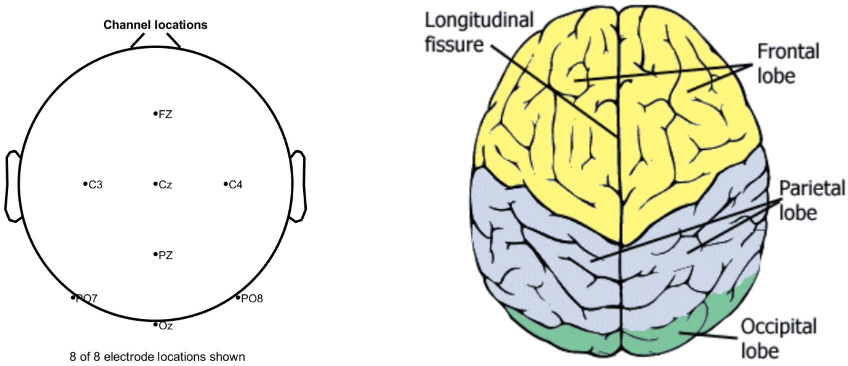


Figure 4.11: Channel Locations and Scalp Map (Chaves et al., 2020)

Table 4.12: Brain Lobe Associations with Electrodes

Electrode Label	Brain Lobe
Fz	Frontal
C3, Cz, C4	Central
Pz	Parietal
PO7, PO8	Parietal / Occipital
Oz	Occipital

Table 4.13: Brain Areas

Brain Areas	Functions
Frontal Lobe	Associated with emotion
Temporal Lobe	Responsible for interpreting sounds and language
Parietal Lobe	Processing sensory information for cognitive purposes
Occipital Lobe	Processing auditory information

Songs were chosen for analysis that had a good split of participant ratings and the signals recorded were decomposed into their ICA components. After rejecting components that represented noise in the data, the first 4 components, ordered by variance, were inspected side by side to identify which areas of the brain are activated when a song is liked or disliked.

Song 10, "Shutdown" by Skepta, had varying participant responses to it, and was chosen to compare topographic plots of ICA components. Figure 4.12 shows the difference in brain activity between positive and negative ratings, with the positive responses in the left column, Participants 3, 4 and 7, and the negative response in the right column, Participants 2, 5 and 9. Participant 4 had 5 of their 8 components rejected as they mainly consisted of noise, however the 3 components displayed depict reliable brain activity.

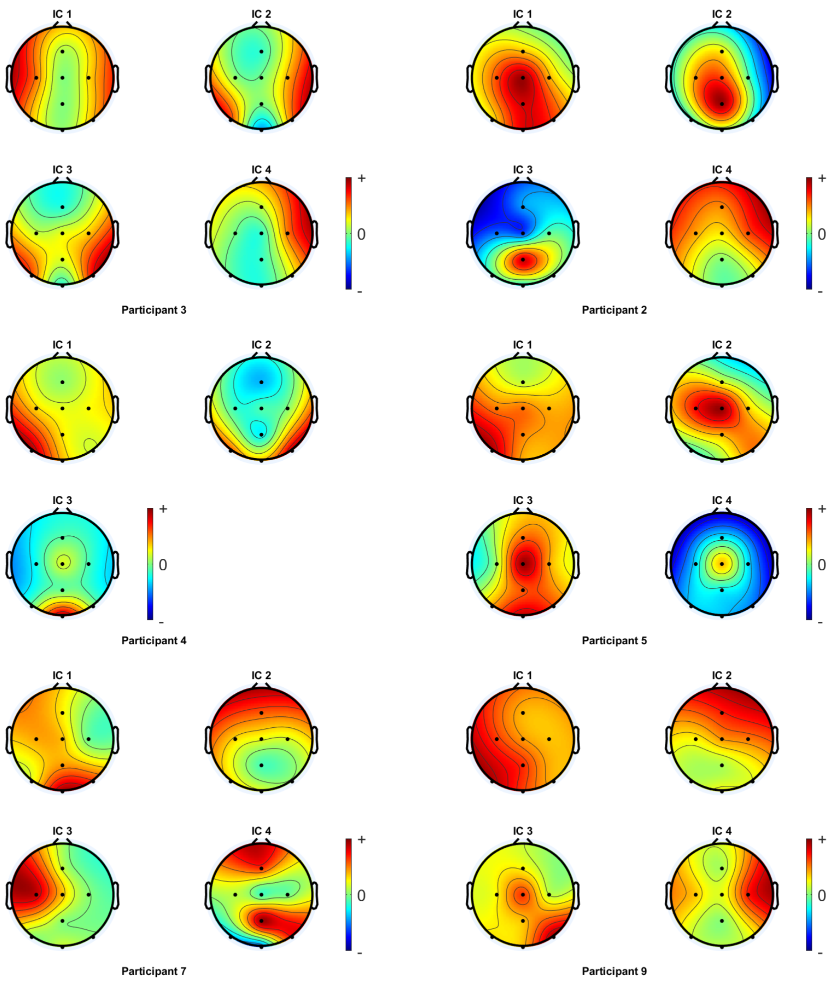


Figure 4.12: Shutdown by Skepta ICA Components

In the negative responses, the ICA components for each participant all feature at least one component where the spectral peak is in the middle of the topographic plot. This electrode is Cz and represents the middle of the central lobe. However, the central lobe is a label and not a physical part of the brain. Looking at Figure **??**, this could be showing activity from the Frontal lobe which is commonly associated with emotional response. This suggests there is a greater emotional response to negative emotions than positive ones, as comparatively there is less activity for positive responses in this area. Another common feature in both positive and negative responses is the activity in the sides and back of the brain, near the occipital lobe. This makes sense as this lobe deals with processing auditory information.

Another divisive song was compared to try and determine whether these patterns were representative of the whole data set or just a one-off occurrence. The song responses shown in Figure 4.13 are for the song "Location" by Dave.

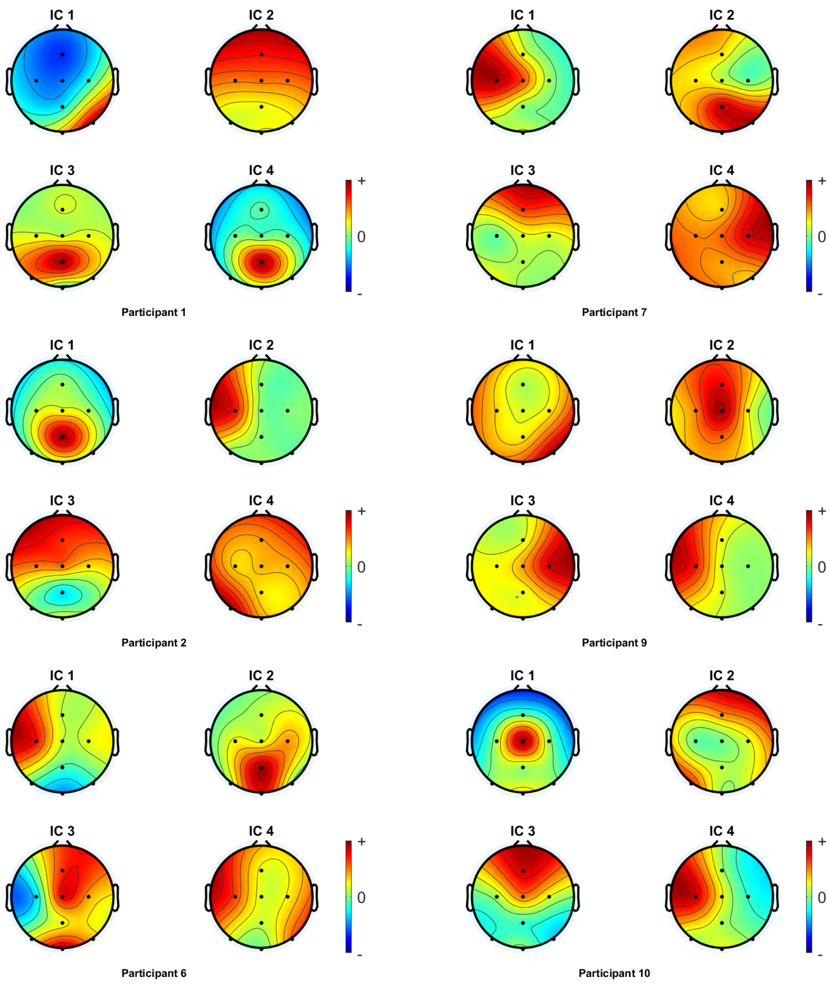


Figure 4.13: Location by Dave ICA Components

In the positive responses for this song, there are spectral peaks at the Pz electrode for all participants, compared to little activity in this area for negative responses from Participants 9 and 10. Looking back to the responses in Figure 4.12, Participant 7 has a positive response and also a peak at this electrode. However, for Participant 2's negative response there is a peak shown in the Pz electrode as well.

In Figure 4.13, there seem to be peaks on the central electrodes for both Participant 6 and 10, however this is not so clear for Participant 7. There is less activity around the central electrode Cz in the positive responses, but not as significantly less as the previous song components.

These results will be discussed further in Chapter 5, exploring the significance of the information obtained and what it may imply.

Chapter 5

Discussion

This chapter contains a discussion of the results obtained in Chapter 4 and what they may imply. It is separated into the three research questions outlined for the study. For each question, the results will be explored, and possible improvements addressed.

5.1 Supervised Classification

The main aim of the study was to determine how effective classification of song preference would be using machine learning algorithms. This involved multiple stages of signal processing and testing of various methods.

The results for preprocessing the data showed that sufficient preprocessing had already been performed during recording, as the SNR values and classification accuracy's made little impact on the quality of the signals, and in some cases made them worse. Since only 2 IIR filters were applied and 1 form of wavelet transform, it may have been beneficial to attempt using different methods or configurations to try improving signal quality, and consequently the classification reliability.

Looking through previous literature, ICA components have been used to remove noise from EEG data. This was performed in this study during the spectral analysis stage, where the software EEGLAB predicted which components were noise and which were brain signals. Components that consisted mainly of noise could then be removed. However, this could have been performed before the signals were classified and may have improved classification performance.

The feature extraction of the relative spectral powers along with normalisation showed to greatly improve classification performance and managed to produce high classification accuracies. The use of PCA had the opposite effect, although combining the two methods may have proved effective.

The model parameters could have been a limiting factor on the quality of classification. Since each participant's EEG data is unique to them, using a generalised model for all data could have limited performance. Despite this, accuracies between 63.5% and 76.6% were obtained when classifying by user rating using the SVM model, with fair to moderate agreement for all results from the Cohen's kappa coefficient. The MLP did not perform as well, achieving accuracies between 55.2% and 65.6% with slight to fair agreement.

Song recognition was also tested using the same models and provided the best classification accuracies for all experiments. The MLP achieved accuracies between 74% to 88.5% with moderate to substantial agreement. This result could suggest that the comparative frequencies produced by the brain when listening to new music are distinctive and can be used to distinguish between known and unknown stimuli. However, using the SVM the accuracies produced were not as effective, classifying at 51% to 60.4% accuracy with slight agreement for all results, the worst performance of all models tested. The varying performance of models makes it difficult to make any clear conclusions.

The high variance for both model results could be an indicator that not enough data was present to test on. The quantity of data used for these tests was greatly reduced after performing feature extraction and as a result increased the variance in performance. The chosen method of extraction was calculating a relative spectral power for each of the 5 bands. This was done for each channel per song, meaning an 8 by 5 array of values was obtained. Performed for each of the 12 songs meant the total number of data rows per participant was 96. This may have improved classification accuracies for some models; however, the consistency of the data could have been affected as well.

Another factoring element to this was the skew in data towards positive ratings of songs. On reflection, this was to be expected as only award-winning music was used, however an effort was made to diversify genres and time periods which did help to produce a range of positive and negative responses. To obtain better data for training, a more dividing playlist of music with more contrasting song pieces could be used instead.

The data used for participant rating was transformed into a binary classification problem, where values of -2, -1 and 0 were classified as dislike (0) and values of +1 and +2 were classified as like (1). Looking at Table 3.5 from Chapter 3, it could have been beneficial to try a different distribution for this binary representation, such as omitting neutral ratings. Doing this would reduce the data set further and may be more detrimental than beneficial.

5.2 Unsupervised K-means Clustering

Clustering was tested for many attributes in the dataset, including song ID, song genre, song rating and song recognition. The quality values received from the clustering show extremely low quality of clusters and high entropy.

The clusters with the highest quality values are found when a binary classifier is used. The best average quality for all participants was found when clusters were classed by recognition of the song, and the highest individual quality was found using binary song ratings. As with the classification accuracy values, there is a high level of variance in the results. When inspecting the results for binary song rating for example, the highest quality value is 60.5%, achieved by Participant 3, and the lowest value is 2.2% obtained by Participant 7. Whilst it seems that supervised classification of the data has potential for training high quality models, this form of unsupervised classification is not as efficient.

After further inspection of the data, there is a bias towards one class for both binary cluster classifications. With song rating, most data points were for positive participant ratings and for song recognition the majority was the song being recognised, which would explain the variance of classification. The accuracy obtained for Participant 3 is so high because most points are classified as positive response, whereas the significantly lower accuracy obtained for Participant 7 is due to a more equal representation of positive and negative responses. This is depicted in Figure 5.1, which shows the points coloured by k-means values on the left and the rating values on the right. The coloured points by binary rating for Participant 3 are shown in Chapter 4, Figure 4.8.

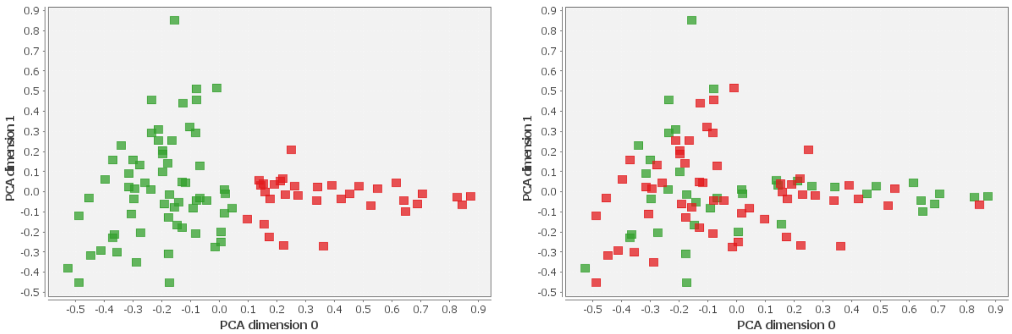


Figure 5.1: Participant 7 Clusters

The results for clustering by song ID and song Genre are more reliable, as the class distribution is more equal. However, the quality of clusters obtained does not suggest any correlation between the input values and the class labels. The data used to assign clusters was the relative bandpowers calculated in the feature extraction stage, however using all the raw EEG data could have been tested as well and may have produced higher quality clusters. It is possible that the data was clustering to a class that had not been recorded in the experiment, but the 4 tried in this study obtained poor accuracy scores.

One possible reason for the low quality of clusters could be due to the curse of dimensionality, which states that as the number of dimensions increases, data becomes sparser. This concept was first proposed by Minsky and Papert (2017) in their book "Perceptrons". A few papers have suggested ways of minimising the negative effects of using multiple dimensions which involve aggregating data and preprocessing (Indyk and Motwani, 1998, Friedman, 1997). One method would be to reduce the number of dimensions using PCA and then performing k-means clustering on that reduced data, which may have improved the cluster quality.

5.3 Spectral Analysis using ICA Components

The aim of the descriptive analysis performed on the ICA components was to identify common areas of activation in the brain when listening to a liked song versus a disliked song. As in the classification section, this analysis was also restricted by the bias towards positive responses to the songs chosen. Because of this, only two songs had enough examples of a participant really liking or really disliking a song, this being represented by a +2 or -2 rating. In both cases, participants who had scored the song either a +1 or -1 were also included for comparison. This may have caused some inaccuracies in the analysis as there were different levels of response to the music present.

Despite these restrictions, peaks at common electrodes were apparent. For both songs there appeared to be peaks on the central electrode for negative responses. This electrode is located around the frontal lobe, which suggests that a greater emotional response is elicited when listening to songs that are not liked. For both negative and positive responses, a peak around the Oz electrode and back of the brain shows activation in the occipital lobe, which could represent the participant processing the auditory stimuli presented to them.

In the positive responses a peak on the Pz electrode is shown, suggesting use of the parietal lobe. Some sources suggest this lobe is activated when a person conducts episodic memory retrieval. Episodic memories consist of memories of a specific time and place that they can recall, for example one might have an episodic memory of their first day at school. A paper by Wagner et al. (2005) makes the conclusion that there are multiple areas of the parietal lobe that are activated during episodic retrieval. In all cases when a positive response is given, the song heard was recognised by the participant. The spectral peaks at the Pz nodes indicating positive responses may instead be signalling the participant remembering the song or a memory associated with the song. In this sense, the recognition of a song therefore can be determined as an important factor in the response experienced. This detail is reinforced by the results shown in the classification of signals, where classification by recognition of song proved to be the most efficient class tested.

The inconsistency of the results and lack of distinct area activation could due to the range of intensity in responses, which in turn is caused by the limiting variety in the dataset obtained. Another potential reason could be that the ICA components were generated over the entire 20 second EEG recording. If the duration of time was concentrated on a peak of brain activity, potentially during the chorus of a song, then the ICA components would be more specific. Patterns were identified and show potential for meaningful findings, however there is a lack of appropriate data to prove their consistency.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

The research in this paper aimed to implement a unique methodology to try and classify participant preference to songs using a single-dimension emotional classification system. The data used for the study was from an original data set recorded from 10 participants using an EEG device with 8 channels. The classification was performed using two machine learning algorithms, a multi-layer perception and support-vector machine, where the target classes were the participants preference and recognition of the songs. Unsupervised machine learning was also investigated on the data, in the form of k-means clustering, and ICA components were created to perform visual analysis of spectral peaks for participants.

Preprocessing and feature extraction proved important steps during signal processing, greatly improving the classification quality of the algorithms implemented. A few different methods were considered and compared, ultimately the most effective ones being chosen for final results.

The accuracies obtained from the classification performed show that binary separation of EEG signals based on like and dislike is possible and given a larger quantity of more contrasting data the classification could have achieved even higher accuracies. A participant's recognition of a song is another promising classifier for EEG signals, proving that the relative band powers produced when a song is known or not are easily separated.

The clustering performance of the EEG bandpowers was not very efficient and suffered from an imbalanced dataset caused by bias responses from participants. Visual analysis of ICA components did suggest spatial patterns of brain activity dependant on song rating. It highlighted the significance of song recognition and the potential link between the activation of the parietal lobe when a participant recognises a song. Taking these ideas further was also restricted due to the imbalance present in the dataset obtained.

6.2 Future Work

In a future study a more diverse playlist of music would be used, and more participants responses recorded. With a larger data set, the classification accuracies would be less varied and more reliable, due to a larger representation of negative participant responses. This also would provide more ICA components to be analysed which could reinforce patterns in the data that were not made obvious during this experiment due to a smaller, less varied dataset being used. A different form of unsupervised machine learning would be applied, as the K-means clustering seemed inefficient for this study. Another form of clustering could be attempted or using different data as input to the K-means algorithm. As an example, the dimensions could be reduced before the data is clustered, which could improve performance.

The feature extraction method of converting raw EEG signals to relative band powers was beneficial for classification accuracy but left a small amount of data to be used. For another study of this nature, a different approach would be considered that resulted in the same representation of the data, but not at the detriment of its quantity. Using the spectral power difference of symmetric channels as another feature would also be considered, as this was used extensively in similar studies and proved to be an effective feature for classification.

With the ICA components, a participant could be asked to choose one of their favourite songs to listen to during recording, as this would provide a strong baseline to compare negative response to. Since the music was chosen for the participants, it was not always ensured that they would listen to a song they like and a song they dislike, so instead giving the participant more control could benefit the findings. This does, however, mean the study could be less consistent as varying stimuli would be used producing varying responses.

Chapter 7

Reflection

This project has been an extremely educational experience, introducing me to a whole new application of computation. Having no prior knowledge in the area of signal processing and brain computer interfaces, I feel I have learnt a great deal in how these studies are conducted and the vast potential they have for technological advancement. I believe I have a greater understanding of the process of data acquisition, preprocessing and feature extraction.

Having to record my own data set proved to be a challenge in many ways. Deciding on what information to record from the participants and distinguishing between useful and redundant data took up a large amount of time. This was most likely due to my lack of prior experience in the area, however I feel that with the help of my supervisor I managed to extract interesting data for the study, even if it was not as good as it could have been. The process of setting up my own experiment and reaching out to people to participate in my study helped me gain confidence in my planning ability and time management.

A lot of time was spent trying to filter the data to improve classification quality, even though the data had already been filtered during recording. I did not consider the possibility that filtering the data further would decrease the quality of the signals for a long time, and consequently a lot of time was spent looking for a solution to a problem that wasn't there.

During the analysis stage of the study, I soon realised that there were some changes that could have been made to the experiment which would have produced more reliable results. In hindsight, it should have been obvious that the music chosen was going to be predominantly liked by the participants, which lead to an imbalanced class distribution for training data. If I were to conduct this experiment again, I would use an equal mixture of typical "good" and "bad" music. The feature extraction performed was effective at producing higher accuracy classification of the data, however these accuracy values were very varied. The datasets were reduced greatly during feature extraction, which I believe was the cause of this variance. On reflection, I would have performed the same method of feature extraction by separating the EEG signals into their respective frequency bands, however I would refrain from reducing the quantity of data available.

One area I would have liked to explore more, but only realised after the experiment had been conducted, was the relationship between participants enjoying songs that made them feel negative emotions. This idea is touched upon in the Chapter 3, stating that the use of only one dimension allows the disassociation of a participant feeling negative emotions with them disliking a song.

References

Al-Fahoum, A. S. and Al-Fraihat, A. A. (2014), 'Methods of eeg signal features extraction using linear analysis in frequency and time-frequency domains', *ISRN neuroscience* **2014**.

Asif, A., Majid, M. and Anwar, S. M. (2019), 'Human stress classification using eeg signals in response to music tracks', *Computers in biology and medicine* **107**, 182–196.

Basterrech, S. and Krömer, P. (2019), 'A nature-inspired biomarker for mental concentration using a single-channel eeg', *Neural Computing and Applications* pp. 1–16.

Bazgir, O., Mohammadi, Z. and Habibi, S. A. H. (2018), Emotion recognition with machine learning using eeg signals, *in* '2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME)', IEEE, pp. 1–5.

Bos, D. O. et al. (2006), 'Eeg-based emotion recognition', *The Influence of Visual and Auditory Stimuli* **56**(3), 1–17.

Chaves, R., oliveira, W., Torres, H., Bachrach, U. and Yaniv, Z. (2020), Human brain decay and alzheimer's disease.

Chinmayi, R., Nair, G. J., Soundarya, M., Poojitha, D. S., Venugopal, G. and Vijayan, J. (2017), Extracting the features of emotion from eeg signals and classify using affective computing, *in* '2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)', IEEE, pp. 2032–2036.

Delorme, A. and Makeig, S. (2004), 'Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis', *Journal of neuroscience methods* **134**(1), 9–21.

Ekman, P. (1999), Basic emotions, *in* 'Handbook of cognition and emotion', Vol. 98, p. 16.

Friedman, J. H. (1997), 'On bias, variance, 0/1—loss, and the curse-of-dimensionality', *Data mining and knowledge discovery* **1**(1), 55–77.

Hadjidimitriou, S. K. and Hadjileontiadis, L. J. (2012), 'Toward an eeg-based recognition of music liking using time-frequency analysis', *IEEE Transactions on Biomedical Engineering* **59**(12), 3498–3510.

Hanich, J., Wagner, V., Shah, M., Jacobsen, T. and Menninghaus, W. (2014), 'Why we like to watch sad films. the pleasure of being moved in aesthetic experiences.', *Psychology of Aesthetics, Creativity, and the Arts* **8**(2), 130.

Indyk, P. and Motwani, R. (1998), Approximate nearest neighbors: towards removing the curse of dimensionality, *in* 'Proceedings of the thirtieth annual ACM symposium on Theory of computing', pp. 604–613.

Ishino, K. and Hagiwara, M. (2003), A feeling estimation system using a simple electroencephalograph, *in* 'SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance (Cat. No. 03CH37483)', Vol. 5, IEEE, pp. 4204–4209.

Li, M. and Lu, B.-L. (2009), Emotion classification based on gamma-band eeg, *in* '2009 Annual International Conference of the IEEE Engineering in medicine and biology society', IEEE, pp. 1223–1226.

Li, M., Xu, H., Liu, X. and Lu, S. (2018), 'Emotion recognition from multichannel eeg signals using k-nearest neighbor classification', *Technology and Health Care* **26**(S1), 509–519.

Lin, Y.-P., Wang, C.-H., Jung, T.-P., Wu, T.-L., Jeng, S.-K., Duann, J.-R. and Chen, J.-H. (2010), 'Eeg-based emotion recognition in music listening', *IEEE Transactions on Biomedical Engineering* **57**(7), 1798–1806.

Lin, Y.-P., Wang, C.-H., Wu, T.-L., Jeng, S.-K. and Chen, J.-H. (2009), Eeg-based emotion recognition in music listening: A comparison of schemes for multiclass support vector machine, *in* '2009 IEEE international conference on acoustics, speech and signal processing', IEEE, pp. 489–492.

Liu, Y. and Sourina, O. (2013), Eeg databases for emotion recognition, *in* '2013 international conference on cyberworlds', IEEE, pp. 302–309.

Liu, Y., Sourina, O. and Nguyen, M. K. (2010), Real-time eeg-based human emotion recognition and visualization, *in* '2010 international conference on cyberworlds', IEEE, pp. 262–269.

Minsky, M. and Papert, S. A. (2017), *Perceptrons: An introduction to computational geometry*, MIT press.

Nathan, K. S., Arun, M. and Kannan, M. S. (2017), Emosic—an emotion based music player for android, *in* '2017 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)', IEEE, pp. 371–276.

Onton, J., Westerfield, M., Townsend, J. and Makeig, S. (2006), 'Imaging human eeg dynamics using independent component analysis', *Neuroscience & biobehavioral reviews* **30**(6), 808–822.

Plutchik, R. (2003), *Emotions and life: Perspectives from psychology, biology, and evolution.*, American Psychological Association.

Rafieyan, V. (2016), 'Relationship between acculturation attitude and translation of culture-bound texts', *Journal of Studies in Education* **6**(2), 144–156.

Ramirez, R., Palencia-Lefler, M., Giraldo, S. and Vamvakousis, Z. (2015), 'Musical neurofeedback for treating depression in elderly people', *Frontiers in neuroscience* **9**, 354.

Ramirez, R., Planas, J., Escude, N., Mercade, J. and Farriols, C. (2018), 'Eeg-based analysis of the emotional effect of music therapy on palliative care cancer patients', *Frontiers in psychology* **9**, 254.

Rozgić, V., Vitaladevuni, S. N. and Prasad, R. (2013), Robust eeg emotion classification using segment level decision fusion, *in* '2013 IEEE international conference on acoustics, speech and signal processing', IEEE, pp. 1286–1290.

Schultz, T., Schaaff, K. and Wand, D.-M. M. (2008), 'Eeg-based emotion recognition', *Universitat Karlsruhe, Institut fur Algorithmen und Kognitive Systeme, Cognitive Systems Laboratory* .

Sourina, O., Kulish, V. and Sourin, A. (2009), Novel tools for quantification of brain responses to music stimuli, *in* '13th International Conference on Biomedical Engineering', Springer, pp. 411–414.

Sourina, O., Liu, Y. and Nguyen, M. K. (2012), 'Real-time eeg-based emotion recognition for music therapy', *Journal on Multimodal User Interfaces* **5**(1-2), 27–35.

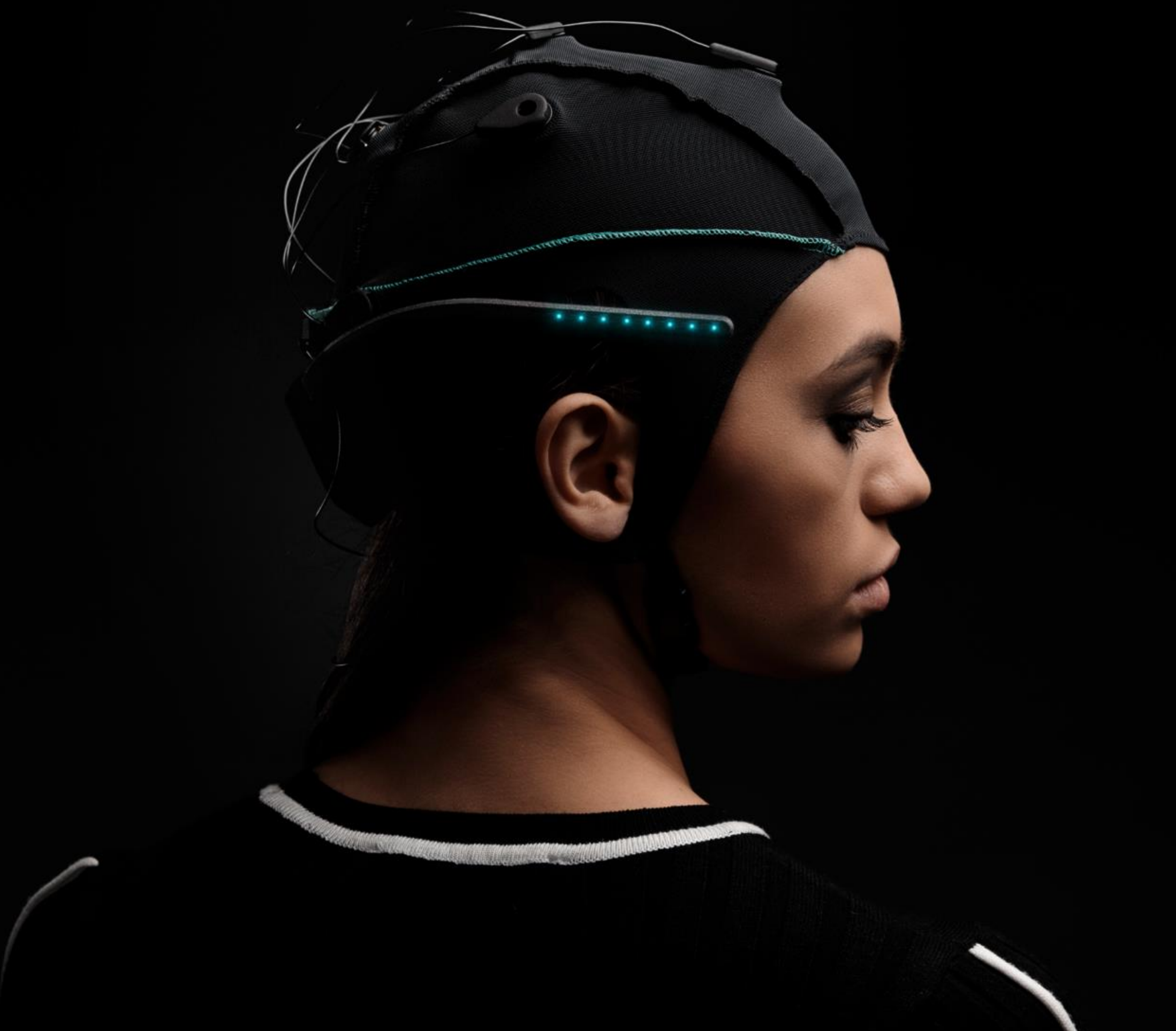
Wagner, A. D., Shannon, B. J., Kahn, I. and Buckner, R. L. (2005), 'Parietal lobe contributions to episodic memory retrieval', *Trends in cognitive sciences* **9**(9), 445–453.

Appendix A

Unicorn Hybrid Black User Manual

unicorn
HYBRID BLACK

USER MANUAL



2. INTRODUCTION

The Unicorn Brain Interface is a consumer grade biosignal amplifier kit. It allows developers, artists and makers to integrate signals from the human body within their projects – ranging from simple display of the signals to designing and controlling attached devices and interacting with artistic installations, toys, computer programs or apps and more. The Unicorn Brain Interface acquires the EEG from eight Unicorn Hybrid EEG Electrodes. The Unicorn Brain Interface consists of the Unicorn Brain Interface Hybrid Black, Unicorn C Size M, Unicorn Hybrid EEG Electrodes, Unicorn USB Charging Cable and a Unicorn Bluetooth dongle to acquire data on a computer. The Unicorn Suite is the software environment, consisting of standalone applications and APIs to interface the Unicorn Brain Interface, acquire and process data and to perform BCI paradigms.

2.1. HIGHLIGHTS

- EEG recordings without cable connection via radio signal
- Bluetooth 2.1 interface
- Hybrid electrodes for wet and dry measurements
- 8 DC-coupled analog input channels with 24 Bit resolution
- sampling rate of 250 Hz per channel
- oversampling to achieve a high signal-to-noise ratio
- input sensitivity of ± 750 mV
- 3-axis accelerometer
- 3-axis gyroscope

2.2. INTENDED USE

The Unicorn Brain Interface is intended for use in non-medical environment for non-medical applications. The Unicorn Brain Interface is used by developers, artists, makers and gamers in the user's environment.

2.3. RELEASE NOTES

Version Name	Version Number	Date	Changes
Unicorn Hybrid Black	1.18.00	6/7/19	Initial Release

2.4. CONDITIONS OF USE

2.4.1. OPERATION AND STORAGE

- Temperature: +5 to +40 °C
- Relative humidity: 25 to 80 %, non-condensing
- Atmospheric pressure: 700 to 1060 hPa

7. TECHNICAL SPECIFICATIONS

7.1. UNICORN BRAIN INTERFACE

Technical Specifications	
Model	Unicorn Hybrid Black
Type	8-channel amplifier
Battery	LP-422339-PACK, 350 mAh, IEC 62133 (ed.2)
Rated power consumption	0.3W
Rated DC voltage	3.7 V
Rated current of fuse	Little fuse 0467001.NR (1.0 A)
Rated voltage of fuse	32 V
Manufacturer	g.tec neurotechnology GmbH Sierningstrasse 14 4521 Schiedlberg, Austria

7.2. UNICORN BRAIN INTERFACE SETTINGS

Technical Specifications	
Channels	1 to 8 and R channel
Sensitivity	± 750 mV
Highpass	0 Hz
Lowpass	10.23 kHz
Input impedance	>100 MΩ

7.2.1. ANALOG-DIGITAL-CONVERTER (ADC)

Technical Specifications	
Resolution	24 Bit
Sampling frequency	250 Hz
Number of ADCs	8

7.3. MOTION TRACKING

Technical Specifications	
Acceleration range	± 8 g setting in x, y and z directions
Acceleration bandwidth	44.8 Hz
Gyroscope range	± 1000 °/s setting in x, y and z directions
Gyroscope bandwidth	41 Hz

7.4. RF MODULE

Technical Specifications	
Operating Frequency range	2400.0 … 2483.5 MHz (ISM Band)
Transmit power Max	+3 dBm
Receiver sensitivity	-86 dBm
Compliance	Bluetooth specification, version 2.1 + EDR
Marking	CE, FCC, IC, Japan and South-Korea
FCC ID	QOQWT12
IC	5123A-BGTWT12A
Japan	R 209- J00036