June 28, 2022

Beating Spotify's algorithm: towards an improved emotion label for Billboard songs

Jorrit van der Kooi

6005764

In partial fulfilment of the requirements for the degree of

Master of Science

Business Informatics

Utrecht University

Sharing science, *shaping tomorrow*

Supervisor: Dr. E.L. van den Broek
Second reader: Dr. F. Wiering

Abstract

Spotify's recommendation algorithm tailors music offerings to create a personal listening experience. Though this recommender system performs admirably, there is always room for improvement. It remains unclear if Spotify accurately classifies the affective decomposition triggered by songs. This report tries to improve these emotion labels for Billboard songs. Emotion labels can be determined based on valence and arousal. The Spotify dispositions will be compared to the valence and arousal values derived from audio features, lyrics, audio features and lyrics combined, and a listener panel. These comparisons will provide insights about how emotion labels behave when audio features and lyrics are decomposed or combined. A survey was conducted to validate the results of Spotify. Participants had to rate the most "extreme" songs on valence and arousal inter alia. Results showed that it is necessary to analyse and combine valence and arousal values from the audio signal and lyrics. Based on the valence and arousal values of the mentioned models, significant differences were found compared to valence and arousal values provided by Spotify's algorithm. From the created models, it can be concluded that Spotify applied normalisation to increase the difference between emotion labels. This way, Spotify can provide a better recommendation based on emotion labelling. Compared to a combination of audio signal and lyrics values, Spotify did a fairly accurate job in labelling emotions.

Acknowledgements

Before you lies my master's thesis, representing the end point of a long period of learning and intensive work. This report was only possible because of the great guidance that I had from different parties that supported the execution and achievement of this research project.

First and most important, special thanks to Egon van den Broek for the constant feedback and supervision. You gave me valuable insights which helped me in pursuing the right paths and achieving better results. At the same time, you kept me motivated with a humorous and optimistic attitude, which was much appreciated during my thesis time. You always shared your time with me in moments of need, answering your mail in the weekends or in the late evenings for example. Thank you very much for all your guidance, patience, and good meetings this year. For that, I will be forever grateful.

Second, thanks to Frans Wiering for your insightful feedback during this research project. Although in a smaller proportion, your feedback helped me to achieve a better final report.

Finally, I want to thank my family, friends, fellow table tennis players from VTV and my colleagues at UMCU who participated in my survey.

Jorrit van der Kooi
June 28, 2022

Contents

Chapter 1

Introduction

Music is something that keeps humanity fascinated from all times and belongs to all cultures. It is an integral part of people's daily lives. People listen to music while on the job, on the train, and while exercising. It has been demonstrated that music can affect a person's well-being and mood over time, and as a result, it has an effect on their health. Often, music can be a potent emotional stimulation that alters one's perception of time (Droit-Volet et al., 2013). Listening to pleasant music makes time seem to pass quickly. Therefore, music has been employed to minimise the subjective duration of time spent sitting in waiting rooms (Lai & Amaladoss, 2021). However, music genres are situational, such as playing rock music in a doctor's office waiting area. There are numerous instances in which music is utilised in practise. Several instances are provided to support this claim.

Frequently, music is utilised in the following settings:

1. It is customary for supermarkets to play music to entice customers to stay longer and buy more. The tempo of a song has an effect on the quantity of records sold (Milliman, 1982). Slower-paced song selections were consistently connected with greater sales volume. Customers move around the store more slowly when slow-tempo music is playing. This will increase sales because shoppers tend to purchase additional items during their supermarket visits. Additionally, familiarity with the song influences shopping durations. Customers were less stimulated by new music when compared to familiar music (Yalch & Spangenberg, 2000).

2. Music is used in call centres to minimise caller anger and employee fatigue (Niven, 2015).

3. Additionally, music has practical applications, especially in the field of healthcare. Music intervention considerably decreased anxiety levels (Nilsson, 2008).

Music consists of numerous facets. The preceding points examined the practical applications of music. Nonetheless, music can be described in various ways. The following factors illustrate why music has cultural and economic significance:

1. Technically speaking, music is a stream of information consisting of an audio signal and lyrics. The audience can be told a story or feel an emotion through music. This story may be based on whatever the performer has experienced in the past or present, for instance. In reality, history may teach us a great deal about how the current situation is realised.

2. The music industry is a commercial environment (North & Hargreaves, 2006). Streaming services such as Spotify and Apple Music are well-known for their explosive development, generating revenue through subscription fees and advertising (Vonderau, 2019; Coffey, 2016). This expansion is usually regarded as evidence of its economic and cultural significance. Artists are given the chance to discover new customers in countries where their music may not be generally available. In 2017, Spotify was valued at over $13 billion.

3. Publishing rights are an additional commercial aspect. They are a valued commodity inside the music industry. When a song is performed, covered, or aired, its owners receive compensation. Bob Dylan reportedly sold his repertoire of songs to Universal Music for $300 million (Steffes-Halmer, n.d.). Bruce Springsteen's complete discography was acquired by Sony Music for approximately $500 million.

The use of music and the manner in which it has been presented to the audience have evolved dramatically over time. Technology was a crucial contributor to this evolution. Historically, individuals could only listen to music when it was performed live. Then, jukeboxes and record players appeared and provided the crowd with extra listening opportunities. Today, consumers may essentially listen to any song on demand. This was made possible by mobile devices, the internet, and music streaming services. Popular streaming services, such as Spotify, have a song library of over one million tracks. People's musical tastes are extremely personal and may even be influenced by their personality (Vuoskoski & Eerola, 2011). Moreover,

the reasons people listen to music and when they do so are influenced by their own preferences (North et al., 2004).

1.1 Problem Statement

Over the past decades, the way music has been developed and consumed has substantially changed (Schedl, 2019). It has been observed that different generations prefer distinct musical genres (Interiano et al., 2018). The music industry has been confronted with a range of emerging music genres that have evolved, flourished, and then faded. This huge transformation comes with some extra ramifications and problems that need to be researched.

A vast quantity of music-related data and meta-data is readily accessible in the present day (T. Li et al., 2011). This information can be located and collected from the websites of musicians and record labels. In addition, community-developed websites contribute to the collection of data by providing the public with databases. This way, music data and metadata can be collected for personal interests or research purposes, for example. Evaluations of music from various discussion forums and blogs could potentially serve as a source of information for study. As the quantity of available music-related data expands, the music industry may find the problems of organising and analysing this data intriguing. Due to this huge growth in the sector, it has become hard for humans to keep track of the changes during the previous decades. Nonetheless, with the assistance of data mining, it is easier to research these vast amounts of data (Fayyad et al., 1996). Data mining is the application of data analysis and discovery methods that, within acceptable limits of computer efficiency, provide a certain enumeration of patterns across the data. With the aid of data mining techniques, this exploratory report focuses on identifying patterns in the data.

Many academics are interested in music emotion recognition (abbreviated as MER) due to the fact that people's conduct in consuming music is changing (Sangnark et al., 2018; Soleymani et al., 2014). Recommender systems on a variety of music streaming sites alter the way people choose the music they want to listen to. Therefore, the results of MER research might allow listeners to choose music that matches their mood and to provide a way to design or recommend a special playlist for the listener. However, recommendations are based on similarities by comparing what others with similar tastes have listened to before (collaborative filtering), or by matching attributes (e.g., genre, artist) of the music pieces (content-based filtering). This can result in recommendations being supplied may fit the user's taste,

but may not match the user's actual needs at that moment (Ferwerda & Schedl, 2014).

Thus, recommender systems have certain limitations and therefore could be improved on the following aspects:

1. Spotify offers a collection of playlists under the "Mood" section (Amini et al., 2019). However, these playlists are based more on context than mood, such as activity- or time-based. In addition, these playlists are generalised, meaning they are personalised for a particular consumer. They do not customise their recommendations based on the user's present or desired mood. On the basis of music, it is possible to develop a mood-inducing recommender system. The study of Garrido & Schubert (2011) demonstrates that non-personalised music is less effective in changing a desired mood compared to personalised songs.

2. Technology has also made it possible for users to receive recommendations based on their listening habits. The purpose of a recommender system is to aid users in content discovery and exploration by offering items that match their own tastes (Bollen et al., 2010). It intends to minimise the user's option overload by suggesting songs that match their musical preferences and listening objectives. However, the majority of these suggestions are similar tunes. It is unlikely that users will receive recommendations that they are unfamiliar with, resulting in frequent exposure to the same musical genres. In contrast to the employment of jukeboxes in the past, the exploration of new musical genres is restricted by this method.

3. Recommender systems rely on users' consumption patterns to be able to recommend a new song (Panda et al., 2021). This means that a problem appears, if a new listener makes use of the system. This will mean that the recommender system is unable to recommend new and unpopular songs to the listener given the lack of listening data. This is known as the cold-start problem. Providing a listener with songs that fit the user's taste, for example, based on the desired mood, would overcome this problem. Labelling emotions correctly would be of great use.

4. People's behaviour is altered by the employment of persuasive technologies. Once record labels realise which new artists/songs were successful, they may adapt to the market more swiftly. Keeping up with current trends allows music recommendation algorithms to provide more accurate recommendations.

Understanding music is of core value to many industries. Music is used in online advertisements, in supermarkets, in hospitals, by call centres, and represents value itself. These cases share the essence of music: it influences its listeners' affective state (e.g., emotions). This graduation project aims to analyse music features in relation to experienced emotions.

Computational models of music are well-studied, as is even evident in the department of Utrecht University. However, the relationship between music and emotions is understudied, and computational models based on this relationship are scarce. Besides, music is often treated from a single data perspective (e.g., only audio analysis or lyric text analysis), where multiple perspectives are relevant and complementary. This project aims to combine audio analysis with lyric text analysis.

1.2 Research Aim

This research is aimed at exploring a self-crafted dataset filled with data that contains songs from the Billboard Year-End chart. The songs are from the years ranging between 1956 and 2020.

This dataset was chosen due to its advantages over other music datasets for a number of reasons:

1. Annually, the Billboard Year-End chart ranks the most popular songs in the United States. This way, it represents data that can be used for time series analysis as the number of songs remained consistent over the years that appeared in the chart (van Balen et al., 2013). It can be expected that this dataset will reflect important commonalities and significant trends in popular music.

2. It represents the musical tastes of the populace because the songs on the charts are deemed to be the most popular of the year. Popular music can be viewed as a "mirror of society" since it reflects changes in people's demands and tastes within the framework of social and cultural change (Schellenberg & von Scheve, 2012).

3. Popular music is multidimensional, ranging from a recreational activity to an identity marker, a commercial commodity, a means of forming social bonds, and a means of communicating political influence. Billboard has maintained track of chart records for decades, beginning in 1940. A further advantage of selecting the Billboard chart over alternative datasets is that each year contains the same number of songs. Thus, results will not be distorted.

This dataset will be the fundamentals to study the relation between music and emotion. It contains valuable features that can be used to classify what kind of emotion it evokes by the consumer. This graduation project will examine three facets of this relation:

1. The emotional part of the songs and the audio features of a song. Songs can be classified based on the type of emotion they evoke in the consumer using audio features provided and extracted from Spotify itself. These audio features provide critical information about the characteristics of a song, e.g.: valence, instrumentalness, acousticness, tempo, and more. It can be interesting to investigate whether songs have changed over time with regard to these features.

2. In addition to their audio features, lyrics can also influence the emotions of listeners given how they carry the semantic blueprints of a song. Therefore, natural language processing will be employed over song lyrics, to extract their emotional load. This will be accomplished by utilising a lexicon, including the valence and arousal levels of specific words.

3. Music is understood to be culturally and historically contingent. In order to examine the progression of music over time, the Billboard Year-End chart will also be analysed. This will reveal tendencies and enable the prediction of future music trends.

Subsequently, these three elements will be interconnected and compared to Spotify's classifications. Spotify employs an algorithm to ascertain the emotional valence of its tracks. However, it is questionable whether Spotify appropriately categorises these categories. Two metrics are used by Spotify to describe the moods of their music: valence and arousal. It is believed that these two attributes for the majority of streaming services are obtained via rather complex modelling techniques applied to the audio signal (Sherga Jr et al., 2021). Furthermore, it remains unclear influence lyrics has on valence and arousal.

This report's findings should contribute to an in-depth understanding of the effect of music on listeners' emotional states, which has implications for a variety of corporate settings and industries. It would enable the enhancement of music recommendation systems, as stakeholders seek specialised song lists that provoke particular feelings (e.g., in a hospital, a reduction of stress will be preferred). These findings may also provide input for music in practise, given that music is increasingly utilised in practise.

1.3 Research Questions

Based on the current gap in the literature and the current problems with music recommender systems, the following main research question is composed:

[**MRQ**]: *"Does Spotify accurately classify the affective disposition triggered by songs from the Billboard chart?"*

To provide an answer to this research question, the Spotify dispositions will be compared with those derived via:

[**RQ1**]: *Audio features*

Audio features contain valuable information about each song and describe a certain characteristic. These audio features are, e.g., acousticness (measurement of acousticness), speechiness (presence of spoken words in a song) and tempo (estimated beats per minute of a song). Valence and arousal, which are derived from the audio signal, can be used to determine the emotion of a song. These metrics will be explored. Furthermore, based on the other audio signal features, new values of valence and arousal will be composed. These newly developed metrics will be compared to the values of valence and arousal of Spotify.

[**RQ2**]: *Lyrics*

First, a comparison will be made between state-of-the-art methodologies according to the literature for extracting emotions from song lyrics. The strongest methods will be combined to establish the approach that will be used. Second, using an expanded lexicon will be the fundamental way to determine the emotion of words in the lyrics. Consequently, the valence and arousal values of a song will be determined using this vocabulary. Once more, the lyrics values will be compared to the affective disposition values from Spotify. Third, a number of machine learning techniques will be used to predict the emotional content of a song. The performance of these algorithms will be evaluated.

[**RQ3**]: *Audio features and lyrics combined*

The answers to [**RQ1**] and [**RQ2**] will aid in answering this inquiry. Consequently, audio analysis and lyric text analysis will be merged.

[**RQ4**]: *A listener panel*

This research question will measure affective disposition triggered by songs by a listening panel. An online survey will be used to collect responses. Participants have to annotate to what extent they experience valence and arousal for a selected list of songs.

Moreover, I will look into trends in the affective dispositions of songs from the Billboard chart over time.

1.4 Report Structure

The rest of this report is structured as follows:

Chapter 2 contains the literature review. It will discuss the audio features in depth. Next, approaches for the annotation of valence and arousal in lyrics will be outlined. Additionally, several mood models will be described. Finally, the historical context of Billboard is presented, along with a brief description of the evolution of music.

In Chapter 3, a theoretical framework for data mining (KDD) will be explained. This framework is an iterative process where evaluation measures can be enhanced, mining can be refined, new data can be integrated and transformed in order to get different and more appropriate results. It will be used to discover useful knowledge from the data. Also, the experiment setup of the survey will be explained to conclude the chapter.

Chapter 4 contains the results and aims to answer all the research questions. An overview of the data will be presented. Thereafter, several models will be used to compare the differences between the Spotify dispositions and the audio features, lyrics, audio features and lyrics combined, and the verdict of a listener panel.

This report will be concluded in Chapter 5. There, the findings of this thesis will be summarised. Furthermore, the limitations of this report will be discussed and some recommendations will be provided for future research.

An overview of the report outline can be found in Table 1.1.

Table 1.1: Outline overview

Chapters	***RQ1***	***RQ2***	***RQ3***	***RQ4***
Literature research	✓	✓		
Data and method	✓	✓	✓	
Results	✓	✓	✓	✓

Chapter 2

Literature Research

This section will provide a comprehensive analysis of the current state of the art in audio features, lyrics, and emotional models. Following that, different ways of annotating valence and arousal in lyrics will be discussed. Investigating the most reliable emotional models will provide insight into how music might elicit various feelings in listeners. Different moods will also be investigated in terms of their traits and progression. Following that, a brief history of Billboard will be offered in order to emphasise Billboard's importance to music and why data from Billboard was chosen. The Billboard dataset includes these audio attributes.

2.1 Spotify

In 2020, the Recording Industry Association of America (RIAA) determined that music streaming platforms accounted for 83% of today's digital music industry revenue, up from just 5% in 2009 (Sherga Jr et al., 2021). The transition to digital music consumption has also brought with it the technology to propose songs and enhance the listening experience for the listener, including recommendations.

The advent of Spotify and other music-streaming services has facilitated consumers' access to a greater variety of music. This has the unintended consequence of overwhelming users with options. Consequently, music streaming platforms are competing to provide users with the finest music discovery experience. Methods for defining and modelling emotional measures are traditionally drawn from mechanical study of the music itself, i.e. the audio signal. The Spotify model evaluates audio features to offer the recommendation algorithm with characteristics that provide listeners with personalised

listening options. On each tune in their repertoire, Spotify employs a suite of audio analysis algorithms. These extract a dozen high-level acoustic characteristics from the audio, which will be explained in the following sections. According to Panda et al. (2021), Spotify also uses automatic data extraction from songs by web crawling (e.g., metadata, lyrics, reviews) to estimate the audio feature values. Spotify provides two measurements, which can be used combined to determine the emotion of a song: valence and arousal/energy. These values are computed from the audio signal (Panda et al., 2021). These measurements will be discussed more elaborately in the upcoming sections. This reports intends to utilise these Spotify audio features to study how they relate to the actual affective responses of listeners as this remains unknown.

2.2 Music Emotion Recognition

Music Emotion Recognition aims to automatically extract emotional information present in music (Panda et al., 2021). This field combines knowledge from areas such as music theory, machine learning, digital signal processing and psychology. In very broad terms, MER uses a variety of musical data sources, e.g., audio signals, lyrics and scores. These sources will be used to understand the relations between its properties (i.e. (audio) features) and its emotional cues (e.g., annotations).

2.3 Affective music players

Another technique to improve the music selecting process is to use music's emotional power (Janssen et al., 2012). If a music recommender system has emotional intelligence, it could automatically build playlists that energise, relax, or make the listener happier. Such technology could focus on the affective qualities specific activities require. Furthermore, this technology may tune the listener's mood, matching the music to their current or desired affective state (Janssen et al., 2012; van der Zande, 2018).

Such technology can be used for an affective music player. This takes and interprets the affective state (e.g., mood) of the user and basis its output (e.g., a song or a playlist) on the interpretation of the affective state.

The ability to influence mood has implications for cognitive function, health, and well-being. Positive moods boost creativity, improve decision-making, and strengthen social relationships. Furthermore, positive moods

decrease stress, which can have a negative impact on a listener's health and well-being.

(van der Zande, 2018) described the current shortcomings of music affective players. The paper stated that the main shortcoming was that the user's chosen mood was not applied. Instead, the main focus was directed on selecting moods that were extremes from each other.

2.4 Audio features

This section provides research on the evolution of audio features and their characteristics. These audio characteristics will serve as research variables. According to the literature, researching them will provide an understanding of how and why these variables changed.

Loudness

Loudness, which is calculated in decibels by measuring the intensity of audio waves, is a feature of a song that affects the level of energy. In general, louder songs are typically more aggressive and energising, whereas softer songs have a calming effect. Our perception of sound amplitude or volume correlates with loudness (Serrà et al., 2012). This variable refers to the inherent loudness of an audio recording, not the loudness a listener can adjust using buttons and sliders. Serrà et al. (2012) researched groundbreaking patterns of pitch, timbre, and loudness usage in popular western music from 1955 to 2010, including over 400.000 distinct recordings.

To comprehend the dynamics of musical tone, pitch and timbre are as important as volume (Fabiani & Friberg, 2011). Pitch is the harmonic content of the piece, including its chords, melody, and tonal arrangements, whereas timbre is the hue, texture, or quality of the sound. Timbre can be linked to instrument types, recording techniques, and expressive performance assets.

According to the study of Serrà et al. (2012), they observed three major tendencies in the evolution of modern western popular music. The first tendency was the limiting of pitch sequences, which was supported by research indicating that pitch progressions were becoming less diverse. Second, the timbral palette became more uniform, as represented by frequent timbres becoming more frequent. The last trend indicated a growing average level of loudness. These findings suggest simpler pitch sequences, trendy timbral combinations, and louder loudness. Figure 2.1 depicts the growth of loudness according to the study of Serrà et al. (2012). In 60 years, the median

value of the loudness variable decreased from -22 dB to -13 dB, illustrating the so-called "loudness war" (Vickers, 2010).

Watanabe et al. (2020) asserts that the increase in loudness is attributable, in part, to the rapid growth of digital technology. The introduction of the CD substantially altered the music industry, as the digital audio format enabled an expansion in dynamic range that analogue forms did not provide (Devine, 2013).

It was also claimed that the increase in loudness could be the result of a conscious decision by the artists, as it provides a competitive edge in terms of attracting attention (Hove et al., 2019). The growth in loudness can also be perceived as a disadvantage since the loss of sound quality and musical emotions, such as excitement or emotional components, may result from the employment of techniques that maximise sound loudness.

Hove et al. (2019) conducted a similar analysis, focusing instead on the Billboard Hot 100 Chart. The same trend appears to hold true for this study's increase in volume. A third study similarly investigated the loudness variable but for Japanese popular music (Watanabe et al., 2020). Nonetheless, the same tendency, shown in Figure 2.2, was seen.

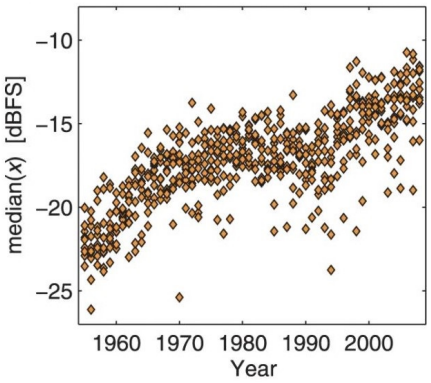


Figure 2.1: Distribution of the empiric median. The loudness values of x grows from $-22dB_{FS}$ to $-13dB_{FS}$ over the years 1955-2010. Decibels can be used to measure the sound intensity (Serrà et al., 2012).

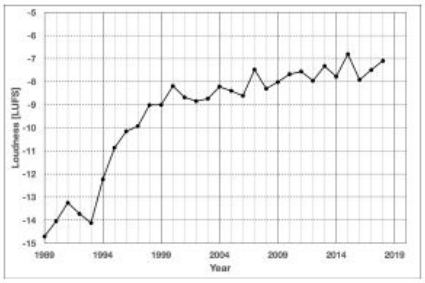


Figure 2.2: Loudness transition in LUFS annually. The start value was -15 LUFS in 1989 and increased to -8 LUFS in 2018. LUFS is a standard loudness measurement unit and considers the human listening experience (Watanabe et al., 2020).

Danceability

The danceability of a song is determined by a mixture of musical elements, such as tempo, rhythm stability, beat strength, and overall regularity. It evaluates the rhythmic salience of a song, which might be viewed as a variable with a high level of arousal (Liew et al., 2020). Christensen et al. (2014) investigated the effects of valence and arousal in boosting emotional experiences through dancing. The research indicated that the valence of dancing moves was significant. Sad music increased the ratings for sad dancing movements, whereas joyful music did not increase the scores for happy dancing motions. This could explain a probable association between *danceability* and *valence*, and, *danceability* and *energy* given that *danceability* is a high-arousal activity.

In recent decades, the number of genres that induce movement has risen (Interiano et al., 2018). This may be a reason for the increase in danceability. Furthermore, popular music tends to place a higher value on danceability because it could lead to a higher chart position (Askin & Mauskapf, 2017). The ease of moving with music can be linked to bass frequencies on various levels (Hove et al., 2019). Increasing the bass frequency is a frequent method for increasing engagement and contributing to the chart success of a song. Additionally, it encourages bodily movement, which might be characterised as danceability.

According to the research of (Interiano et al., 2018), there is a difference in danceability between the top 100 songs and non-top songs, presented in Figure 2.3. They analysed over 500,000 songs released in the United Kingdom between 1985 and 2015 in an effort to comprehend the success dynamics. Regarding the acoustic characteristics of songs that made it to the top of the charts, they wished to determine how this occurred.

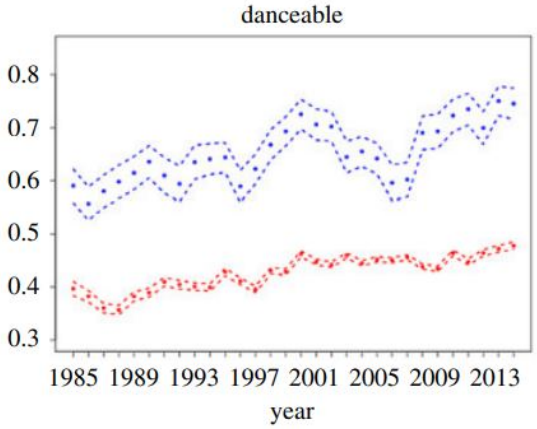


Figure 2.3: Danceability of the top 100 songs (blue dots) and non-top 100 songs (red dots) from 1985 to 2014, along with their t-distribution-based 95% confidence intervals. (Interiano et al., 2018).

Tempo

Tempo, or beats per minute, could serve as a defining characteristic of a song. It reflects the overall tempo of a piece of music. This characteristic is a reliable predictor of the velocity and intensity of the emotion linked with a song (Jamdar et al., 2015). In general, a song with a higher BPM is considered more energetic than one with a lower BPM.

Since the 1960s, Schellenberg & von Scheve (2012) has investigated the structural changes in popular music, focusing particularly on tempo and mode. It has been demonstrated that certain musical features, which are inherent properties of music's structure, influence the emotions of listeners (Juslin & Sloboda, 2011). Russell's approach can be viewed as a predictor of whether songs are judged as happy or sad. However, pace may also be a valid indicator for mode conformity. Music that evokes a happy feeling is typically composed in a rapid tempo and major mode, whereas music that evokes a sad feeling is typically composed in a slower tempo and minor mode (Schellenberg & von Scheve, 2012).

Tempo may be closely associated to arousal according with the valence-arousal model, as tempo is believed to influence a variety of emotional expressions, including happiness, surprise, pleasantness, anger, and fear (Van der Zwaag et al., 2011; Gabrielsson & Lindström, 2010). Regarding the valence-arousal model, pace is related to the dimension of energetic arousal (Gabrielsson & Lindström, 2010). A slow tempo is often connected

with low-arousal (sad) music, whereas a fast tempo is often associated with high-arousal (happy) music (Schellenberg et al., 2000; Webster & Weir, 2005; Chen et al., 2016).

Regarding the research of Webster & Weir (2005), they investigated the impact of *tempo* and *mode* on whether or not a song is viewed as happy or sad. Figures 2.4 and 2.5 represent their findings. It turns out that music in the minor mode is seen as sadder or less joyful than music in the major mode. Additionally, songs tend to get happier as the tempo increases. This result suggested that major and minor phrases have different critical speeds at which the influence of tempo on mood is either weakened or strengthened.

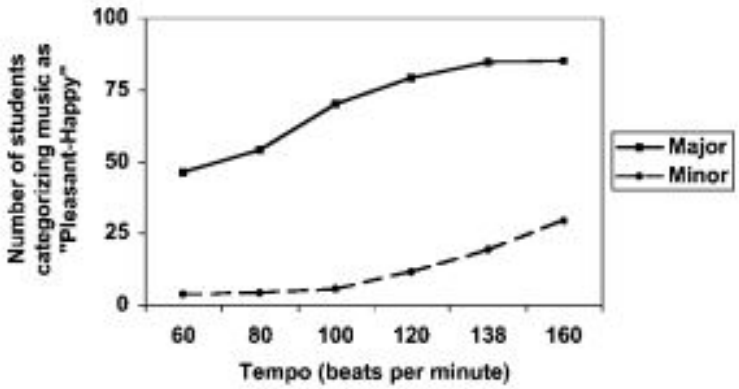


Figure 2.4: 83 participants assessed five musical phrases as "Pleasant-Happy" as a function of tempo & mode. Tempo has a decelerating influence on the major mode, but it has an accelerating effect on the minor mode (Webster & Weir, 2005).

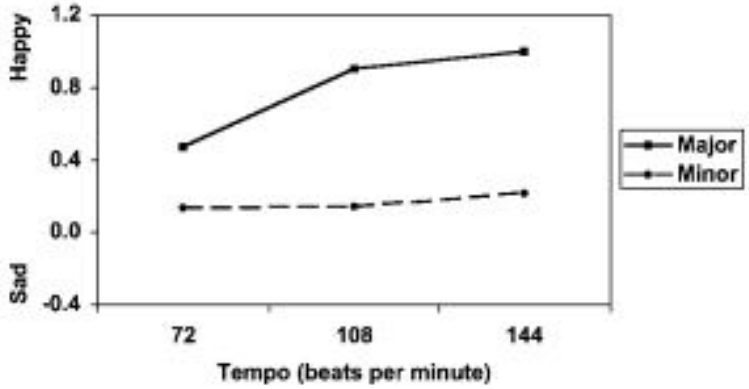


Figure 2.5: Emotional responses to a variety of musical stimuli with variable tempo (slow versus fast) and mode (major versus minor). The y-axis depicts happy-sad ratings as a function of mode and tempo. Positive ratings indicate increasing happiness, while negative ratings indicate increasing sadness. Tempo has a slight decelerating effect on major music and a slight accelerating effect on minor music. According to the ratings, however, faster tempos were positively associated with happiness and negatively associated with sadness (Webster & Weir, 2005).

According to Husain et al. (2002), altering mode affected mood but not arousal, whereas manipulating tempo affected arousal but not mood. In a tonal context, the valence dimension of Russell's model can be related to mode. As the Figures show, the minor mode is linked to sadness, while the major mode is linked to happiness. There are, however, contradictory indications that pace promotes arousal. This indicates that the valence and arousal axes of the valence-arousal model cannot be completely adjusted independently by tempo and mode (Van der Zwaag et al., 2011).

Additionally, Schellenberg & von Scheve (2012) examined the Billboard Year-End chart and assessed tempo and mode, albeit using a lower sample size. The study showed that pace decreases linearly over time. However, the 1990s saw the slowest tempo recordings, and since then the tempo has grown. This may indicate that the declining trend has levelled out and is beginning to reverse.

Duration

The duration of a song is another aspect of popular music that will be studied. Pettijohn & Sacco Jr (2009) examined Billboard number one songs

from 1955 to 2003 in relation to changes in US social and economic circumstances and song features. They utilised an indicator that represents a standardised, global measure comprised of the US employment rate, the change in disposable personal income, the change in the consumer price index, the death rate, the birth rate, the marriage rate, the divorce rate, the suicide rate, and the homicide rate. There was apparently a link between song duration, song ratios, and the GHTM. Specifically, when social and economic situations in the United States were relatively bad, longer songs were popular. Other criteria that contributed to a song's popularity were a higher level of significance in its content, a more comfortable tone, a more romantic tone, and a slower tempo. In general, the duration of songs has increased throughout time, however the pattern fluctuates regularly.

Schellenberg & von Scheve (2012) examined the Billboard Year-End chart to determine the relationship between song length and pace. Their investigation demonstrated a strong negative correlation between tempo and duration, leading to the following conclusion: longer recordings tend to have a slower speed. To illustrate this assumption, a ten-second increase in duration was accompanied by an average decrease of 1.7 beats per minute. Additionally, there was a strong correlation between recording year and duration. On average, the duration of songs increased by 1.3 seconds every year over time. The average song length rose from less than three minutes between 1965 and 1969 to about four minutes between 2005 and 2009. Duration peaked in the 1980s and thereafter declined.

Furthermore, technology may have affected the duration of songs. Shortly after the introduction of the compact disc in 1982, sample duration peaked in the late 1980s. Additionally, the increased storage capacity of compact discs enabled musicians to keep lengthier songs on them, allowing them to produce longer songs and still sell them in stores.

The lengthening of songs could be seen as another indication that music is becoming more intricate. When the pace was held constant, the correlation between duration and recording year remained unambiguous. This indicates that the overall number of beats in a song has increased.

Acousticness

Acousticness corresponds to the acousticness value of a song. Acousticness is caused primarily by instruments, such as an acoustic guitar or a piano. A low acoustic value indicates that a song consists primarily of electric sounds (Panda et al., 2021). Electric guitars, synthesisers, and auto-tuned vocals generate these electric sounds. Panda et al. (2021) gathered audio clips and

investigated their audio characteristics. According to their research, low acousticness is typical of joyful music. Music that was very loud or angry also had a low acoustic value.

Second, the study revealed a strong inverse relationship between acousticness and energy. Therefore, when acousticness is high, energy is low, and vice versa.

On the basis of these findings, it can be assumed that acousticness decreased over time as angry music and energy increased.

Energy/Arousal

The energy feature can be defined as a measurement indicating the intensity and activity of the song. Intensity, also known as dynamics, gives an indication of the degree of loudness or peacefulness of a music piece (Bhat et al., 2014). In reality, energy is not exactly arousal, but serves as its surrogate (Panda et al., 2021).

Furthermore, it is a reference to measure stress in the composition of music. Hove et al. (2019) investigated how energy levels in popular songs changed between 1955 and 2016. In the study, energy/intensity is described as root mean squared energy. To evaluate the intensity of a song, the root mean squared energy can be utilised (Bhat et al., 2014). *RMSE* gives the global energy of an audio signal's waveform and can be used to identify music (Girmal et al., 2018).

In general, peaceful music should have a low intensity level, and joyous music should have a medium intensity level. The study of Hove et al. (2019) demonstrates a growth in energy throughout recent decades. This rise can be caused by an increase in bass frequencies. A recent analysis of songs on German charts from 1965 to 2013 found that the ratio of low-to-high frequency energy has grown over time (Oehler et al., 2015; Hove et al., 2019). This shows that there is a rise in bass levels, which might be related to greater amounts of RMS energy. Bass frequencies appear to be especially important in inducing and stabilising movement to music (Burger et al., 2013). When music is broadcast over a sub-woofer, it should stimulate the body more compared to earbuds merely, because a sub-woofer has more bass (Hove et al., 2020). The study of Hove et al. (2020) demonstrated that respondents moved more and regarded songs as more pleasurable under these circumstances, in essence: listening to songs presented via a sub-woofer.

However, these sub-woofer impacts were stronger for more contemporary songs (2004-2015) compared to earlier songs (1972-1986). (1972-1986). It suggests that newer songs had significantly more spectral flux in the bass

range (0–100 Hz) than older songs. Spectral flux defines the fluctuations in the frequency content of an audio signal. More spectral flux will result in more bass.

The growth of bass frequencies could be linked to technological advancement. The upright acoustic bass, used for rock and roll in the 1950s, prepared the way for the more prominent electric bass in the 1960s Perone (2009); Hove et al. (2019). Bass grew much louder and lower with the development of drum machines in the 1980s. Low-end bass levels continued to climb as portable boomboxes in the 1980s became popular. Eventually, subwoofers functioned as a substitute for the boombox a decade later. On top of that, recording and editing equipment got better, which made it possible to reach deeper bass levels.

Interestingly, genres that originated in the 1980s/1990s, such as hip-hop/rap and electronic dance genres, are typified by their significant low-end bass frequencies (Mauch et al., 2015). This shows that technological improvement might have influenced the evolution of musical styles, genres, and listening preferences.

Valence

The psychological valence scale can be explained using the phrase "happy-unhappy." Valence is the melodic positivity expressed by a song. At one end of this scale, respondents feel happy, pleased, and satisfied, whereas at the other end, they feel unhappy, angry, and dissatisfied. The average psychological valence scores for the word were derived by the ANEW study (Bradley & Lang, 1999). This allows the valence of song lyrics to be measured.

Liveness

Performing live music provides numerous sector-specific benefits. According to the findings of Holmes et al. (2006) , live, interactive music is an effective short-term treatment for apathy in individuals with moderate to severe dementia. In addition, live music is superior to recorded music for short-term therapy. There was no substantial short-term efficacy of prerecorded music in the treatment of apathy in participants with severe dementia.

Furthermore, the category of *live performance* has received a growing amount of scholarly attention recently (Holt, 2010). This interest is mostly driven by the expansion of the live music industry. However, there is a gradual reduction in live performances in the United Kingdom (Frith, 2007). The study suggested that as more time was spent listening to music at home,

less time was spent attending live performances. The study by Brown & Knox (2017)contradicts this statement. There has been an unparalleled increase in the economic value of live music, such as festivals and concerts. According to the study by Connolly & Krueger (2006), the primary source of income for artists is live music, not the sales or streams of their recorded music. Between 2001 and 2010, ticket prices increased by an average of 39%. The fact that ticket prices have skyrocketed in recent years indicates a persistent willingness to pay above the retail price for live events. Ticket scalping has created an entirely new industry. (Black et al., 2007).

Experiencing a live performance is distinct from listening to the radio (Brown & Knox, 2017). Live, musicians frequently improvise or alter their compositions, which can be difficult to capture on a recording. Additionally, the atmosphere at a live concert is unique. Multiple factors contributed to the audience's decision to attend a live concert as opposed to listening to recorded music.

Speechiness

The *speechiness* feature recognises spoken words in a track. According to research, music can be used to foster an environment conducive to learning (Lems, 2001). The spoken words in a recording could assist its subjects in expanding their vocabulary.

2.5 Lyrics

This section will describe two lyric segments. First, the effect of lyrics on eliciting emotion will be examined. On the basis of current research, various methods for assigning the proper value of valence and arousal to each word will then be described. Finally, this section will provide an answer to the first part of RQ2 described in section 1.3.

Lyrics can play a crucial role in determining a song's emotion (Y. Hu et al., 2009). Listeners will hear and comprehend the lyrics, which means that distinct emotions can be created by understanding what is being said. Consequently, identifying the emotion of lyrics can successfully aid in determining the sentiment of a song. In addition to the audio signal, the lyrics of a song are inherently meaningful and have a deep influence on the human perception of songs (Chen et al., 2016). Lyrics comprise words and phrases that describe the human situation. The situations expressed by lyrics typically consist of locations and events that trigger human memory to produce feelings that make it seem as if an event was just experienced. However,

the current state of the art is devoid of actual ways for detecting emotions based on music lyrics. This claim is based on the fact that different articles use different methods and don't always have good evidence.

Lyrics in music have been considered to serve multiple different purposes (Barradas & Sakka, 2021). In certain situations, lyrics can facilitate:

1. Exploring feelings, problems, and difficulties.
2. Influencing people's behaviour.
3. Helping overcome everyday problems.

Thus, when selecting music to listen to, listeners may look for lyrics that are deep, insightful, convincing, and emotionally charged.

The majority of research on the connection between music and emotion has concentrated on instrumental music (Zentner et al., 2008). However, research on the relationship between lyrics and emotions is scant (Mori & Iwanaga, 2014). The impact of lyrical content on the emotions evoked by songs should not be overlooked, given that the vast majority of music has words.

Stratton & Zalanowski (1994) revealed that when comparing emotional responses to music alone versus a combination of music and lyrics, the addition of lyrics to music had a greater impact on mood. Moreover, the authors hypothesised that when music and lyrics convey contradictory affective information, the lyrics will influence the induced mood more than the music. However, these results were not supported by Sousou (1997), who regarded the opposing findings as the result of a difference in stimulus selection. The results may indicate the response to emotionally loaded words in combination with uncertain-sounding music and unclear musical selection. The auditory stimuli employed in Sousou's study were, however, selected based on joyful and sad music, as determined by two independent expert musicians. Thus, Sousou believes that when the music is not emotionally ambiguous, the effect of the music is greater than the effect of the lyrics on emotion induction.

Ali & Peynircioğlu (2006) tested the effects of instrumental music versus music with lyrics on the perception of four emotions (happiness, sadness, calmness, and anger). When participants listened to sad music, the presence of lyrics increased their sense of sadness, and when they listened to angry music, the presence of lyrics increased their experience of anger. When happy or calm music was performed, however, the lyrics had no effect. These results imply that lyrics can affect how sad or angry a piece of music is perceived to be.

Even though a great deal of research has been conducted on emotion analysis, nearly all of these studies employ a one-dimensional model of emotions, such as positive-negative, which cannot be used to represent the feelings elicited by lyrics. This indicates that the intensity, or arousal, is disregarded. In this report's methodology, Russell's two-dimensional model of valence and arousal will be utilised.

2.5.1 Annotation

The method of assigning each word or sentence an appropriate valence and arousal value is flexible in contemporary research and could be the subject of multiple articles. On the basis of their rationale, the employed procedures may be questioned.

Y. Hu et al. (2009) examined the emotions in Chinese song lyrics. Despite the differing terminology, the approach could be effective for analysis. They utilised a vocabulary that was constructed from the state-of-the-art, Chinese-specific ANEW dictionary. This new lexicon is based on the translation of the emotional words from the ANEW dictionary. In order to prepare the lyrics for analysis, they employed a tool for natural language processing to perform word segmentation, part-of-speech annotation, and named entity recognition. After removing stop words, the remaining words of a phrase were analysed to determine if they appeared in the lexicon. Each word in the lexicon represents a unit of emotion. Once an adverb that modifies or negates an emotion term has been identified, it is included in the associated emotion units. Using an NLP tool, these modifiers were identified. The name of this NLP tool was not disclosed in the publication.

Calculating the emotion of an emotion unit is as follows:

- $\mathrm{v}u = \mathrm{v}\,\mathit{Word(u)}\ \mathrm{x}\ \mathrm{m}\mathit{Modifier(u),v}$
- $\mathrm{a}u = \mathrm{a}\,\mathit{Word(u)}\ \mathrm{x}\ \mathrm{m}\mathit{Modifier(u),a}$

In this formula, the valence and arousal values of the emotion units are shown by $\mathrm{v}\,\mathit{Word(u)}$ and $\mathrm{a}\,\mathit{Word(u)}$, respectively. These values were obtained by looking them up in a dictionary that had been compiled. The modifiers denote factors that represent the effect of the modifier on the valence and arousal of the emotion unit. Sentences without an emotion unit were eliminated.

Individual modifiers were gathered and assigned a value based on the polarity and intensity with which they influence the emotions of an emotion unit. Modifiers acquired modifying factors on valence and arousal. A negative modifier adverb was assigned a value between -1.5 and 0 and a positive

modifier adverb was assigned a value between 0 and 1.5. The paper did not make it clear how they valued polarity and influence.

In addition to proposing that the tense of a sentence influences the emotion, the authors accounted for this in their calculations. One of the examples they provided was that some sentences literally depict a happy life or recount romantic stories in one's memory, whereas the lyric actually convey the sentiment of longing for the happiness or romances of the past. This example appears to be more speculative and is unsupported by the literature.

Jamdar et al. (2015) used the following calculation:

- $\sum \frac{Valence(sentence)xWeight(sentence)}{Weight(total)}$
- $\sum \frac{Arousal(sentence)xWeight(sentence)}{Weight(total)}$
- where, weight(total) = Total weight of all sentences

Furthermore, the weight of a sentence is determined by whether it appears in a verse segment or a chorus segment. Chorus segments were allocated a greater value, which offered better outcomes. According to the study, it also led to the assumption that repeated words have stronger emotional power over the user.

The formula and the computation of the weight of a sentence are controversial. First of all, there is no explanation behind the weight of a sentence. By assigning chorus segments a higher weight, the results will be skewed and unreliable. There is also no explanation regarding the degree of increased weight of the sentences in the chorus passages. Furthermore, there was no literature used about the "concept that repeated words have stronger emotional power upon the user." This assumption prepared partly the way for the weight of each sentence.

Solitary words might have a different meaning than words in context. To guarantee that the right context of each sentence was taken into account, this paper applied a pair of association criteria while calculating the AV characteristics. These guidelines were founded on core linguistic principles of the English language and allow modifiers (verbs, adjectives, and negation words) to modify the valence and arousal levels.

According to the study, verbs can operate as modifiers for entire sentences, but an adjective acts as a modifier for a specific noun that it is related to. Negation words can affect the meaning of the verbs and adjectives themselves, which can further influence connected nouns or phrases.

The following principles of association were established for adjectives serving as local modifiers in a sentence:

- When two adjectives appear together, their effects are merged into a single adjective.
- When an adjective is immediately before a noun, it is connected with that noun.
- An adjective is paired with the preceding noun that occurs closest to it, unless that noun is already paired with another adjective.
- An adjective is related to the noun that follows it most closely. In this instance, if two adjectives compete for the same noun, the adjective closest to the noun is used.

Negation words, such as "not" and "never", are capable of inverting the meaning of the related verb or adjective. For example: "the boy was unhappy." The word "not" eliminates the high valence value of the word "happy." Verbs can also entirely alter the meaning of a statement. For example, "kill the happy rabbit." The word "kill" has a negative connotation, hence the statement should have a negative valence despite the positive connotation of the phrase "happy rabbit." The valence rating of "happy rabbit" should therefore be disregarded.

Çano & Morisio (2017) used a different method compared to Jamdar et al. (2015)'s study. Without using any association rules or calculations, they simply assigned each word a value depending on its dictionary value. Importantly, in their research, lyrics that were not part of the lexicon were disregarded, and lyrics containing fewer than 10 lexicon words were removed. The latter is debatable because it is difficult to specify where to draw the line. Ten-word-valued lyrics are taken into account, but one-word-valued lyrics are not. The line could potentially be deduced based on the proportion of words in the song.

Malheiro et al. (2016) conducted a similar study, preprocessing and annotating lyrics with valence and arousal values. However, participant annotation of the data was conducted. The subjects were required to read the songs and then identify the major emotion represented. The individuals were then required to assign valence and arousal ratings on a scale ranging from -4 to 4. Even though it is a scientific method for putting subjects in a position to assign a value to the lyrics, the scientists had to assign a value to the lyrics based on the available literature before putting subjects in a position to do so. This allowed them to compare the numbers and calculate the differences statistically. The article also showed how the individuals' familiarity with the songs could be a problem. If the lyrics appear familiar,

the listener may assign them a different value. However, only 13% of those polled claimed to be familiar with an average of 12% of the lyrics.

In an experiment, Chi et al. (2009) examined a dataset of lyrics and assigned subjects three distinct tasks. On a scale from -2 to 2, they were required to score the valence and arousal of lyrics, audio, and a combination of both. In estimating ratings for either valence or arousal, the model presented using features from both the lyrics text and the audio track performed better than models with characteristics from either the lyrics text alone or the audio track alone.

As extra information to increase the correctness of ANEW,Pambudi et al. (2018), attempted to determine the accurate definition of every word. In doing so, they calculated the average valence and arousal of a paragraph using the following formula:

- $\mathrm{v}text = \frac{\sum_{i=1}^{n} vifi}{\sum_{i=1}^{n} fi}$
- $\mathrm{a}text = \frac{\sum_{i=1}^{n} aifi}{\sum_{i=1}^{n} fi}$

In this instance, vi and ai represent the valence and arousal value of each word in ANEW. Fi indicates the frequency with which certain terms appear in the text. Notably, this technique was not established specifically for lyrics.

Based on the aforementioned literature, it may be stated that there is currently insufficient study on the correct and scientific annotation of valence and arousal. However, the most recent studies provide subtle recommendations on how to annotate the values. There is a distinct distinction between the investigations. Some studies employ their own techniques for coding valence and arousal (sometimes based on assumptions). Other studies determine the values for valence and arousal using volunteers. The values of valence and arousal will be tagged to the words using a combination of the approaches employed in the publications.

In summary, the association rules of Jamdar et al. (2015)'s paper will be applied because they could be fair and beneficial to assign a value to a piece of text considering its meaning. This work provides the strongest basis in linguistics in comparison to the other discussed publications. It makes sense that adjectives have an effect on the noun as they modify the meaning of the

words. These association rules are also partly stated in the paper of Y. Hu et al. (2009).

Furthermore, each phrase will be treated the same whether it is found in the "chorus" or "verse" segment compared with Jamdar et al. (2015)'s paper because treating a verse or a chorus in a separate way does not seem to be suitable based on the current literature. Last, sentences will be discarded if they do not contain words with emotional value.

2.5.2 ANEW dictionary

Providing words in lyrics with appropriate valence and arousal values remains a challenge at present. Numerous studies have examined the sentiment analysis of lyric content. The majority of the aforementioned publications employ an emotional vocabulary known as the "ANEW dictionary". This vocabulary contains English words that have been assigned arousal, valence, and dominance values. The ANEW dictionary, however, comprises just 2476 terms. This could imply that not every word in the lyrics could be assigned a value, resulting in inconsistent and untrustworthy outcomes. Consequently, additional researchers have been using WordNet to enlarge the NEW dictionary (Jamdar et al., 2015; Çano & Morisio, 2017). WordNet is a significantly larger and more general English dictionary than the ANEW dictionary (Miller, 1995).

It comprises almost 166.000 word pairings, indicating that the words have a semantic relationship. The dictionary is brimming with synonyms, antonyms, and other related terms. Creating a link between the ANEW lexicon and WordNet would increase the number of terms with valence and arousal levels. Once a word is present in both dictionaries, its synonyms may be assigned the same value as the word in the ANEW dictionary. Antonyms may be assigned opposing values.

2.6 Mood models

This section will present a literature study of three mood models in music research. A mood model is necessary for this research to classify songs into distinct mood groups. This will result in an examination of how popular music's mood has altered over the years, and so, what the recipients wish to listen to. The mood models will be debated and compared with each other, resulting in choosing the model that would be most suitable. The best suitable model needs to identify emotions based on available variables, so that by using the data, the songs can be categorised. The existent theoretical models can be clearly differentiated into two basic approaches: categorical and dimensional (Panda & Paiva, 2011).

2.7 Musical mood

Music can be viewed as a structured sound composed of various acoustic components. However, the primary reason that humans listen to music is the ability that it has to convey and evoke strong feelings and emotions in the listener (Salakka et al., 2021). Songs can be classified as "happy," "sad," "angry," or "relaxed," for example (Song et al., 2012).

In everyday life, music is most commonly utilised in emotional self-regulation (Saarikallio, 2011). This can be described as one of the core human abilities related to emotions. Emotional self-regulation refers to the process of modifying various aspects such as valence and intensity of emotions (Cole et al., 2004).

This research question will be the basis for the investigation into the change of emotion in a song over time. In addition, the emotional influence of music in various contexts will be examined, as this is a crucial factor in determining if a song has a positive or negative valence. Previous research has demonstrated that musical mood is linked to characteristics based on rhythm, timbre, spectrum, and lyrics (Jamdar et al., 2015). The study of these characteristics will reveal how the musical mood has evolved over time.

According to the research of Russell & Barrett (1999) emotions can be depicted as either a core affect or a prototypical emotional episode. The core affect can be characterised as conscious, accessible elemental processes of pleasure and activation, which are always present. This concept is also referred to as valence and arousal.

A prototypical emotional episode, on the other hand, refers to a complex process that develops over time. This includes sub-events, such as

antecedent, appraisal, and self-categorisation, that are interconnected. Although the two notions are similar, they are distinct in many respects. For instance, the intensity of core affective experiences can vary, although the intensity of prototypical emotional episodes cannot, and as previously indicated, humans are always experiencing core affect. There are no sub-events in music songs, as there are in prototypical emotional episodes, so the main emotional effect will be considered.

2.7.1 Thayer's model

Thayer's model can be viewed as a circumplex model of music, presented in 2.6, incorporating two variables: energy and stress (Seo & Huh, 2019; Thayer, 1990). Using two dimensions, this model can thus be considered dimensional. Thayer's model can be used to classify parts of music based on how much energy and stress they have, which is basically a way to classify mood.

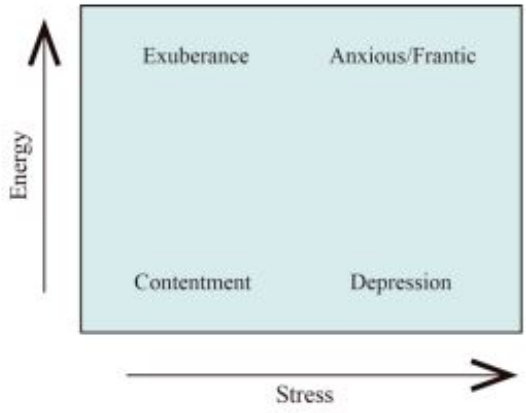


Figure 2.6: Thayer's model. The model represents emotional states with energy and stress. *Energy* refers to the volume or intensity of sound in music and *Stress* refers to the tonality and tempo of music. There are four mood clusters: calm-energy (e.g., Exuberance), calm-tiredness (e.g., Contentment), tense-energy (e.g., Frantic), and tense-tiredness (e.g., depression). Using Thayer's approach, it is possible to characterise musical passages based on the energy and stress dimensions (Seo & Huh, 2019).

2.7.2 Hevner's model

Hevner's model is the earliest and currently best known systematic attempt at constructing a music mood taxonomy in music psychology (X. Hu, 2010). It is defined by its mood clusters and is presented in Figure 2.7 (Laurier et al., 2009).

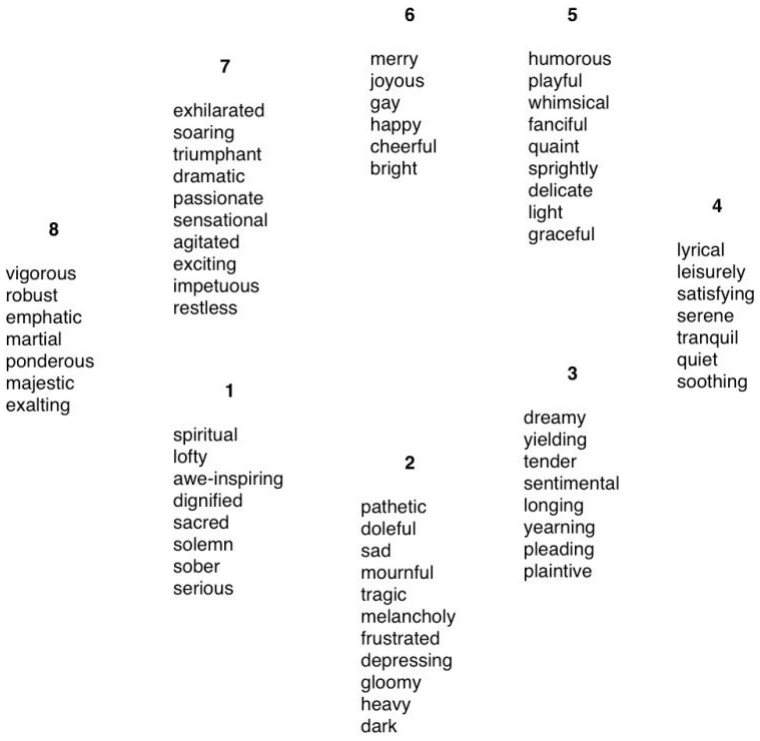


Figure 2.7: Hevner's model, a a circle of eight clusters comprising mood-related adjectives. It is a categorical model because emotions are structured into distinct categories. These clusters contain moods that are close in meaning. The meanings of adjacent clusters would only slightly alter. However, opposite clusters are their counterparts. Cluster two and cluster six, for example, comprise emotions such as happiness and sadness and can be called opposing mood clusters (Laurier et al., 2009).

2.7.3 Russell's model

The paradigm of Russell circumplex allows for the classification of emotions and mood (Russell, 1980). This model is frequently used for music emotion recognition, human emotion recognition, and other psychological investigations (Kim et al., 2010; Panda et al., 2018). In addition, the studies of Juslin & Sloboda (2001) and Laurier et al. (2009) support the usefulness of this two-dimensional model for representing emotions across a broad range of musical styles.

In addition to mood clusters, Downie et al. (2008) proposes mood clusters in their investigation of audio mood classification. The automatic recognition of emotions or moods in music is still in its infancy, despite increased interest in recent years. Based on valence and arousal/energy, Russell's two-dimensional model assigns emotions. This model is presented in Figure 2.8. Arousal is a subjective state of feeling activated or deactivated, whereas valence is a subjective feeling of pleasantness or unpleasantness.

Consequently, according to Russell's model, a song can be categorised along the following two axes:

- Songs that evoke positive emotions as opposed to those that evoke negative emotions
- Songs that are highly energetic as opposed to those that are lowly energetic

Even though Russell's model has been widely accepted in numerous researches, the model has its limitations (Van der Zwaag et al., 2011). Nowadays, it is still a topic of debate whether other dimensions should be included to capture emotions (Eerola & Vuoskoski, 2011). Several researchers proposed a third dimension which may be added to the model. Tension could be one of these dimensions, ranging from restless/under tension to calm/relaxed.

Tension was experienced by listeners while listening to music (Krumhansl, 1997). Different groups of subjects needed to rate a short music fragment on continuous scales of sadness, fear, happiness, and tension. An outcome of this study was, that subjects frequently experienced tension during listening, next to possible emotions such as sadness, fear and happiness. Tension can be experienced in music, for example, when an unexpected pitch is included in a chord (Steinbeis et al., 2006).

Also, music written for films often incorporates methods of producing tension to underline feelings of tension in the film (Van der Zwaag et al., 2011). These ideas make tension a third dimension of interest in music study.

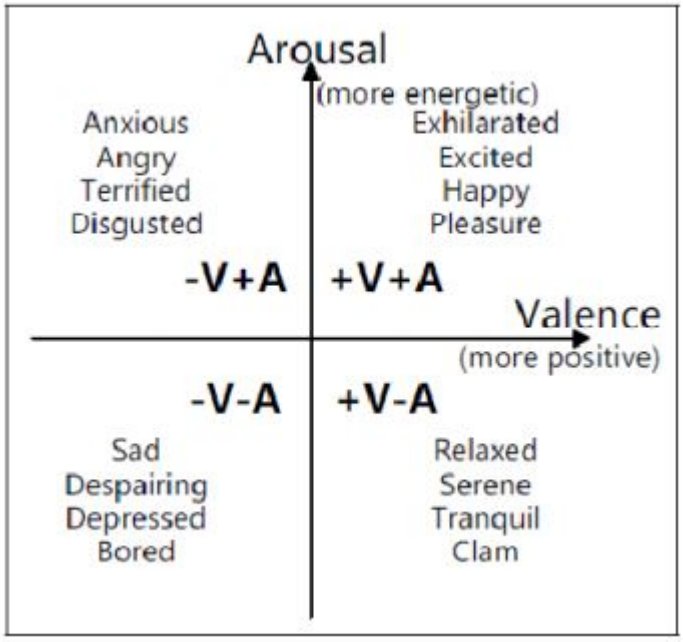


Figure 2.8: Russell's circumplex model of affect. In the model, emotions are distributed in a two-dimensional plane. The x-axis represents valence and the y-axis represents arousal. Arousal refers to a subjective state of feeling activated or deactivated, whereas valence refers to a subjective feeling of pleasantness or unpleasantness. Using the circumplex model, emotional states can be presented at any level of valence and arousal (Jamdar et al., 2015).

Another limitation of the model in music research is that some human emotions that occur in daily life are not represented in the model. Contempt or jealousy are emotions that are difficult to elicit through music (Juslin & Sloboda, 2011). Leaving emotions out of the model will put more emphasis on the emotions that are chosen.

The extent to which music may be able to convey mixed emotions is unclear in the model. It is unclear how to interpret the model when the subject is experiencing both happy and sad emotions simultaneously. Although a solution for the occurrence of mixed sentiments has been proposed, the method has not yet been widely applied in research (van den Broek & Westerink, 2009). The method implies separate scales/dimensions for positive and negative valence.

As it comprises variables that are already present in the dataset, Russell's model will be the most desirable model for this investigation. Due to the lack of mood data, Hevner's model, which utilises mood clusters, cannot be used. Russell's and Thayer's models are relatively similar, but Thayer's model employs stress rather than valence, which are distinct concepts.

Based on the values of "valence" and "energy", it will be possible to create a new categorical variable in the dataset titled "Emotion" by applying Russell's model to the data. The "Emotion" variable can take on the following values based on the quadrant's four categories:

1. Happy/Joyful
2. Turbulent/Angry
3. Chill/Peaceful
4. Sad/Depressing

In the subsections that follow, these four types of emotions will be described. Each type has different effects on humans. Music has a unique capacity to evoke a wide range of emotions compared to other art forms.

Happy/Joyful

In general, happy music is characterised by its medium intensity, medium timbre, very high pitch, and very high rhythm (Bhat et al., 2014). A fast tempo or rhythm is frequently associated with joy or anger. A piece of music with a high pitch connotes light, joyful, carefree, and humorous moods.

Turbulent/Angry

Turbulent/angry music is marked by its extreme intensity, very high timbre, low pitch, and very high beat (Bhat et al., 2014). As these traits are related to energy, it may be stated that it seems that turbulent or angry music has a high degree of energy.

Music and culture share a profound interaction with each other that has been left unexplored: music acts as a cultural product that documents changes in U.S. culture across time. People have an inherent need for social interaction. However, feelings of loneliness and social isolation increased by 250 percent in the United States between 1985 and 2004 (McPherson et al., 2006). Songs in America have also become increasingly self-focused.

Raskin & Shaw (1988) revealed that scores on the Narcissistic Personality Inventory were linked positively with the number of first-person singular pronouns and negatively with the number of first-person plural pronouns used. Another study also found evidence that there was a considerable increase in narcissistic tendencies across the generations (Twenge & Foster, 2010). Narcissism is related to the usage of self-focus terms. As indicated above, social connectedness is also declining over time. Both these ideas are associated to an increase in anger and antisocial behaviour (Bushman & Baumeister, 1998; Twenge & Campbell, 2003).

In popular American music, the usage of swear-and anti-social phrases did increase through time (DeWall et al., 2011). Thus, popular songs' lyrics have gotten increasingly aggressive and antagonistic throughout time. The study measured the percentage of words that appeared in lyrics. There was a definite increase in the use of first-person singular pronouns, while first-person plural pronouns declined. Changes in popular music's lyrics reflect a growth in narcissism in society, with musical lyrics becoming ever more self-focused throughout time.

Chill/Peaceful

Compared to other types of emotions, the acoustic characteristics of music that evokes calm and serenity are very distinct. It has a very low volume, timbre, pitch, and rhythm (Bhat et al., 2014).

Sad/Depressing

In daily life, people attempt to avoid experiencing negative emotions. However, many people derive immense pleasure from listening to sad music (Vuoskoski et al., 2011). Sadness is generally regarded as a negative emotion, and it is frequently a reaction to distressing and unfavourable circumstances (Sachs et al., 2015). "Why do people enjoy listening to music that evokes sadness?" is one of the most intriguing questions in the history of music. Human survival depends on avoiding painful and sad experiences. However, music is frequently used to alleviate mental suffering.

Many contemporary American pop songs contain sad-sounding motifs, which can evoke sadness (Schellenberg & von Scheve, 2012). Sadness is a complex bodily and neural state that results in low energy, social withdrawal, low self-worth, and a sense of a limited future horizon. There are two ways to describe sad music: objectively and subjectively.

On an objective level, sad music can be identified based on its acoustic characteristics. The characteristics of music that are commonly associated with melancholy are a lower overall pitch, a narrower pitch range, a slower tempo, the use of the minor mode, dull and dark timbres, softer and lower sound levels, legato articulation, and less energetic execution. Darker timbres tend to have a calming effect on human emotions (Bhat et al., 2014). In addition, the emotional content of sad music can be described in a two-dimensional space of valence and arousal. In accordance with Russel's model, it is defined as music with low valence and low arousal (Trost et al., 2012).

The subjective perspective is based on the listener's interpretation of the presumed feeling that the composer intended to communicate. The subjective classification of emotions is largely decided by asking participants which emotion they believe the song expresses or which emotion they are experiencing at that moment while listening to music. A song's lyrics can have a significant influence in classifying the mood it generates. Themes like sorrow and lost love can cause the listener's imagination to link the music with melancholy.

Recent advances in cognitive science and neuroscience have made it possible to investigate the relationship between perceived melancholy and positive affect in music. By examining how the brain reacts to music listening, aesthetic judgement, and the processing of emotions, it is possible to acquire a deeper understanding of how and why specific auditory stimuli elicit a pleasurable reaction (Sachs et al., 2015). The paradoxical notion that humans desire to minimise misery in their lives, yet find it enjoyable to listen to sad music, is known as the "tragedy paradox" (Xu et al., 2021). However, not everyone finds listening to melancholy music pleasurable. Several studies have documented the psychological consequences of individual differences.

Personality, mood, and the surrounding social context are significant determinants of a person's enjoyment of sad music (Sachs et al., 2015). Among the characteristics associated with the preference for sad music include absorption, as measured by the Tellegen Absorption Scale, and scores on subscales of the Interpersonal Reactivity Index, such as fantasy and sympathetic concern (Garrido & Schubert, 2011). According to the Big Five Model of personality traits, additional indicators include higher openness to experience and lower extraversion scores.

The current mood appears to play a factor in preferences for sad music as well, according to the following study (Hunter et al., 2011). Regarding the experiment of Hunter et al. (2011), they discovered individuals' enjoyment of sad-sounding music rose after creating sad moods in subjects. On the

contrary, Taruffi & Koelsch (2014) discovered in a slightly similar study that people prefer sad music when they are sad. However, others preferred to listen to joyful music. This suggests that individuality in the liking of sad music is still important, which needs to be taken into consideration in sad music recommendation systems.

The last crucial factor in picking and liking sad music is the current situation of the individual. According to the research of Taruffi & Koelsch (2014), listening to sad music can be done for multiple reasons and is grouped in Table 2.1.

Table 2.1: Situations in which individuals engage with sad music

Situation category	***Description***	***Function***
Emotional distress	Death, failure, frustration, love-sickness	Emotional: mood enhancement
Social	Homesickness, feeling lonely	Social and emotional: consolation due to mood-sharing and context
Memory	Retrieving memories of valued past events	Memory trigger
Relaxation and arousal	Relaxing, getting new energy, quieting down	Emotional: mood and arousal regulation
Nature	Travelling	As a reflection of the environment
Musical features	Engaging for its musical features	Aesthetic
Introspection	Contemplating, reappraising personal experiences	Cognitive: improve personal introspection
Background	Parallel activity: driving, reading, working	Pleasant background music
Fantasy	Creative thinking, looking for inspiration	Cognitive: engage creative thinking
Intense emotion	Seeking a touching emotional experience	Emotional: experience intense emotions

Continued on next page

Table 2.1 – continued from previous page

Positive mood	When being in a positive mood or emotional state	Emotional: mood control
Cognitive	Improving rational thinking, obtaining a better focus	Cognitive: engage rational thinking

People who took part in this study indicated that when they were feeling down, they listened to sad music (470 nominations, compared to 184 or fewer nominations for the other categories).

According to the findings of Napier & Shamir (2018), who analysed the Billboard Hot 100, there was a slight increase in the popularity of sad music in the late 1980s, which peaked in the early 21st century.

The conclusion of Schellenberg & von Scheve (2012) emotional's study is that music has become more melancholy over time. Primarily as a result of a decrease in tempo and an increase in the use of the minor mode.

Multiple studies demonstrated that, song lyrics, like language can express and elicit emotions (Juslin & Sloboda, 2001; Ali & Peynircioğlu, 2006). It appears that lyrics can influence the overall emotional valence of emotions, making it easier for music to convey positive emotions when they are present or absent.

DeWall et al. (2011) studied how lyrics alter the induction of emotions by music over time. Validating their premise, song lyrics have gotten worse throughout time. The year and the number of words related to pleasant emotions were negatively associated, obviously indicating that lyrics contain a lower percentage of positive emotion phrases throughout time.

2.8 Historical background - Billboard

The Billboard Hot 100 used to be a magazine in the past (Giles, 2007). Today, Billboard offers a variety of popular music charts for North America. Currently, a new chart will be posted on their website every Saturday. The "Hot 100" chart, which comprises the one hundred most popular songs of a given week and is updated weekly, is their most famous. Billboard also publishes the weekly "Billboard 200", "Artist 100", "Songs Of The Summer", "Billboard Global 200", and "Billboard Global Excl. US" charts. Billboard also publishes their "Year-End Hot 100 Songs" at the end of each year, which

comprises the one hundred most popular songs played throughout the previous year. Billboard then publishes annual charts of the top 200 albums and box scores.

The magazine *Billboard* was founded in 1894 by William H. Donaldson and James H. Hennegan of Cincinnati (Ohio) with the intention of becoming the leading trade publication for the advertising business (Kwame Harrison & Arthur, 2011; Anand, 2005). Despite many alterations over the past 127 years, its original essence has remained unchanged. Billboard is still an industry-focused journal that is consumed by stakeholders in all sectors of the music business. Billboard is in the business of physical billboard advertising, live performances, recording music, and, broadcasting radio and television. Briefly, Billboard is the preeminent trade publication for the music industry (Kwame Harrison & Arthur, 2011).

In its early years, Billboard swiftly expanded into a firm that documents, promotes, and disseminates information about touring musical and theatrical performances. Beginning in the early 20th century, the entertainment business developed rapidly. This brought a change at the Billboard company. From publishing the routes of touring performers to chronicling the booming recorded music industry, Billboard has become an indispensable business resource for individuals with a financial stake in the music industry. In 1955, Billboard created the "Top 100" singles chart, which remains one of Billboard's greatest ideas under a new name today. Billboard's editor-in-chief, Tom Nonnan, renamed this chart "Hot 100" in 1958, after competitors had copied its format and moniker and trademarked the brand. Now, the business is more focused on album charts than singles charts. This is due to the fact that recording firms typically lose money on singles. Radio airplay is primarily a promotional tool (Anand, 2005). As revenues are generated mostly through album sales, the album chart can be viewed as an indicator of a record label's financial performance.

Billboard has been publishing weekly for nearly a century, informing the music industry of all major events. This has proven vital to the development of the commercial music industry. Its news, information, gossip, advertising, opinion, and music charts keep the music industry abreast of all recent and forthcoming happenings.

Billboard magazine has influenced the evolution of the commercial music industry in two significant ways. First, Billboard played a crucial role in connecting the commercial music industry. By providing vital market activity information to its customers, Billboard helped the industry connect and grow. Second, by releasing a well-known music chart every week, it established the path for how field participants hear music and how they

respond to the commercial music market. Participants in the market have grown to rely on the charts in order to determine what is popular with their audience. Anand (2005) defines the Billboard chart as a "frequently updated chart that informs about market activity provided by an independent source, published in a predictable style with consistent frequency and available to all interested parties for a modest fee."

It is essential to note how Billboard ranks their music. According to (McAuslan & Waung, 2018; Haampland, 2017), the "Hot 100" is determined by radio airplay, audience impressions as measured by Nielsen BDS, retail and digital sales, and streaming from online music sources like Spotify. Nielsen BDS is a service that tracks monitored radio, television, and internet play of songs by counting spins and detections. Paid digital downloads were factored into the Billboard Hot 100 chart at the start of 2005. As technology and formats have advanced, new measurement parameters, such as streaming data and on-demand services, have been included beginning in 2007. In addition, YouTube views have been factored into the ranking of songs since 2013, when they became an additional factor. This contentious choice had an immediate effect (Kellogg, 2013). The number-one song on the chart was "Harlem Shake" by producer Baaur, a viral success. Considering that only 262,000 copies of the record were sold that week, this song would not ordinarily rank at the top. Including YouTube views, however, this single rocketed to the top of the chart. In anticipation of subsequent graphs, it remains to be seen what this decision's long-term effects will be. Historically, before online services played a significant role, Billboard's chart rankings were determined by two key elements that could be measured: radio play and record sales. Thus, the Billboard ranking underwent a transformation: from a system based solely on radio play and record sales to a system that includes streaming data and other metrics. A brief summary of the historical changes in ranking indicators can be created using the Nishina (2017) and the McAuslan & Waung (2018).

The year 1991, marked the introduction of the first modification to Billboard's ranking methodology. Actual point-of-sale data from *SoundScan* will be utilised in determining the Top Album and Hot 100 Singles chart rankings(Kellogg, 2013). This relationship resulted in an instant boost in the sales, quantity, and programming frequency of urban music on top 40 radio. Michael Shalett and Michael Fine founded SoundScan in 1991 as a computerised music retail sales monitoring data resource that validates each sale when the bar-code of an album or single is scanned at retail outlets. This modification to the approach improved the accuracy, fairness, and data-driven nature of the ranking. Table 2.2 provides a historical summary

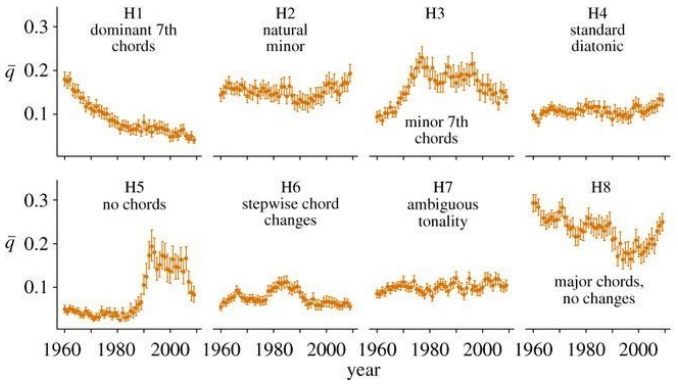
of the ranking's fluctuations.

Table 2.2: Historical ranking metrics Billboard

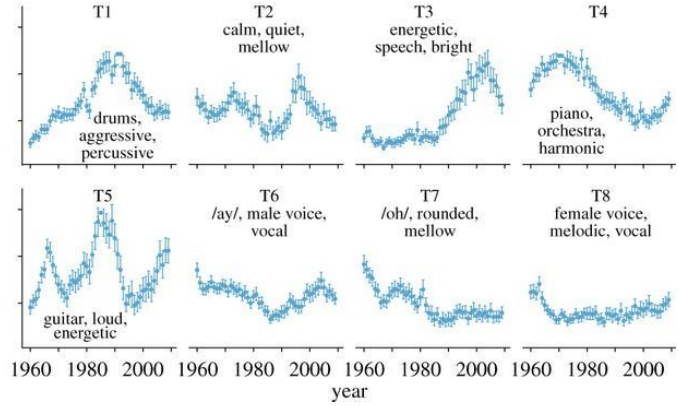
Year(s)	***Description***
1958 - 1991	Ranking based on singles sales and radio airplay.
1991	Billboard began collecting sales data digitally (using Nielsen SoundScan) in 1991, allowing for faster and more precise charts.
1998	Billboard eliminated the requirement that a song be released as a single in order to be on its charts.
2005	Billboard incorporated iTunes digital downloads into their ranking.
2007	Streaming data and on-demand services were added (Spotify).
2013	The number of music video views on YouTube in 2013 is contained.

2.9 Musical Evolution

Philosophers, sociologists, and journalists have long argued about the history of music. However, robust assessments of unambiguous hypotheses grounded in quantitative facts and statistics are lacking (Mauch et al., 2015). The absence of data was also a significant obstacle to music study. Recently, this has altered with the introduction of extensive digitalised archives of audio recordings, musical scores, and lyrics. In this manner, music can be quantified. Mauch et al. (2015) analysed the harmonic and timbral qualities of 17000 records that appeared on the Billboard Hot 100 charts between 1960 and 2010 in an effort to demonstrate quantitative trends. In addition, they investigated the emergence of musical variety. This corpus of digitised music was utilised to study the history and development of American popular music. Figure 2.9 depicts the evolution of their data's themes.



(a) Changes in harmony are shown by the yellow lines.



(b) Changes in timbre are shown by the blue lines.

Figure 2.9: Evolution of musical topics in the Billboard Hot 100 over the years (Mauch et al., 2015).

Both the harmonic and timbral charts can reflect particular music genres because these musical qualities can define a genre. *H1*, for instance, illustrates the use of dominant-seventh chords, which are frequently employed in jazz music to build tensions that are later resolved by consonant chords. Additionally, dominant seventh chords are utilised in blues music. Numerous blues and jazz-tagged tracks have a high frequency of H1, indicating a quick drop of these topics. Another musical style, H3, employs minor-seventh chords, which may be traced back to funk, disco, and soul music for their use of harmonic colour. In the first decade, the mean frequency of this style virtually doubled. Among all the themes, H5 is the only one devoid of chords, which seems extremely odd given that songs should have easily identifiable chords. Nonetheless, this style conveys their absence and

represents hip hop and rap. The frequency of these genres increased in the late 1980s.

The timbral charts appear to be more dynamic than the harmonic charts, as many of them display quick up-and-down motions. T3 is the timbral descriptor for "energetic, speech, and bright." It exhibits the same characteristics as H5 and may also be associated with the growth of genres associated with hip-hop. Instruments such as the guitar and the piano are described using timbral terms such as T1, T4, and T5. It appears that these timbral subjects rise and fall periodically, indicating a periodic repetition of instrumentation. The rise of T1 can be explained by the introduction of the drum machine, a new percussion technology. The drum machine's popularity rose steadily until 1990, when it reached its peak. After 1990, the incidence of T1 decreased, indicating the end of the drum machine's rule.

Chapter 3

Data and Method

Regarding the data and the approach, this part will describe what method was used to analyse the data. In addition, it will be revealed precisely what data has been selected and what adjustments have been made. Because they need a separate methodology, a clear separation has been drawn between Spotify's audio features and its lyrics.

In the realm of music data mining, data is being gathered and amassed at an accelerated rate. There are two plausible techniques for mining musical data. Either the more conventional method, Knowledge Discovery in Databases, or the methodology presented in the book by T. Li et al. (2011) could be implemented. In this part, the two frameworks will be described and contrasted. Knowledge Discovery in Databases (abbreviated, KDD) refers to the process of discovering knowledge in data and emphasises the high-level application of certain data mining methodologies (Fayyad et al., 1996). The interpretation of data after the execution of a set of actions is one of the most crucial components of this. KDD in databases is the nontrivial process of detecting legitimate, innovative, possibly helpful, and ultimately intelligible patterns in data, as mentioned in the work cited in Fayyad et al. (1996). This framework consists of multiple phases, employs data mining techniques for pattern identification, and attempts to address one of the most pressing issues of the digital information age: "data overload." Historically, data analysis was conducted in a more conventional manner. It was common practise in the healthcare industry for medical specialists to manually analyse trends and changes in healthcare data on a regular basis. However, this way of manually analysing a dataset is time-consuming, subjective, and costly. Due to the increase in data, this form of manual

analysis is becoming difficult in many fields, including healthcare and business. The KDD technique attempts to resolve this issue. The name "KDD" was coined at the KDD workshop in 1989 to emphasise that knowledge is the ultimate result of data-driven discovery, as stated in Piatetsky-Shapiro (1990)'s paper. According to the work by Fayyad et al. (1996), KDD refers to the general process of discovering knowledge from data, whereas data mining refers to a specific phase inside this process. The ultimate objective of the KDD framework is to provide tools to automate as much as possible the entire data analysis process and the statistician's art of hypothesis selection. All KDD processes are displayed in the preceding diagram. In the following section, all of these procedures will be described in relation to this report (Mariscal et al., 2010; Fayyad et al., 1996).

3.1 Phases of the KDD model

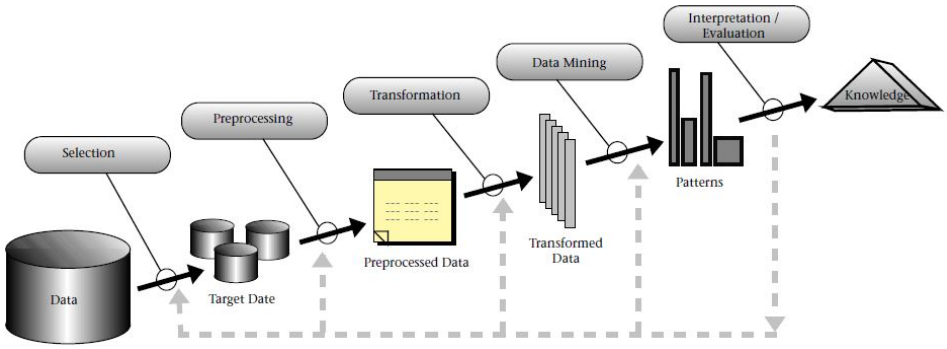


Figure 3.1: KDD framework for data mining. It contains multiple phases. Each phase of the KDD process includes input and output. This input and output will be modified by a command, such as "Selection" (Rodriguez & Dolado, 2018).

Figure 3.1 presents an overview of the steps constituting the KDD process. In this part, the input and output will be discussed in conjunction with their respective actions.

3.2 Goal-setting and understanding

This framework begins with data, which is a collection of facts gathered from databases, for instance. These facts could be any type of information. The

most significant components of this phase are building an understanding of the data and establishing objectives for what you wish to accomplish with the data. This will ultimately aid in the discovery of answers to the research questions. The following objectives are formulated:

- To provide a comprehensive understanding of the data, time series analysis of the audio features will be performed. This will be performed to investigate the evolution through time.
- To assign each song a new valence and arousal value (using lyrics) and then compare these newly determined values with Spotify's features. Words with matching entries in the ANEW lexicon are annotated with new valence and arousal levels. Comparing these Figures will aid in determining the algorithm's dependability. Additionally, this may be a way to improve Spotify's algorithm.

3.3 Data selection and integration

The subsequent step involves the creation of a target dataset, which will be chosen beforehand. On this particular subset of the data set, discovery will be conducted. It is also fairly typical to select a subset of variables from the dataset, as opposed to focusing on the entirety of the variables. Selecting your data should be done with care because it is difficult to modify as you progress through the framework's later stages. Changing your choice in the last step of the framework will take a lot of time because you have to go through all the steps again.

RStudio will be utilised to develop and analyse the dataset. The Integrated Development Environment (IDE) for the computer language "R" is RStudio (Allaire, 2012).

Regarding the collection of data, RStudio will use and get Spotify audio lists. These Spotify playlists comprise around one hundred songs from the years 1956 through 2020. Due to the fact that these tracklists are compiled by individuals, each list must be validated. This procedure will be carried out by scraping the charts from the Billboard website. This makes it possible to check that the Spotify lists are complete and contain the right music.

Using the *rvest* package in RStudio, all billboard.com pages containing tracks from the previously mentioned years were extracted and compared to the Spotify playlists. This package includes several functions aimed at accomplishing web scraping (citeprvest). After a comparison, no errors were

discovered, indicating that the Spotify lists are nearly ready for *data preparation* and *data cleaning*.

Next, the *spotifyr* package is necessary to import Spotify playlists into RStudio (Thompson et al., n.d.). This package provides an R wrapper for retrieving track audio features and other data from Spotify's Web API in bulk. In computer programming, a wrapper calls another function, in this example, the Spotify Web API, which performs the actual job. Thus, RStudio encapsulates the additional software component, the Spotify Web API. The primary function of the wrapper is to provide an alternative method of using the wrapped object. Using the *spotifyr* package, it is now possible to access the Spotify Web API from within the R environment. The Spotify Web API communicates only with the R wrapper, which transmits commands to the wrapped programme and delivers results. In this scenario, the only component that interacts directly with both programme components is the R wrapper. Thus, it was possible to retrieve and extract playlists via an API request that utilised automatic batching. In order to use the Spotify Web API, a developer account was required to obtain a Client ID and Client Secret. These credentials are required in order to access an access token. Thus, the R environment can establish a connection with the Spotify Web API.

The entirety of the collection includes publications from 1956 to 2020. Each year has one hundred songs. This time period was chosen since year-end charts prior to 1956 only contained thirty songs, as opposed to one hundred. This could lead to distorted findings as the number of songs released in these years varies. Thus, the result would be skewed and erroneous. The data frame contains a total of 6481 songs. Some tracks were not available on Spotify, as shown by their greyed-out status. This prevented the Spotify API from using these tracks, resulting in the omission of some songs from the dataset. These may result from one of the above causes (Lee, n.d.):

1. **Country restrictions**: It is possible that certain songs are unavailable in specific places of the world due to "country restrictions." This may be determined by whoever holds the song's copyright (often the label that created the song or the album), but the song's availability may alter over time.

2. **License expirations**: Spotify must negotiate a licence agreement with music labels and other copyright holders so that artists' songs can be streamed on its platform. These licensing agreements have defined expiration dates; if they are not renewed, the songs cannot be

streamed. Unfortunately, Spotify has no control over this as it is the responsibility of individual music companies.

3. **Network issues**: Although unlikely, it is possible that a network connection issue could cause music to be unavailable.

To utilise the Spotify Web API, however, an application must be created on Spotify's developer platform (Sciandra & Spera, 2020). Then, the application was given a name, the reason for its creation was described, and agreement on the terms of condition was requested. Following this approach, the 'Client ID' (username) and the 'Client Secret' (password) were generated, which are required to receive the access token in order to link Spotify with the R environment. The Spotify Web API enables users to extract certain audio characteristics from songs. In addition, the user can generate data, such as recently played songs and all-time favourite musicians. Additionally, you can add songs to your playlists and produce new releases.

Regardless, the *spotifyr* package supports a function named "get_playlists _audio_features" that imports automatically the audio features of the tracks. Thus, the data was collected and prepared for the following phase. These auditory characteristics are the variables outlined in Table A.1.

Now that the data set's foundational elements (artist, title and audio attributes) have been collected, the dataset can be completed. Next, lyrics were retrieved via the *lyrics.ovh* API. Using a looping request issued to the API, a data frame containing song lyrics was generated and saved. This situation requires the use of the *httr* package, which provides the functions to request lyrics from the API.

3.4 Data cleaning

The framework's third phase covers data cleansing and preprocessing. This phase focuses on detecting inconsistencies in the data set (Rahm & Do, 2000). RGetting rid of these so-called "errors" in the data will improve the data. In this step, however, it is crucial to formulate a plan for addressing these inconsistencies. To preserve data quality, you must determine how to manage missing data fields (Fayyad et al., 1996). Using multiple R packages, inconsistencies in the data will be identified. Regarding missing values, only two songs in the data contained audio features with the value "Not Available" (NA), which indicates that these rows lack data. All rows have been inspected for outliers and values that are inapplicable to the dataset,

such as valence values that are larger than one or less than zero. Fortunately, only a few inconsistencies were found in the dataset.

3.5 Data preparation/Data transformation

After data has been preprocessed and cleansed, it can be altered and readied for data mining techniques. This phenomenon is also known as data wrangling among data scientists, which is technically described as an iterative process of data exploration and transformation that permits analysis (Kandel et al., 2011). This phase's objective is to use various data reduction and transformation techniques to the dataset (Shafique & Qaiser, 2014; Lara et al., 2014). Data preparation is an important phase in this architecture for the following reasons:

- Real-world data may be corrupted, for example, it may be insufficient, loud, or inconsistent. This could result in unworkable models.
- Data preparation can decrease the dimensionality of the original data and increase the efficiency of data mining techniques.
- Preparing data yields high-quality information that can be utilised to construct useful and representative models.

This phase employs a variety of data analysis techniques capable of enhancing data quality, allowing knowledge extraction procedures to ultimately produce more and better information (Lara et al., 2014).

As the original dataset comprised numerous meaningless variables, such as the playlist_id, playlist_img, track album type etc., these variables have been eliminated. In order to remove a nested variable (a list-column of a data frame) from the dataset, it has to be denested. The "playlist name" field included the song's release year. These years had to be separated from the "playlist name" column in order to analyse solely the year of the song. Some columns were renamed or modified to enhance the overall data consistency. Previously, the track duration column was presented in milliseconds; now, it is displayed in seconds, which is more legible for the reader.

In addition to studying the retrieved audio elements, the focus of this research will be on the use of lyrics. There are various advantages to studying lyrics over studying common audio. First of all, music is typically copyrighted and prohibited, as opposed to lyrics, which may be retrieved without restriction from the Internet (Çano & Morisio, 2017). Moreover, lyrics are rich in high-level semantic characteristics. Audio, on the other hand,

possesses only low-level semantic characteristics, resulting in the so-called "semantic gap" (Celma, 2006). This semantic gap is the difference between low-level descriptors and the concepts used by music listeners to relate to music collections. Regardless, the fundamentals of music are low-level descriptors, which are signal characteristics such as *energy* and *frequency*. Content objects are concepts that music listeners link most frequently to music. These items' lyrics are more profound than their signal aspects.

However, song lyrics differ from other text documents in that they provide unique challenges and must be formatted in a certain manner. They are frequently brief and have a limited vocabulary. Moreover, metaphorical language in lyrics might result in ambiguity, making mood assessment difficult.

The *lyrics.ovh* API was utilised to retrieve lyrics. The core request includes a link consisting of the base URL of the API, the artist's name, and the song's name. The acquired artist and song names were preprocessed in order to improve their quality in the following areas:

- Elimination of text unrelated to the song name (e.g. words such as "Remastered" and "Version")
- Elimination of punctuation to recover a cleaned link
- Elimination of years at the end of a song's name

Once the lyrics have been extracted, the second phase of data preparation may commence. Many lyrics contain ambiguous words. Additionally, repetitive textual patterns must be avoided in the lyrics. Notably, the link to retrieve the lyrics from the API did not function for all songs. This indicates that the size of the lyric sample is smaller than that of the Spotify sample. Thus, the lyrics were prepared for analysis by performing the following tasks:

- Elimination of text unrelated to lyrics (e.g., names of the artists, composers, instruments)
- Elimination of frequent textual patterns in lyrics such as [Chorus x2], [Vers 1x2]
- Elimination of punctuation in order to retain only words
- Elimination of English stop words.

The ANEW (Affective Norms for English Words) vocabulary was utilised to supply all the words in the lyrics with a relevant valence and arousal score (Bradley & Lang, 1999). Çano & Morisio (2017) and Jamdar et al. (2015) are examples of studies that utilised this dictionary for the sentiment analysis of lyrics. This lexicon gives a standard set of emotional ratings for 2,471 unique English words. During a psycho-linguistic experiment, respondents judged them based on their *Valence*, *Arousal* and *Dominance* qualities. Because *Dominance* was not used in this study, it was omitted from the data.

This lexicon was augmented by the addition of the WordNet dictionary. WordNet offers synonyms for ANEW-lexicon entries. All of the words in the ANEW dictionary have been rendered through the WordNet lexicon in order to annotate WordNet synonyms with the same valence and arousal levels as the ANEW words.

In addition, for the words in the songs to possess valence and arousal, they must be an exact match with the words in the lexicon. This means that if a word in the lyrics is in its plural form and the dictionary only has its singular version, the word will not be assigned a value.

Part-of-speech tagging has been employed to circumvent this issue. POS tagging identifies the form of each word in a phrase (Goh et al., 2022). This report uses nouns, verbs, adjectives, and adverbs, among other types. This was established in RStudio using the *udpipe* package, which is designed specifically for NLP-related applications such as POS. The ability to change singular nouns to plural nouns and vice versa was made feasible by the linguistic norms of the English language. The modified words retained the same valence and arousal levels as their initial form. In addition to being changed and added to the lexicon in the forms of its present participle (root + -ing) and past tense, the words that were labelled as *verbs* were also given the forms of their present participle (root + -ing) and past tense. Also, *adverbs* were converted to *adjectives* and vice versa. Once a word appeared several times in the dictionary, the average valence and arousal values were calculated and assigned to the word, with duplicates removed.

After all changes were made, the lexicon had 28,692 different terms, whereas it originally contained only 2,472 words.

Where possible, each word in the lyrics was assigned a value based on the valence and arousal values found in the dictionary. This was accomplished by employing the association rules described at the end of the lyrics in the literature section. On the basis of the number of words and their values, a song's average valence and arousal level were calculated. So, words that weren't in the lexicon weren't taken into account, and they weren't given a value.

3.6 Data mining

Next, various data mining techniques were used, such as classification, regression, and clustering, to search for patterns in the data (Olson & Delen, 2008). Data mining plays a crucial role in the KDD procedure, employing specific algorithms to extract valuable knowledge or intriguing patterns from the (subset of) dataset. This phase consists of exploratory data analysis utilising the selected data mining algorithms and maybe additional ways to identify data patterns. According to the findings of Hand (2007), data mining can be separated into two primary tool classes: model construction and pattern finding. Model construction can be regarded as a high-level global descriptive description of datasets, which in modern statistics includes regression models, cluster decomposition, and Bayesian networks. These models will describe the data's overall structure. According to Wong & Wang (2003), pattern discovery can be stated as an optimisation problem that recursively splits the sample space for the optimal set of relevant events.

Mood and emotion classification is one of the data mining techniques that have been employed. This technique will detect the emotional significance of a song using the valence and energy levels of Spotify's algorithm. Next, *text mining* was applied to the lyrics in order to assign each word a suitable value.

Several machine learning algorithms have also been used to predict the emotional category of songs based on their lyrics and newly created attributes.

3.7 Knowledge discovery

The final step is determining how to use the newly acquired knowledge. This can be interpreted in a variety of ways, including utilising the knowledge directly, merging it into another system for additional analysis, or merely documenting it and communicating it to relevant parties/stakeholders. After the information has been correctly interpreted, the discovered data must be evaluated. This involves identifying and resolving any potential contradictions with previously held beliefs or extracted information. The outcomes will be analysed and debated. In addition, a section with suggestions for further study on this topic will be created.

3.8 Music data mining

In addition to the more general KDD framework, the book T. Li et al. (2011)proposes a second framework for particular music data mining tasks. This book focuses on a four-concept data mining approach. According to the work of T. Li et al. (2011), music data mining is a broad study field that includes approaches for a variety of applications, such as genre classification, artist/singer identification, mood/emotion detection, instrumentation recognition, and music summarization, among others. As it is not yet widely adopted, it was decided to use the KDD approach instead. Appendix B contains a comprehensive discussion of the procedure.

3.9 Experiment setup

A third (external) party was enlisted to generate a more accurate comparison to the Spotify features. The survey was completed by 48 participants. None of them had hearing issues or other known disabilities that prevented them from participating in the survey. To prevent bias, participants were not informed of the objectives of the study. The goal of this survey is to measure the affective disposition triggered by songs judged by a listening panel, i.e. [**RQ4**].

An online survey was created with the Qualtrics platform, a survey tool offered by Utrecht University (Mathur & Reichling, 2019).

The participants were first given a concise introduction to the research. They were then required to fill out a consent form. Even though no sensitive information was collected or retained, they were required to provide their country of origin and gender. The concepts *valence* and *arousal* were then described. Participants were instructed to evaluate 46 audio snippets. It took approximately thirty minutes to complete the survey.

As the dataset spans the years 1956 to 2020, it has been separated into six decades. The years 1956 - 1959 have been added to the decade 1960, while the year 2020 has been added to the decade 2010. This was done to restrict the number of decades and, consequently, the quantity of audio pieces. Two songs from each decade's emotional categories were selected. It was determined that the "most extreme" songs in each category would best illustrate the disparities between the categories. Songs that have the shortest distance to the corners of Russell's circumplex model of affect are characterised as "extreme."

The years 1956-1959 and 1960 did not feature any songs that fell into

the "turbulent/angry" emotional category, resulting in 46 audio fragments instead of 48.

Each audio piece contained the refrain, as this is a recurring element of the song. It took between 20 and 30 seconds to listen to. Audacity, a free, open-source audio capture and editing programme, was used to record and modify the audio snippets (Audacity, 2014). In addition, a song's introduction could only feature an instrumental sound. Excluding vocals from audio fragments could affect participants' responses, as it has been demonstrated that lyrics play a significant role in eliciting emotions in people.

The subjects were instructed to assign valence and arousal ratings to each song using a digital rating system. Using radio buttons, this was displayed on a Likert scale ranging from 0 to 10 points. This scale offers the versatility of a central point. Comparing the Spotify features to the findings from the lyrics is also beneficial. Both were graded on a scale of zero to one. This will allow the scores to be normalised/scaled without any obstacles.

The scales 0, 5, and 10 were augmented with three Self-Assessment Mannequin (SAM) images (Van den Broek, 2013; Bradley & Lang, 1994). With these visuals, the level of valence and arousal experienced by the participants may be measured. A screenshot of these augmentations is shown in Figure 3.2. Additional screenshots of the survey can be found in Appendix C.

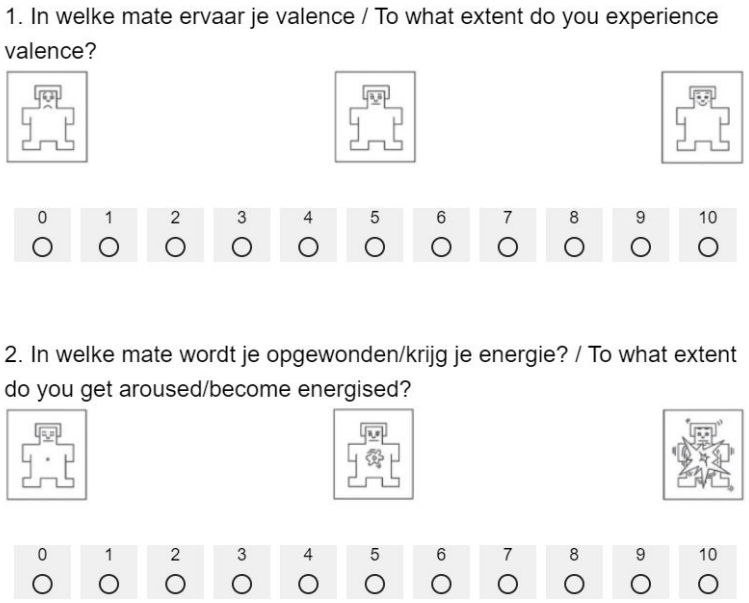


Figure 3.2: A screenshot of the digital rating system used in this survey. It contains an 11-point (0-10) Likert scale with radio buttons augmented with three Self Assessment Mannequin images. With these images, the experienced valence and arousal were assessed. The visuals indicated the minimum, median, and maximum values connected with each point. They assisted respondents in forming an intuitive and validated subjective evaluation of their emotions.

The individual was then asked to annotate the words that evoked strong emotions. They could enter multiple words as a response. This inquiry was posed to determine which words elicited an emotional response from the individual. To aid the subject in annotating the words, the matching audio fragments' lyrics have been provided.

Validation questions comprised the final two questions. Asking a person if they are familiar with a song and if they like it can help discover possible trends and biases.

Except for the third question, all the questions were obligatory (annotating specific words that evoked strong feelings). It is possible that none of the lyrics created any emotion in the subject. However, if the user neglected to react, they were requested to do so before proceeding to the next question. They were not required to leave a message.

Participants were instructed to conduct the survey in a calm setting so that they could listen to the audio samples with the greatest clarity. Because the first impression is so important in emotion research, there were no back buttons on the questionnaire.

Chapter 4

Results

The first section of this chapter will provide an overview of the data. Delving deeper into the audio features and metrics of the dataset will provide a thorough understanding of the available data.

Next, the emotion labels from the Spotify dataset will be discussed. These emotion descriptors may be time dependent, thus they will be investigated throughout time.

The third section will present a decomposition approach for creating new values for *valence* and *arousal* based on the audio features from the audio signal input. This will generate emotion labels exclusively based on audio signal features. These labels will be compared to Spotify's labels, obtaining an answer to [**RQ1**].

Following that, the fourth section will provide a detailed examination of the lyrics from the Billboard chart. To begin, the *valence* and *arousal* values will be compared throughout time. Violin plots will be utilised to provide further insights. Again, the emotion labels generated for the lyrics will be compared to Spotify's labels using Russell's circumplex model of affect. This will provide an answer to [**RQ2**]. A subset of this section will employ machine learning classification algorithms. These will be applied to develop emotional prediction models. These models will be based on lexical features.

The fifth section will combine the *valence* and *arousal* values calculated from the audio signal features and the lyrics. These values will be used to classify emotion labels. These labels will be matched to Spotify's labels once more, acquiring in an answer to [**RQ3**].

Concluding, the last section will discuss the survey results. Participants rated certain songs on valence and arousal. Based on these findings, Spotify's

emotion labels will be analysed and compared to the survey results, resulting in an answer to [**RQ4**].

4.1 Data analysis

4.1.1 Overview Data

There are nine audio features extracted from Spotify. In this report, *valence* and *arousal* will count as metrics. The other seven audio features will be categorised as audio signal features. This differentiation was made as *valence* and *arousal* are computed based on other features. Machine learning models were also used to calculate these two features. Supplementary information of all the features can be found in Tables A.1 and A.2. The audio features are determined for 6481 songs. Based on these features, the graphs in 4.1, 4.2, and, 4.3 are generated.

1. *Valence:* according to the data, the valence of songs did slightly decrease over time. This means that the musical positiveness conveyed by tracks reduced. Over the years, more emotionally negative songs got in the *Year End Chart*. This will be explained in Section 4.2

2. *Energy/Arousal:* over the years, songs got more energetic, meaning that the perceptual measure of intensity and activity increased. According to the literature research (Section 2.4), energy also increased. An increase of this feature, will have a direct impact of which emotional categories have evolved. It is plausible that happy/joyful and turbulent/angry songs increased because these songs have a high energy level.

3. *Danceability:* the danceability of the tracks increased also overall. Danceability is mainly based on a combination of musical elements, such as tempo, rhythm stability, beat strength, and overall regularity.

4. *Acousticness:* this variable did undergo the biggest transformation amongst all the other variables as it decreased from 0.8 to 0.2. As been mentioned in the literature research, songs with a low acousticness will mainly evoke happiness or anger. These emotional categories might have increased.

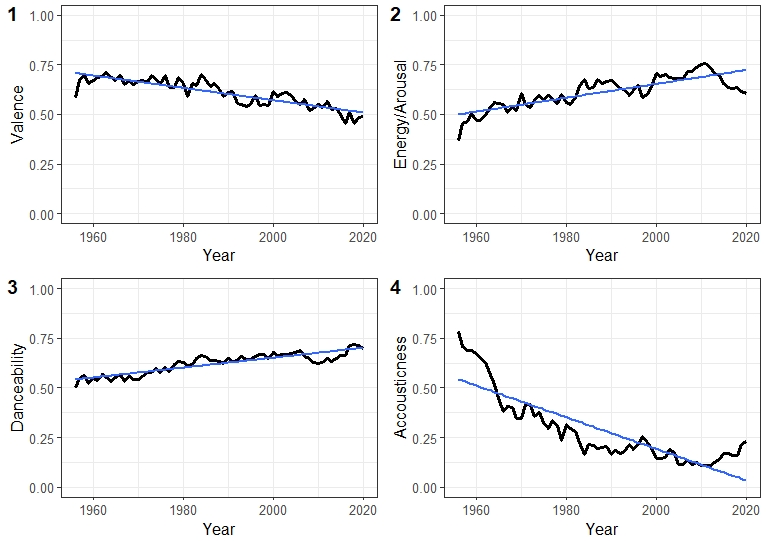


Figure 4.1: Time series analysis: audio features 1. Valence, 2. Energy, 3. Danceability, 4. Acousticness. The blue line represents a linear trend line.

5. *Speechiness:* the presence of spoken words in a track doubled over the years. This may be the result of upcoming genres. However, this report will not discuss that.

6. *Loudness:* the loudness of commercially succeeded songs are measured in LUFS (Watanabe et al., 2020). Loud songs tend to be around -11 dB LUFS, as quiet songs are approximately valued at -23 dB LUFS. The standard according to Spotify is -14 db LUFS. In this case, songs got louder over the past years.

7. *Liveness:* Liveness decreased a bit over the years, but as can be seen in the graph, it fluctuates.

8. *Tempo:* the amount of beats per minute fluctuates heavily.

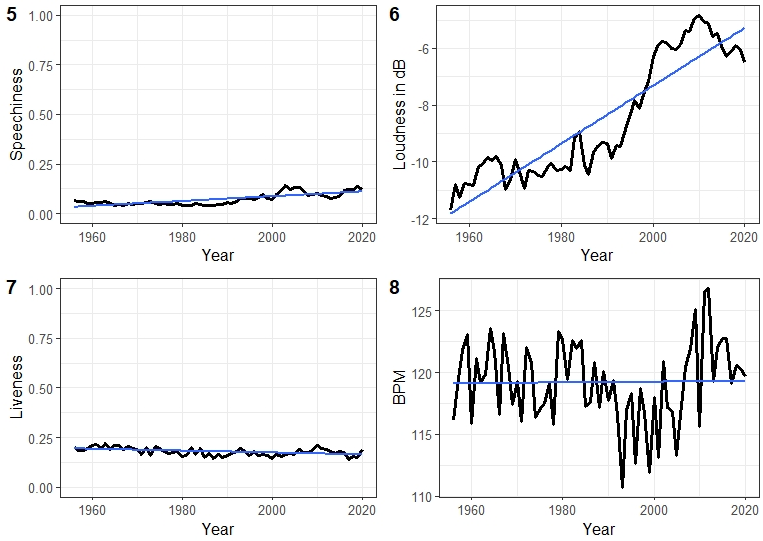


Figure 4.2: Time series analysis: audio features 5. Speechiness, 6. Loudness, 7. Liveness, 8. Tempo. The blue line represents a linear trend line. Please note that the y-axis of 6. Loudness and 8. Tempo are on a different range.

9. *Track duration:* the duration of songs increased in a linear fashion for over thirty years, but since 1990 there is a downtrend.

The goal of these graphs is to visualise trends in the data. Furthermore, it provides information of the values of the data. Interesting to note, *speechiness* and *liveness* contain relatively small values compared to the other audio features on a scale from zero to one.

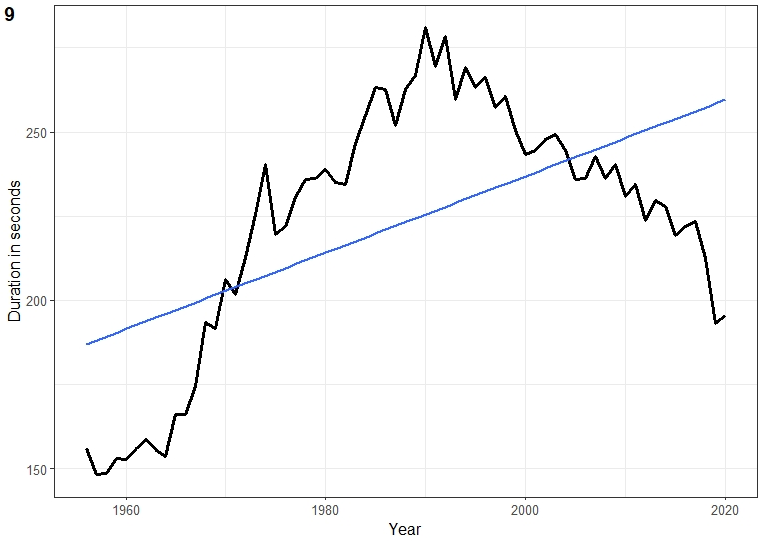


Figure 4.3: Time series analysis: audio feature 9. Track duration. The blue line represents a linear trend line. Please note that the y-axis is on a different range.

4.2 Emotional categories

According to the literature study, Russell's circumplex model classifies emotion labels in four quadrants based on *valence* and *arousal* (figure 2.8. The dataset contains data of the metrics *valence* and *arousal* with values ranging from zero to one. Russell's Model can be recreated and applied to this dataset. Songs will be classified in the following manner:

Table 4.1: Russell's classification

Class	***Valence range***	***Arousal range***
Sad/Depressing	0.00 - 0.50	0.00 - 0.50
Chill/Peaceful	0.50 - 1.00	0.00 - 0.50
Turbulent/Angry	0.00- 0.50	0.50 - 1.00
Happy/Joyful	0.50 - 1.00	0.50 - 1.00

Figure 4.4 represents Russell's Model recreated for the Billboard Year-End Chart data. The x-axis depicts the variable *valence* and the y-axis depicts the variable *arousal*. The data points represent the songs from the

Billboard Year-End Chart and are coloured by year. As can be noticed from the Figure, the quadrant of *Turbulent/Angry* is overwhelmed with data points from the year 2020. On the other hand, the quadrant of *Chill/Peaceful* is filled with data points from the 1960s. The results of Figure 4.4 will become more clear in Figure 4.5.

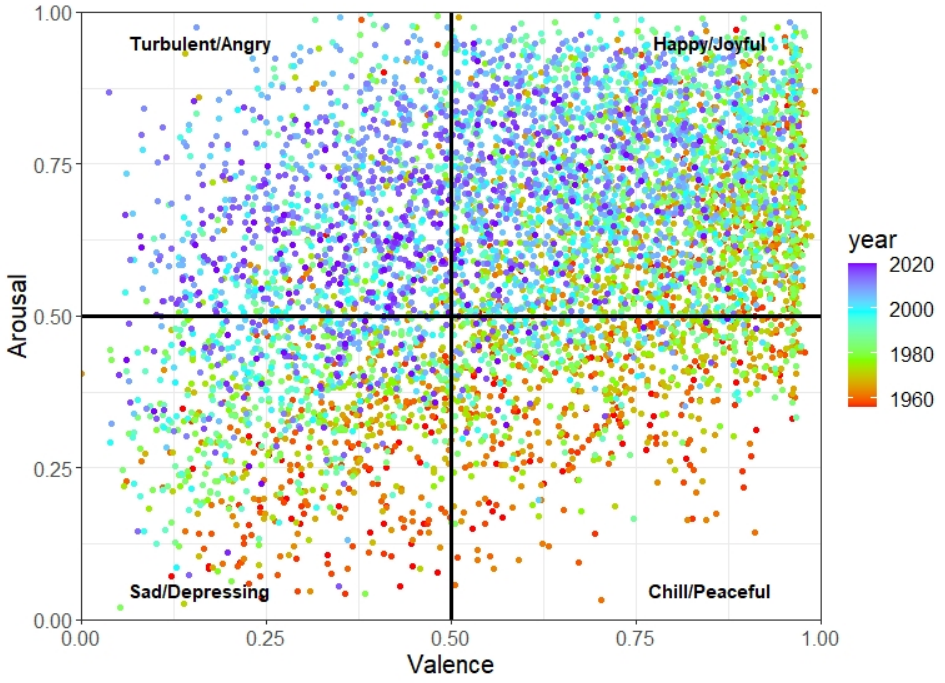


Figure 4.4: Russel's Model applied to the Billboard Year-End Chart data. Four emotion labels are used. Valence is represented on the x-axis, whereas energy/arousal is represented on the y-axis. The colour of the points indicate the year.

Figure 4.5 represents four trend lines of Russell's model. The goal of this model is to show the evolution of the four different emotion labels throughout the years of the Billboard data. According to this Figure, songs in the *Billboard Year-End Chart* got more *Turbulent/Angry* as time progresses. On the contrary, *Chill/Peaceful* songs are decreasing over time. This can primarily explained by the decrease of *valence* and the increase of *arousal* over time, demonstrated in Figure 4.1. The data for this Figure is presented in Table A.4.

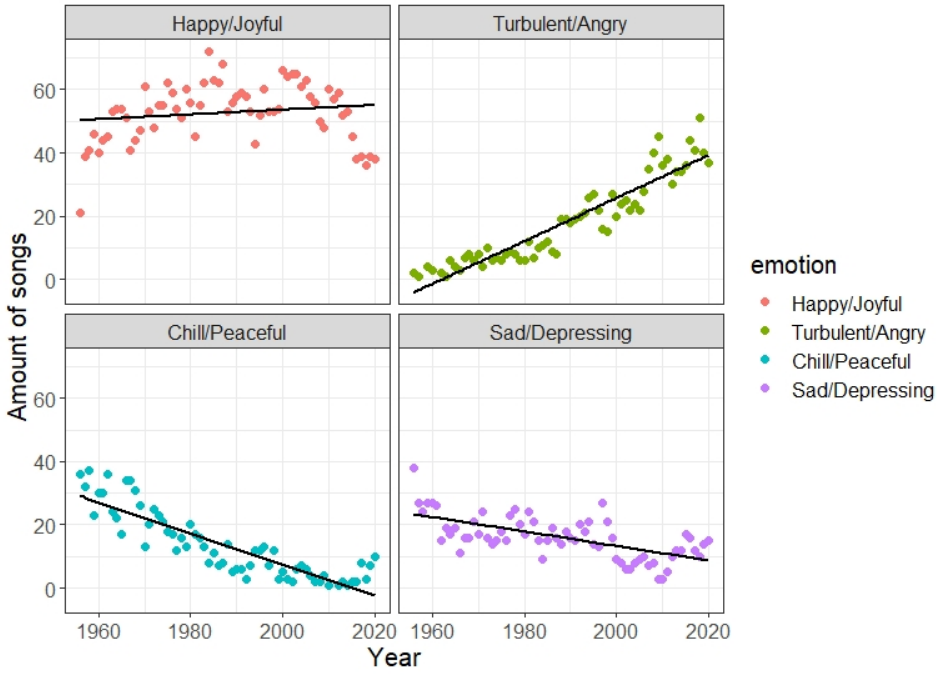


Figure 4.5: This supplementary Figure presents four trend lines of the data presented in Figure 4.4. The amount of songs per emotion label are presented on the y-axis. The x-axis contains the years, ranging from 1956-2020. The black line represents a linear trend line.

4.3 Audio feature selection

Valence and arousal can be computed from the audio signal. The dataset contains seven audio attributes (acousticness, danceability, instrumentalness, liveness, loudness, speechiness and tempo) which are metrics calculated from the audio signal. These features could be utilised to determine new values of valence and arousal from the audio signal.

In this report, the seven mentioned audio features will be combined in an item set. This way the values of *valence* and *arousal* can be determined. However, not each feature will have the same impact on the variables valence and arousal.

To determine the new values of valence and arousal, the contribution of the audio features have to be calculated, called *feature importance*. This refers to a technique that assigns a score to input features (the seven audio features) based on how useful they are at predicting target variables (valence and arousal from Spotify). Initial tests using a reduced set of features lower the model's performance significantly. Therefore I have chosen to use the complete set of features.

Extreme Gradient Boosting (XGBoost), a machine learning algorithm, will be utilised to determine the feature importance. XGBoost is a gradient boosting decision tree (Wang & Ni, 2019). In XGBoost, the feature relative importance can be measured by *average gain* (Shi et al., 2019). Gain represents the relative contribution of the corresponding feature to the model calculated by taking each feature's contribution for each tree in the model.

Results from Section 4.1 indicate that not all the variables act on the same scale. Audio features such as *speechiness* and *liveness* have a lower value overall compared to other audio features such as *valence*. Besides, *tempo* and *loudness* are computed on a different scale. To tackle the differences in the audio signal features, normalisation was employed. This scaling technique will rescale values so that they end up ranging between zero and one. Normalisation has been employed before the features were used in the model. The following formula was used:

- Feature_normalised = (Feature - min(Feature)) / (max(Feature) - min(Feature))

After normalisation, the dataset has been divided in two different sets, i.e., the training set and the test set. The training set (80% of the data) is used to build the model and the test set (20% of the data) is used to evaluate its predictive accuracy. The results of sub tables A and B of Table 4.2 will represent the audio features used for each variable and their feature

importance. These results are based on the test set. The audio attributes are used from 6481 songs.

Table 4.2: Feature importance for *valence* and *arousal*

Table A: Valence

Audio feature	***Importance***
Danceability	0.334
Speechiness	0.151
Liveness	0.145
Acousticness	0.108
Tempo	0.099
Loudness	0.097
Instrumentalness	0.067

Table B: Arousal

Audio feature	***Importance***
Loudness	0.561
Acousticness	0.181
Speechiness	0.081
Danceability	0.057
Instrumentalness	0.046
Tempo	0.041
Liveness	0.031

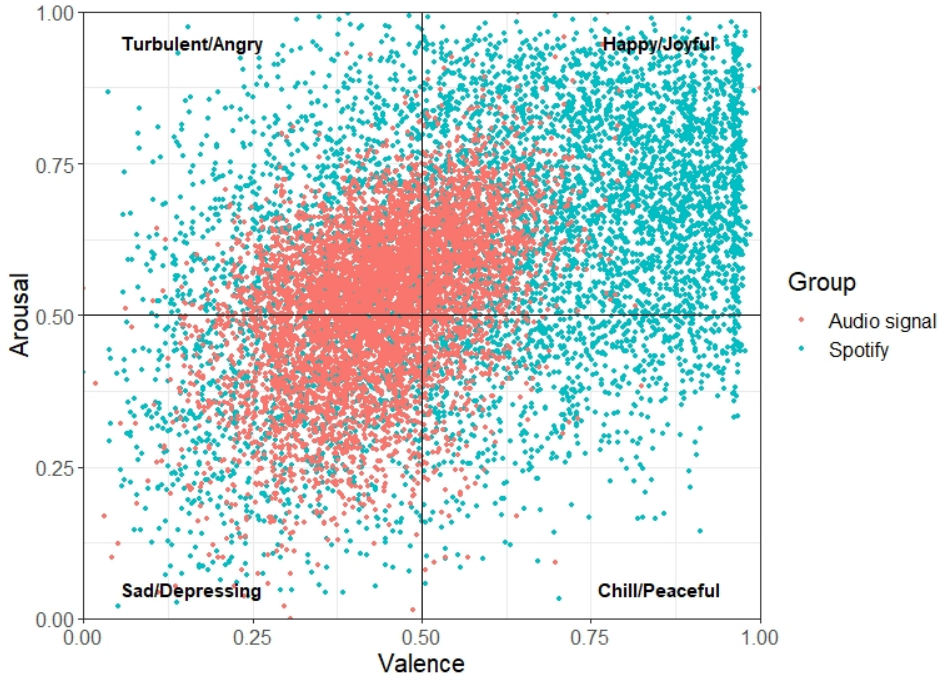


Figure 4.6: Emotion label comparison of audio features vs. Spotify, based on Russell's model. The red dots represent emotions labels calculated by the audio signal, whereas the blue dots represent emotions labels determined by Spotify. The number of data points are equal for both groups.

The results of Table 4.2 show the feature importance for each feature. As mentioned, this importance is based on *gain* of the XGBoost algorithm. The *gain* implies the relative contribution of the corresponding feature to the model calculated by taking each feature's contribution for each tree in the model. A higher value of this metric when compared to another feature implies it is more important for generating a prediction.

The prediction for a new *valence* score is primarily based on *danceability* according to the Table. An importance of 0.334 means a large contribution in generating a new *valence* value. This suggest that *danceability* is an important feature to determine *valence*. Loudness contributes for over 50% for generating a prediction of *arousal*. This suggests that *arousal* is mostly predicted by *loudness*.

New values for *valence* and *arousal* were calculated, using the importance factors supplied in Table 4.2. Thus, the normalised values of the seven audio signal features were multiplied with the importance factor. This was done for each row, resulting in new values for *valence* and *arousal*. These values were used as input for Figure 4.6.

As can be viewed from Figure 4.6, the values generated by the *audio signal* are primarily on the left side of the model. This indicates that the audio signal does not contain high values of *valence* overall and has a right-skewed distribution. This also means that the audio signal does not evoke a lot of musical positiveness because of the low-valence scores on average.

4.3.1 Statistical test - Audio signal vs. Spotify

Once having multiple dependent variables, a regular ANOVA test cannot identify patterns in multiple dependent variables whereas a MANOVA test can. This restriction can be very problematic in certain cases where a typical ANOVA would not be able to produce statistically significant results. This reduces the type I error which can be inflated by performing separate univariate ANOVA tests for each dependent variable. To determine if two groups' mean scores differ on multiple dependent variables, MANOVA was most appropriate. The effect size was reported using η^2 (Warner, 2012). A regular ANOVA test can only assess one dependent variable at a time.

Therefore, a MANOVA was conducted to test the hypothesis that there would be one or more mean differences between the audio signal and Spotify and their scores for *valence* and *arousal*.*Valence* and *arousal* were examined as dependent variables, and audio signal and Spotify were examined as independent variables. Both independent variables had an N value of 6481. The mean values and their standard deviation are summarised in Table 4.3.

The significance level (α) of 0.05 has been used. This indicates a 5% risk of concluding that an association exists when there is no actual association. A statistically significant MANOVA effect was obtained, $F(2,12959) = 2331.452$, $p << 0.001$. The multivariate effect size (η^2) was estimated at .173, which implies that 17.3% of the variance in the dependent variables was accounted for by the group. The p-value is also lower than the α of 0.05, indicating that there is a statistically significant different in scores of *valence* and *arousal* between audio signal and Spotify.

Next, the univariate effects on the dependent variables were analysed. It was found that all dependent variables were affected by the group they were in. A statistically significant univariate test effect was obtained for *arousal*, $F(1,12960) = 900.681$, $p < .001$. The mean of *arousal* was significantly affected by the group based on these results. The effect size (η_p^2) was estimated at .065, which implies that 6.5% of the variance derived arousal scores was accounted for by group level. For *valence*, $F(1,12960) = 2573.706$, $p << 0.001$. The multivariate effect size (η_p^2) was estimated at .166, which implies that 16.6% of the variance in the valence scores was accounted for by group. This means that there is a large practical significance between the independent variables.

Concluding, based on the audio signal, songs evoke less arousal on average compared to Spotify (see Table 4.3). This difference is significant, but not large. Besides, based on the audio signal, songs evoke less valence on average compared to Spotify. This difference is significant and quite large. This means that the analysis of the audio signal and Spotify provided similar arousal values, but Spotify provided substantially higher values of valence than the audio signal analysis.

Table 4.3: Descriptive statistics - Audio signal vs. Spotify ($N = 6481$ for both)

	Group	***Mean***	***Std. Deviation***
Arousal	Audio signal	.523	.132
	Spotify	.611	.196
	Total	.567	.173
Valence	Audio signal	.439	.116
	Spotify	.608	.241
	Total	.523	.207

4.4 Lyrics

Analysing lyrics can be very labour-intensive as it consists of many ambiguous words. On the contrary, lyrics are a very rich resource and many types of textual features can be extracted from them. As already been mentioned, the lyrics are only collected for 3605 songs as not all API requests functioned. Based on the expansion of ANEW, mentioned in sections 2.5 and 3.5, values of valence and arousal can be annotated to words. However, not all the words were matched as the expansion did not contain all the words that were present in the lyrics. Ambiguous words such as "Yeah" or "doo" are examples of words that were not annotated any value.

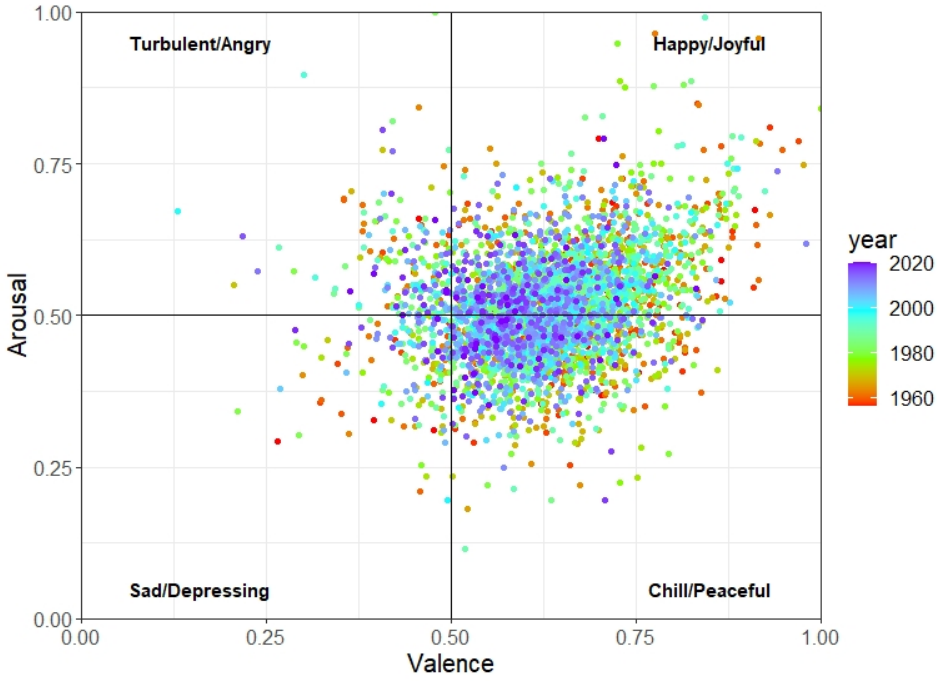


Figure 4.7: Russell's Model applied to the dataset of the normalised lyrics. It is divided into four emotional categories. Valence is represented on the x-axis, whereas energy/arousal is represented on the y-axis. The colour of the points indicate the year.

Figure 4.7 illustrates the values of valence and arousal of the lyrics after normalisation. Initial results showed a high kurtosis in the data of *arousal*. The kurtosis of *valence* was also slightly above the threshold. This means

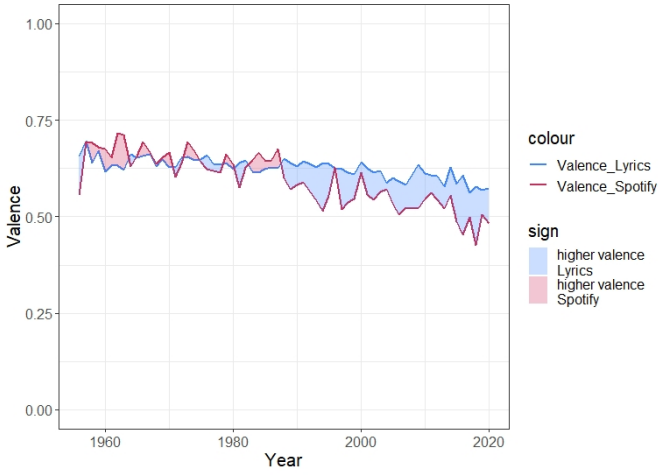
that the arousal values are not distributed evenly. To tackle this impediment, normalisation has been employed. At first, the range of the values on the y-axis (arousal) were limited. Practically all the arousal values were in a range between [0.45, 0.55]. Employing a normalisation, called "min-max feature scaling" (linear scaling), will bring all the values into the range of [0,1] (Singh et al., 2015). All the values have been normalised using the following formula to stretch the data:

- A_normalised = (Arousal - min(Arousal)) / (max(Arousal) - min(Arousal))
- V_normalised - (Valence - min(Valence)) / (max(Valence) - min(Valence))

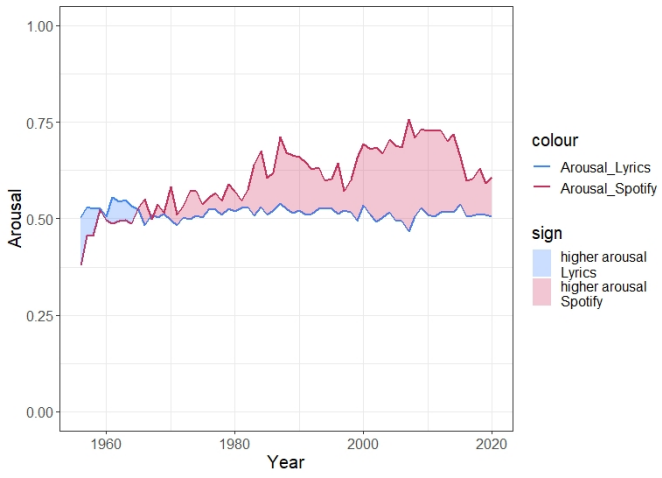
As can be seen, the majority of data points stay in the right quadrant. This shows that the distribution of the valence variable is left-skewed.

4.4.1 Comparison

Figures 4.8a and 4.8b are composed to exploit the differences between the values of valence and arousal of the lyrics and Spotify. Both figures represent the mean values over the year. As can be seen in Figure 4.8a, the mean valence value of the lyrics tends to be higher than the mean valence value of Spotify. This could indicate that lyrics convey more musical positiveness, i.e. sound more happier. Figure 4.8b visualises the differences in arousal over the years. Whereas arousal of the lyrics remains consistent, arousal of Spotify songs prospers and fluctuates. This could indicate that lyrics evoke a less aroused state.



(a)



(b)

Figure 4.8: Valence (a) and arousal (b) values from the lyrics compared to Spotify. The blue line represents valence (a) and arousal (b) of the lyrics, whilst the red line represents valence (a) oandr arousal (b) of Spotify. To emphasise the distinctions, the area's colour indicates whether the valence of lyrics or Spotify has a greater value. (a): The later decades of the plot are dominated by the valence of the songs. Overall, the valence of lyrics remains stable, whereas Spotify's valence changes substantially and falls steadily. (b): Important to note is that Spotify's arousal values are significantly higher with time than lyrics' values. Again, the values of the lyrics remain constant.

To provide a more in-depth analysis of the data in Figures 4.8a and 4.8b, violin plots, presented in figures 4.9 and 4.10, were created. A violin plot is a hybrid of a box plot and a kernel density plot. A box plot displays a five-number summary of a set of data. This contains the minimum value, the first quartile (25% of the data), the median (the vertical line which goes through the box), the third quartile (75% of the data), and the maximum. The median represents the exact middle point in the data. The other aspect of the violin plot, the kernel density plot, shows the distribution of the values.

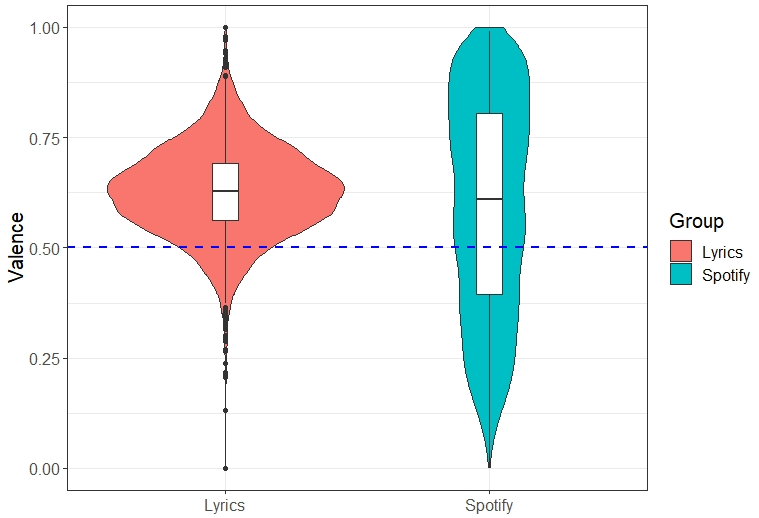


Figure 4.9: A violin plot (i.e., boxplot + kernel density estimation), which shows that the valence of the lyrics does not follow a normal distribution because the kurtosis is too high. Valence determined by Spotify has an extraordinary distribution. All the values are evenly spread, which is remarkable. The amount of positive as negative extreme values are evenly spread. which suggests that this distribution is not credible. This means that the valence of Spotify does not follow a normal distribution. The blue dotted line ($y = 0.5$) separates the positive valence values from the negatives. Both valence determined via lyrics (on the left) and Spotify (on the right) show roughly more positive values than negative values as is well-illustrated by the symmetry shown in the violin plots.

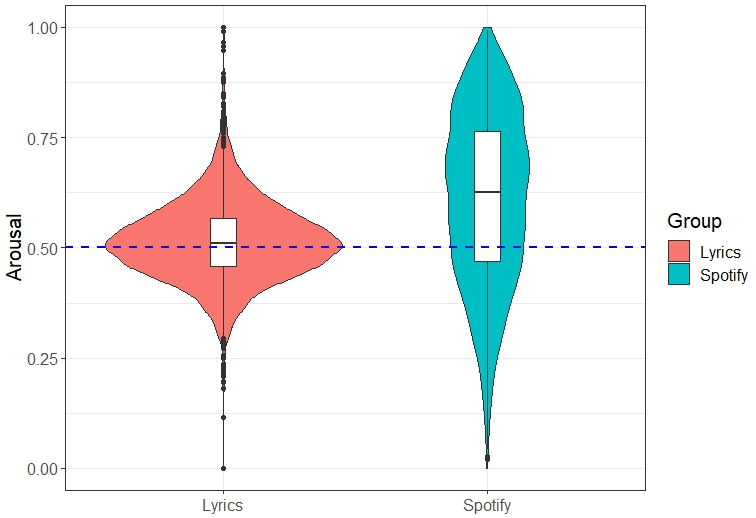


Figure 4.10: A violin plot (i.e., boxplot + kernel density estimation), which shows that the arousal of the lyrics roughly follows a normal distribution, where the arousal determined by Spotify has a lower kurtosis and thus not follow a normal distribution. The blue dotted line (y = 0.5) separates the positive arousal values from the negatives. Both arousal determined via lyrics (on the left) and Spotify (on the right) show roughly as many negative as positive values as is well-illustrated by the symmetry shown in the violin plots.

The violin plots provide critical knowledge to explain Figures 4.8a and 4.8b. The violin plots of the lyrics show a relatively high kurtosis, which means that many values are centred the mean. This could explain the stable transition over time. On the contrary, the violin plots of Spotify do not show a normal distribution as the data is spread evenly. This means that there are many outliers (extreme values, low and high). This could explain the fluctuations in the data as a year could contain more low or high outliers.

Figure 4.11 shows the emotion labels of the lyrics and Spotify's emotional markings. The number of data points are equal for both groups. Notably, the values of Spotify are more spread in the figure than the values of the lyrics. This could suggest that Spotify overestimates their values as the lyrics provide less extreme values and are more centred around the range of [0.3-0.7]. Furthermore, the values generated by the lyrics are primarily on the right side of the model. This indicates that the lyrics do not contain low values of *valence* overall. The distribution is left-skewed.

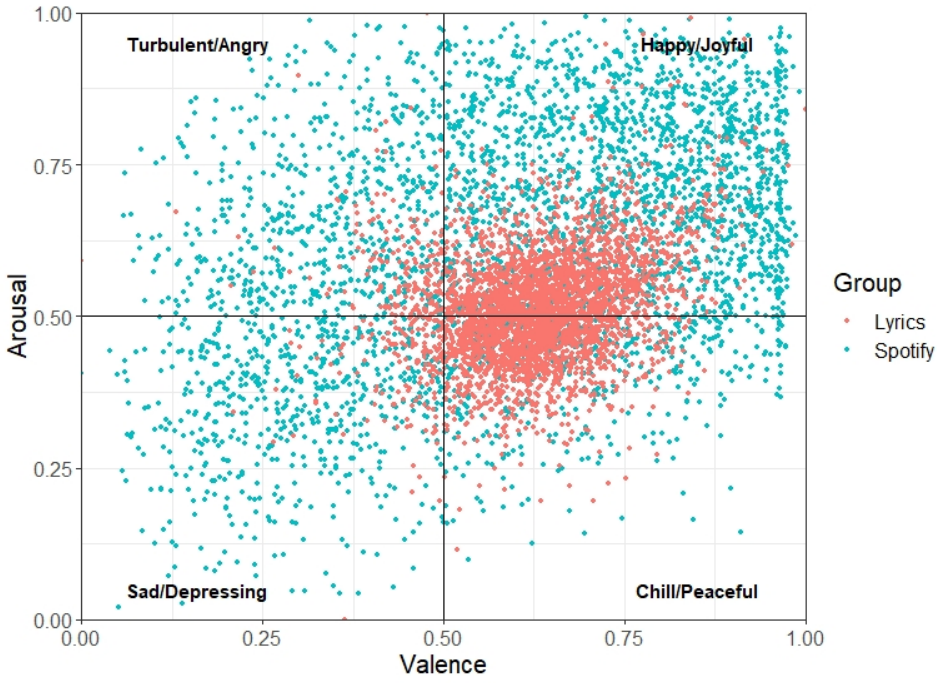


Figure 4.11: Emotion label comparison of lyrics vs. Spotify, based on Russell's model. The red dots represent emotions classified by lyrics, whereas the blue dots represent emotions classified by Spotify. The number of data points are equal for both groups.

4.4.2 Statistical test - Lyrics vs. Spotify

To formally test that there would be one or more mean differences between the lyrics and Spotify and their scores for *valence* and *arousal*, a MANOVA was conducted. *Valence* and *arousal* were examined as dependent variables, and lyrics and Spotify were examined as independent variables. Both independent variables had an N value of 3605. The mean values and their standard deviation are summarised in Table 4.4. The significance level (α) of 0.05 has been used. This indicates a 5% risk of concluding that an association exists when there is no actual association. A statistically significant MANOVA effect was obtained, $F(2{,}7207) = 530.202$, $p < .001$. The multivariate effect size (η^2) was estimated at .128, which implies that 12.8% of the variance in the dependent variables was accounted for by the group. The p-value is also lower than the α of 0.05, indicating that there is a statistically significant different in scores of *valence* and *arousal* between lyrics

and Spotify.

Next, the univariate effects on the dependent variables were analysed. It was found that all dependent variables were affected by the group they were put. A statistically significant univariate test effect was obtained for *arousal*, $F(1,7208) = 714.682$, $p < .001$. The mean of *arousal* was significantly affected by the group based on these results. The effect size (η_p^2) was estimated at .090, which implies that 9.0% of the variance derived arousal scores was accounted for by group level. For *valence*, $F(1,7208) = 56.254$, $p < .001$ was determined. The multivariate effect size (η_p^2) was estimated at .008, which implies that 0.8% of the variance in the valence scores was accounted for by group. It can be concluded that based on the lyrics, songs evoke less arousal compared to Spotify (see Table 4.4). This difference is significant and relatively medium-sized. Also, based on the lyrics, songs evoke more valence compared to Spotify. This difference is significant, but very small.

The analysis of the lyrics provide relatively similar valence and arousal values compared to Spotify. However, in the arousal dimension does Spotify provide a substantial larger value than the lyrics analysis.

Table 4.4: Descriptive statistics - Lyrics vs. Spotify ($N = 3605$ for both)

	Group	***Mean***	***Std. Deviation***
Arousal	Lyrics	.515	.092
	Spotify	.611	.196
	Total	.563	.160
Valence	Lyrics	.626	.103
	Spotify	.593	.244
	Total	.610	.188

4.4.3 Feature engineering

A necessary yet labour-intensive component of machine learning is feature engineering. In order for the model to utilise feature engineering, the input data must be preprocessed by adding new features based on existing features. These are the new features:

- Emotion: The emotion of a song as determined by the valence and arousal values of the lyrics. According to Russell's model, the emotion will be divided into four categories. In this instance, the emotion is

the predictor, as the machine learning model will attempt to predict the emotion of a song based on the characteristics described below.

- Word frequency: the number of words per song (Choi, 2018).
- Lexical diversity: the number of unique words per song.
- Lexical density: measurement of the complexity of a song. Has been calculated by dividing the *lexical diversity* by *word frequency*.
- Average word count: measurement of the average number of words in a song.
- Large word count: counter for the amount of words with more than seven characters.
- Small word count: counter for the amount of words with less than three characters.

To distinguish the different types of emotions, a feature has been developed that counts the amount of emotion-specific words (Kamalnathan et al., 2019). These words characterise each emotion and are words that have the most appearance according to its percentage in the songs. Each emotion has its feature, meaning that the amount of happy/joyful words etc. will be counted for each song. This way, they will function as a contextual predictor.

4.4.4 Lexical diversity

The more varied a vocabulary a text possesses, the higher its lexical diversity. Song vocabulary reflects the number of unique words present in the lyrics of a song. In short, it represents the lyrics' vocabulary variation (Kamalnathan et al., 2019).

Figure 4.12 shows that overall, there was an upward trend, nearly multiplying the number of unique words with two. In the end, there was a slight downward trend. However, this trend is mainly influenced by the word frequency of a song. This suggests that songs got more complex over time, as the contextual amount increases. The small amount of unique words in a song can be explained because English stop words and other unnecessary words are removed from the lyrics.

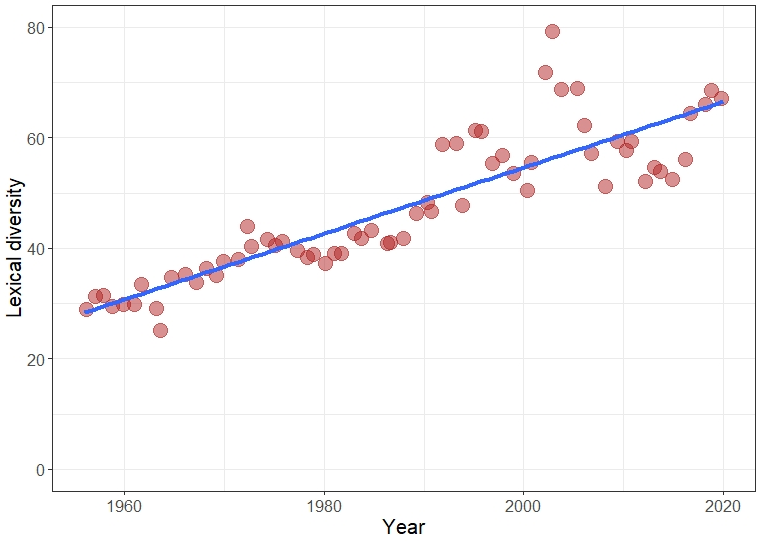


Figure 4.12: Lexical diversity over the years. The x-axis represent the number of unique words in a song, whilst the y-axis represents years. The blue line represents a linear trend line. The upward trend could suggest that songs got more complex as the amount of unique words doubled in the end.

4.4.5 Lexical density

Lexical density can be described as an indicator of word repetition. It can be calculated by dividing the number of unique words with the total number of words occurring in that specific song (Kahraman, 2020). As lexical density decreases, repetition increases. Note that this does not imply sequential repetition, even though it could occur that words get repeated after each other, lexical density is not intended to calculate this.

Figure 4.13 represents a downtrend, insinuating that the amount of unique words in a song decreases. However, 4.12 contradicts this because the amount of unique words increases. Thus, this means that the total number of words in a song over time increases. This results in an increase of lexical complexity because songs tend to have more words.

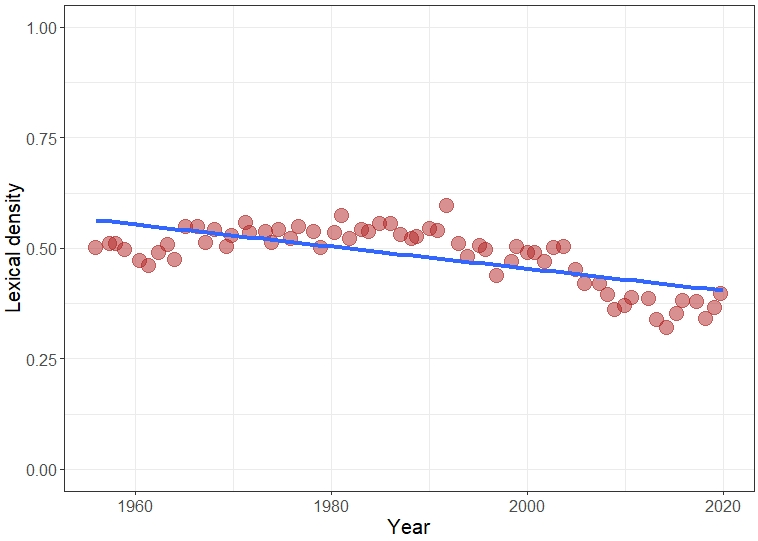


Figure 4.13: Lexical density can be calculated by dividing the amount of unique words per song (lexical diversity) by the number of words per song (word frequency). It shows the proportion of word repetition. As lexical density decreases, repetition increases as the amount of unique words overall is lower.

Lexical density and diversity are both indicators for lexical complexity. Concluding, lyrics got more complex over time according to these indicators. More unique words were used, which suggest an increase in variety of song vocabulary. The downtrend of lexical density shows an increase of more used words in lyrics. However this could also suggest that word repetition increases, i.e. more words are used with less context.

4.4.6 Emotion prediction

Previously, the emotion label of music is purely defined by its valence and arousal ratings. Now, since the lyrics have been pre-processed, machine learning models can be created to predict the emotion of a song based on words. A number of machine learning classification techniques will help develop these models.

Lyrics-based analysis could provide benefits to the music industry by automatically tagging the emotion labels of a song published by an artist. This will improve the user's experience when searching for songs (S. Li et al., 2017). Furthermore, recommender systems can make advantage of emotion

prediction. Based on the users' listening behaviour/history, the preferred emotion label can get recommended.

4.4.7 Machine learning algorithms

As the predictor consists of many classes (four categories of emotions), the classification job encloses more than two classes, which likely to be multi-class classification. This form of classification makes the premise that each sample is given to one and only one label, e.g. an animal can be either an elephant or a horse, but not both at the same time.

A wide range of machine learning classification approaches have been chosen that fulfils the above-mentioned requirements. As there is no single completely appropriate solution for a multi-class classification problem, many techniques were applied and assessed on its model correctness. This statistic is used to evaluate classification models. Informally, accuracy is the fraction of predictions that the model predicted accurately.

The following machine learning techniques were picked based on popularity and usage in another multi-class classification study (Liu et al., 2017):

1. Random forest
2. Recursive partitioning (Rpart)
3. XGBoost (extreme gradient boosting)
4. K-nearest neighbours
5. Linear discriminant analysis
6. Support vector machines
7. PART (decision trees)
8. Naive Bayes
9. Neural network

4.4.8 Validity

The validity of the machine learning classification algorithms has been assessed by dividing the dataset in a training-and test set (split: 80% training - 20% test) (split: 80% training - 20% test). Furthermore, the model has been tested using multiple machine learning methods.

To evaluate the model locally and to avoid the model from overfitting, K-fold cross validation has been used. *K*-cv is a re-sampling process and assesses the prediction error. In *k*-cv, the train data set is randomly divided into k-folds. Then, a model is trained K times on all the data except for a single fold. K-fold cross-validation offers the possibility to train on many train-test splits instead of using a single holdout set. This provides a better indicator of how well the model will perform on unknown data. Common values of k folds are three, five or ten. There is a trade-off, utilising more folds is computationally more expensive because the number of fitting and prediction rises. In theory, a higher number of k-folds should lead to a reduced prediction error, so k=10 has been chosen.

4.4.9 Normalisation

The train set and test set data have been normalised to account for the difference in value range of the features. This technique transformed data values between zero and one. If some variables/features are significantly greater in value and on a different scale than the others, the model will be thrown off by giving those variables more weight. Normalisation eliminates this impediment.

4.4.10 Machine learning models

To quantify the quality of the models, accuracy will be the major metric employed for this. The features that were explained in the paragraph "Feature engineering" will be used. Furthermore, to illustrate the importance of emotion-specific language characteristics (explained at the end of the end of the paragraph "Feature engineering"), two machine learning models were developed.

1. A machine learning model including the contextual features based on the training data.

2. A machine learning model excluding the contextual features based on the training data.

These models highlight the importance of the emotion-specific language characteristics. The training and test data sets contain all the features, although there is also an alternative model removing the emotion-specific terms characteristics. As can be observed in figure 4.14 on the x-axis, the machine learning model performs better with these additional features, having a greater accuracy.

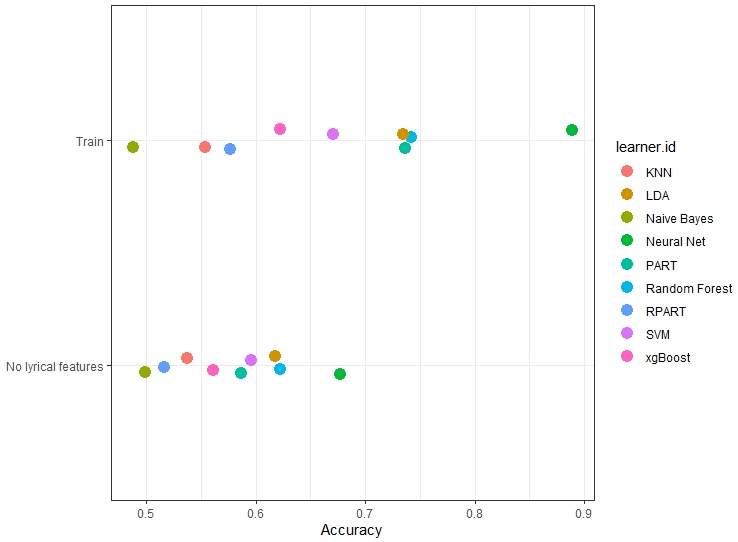


Figure 4.14: Visualisation of the differences in machine learning model accuracy of a machine learning model including contextual features and a machine learning model excluding contextual features. The upper dots on the y-axis represent the accuracy (x-axis) of machine learning algorithms (specified in the legend). These algorithms were applied to a machine learning model including contextual features. The lower dots represent the accuracy of the algorithms applied to a model excluding contextual features.

As can be observed from Table 4.5, the *neural net* had the best performance, implying that this machine learning algorithm had the best outcomes training the model and predicting the proper emotions. Having an accuracy of 0.883, it sets a very high accuracy level, outperforming the other models.

Table 4.5: Train data - Accuracy statistics

Model type	***Accuracy***
Neural Net	0.893
Random Forest	0.746
PART	0.736
LDA	0.735
SVM	0.670
xgBoost	0.622
RPART	0.576
KNN	0.554
Naive Bayes	0.488

The columns of Table 4.6 show the predicted values and the rows show the true values. The table indicates that predicting a "Chill" song get often predicted wrong as a "Happy" song.

Table 4.6: Confusion matrix

	Chill	***Happy***	***Sad***	***Turbulent***	***-err.-***
Chill	1085	152	8	10	170
Happy	66	1492	7	9	82
Sad	4	20	102	7	31
Turbulent	17	27	8	130	52
-err.-	87	199	23	26	335

To determine whether the performance of the models meets expectations, the accuracy of the top three machine learning algorithms were determined of the test set. The accuracy scores can be found in Table 4.7.

Table 4.7: Test data - Accuracy statistics

Model type	***Accuracy***
Neural Net	0.813
Random Forest	0.786
PART	0.695

The predicted results from the best performing machine learning algorithm are shown in Table 4.8. Results from the table indicate that predicting

a "Chill" song get often predicted wrong as a "Happy" song. However, a "Happy" song does not get wrongly predicted as a "Chill" song.

Table 4.8: Confusion matrix

	Chill	***Happy***	***Sad***	***Turbulent***	***-err.-***
Chill	179	122	1	2	125
Happy	0	389	1	4	5
Sad	0	7	27	0	7
Turbulent	0	8	0	37	8
-err.-	0	137	2	6	145

An important finding in machine learning is that no single algorithm works best across all possible scenarios or cases (Roßbach, 2018). Thus, no algorithm strictly dominates in all applications; the performance of machine learning algorithms varies widely depending, for example, on the application and the dimensionality of the dataset. Accordingly, a smart practise is to compare the performance of several learning algorithms to discover the best one for the particular situation. In some circumstances, it is also desirable to form ensembles of numerous models built with different methodologies to combine strength and eliminate weaknesses. However, often there is not enough time and/or money to test and improve any algorithm in order to its quality in a certain environment.

Comparing both outcomes, the test accuracy for neural net is slightly lower than on the training set. Neural nets are generally flexible models, resulting in overfitting the training set occasionally (Roßbach, 2018). Random forests can be trained with a relative small amount of data, but neural nets normally need more data to reach the same degree of accuracy. However, random forests have little performance increase when a specific quantity of data is reached, while neural nets normally benefit from enormous amounts of data and continuously improve the accuracy. This is primarily owing to its construction.

The random forest model outperforms its mutual part on the test data set, while PART (decision tree) performed marginally worse. Random forest takes advantage of the power of multiple decision trees (Prajwala, 2015). In addition, it does not rely on the feature importance assigned by a single decision tree, which tends to be highly dependent on a collection of features. Random forests select features arbitrarily during training.

4.4.11 Statistical test - Audio signal vs. Lyrics

Before combining the audio signal values with the lyrics values, a statistical test had to be executed to test on significant differences. A model of the comparison was not composed because this report intends to compare its models to Spotify.

A MANOVA was conducted to test the hypothesis that there would be one or more mean differences between the audio signal and the lyrics and their scores for *valence* and *arousal*.*Valence* and *arousal* were examined as dependent variables, and audio signal and lyrics were examined as independent variables. Both variables had an N value of 3605. The mean values and their standard deviation are summarised in Table 4.9. The significance level (α) of 0.05 has been used. This indicates a 5% risk of concluding that an association exists when there is no actual association. A statistically significant MANOVA effect was obtained, $F(2,7207) = 3635.509$, $p << .001$. The multivariate effect size (η^2) was estimated at .502, which implies that 50.2% of the variance in the dependent variables was accounted for by the group. The p-value is also lower than the α of 0.05, indicating that there is a statistically significant different in scores of *valence* and *arousal* between audio signal + lyrics and Spotify.

Next, the univariate effects on the dependent variables were analysed. It was found that all dependent variables were affected by the group they were in. A statistically non-significant univariate test effect was obtained for *arousal*, $F(1,7208) = 0.659$, $p = .417$. The mean of *arousal* was significantly affected by the group based on these results. The effect size (η_p^2) was logically estimated at 0. For *valence* on the contrary, a statistically significant test effect was obtained, $F(1,7208) = 6276.448$, $p << .001$. The multivariate effect size (η_p^2) was estimated at .465, which implies that 46.5% of the variance in the valence scores was accounted for by group.

Concluding, based on the audio signal, songs evoke less valence on average compared to the lyrics (see Table 4.9). This difference is significant and large, showing a strong complementary difference in valence. This means that the analysis of the audio signal provided substantially lower values of values than the lyrics analysis. Finally, based on the audio signal, songs evoke more arousal on average compared to the lyrics. However, this difference is insignificant, meaning that the data cannot tell whether there is a difference or not.

Table 4.9: Descriptive statistics - Audio signal vs Lyrics ($N = 3605$ for both)

	Group	***Mean***	***Std. Deviation***
Arousal	Audio signal	.517	.131
	Lyrics	.515	.092
	Total	.516	.113
Valence	Audio signal	.424	.113
	Lyrics	.626	.103
	Total	.525	.148

4.5 Audio feature selection + lyrics

Figure 4.15 demonstrates the combination of audio analysis 4.6 and lyric text analysis 4.11. As can be noted, the affective decomposition of Spotify is widely spread. On the contrary, the affective decomposition of audio analysis and lyric text analysis combined is more centred around the middle area. Analysing these results, it could suggest that Spotify normalised all their scores. This could be done for the reason of increasing the Δ (difference in principle between parameters). Enlarging the Δ will give Spotify an easier job of differentiating emotion categories.

The new values of valence and lyrics have calculated by adding the valence values of the audio analysis to the valence values of the lyrics and divide them by two. This resulted in the combined value of valence. The same principle was used for the arousal values.

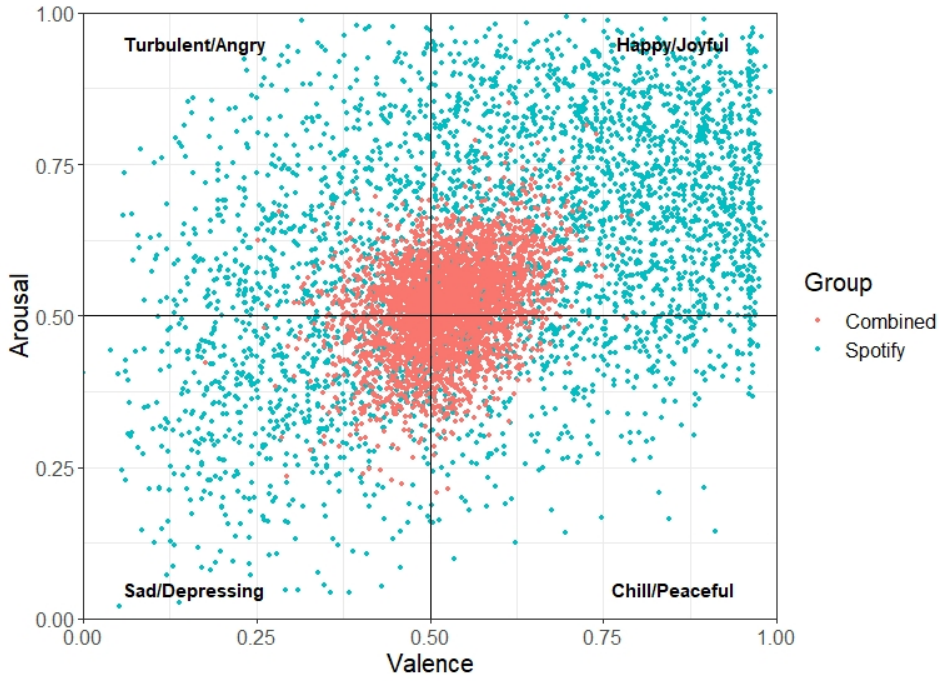


Figure 4.15: Emotion label comparison of audio features + lyrics vs. Spotify, based on Russell's model. The red dots represent emotions classified by the audio signal and lyrics combined, whereas the blue dots represent emotions classified by Spotify. The number of data points are equal for both groups.

4.5.1 Final result

The final model, presented in Figure 4.16 contains normalised values of audio features + lyrics. The combined values are more spread, however still do not exactly match Spotify's.

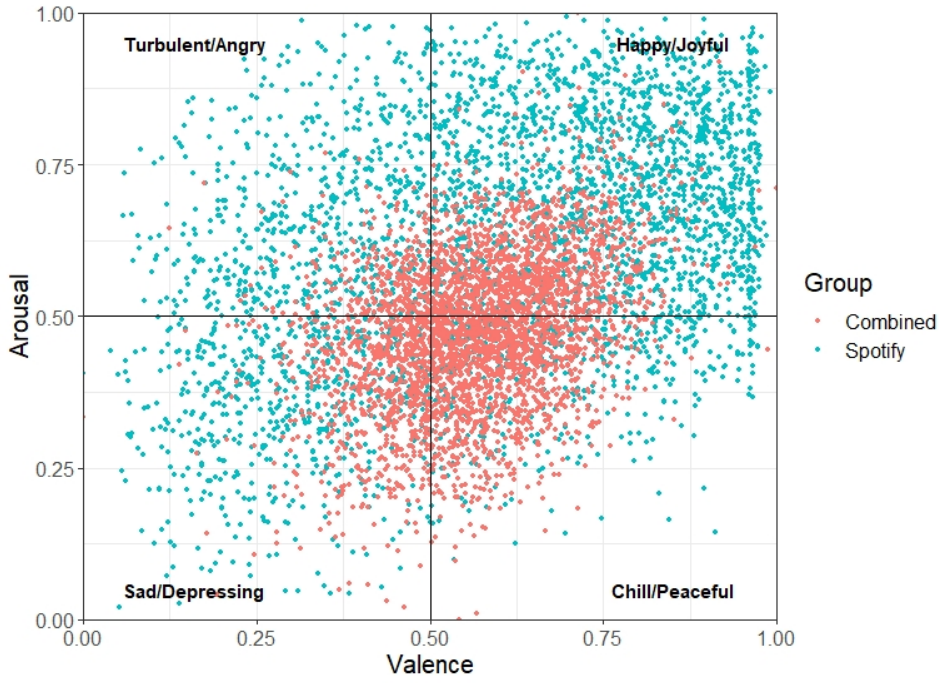


Figure 4.16: Emotion label comparison of audio features + lyrics normalised vs. Spotify, based on Russell's model. The red dots represent emotions classified by the audio signal and lyrics combined, whereas the blue dots represent emotions classified by Spotify. The number of data points are equal for both groups.

4.5.2 Statistical test - Audio signal + Lyrics vs. Spotify

A MANOVA was conducted to test the hypothesis that there would be one or more mean differences between the audio signal + lyrics and Spotify and their scores for *valence* and *arousal*.*Valence* and *arousal* were examined as dependent variables, and audio signal + lyrics (combined) and Spotify were examined as independent variables. Both independent variables had an N value of 3605. The mean values and their standard deviation are summarised in Table 4.10. The significance level (α) of 0.05 has been used. This indi-

cates a 5% risk of concluding that an association exists when there is no actual association. A statistically significant MANOVA effect was obtained, $F(2,7207) = 625.055$, $p < .001$. The multivariate effect size (η^2) was estimated at .148, which implies that 14.8% of the variance in the dependent variables was accounted for by the group. The p-value is also lower than the α of 0.05, indicating that there is a statistically significant different in scores of *valence* and *arousal* between audio signal + lyrics and Spotify.

Next, the univariate effects on the dependent variables were analysed. It was found that all dependent variables were affected by the group they were in. A statistically significant univariate test effect was obtained for *arousal*, $F(1,7208) = 1186.986$, $p < .001$. The mean of *arousal* was significantly affected by the group based on these results. The effect size (η_p^2) was estimated at .141, which implies that 14.1% of the variance derived arousal scores was accounted for by group level. For *valence*, $F(1,7208) = 31.241$, $p < .001$. The multivariate effect size (η_p^2) was estimated at .004, which implies that 0.4% of the variance in the valence scores was accounted for by group.

Concluding, based on the audio signal + lyrics, songs evoke less arousal on average compared to Spotify (see Table 4.10). The difference is significant and also large. Furthermore, based on the the audio signal + lyrics, songs evoke less valence compared to Spotify. However, this difference is very small.

This means that the analysis of the audio signal + lyrics provided similar valence values, but Spotify provided substantially higher values of arousal than the audio signal and lyrics combined.

Table 4.10: Descriptive statistics - Audio signal + Lyrics vs. Spotify (N = 3605 for both)

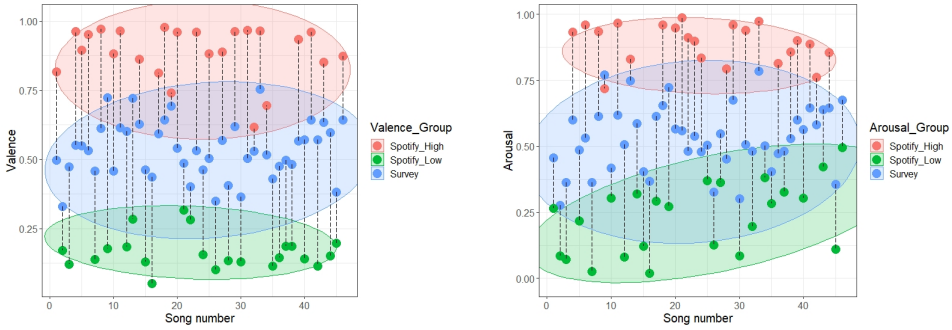
	Group	***Mean***	***Std. Deviation***
Arousal	Audio signal + lyrics	.478	.124
	Spotify	.611	.196
	Total	.545	.177
Valence	Audio signal + lyrics	.568	.120
	Spotify	.593	.243
	Total	.581	.193

4.6 Survey results

This section discusses the processing of all survey data. In addition, the relevance of the results will be evaluated to establish the results' dependability. This section's objective is to compare the results of the listener panel with Spotify's classification and will provide an answer to **[RQ4]**.

An overview of the experiment setup can be found in section 3.9. Additional screenshots of the survey are provided in Appendix C.

In total, the listeners were required to hear 46 audio fragments. The following charts compare valence and arousal results to Spotify's classification. As the most "extreme" songs have been selected, the Spotify values primarily consist of high and low values. Consequently, these values have been divided into two groups: "Spotify_High" and "Spotify_ Low" (the red and green ellipses respectively). The survey results are displayed within the blue ellipses and represent mean values. The space between the dashed lines represents the difference in valence and arousal between each song. Figures 4.17a and 4.17b represent these results.



(a) Comparative analysis of the Survey and Spotify's valence values

(b) Comparative analysis of the Survey and Spotify's arousal values

Figure 4.17: The songs used in the survey are on the x-axis and the value of valence or arousal is presented on the y-axis respectively. The red area represents Spotify values with high valence or arousal, whilst the green area represents values with low valence or arousal. The blue area comprises survey points with valence or arousal values. The lines between the points show the distances between the value for each song.

Observe that the survey values for both Figures fall within the range [0.25-0.75]. In addition, the majority of the survey responses cluster around the median of 0.5. Moreover, the gaps between the values are remarkable.

This could mean that Spotify does not accurately classify the affective disposition triggered by songs as the differences are tremendous.

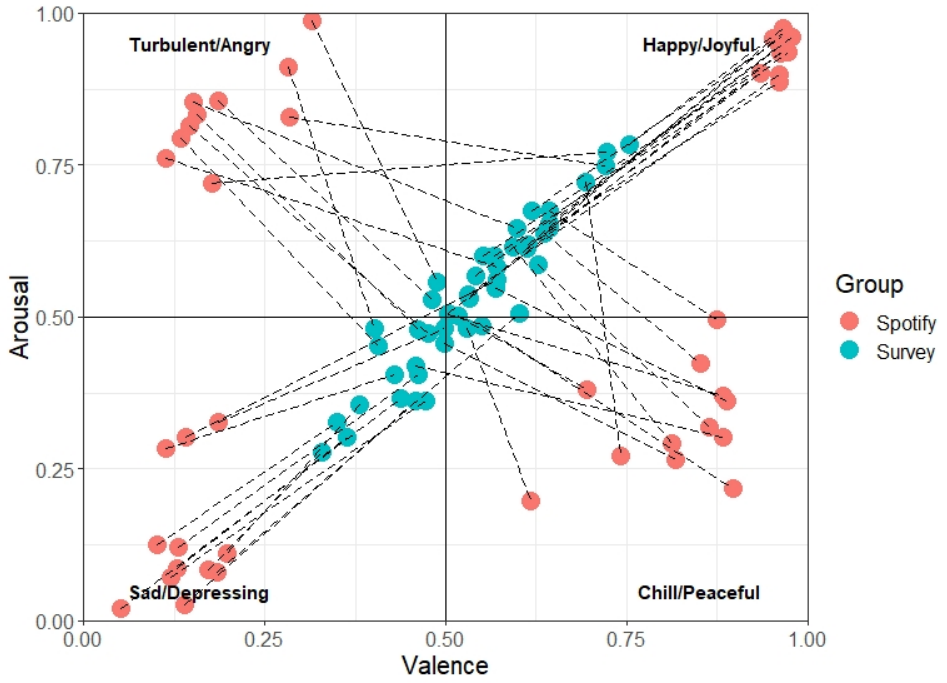


Figure 4.18: Emotion label comparison of the survey vs. Spotify, based on Russell's model. The red dots represent emotions classified by Spotify, whereas the blue dots represent emotions classified by the listeners panel. The dashed lines represent the distance between the red and blue dots. This annotates the current gap between the values.

Figure 4.18 represents Russell's model, in this case, containing the results presented in Figures 4.17a and 4.17b. This model clarifies the distinction even further. Even though, the results of the survey are not as extreme as Spotify's, songs classified as "Happy/Joyful" and "Sad/Depressing" are primarily corresponding. The other emotion categories of the quadrant are rarely similar. However, this model suggest that Spotify overestimates their values as the most "extreme" songs of each category are not rated similarly by the listeners panel.

4.6.1 Statistical test - Survey vs. Spotify

A MANOVA was conducted to test the hypothesis that there would be one or more mean differences between whether a subject was familiar with the song or not and their scores for *valence*, *arousal* and *verdict* (score of likeliness, question five of the survey). *Valence*, *arousal* and *verdict* were examined as dependent variables, and *familiarity* was examined as independent variable. Familiar answered with *yes* had an N of 909, and familiar answered with *no* had an N of 1299. The mean values and their standard deviations are summarised in Table 4.11. The significance level (α) of 0.05 had been used. This indicates a 5% risk of concluding that an association exists when there is no actual association. A statistically significant MANOVA effect was obtained, $F(3,2204) = 223.275$, $p < .001$. The multivariate effect size (η^2) was estimated at .233, which implies that 23.3% of the variance in the dependent variables was accounted for by the group. The p-value is also lower than the α of 0.05, indicating that there is a statistically significant different in scores of *valence*, *arousal* and *verdict* between whether a participant was familiar with the song.

Next, the univariate effects on the dependent variables were analysed. It was found that all dependent variables were affected by the group they were in. A statistically significant univariate test effect was obtained for *arousal*, $F(1,2006) = 379.087$, $p < .001$. The mean of *arousal* was significantly affected by the group based on these results. The effect size (η_p^2) was estimated at .147, which implies that 14.7% of the variance derived arousal scores was accounted for by group level. For *valence*, $F(1,2206) = 310.316$, $p << .001$. The multivariate effect size (η_p^2) was estimated at .123, which implies that 12.3% of the variance in the valence scores was accounted for by group. For *verdict*, $F(1,2206) = 667.156$, $p << .001$. The multivariate effect size (η_p^2) was estimated at .232, which implies that 23.2% of the variance in the valence scores was accounted for by group.

Concluding, based on the results of the survey, if a subject was not familiar with a song, songs evoked less *valence* and *arousal* compared to if a subject was familiar with the song. These differences were significant and large. Furthermore, the extent of likeliness, i.e. verdict was much greater when a subject was familiar with a song compared to not familiar. This difference was also significant and was substantially large. This provided evidence that if a subject is familiar with the song, it will experience more valence, become more aroused, and will annotate a substantially higher score.

Table 4.11: Descriptive statistics - Survey vs. Spotify (N(no) = 1299, N(yes)=909)

Dependent variable	***Familiarity (independent variable)***	***Mean***	***Std. Deviation***
Valence	No	4.68	2.092
Valence	Yes	6.32	2.238
Valence	Total	5.35	2.299
Arousal	No	4.52	2.242
Arousal	Yes	6.42	2.265
Arousal	Total	5.30	2.265
Verdict	No	4.35	1.966
Verdict	Yes	6.56	2.009
Verdict	Total	5.26	2.263

Chapter 5

Discussion

The objective of this research project was to analyse if Spotify correctly assessed its affective dispositions triggered by songs. This report offered numerous parallels to the emotion labels generated by Spotify. This also offered insights into the relationship between music and emotion labels. Analysing the impact of audio signal features and lyrics separately provided evidence in how they influence emotion labels. Moreover, creating appropriate emotion labels can aid in the improvement of recommender systems. Current recommender systems have significant flaws, as outlined in Chapter 1. Providing a suitable emotion label will assist the listener in receiving better recommendations. Besides, correct emotion labels could supply the listener with a song that matches the listener's actual needs at a certain moment.

The findings of this report could reveal how Spotify produces its emotion labels. Spotify presumably combined values of valence and arousal from audio signal analysis and lyrics analysis. Then, Spotify normalised the data to stretch it. This provided a significant gap between the emotion labels.

In section 1.3, four research questions were formulated. Based on the literature review and the results, an answer will be provided for these research questions:

[**RQ1**]: *Audio features*

Concluding from the audio features, results show that values generated by the audio signal are primarily on the left side of Russell's model. This indicates that the audio signal does not convey much musical positiveness because of the (on average) low value of valence. However, the emotion labels generated by the audio signal are widely spread over the y-axis, indi-

cating that the audio signal can vary in arousal. The statistical test proved that Spotify provided substantially higher values of valence than the audio signal analysis.

[**RQ2**]: *Lyrics*

It can be concluded ANEW does not provide a distinctive value for arousal. The literature research provided several methods in determining the valence and arousal values from lyrics. Initial results showed that the average arousal scores where within a range of 0.45 to 0.55. Therefore, normalisation had to be applied. On average, lyrics have a high value of valence, indicating that lyrics convey more musical positiveness. Furthermore, lexical complexity of the lyrics increased over time according to the measurements of lexical density and lexical diversity. This indicated that songs contained more words and unique words over time. From the measure learning model, it can be concluded that lexical features play an important role in accurately predicting the correct emotion label.

[**RQ3**]: *Audio features and lyrics combined*

Analysing both can be seen as a useful process because audio features and lyrics both contributed to the final model. Most important to note is that both models acted in a controversial way when separated. Combining both models contribute to a model where the emotion labels are spread evely across Russell's model, presented in Figure 4.15. Emotion labels generated by the audio signal were for the most part on the left side of Russell's model, whereas the emotion labels generated by the lyrics were on the right side of the model. This resulted in a stabilised result. Furthermore, it can be concluded that Spotify applies normalisation to its scores. Enlarging the difference between emotion labels will provide Spotify with an easier job of differentiating emotion labels.

[**RQ4**]: *A listener panel*

The results of the survey showed a different distribution of emotion labels compared to Spotify's. It can be concluded that Spotify overestimates the values of their labels because there is a substantial distance between the emotional markers. Results of the survey are not as extreme as Spotify's, songs classified as "Happy/Joyful" and "Sad/Depressing" are primarily corresponding. The other emotion categories of the quadrants are rarely similar.

This suggests that subjects do not experience the same level of valence and arousal as Spotify would suggest.

[**MRQ**]: *"Does Spotify accurately classify the affective disposition triggered by songs from the Billboard chart?"*

Assuming that Spotify normalised all their emotion labels, Spotify did a fairly accurate job of classifying. Increasing the difference between the emotion labels (i.e. increasing the delta) can be seen as an intelligent choice made by Spotify. This provides an easier way to differentiate the emotion labels.

An analysis of the results reveals the following findings:

A significant point must be made regarding the Spotify audio features. They are generally simple to extract in quantity and can provide valuable information, but they have limits that may affect the research outcomes. For instance, the features are the average song metrics. This indicates that variation in musical segments will not be considered. The metrics may fluctuate throughout a song.

One of the most difficult aspects of lyrical analysis is coping with ambiguity. Words can have numerous meanings depending on their context and sentence construction. This is also one of the limitations of the ANEW-dictionary, as it does not account for polysemy. According to ANEW, the emotion of each word represents how the typical person interprets that phrase in isolation. Therefore, it lacks sufficient information to distinguish between several meanings of a word or to determine its possible meaning in a phrase. To address this specific issue, a new vocabulary may be created, for instance, by conducting a new survey with more distinctive words. In addition, lyrics may contain "shouting" terms, such as "yeah" and "ya," which ANEW disregards.

All the songs have an average *arousal* score between 0.45 and 0.55, indicating that ANEW has no distinctive *arousal* value. Therefore, normalisation had been applied to stretch the data.

Finally, combining the values of valence of the audio signal analysis with the values of valence of the lyrics analysis resulted in a combined valence value for Figure 4.15. Values calculated by the audio signal analysis and the lyrics analysis both received a weight of 0.5. The same procedure was used for arousal. However, this weight of 0.5 could raise questions. Determining the importance of the audio signal and the lyrics was beyond the scope of this research project.

Possible suggestions for future research include: Genres could be used to expand upon this research. It could be of interest to recommender systems to examine genres based on their emotional categories.

Another dataset from a different region could be utilised. It could be fascinating to examine and compare the development of music in various nations. Recommendation systems could accommodate regional preferences.

The order of the songs that participants listened to during the survey was completely random. Consequently, various or identical emotion categories could be played sequentially. It would be intriguing if they could self-regulate their emotions by selecting their own emotional categories.

Taken together, the findings presented in this report are not only relevant for affective computing but also for other fields. Listeners' subjective judgements of music fragments demonstrated that in the real-world, music can be experienced differently. A good overview has been provided of how Spotify determines its emotion labels. Finally, the decomposition of audio and lyrics analysis provided a complementary difference in valence values. In turn, such research can help science to find its way to technology, hopefully resulting in many meaningful innovations in recommender systems.

Appendix A

Data variables

The following audio features have been captured from the Spotify API Documentation. These metrics are key and will be researched. Results of this research will be shown in the next chapter (Spotify, n.d.). These audio features are part of the dataset. Spotify runs a suite of audio analysis algorithms on every track to estimate the audio features (Skidén, n.d.). These algorithms extract about a dozen high-level acoustic attributes from the audio. In 2014, Spotify acquired company "The Echo Nest", which can be called a music intelligence service (Panda et al., 2021). They provided automatic data extraction from songs by web crawling (e.g., metadata, lyrics, reviews). Furthermore, they provided the service of digital signal processing techniques on the audio signal itself. Combined with the knowledge of The Echo Nest, Spotify developed these audio analysis algorithms to provide knowledge of a song's characteristics.

Table A.1: Audio features

Variable	Description
Accousticness	A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
Danceability	Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is the most danceable.
	Continued on next page

Table A.1 – continued from previous page

Variable	Description
Duration	The duration of the track in seconds.
Energy/Arousal	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.
Instrumentalness	Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.
Liveness	Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides a strong likelihood that the track is live.
Loudness	The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing the relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typical range between -60 and 0 db.
Mode	Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.

Continued on next page

Table A.1 – continued from previous page

Variable	Description
Speechiness	Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audiobook, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.
Tempo	The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
Valence	A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

The following table contains descriptive statistics of the audio features:

Table A.2: Descriptive statistics

Variable	Mean	Standard deviation
Acousticness	0.287	0.279
Danceability	0.621	0.152
Duration	223.45	57.453
Energy	0.611	0.196
Instrumentalness	0.027	0.127
Liveness	0.178	0.146
Loudness	-8.535	3.561
Mode	0.706	0.456

Continued on next page

Table A.2 – continued from previous page

Speechiness	0.072	0.078
Tempo	119.19	28.0
Valence	0.608	0.241
Key	5.3	3.574

The following variables are also part of the dataset. These tables are split in parts to emphasise on the difference between the variables. Table A.1 represents the (measurable) audio features and Table A.3 represents the (non-measurable) variables.

Table A.3: Variables

Variable	***Description***
Year	The year of which the track appears in the Billboard Year-End Chart "Hot 100 Songs". Songs could occur in multiple years. In most cases, this is the year after the first year the song has entered the Year-End chart as the song remained popular the year after.
Artist	The artist that has performed the song. Multiple artists are listed when they also worked on the song.
Track name	The name of the song/track
Emotion	The emotion of a song consists can consist of one of the following emotions: "Sad/Depressing", "Chill/Peaceful" ,"Turbulent/Angry" or "Happy/Joyful". This variable has been created based on the values of valence and energy. Besides, it is developed in line with Russel's model.

Thus, the variables in these tables describe all the variables of the entire dataset.

Table A.4: Russel's model data

Year	Happy	Angry	Chill	Sad	Year	Happy	Angry	Chill	Sad
1956	21	2	36	38	1989	56	19	5	18
1957	39	1	32	27	1990	58	18	6	16
1958	41	0	37	24	1991	59	19	6	15
1959	46	4	23	27	1992	58	20	3	20
1960	40	3	30	27	1993	53	21	7	18
1961	44	0	30	26	1994	43	26	12	21
1962	45	2	36	15	1995	52	27	12	14
1963	53	1	24	19	1996	60	22	13	13
1964	54	6	22	17	1997	53	16	7	27
1965	54	4	17	19	1998	53	15	12	21
1966	51	3	34	11	1999	54	27	3	16
1967	41	7	34	16	2000	66	20	5	9
1968	44	8	31	16	2001	64	24	3	8
1969	47	6	26	21	2002	65	25	2	6
1970	61	8	13	17	2003	65	22	6	6
1971	53	4	20	24	2004	61	24	7	8
1972	48	10	25	16	2005	63	22	6	9
1973	55	6	23	14	2006	58	28	4	10
1974	55	7	21	15	2007	56	35	2	7
1975	62	6	18	18	2008	50	40	2	8
1976	59	8	17	15	2009	48	45	4	3
1977	54	9	12	23	2010	60	36	1	3
1978	51	8	16	25	2011	57	38	0	5
1979	60	6	13	20	2012	59	30	1	10
1980	56	6	20	17	2013	52	34	2	12
1981	45	12	17	24	2014	53	34	1	12
1982	55	7	16	21	2015	45	36	2	17
1983	62	10	13	15	2016	38	44	2	16
1984	72	11	8	9	2017	39	41	8	12
1985	63	12	11	15	2018	36	51	3	10
1986	62	9	7	19	2019	39	40	7	14
1987	68	8	8	16	2020	38	37	10	15
1988	53	15	12	21	*Unk.*	*Unk.*	*Unk.*	*Unk.*	*Unk.*

Appendix B

Music Data Mining

The book discusses an alternative general data mining process in the following four steps (T. Li et al., 2002, 2011):

1. Data management: concerns the specific mechanisms and structures for how data are accessed, stored and managed. Data management is closely related to the implementation of data mining systems.

2. Data preprocessing: ensures that the data format and quality. This step tries to improve the efficiency and ease of the mining process. Raw data can contain incomplete, noisy and/or inconsistent data. To ensure the data quality, the data will be cleaned to remove the outliers and noisy data. Furthermore, this step includes data integration to integrate data from multiple sources if necessary and data reduction to reduce the dimensionality and complexity of the data. Last, data transformation will be applied to convert the data into a suitable format for the next step: mining.

3. Mining: contains several tasks, e.g.: visualisation, classification, clustering, regression and content retrieval. There are many algorithms to carry out one of these tasks such as Neural Network analysis and Principal Component analysis. This step of the framework is essential for knowledge discovery, applying the mentioned algorithms/machine learning techniques.

4. Post-preprocessing: refines and evaluates the knowledge derived from the mining procedure. This can be expressed in simplifying the extracted knowledge. Evaluating the knowledge is also important, because the knowledge needs to be understandable for the end-user. Creating some additional visualisations could be a task for the evaluation

part. Besides, documentation could also be a step in this phase in order to make the data understandable for the end-user.

Next to the previously described data mining framework, the book explains music data mining resulting in the following steps (T. Li et al., 2011):

1. Music Data Management: contains the gathering and storing of music data and music metadata. Transferring musical data from their originally recorded format to computer-accessible formats, such as MP3 files. A digital library of music data supports effective interaction between knowledge producers, librarians, and information and knowledge discoverers. Storing and arranging music data records effectively is a problem of a digital library. Users need to quickly find music resources of their interest for analysis. One of the big challenges of Music Data Management is *Music Indexing*. For a music index, a document can be a song, an album, an artist, a record label etc. Music indexing can be used to calculate music similarity between pairs of songs or to cluster music into automatically created genres. Getting music indexing done in an appropriate way is a hard task to do as a song can be processed in multiple ways.

2. Music visualisation: can be divided into two categories: visualisation of metadata content or acoustic content of single music documents, and, second, visualisation of a complete music collection for showing the correlations among different music pieces, or grouping music pieces into different clusters based on their pair-wise similarities. The first category of visualisation is mainly aimed at capturing the main idea or the music style of music documents, while the second category is aimed at helping the stakeholder find particular songs that they are interested in. Besides, the presence of similarity between songs will help with this process of searching for a specific kind of song within a collection.

3. Music Information Retrieval: emerging research area, that focuses on the fulfilment of users' music information needs. Strategies that enable access to music collections, both new and historical, need to be developed in order to keep up with users' expectations of search and browse functionality (Casey et al., 2008). The stakeholders in this step are the music industry, which records, aggregates and disseminates music, end-users, who want to search for music in a personalised way, and professionals, such as music performers, teachers, and music producers. At present, the most common method of accessing music data is

through textual metadata. This metadata can be rich in content. *MIR* tasks are mainly exploratory and can be, e.g.: detect plagiarism (identify misattribution of musical performances, misappropriation of music intellectual property), copyright monitoring (monitor music broadcast for copyright infringement or royalty collection), mood (find music based on emotional values). In the field of music data mining, *MIR* can be described as the main scope within the data mining task. Music Similarity Search could also be one of these tasks, searching for similar music sound files, given another music file. Based on the provided music feature, e.g. sound, the user should find similar music works using a music search system.

4. Data Mining: association mining is a concept within music data mining. This concept refers to the detection of correlations among different items in a data set. Association mining can be divided into three categories:

   (a) Detecting associations among different acoustic features. For instance, the association between timbre and tempo can be measured to improve the performance of tempo estimation (Xiao et al., 2008).

   (b) Detecting associations among music and other document formats. For instance, the paper of Liao et al. (2009) describes a model that learns and represents association patterns between music and video clips in professional MTV.

   (c) Detecting associations among music features and other music aspects, for example, emotions. An emotion graph could provide insight for music recommendations (Kuo et al., 2005).

   Another branch of this step is *Sequence mining* which aims to detect patterns in sequences, such as chord sequences. Sequence mining is a relatively new research field, which means that it lacks some extensive research. The main research topic in the music area, where sequence mining is being used, is music transcription. Different types of errors might be found when transcribing audio pieces. These errors are checked on segmentation errors, substitution errors and time alignment errors for example.

   One of the most popular data mining techniques in the music research field is *classification*. The most general classification issue focuses on

music genre/style classification. Other researchers turn their attention more towards the classification of music from audio pieces. The following classification areas can be distinguished:

(a) Audio classification: identify and label audio in three different classes: speech, music, and environmental sound. This type of classification can be used to segment videos in multiple parts, and so, decide where to apply automatic speech recognition.

(b) Genre classification: labelling an unknown recording of a song with the correct genre is one of the toughest tasks within music classification as the relevance of different categories is extremely subjective. Besides, genre classification is biased by Western music genres. Generally, genres can be automatically classified based on three different features sets, related to rhythmic, pitch, and timbre features.

(c) Mood and emotion classification: classifies or detects the emotional meaning of a song. Mood/emotion-based research is helpful in music understanding, music search, and some music-related applications. The emotion of a song is based on several factors of a song. Also, culture is involved in people's mood response to music.

(d) Instrument recognition and classification: providing indexes for locating instruments that are included in a musical mixture (the use of multiple instruments during a song).

(e) Artists classification: traditionally, this is based on acoustic features or the singer's voice.

(f) Singer identification: this tool can be used to organise, browse, and retrieve data in large music collections. Music databases can use clustering based on singer similarity to organise their database.

5. Clustering: can be described as the task that separates a collection of data into multiple groups based on criteria. In music data mining, clustering can be applied to genres and artists for example.

6. Music summarization: the increase in the size of digital multimedia data collections, an informative extraction that summarises an original digital content is a challenge in the music industry and is important knowing that this large-scale of information is easily accessible

by users (Shao et al., 2004). Music summarization aims to determine a representative sample of a given music piece that must be recognised instantly. Since a large volume of digital content has been made publicly available in various media, efficient approaches to automatic music summarization are increasingly in demand. From a customer's perspective, before they make a decision whether to purchase an album or not, they would prefer to listen to the highlights of the music. The first music summarization system that was developed was on the MIDI format (Kraft et al., 2001). Unfortunately, this system is different from the sampled audio format, and, therefore, cannot be utilised for real music summarization.

Appendix C

Survey

The following questions were requested from the participant regarding their demographics.

1. What is your age?
2. What is your country of origin?
   Answer possibilities: The Netherlands, Other
3. What is your native language?
   Answer possibilities: Dutch, English, Other
4. How many hours per day do you listen to music?
   Answer possibilities: 0-1, 1-2, 2-3, 3-4, 4-5, 5-6, 6+
5. What kind(s) of music do you listen? (multiple answers possible)
   Answer possibilities: Pop, Rock, R&B and soul, Latin, Jazz, Classical, Dance/Electronic, Hip-hop/Rap, K-Pop, Country, Metal, blues
6. Do you play a musical instrument?
   Answer possibilities: Yes, No
7. What is your gender?
   Answer possibilities: Male, Female, Non-binary/third gender, Prefer not to say

The following screenshots of the survey provide an overview of what questions were asked for each audio fragment.

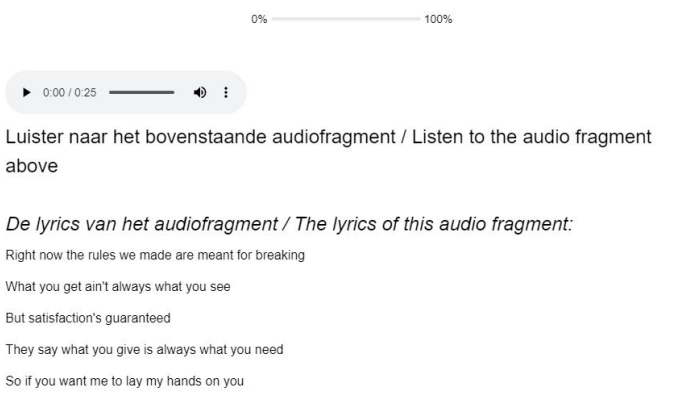


Figure C.1: Survey-audio fragment + lyrics

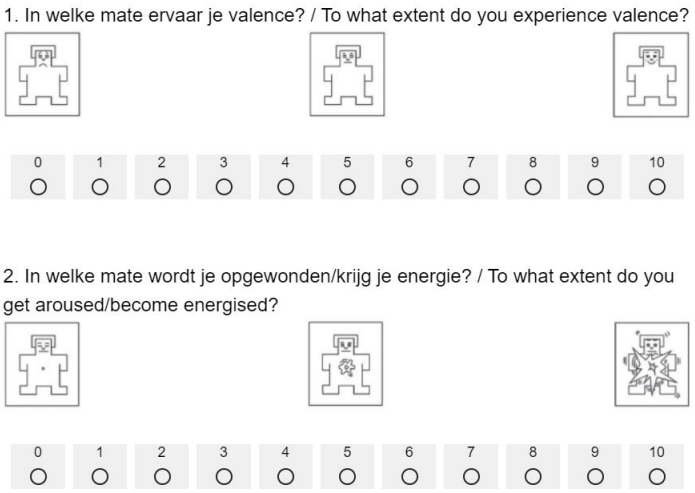


Figure C.2: Survey-question 1, 2

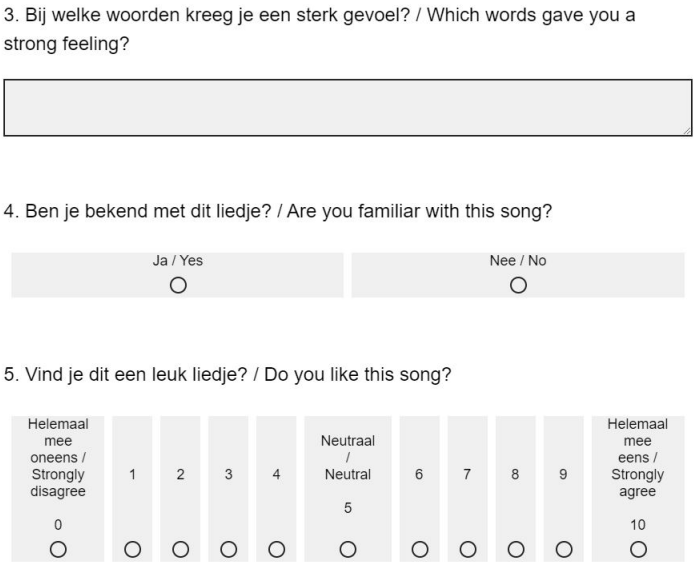


Figure C.3: Survey-question 3, 4, 5

References

Ali, S. O., & Peynircioğlu, Z. F. (2006). Songs and emotions: are lyrics and melodies equal partners? *Psychology of music*, *34*(4), 511–534.

Allaire, J. (2012). Rstudio: integrated development environment for r. *Boston, MA*, *770*(394), 165–171.

Amini, R., Willemsen, M. C., & Graus, M. P. (2019). Affective music recommender system (mrs): Investigating the effectiveness and user satisfaction of different mood inducement strategies.

Anand, N. (2005). Charting the music business: Magazine and the development of the commercial music field1. *The business of culture: Strategic perspectives on entertainment and media*, 139–154.

Askin, N., & Mauskapf, M. (2017). What makes popular culture popular? product features and optimal differentiation in music. *American Sociological Review*, *82*(5), 910–944.

Audacity, T. (2014). *Audacity.* Versão.

Barradas, G. T., & Sakka, L. S. (2021). When words matter: A cross-cultural perspective on lyrics and their relationship to musical emotions. *Psychology of Music*, 03057356211013390.

Bhat, A. S., Amith, V., Prasad, N. S., & Mohan, D. M. (2014). An efficient classification algorithm for music mood detection in western and hindi music using audio feature extraction. In *2014 fifth international conference on signal and image processing* (pp. 359–364).

Black, G. C., Fox, M. A., & Kochanowski, P. (2007). Concert tour success in north america: An examination of the top 100 tours from 1997 to 2005. *Popular Music and society*, *30*(2), 149–172.

Bollen, D., Knijnenburg, B. P., Willemsen, M. C., & Graus, M. (2010). Understanding choice overload in recommender systems. In *Proceedings of the fourth acm conference on recommender systems* (pp. 63–70).

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, *25*(1), 49–59.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for english words (anew): Instruction manual and affective ratings* (Tech. Rep.). Technical report C-1, the center for research in psychophysiology ....

Brown, S. C., & Knox, D. (2017). Why go to pop concerts? the motivations behind live music attendance. *Musicae Scientiae*, *21*(3), 233–249.

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., & Toiviainen, P. (2013). Influences of rhythm-and timbre-related musical features on characteristics of music-induced movement. *Frontiers in psychology*, *4*, 183.

Bushman, B. J., & Baumeister, R. F. (1998). Threatened egotism, narcissism, self-esteem, and direct and displaced aggression: Does self-love or self-hate lead to violence? *Journal of personality and social psychology*, *75*(1), 219.

Çano, E., & Morisio, M. (2017). Moodylyrics: A sentiment annotated lyrics dataset. In *Proceedings of the 2017 international conference on intelligent systems, metaheuristics & swarm intelligence* (pp. 118–124).

Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M. (2008). Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, *96*(4), 668–696.

Celma, O. (2006). Foafing the music: Bridging the semantic gap in music recommendation. In *International semantic web conference* (pp. 927–934).

Chen, Y.-S., Cheng, C.-H., Chen, D.-R., & Lai, C.-H. (2016). A mood-and situation-based model for developing intuitive pop music recommendation systems. *Expert Systems*, *33*(1), 77–91.

Chi, C.-Y., Wu, Y.-S., Chu, W.-r., Wu, D. C., Hsu, J. Y.-j., & Tsai, R. T.-H. (2009). The power of words: Enhancing music mood estimation with textual input of lyrics. In *2009 3rd international conference on affective computing and intelligent interaction and workshops* (pp. 1–6).

Choi, K. (2018). *Computational lyricology: quantitative approaches to understanding song lyrics and their interpretations* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Christensen, J. F., Gaigg, S. B., Gomila, A., Oke, P., & Calvo-Merino, B. (2014). Enhancing emotional experiences to dance through music: the role of valence and arousal in the cross-modal bias. *Frontiers in human neuroscience*, *8*, 757.

Coffey, A. (2016). The impact that music streaming services such as spotify, tidal and apple music have had on consumers, artists and the music industry itself. *Interactive Digital Media. University of Dublin*.

Cole, P. M., Martin, S. E., & Dennis, T. A. (2004). Emotion regulation as a scientific construct: Methodological challenges and directions for child development research. *Child development*, *75*(2), 317–333.

Connolly, M., & Krueger, A. B. (2006). Rockonomics: The economics of popular music. *Handbook of the Economics of Art and Culture*, *1*, 667–719.

Devine, K. (2013). Imperfect sound forever: loudness wars, listening formations and the history of sound reproduction. *Popular Music*, *32*(2), 159–176.

DeWall, C. N., Pond Jr, R. S., Campbell, W. K., & Twenge, J. M. (2011). Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular us song lyrics. *Psychology of Aesthetics, Creativity, and the Arts*, *5*(3), 200.

Downie, X., Laurier, C., & Ehmann, M. (2008). The 2007 mirex audio mood classification task: Lessons learned. In *Proc. 9th int. conf. music inf. retrieval* (pp. 462–467).

Droit-Volet, S., Bueno, L. J., Bigand, E., et al. (2013). Music, emotion, and time perception: the influence of subjective emotional valence and arousal? *Frontiers in Psychology*, *4*, 417.

Eerola, T., & Vuoskoski, J. K. (2011). A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, *39*(1), 18–49.

Fabiani, M., & Friberg, A. (2011). Influence of pitch, loudness, and timbre on the perception of instrument dynamics. *The Journal of the Acoustical Society of America*, *130*(4), EL193–EL199.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., et al. (1996). Knowledge discovery and data mining: Towards a unifying framework. In *Kdd* (Vol. 96, pp. 82–88).

Ferwerda, B., & Schedl, M. (2014). Enhancing music recommender systems with personality information and emotional states: A proposal. In *Umap workshops.*

Frith, S. (2007). Live music matters. *Scottish music review*, *1*(1).

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. *Handbook of music and emotion: Theory, research, applications*, *367400*.

Garrido, S., & Schubert, E. (2011). Negative emotion in music: What is the attraction? a qualitative study.

Giles, D. E. (2007). Survival of the hippest: Life at the top of the hot 100. *Applied Economics*, *39*(15), 1877–1887.

Girmal, R., KAMAT, J., MARTAL, S., NAIR, P., & Guide, S. C. (2018). *Music emotion recognition using acoustic gaussian mixture model* (Tech. Rep.). ST. FRANCIS INSTITUTE OF TECHOLOGY.

Goh, T. T., Jamaludin, N. A. A., Mohamed, H., Ismail, M. N., & Chua, H. S. (2022). A comparative study on part-of-speech taggers' performance on examination questions classification according to bloom's taxonomy. In *Journal of physics: Conference series* (Vol. 2224, p. 012001).

Haampland, O. (2017). Power laws and market shares: cumulative advantage and the billboard hot 100. *Journal of New Music Research*, *46*(4), 356–380.

Hand, D. J. (2007). Principles of data mining. *Drug safety*, *30*(7), 621–622.

Holmes, C., Knights, A., Dean, C., Hodkinson, S., & Hopkins, V. (2006). Keep music live: music and the alleviation of apathy in dementia subjects. *International Psychogeriatrics*, *18*(4), 623–630.

Holt, F. (2010). The economy of live music in the digital age. *European Journal of Cultural Studies*, *13*(2), 243–261.

Hove, M. J., Martinez, S. A., & Stupacher, J. (2020). Feel the bass: Music presented to tactile and auditory modalities increases aesthetic appreciation and body movement. *Journal of Experimental Psychology: General*, *149*(6), 1137.

Hove, M. J., Vuust, P., & Stupacher, J. (2019). Increased levels of bass in popular music recordings 1955–2016 and their relation to loudness. *The Journal of the Acoustical Society of America*, *145*(4), 2247–2253.

Hu, X. (2010). Music and mood: Where theory and reality meet.

Hu, Y., Chen, X., & Yang, D. (2009). Lyric-based song emotion detection with affective lexicon and fuzzy clustering method. In *Ismir* (pp. 123–128).

Hunter, P. G., Schellenberg, E. G., & Griffith, A. T. (2011). Misery loves company: Mood-congruent emotional responding to music. *Emotion*, *11*(5), 1068.

Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music perception*, *20*(2), 151–171.

Interiano, M., Kazemi, K., Wang, L., Yang, J., Yu, Z., & Komarova, N. L. (2018). Musical trends and predictability of success in contemporary songs in and out of the top charts. *Royal Society open science*, *5*(5), 171274.

Jamdar, A., Abraham, J., Khanna, K., & Dubey, R. (2015). Emotion analysis of songs based on lyrical and audio features. *arXiv preprint arXiv:1506.05012*.

Janssen, J. H., Van Den Broek, E. L., & Westerink, J. H. (2012). Tune in to your emotions: a robust personalized affective music player. *User Modeling and User-Adapted Interaction*, *22*(3), 255–279.

Juslin, P. N., & Sloboda, J. (2011). *Handbook of music and emotion: Theory, research, applications*. Oxford University Press.

Juslin, P. N., & Sloboda, J. A. (2001). *Music and emotion: Theory and research.* Oxford University Press.

Kahraman, V. (2020). *A comparative analysis of metal subgenres in terms of lexical richness and keyness* (Unpublished doctoral dissertation). lmu.

Kamalnathan, S., Mishra, Y., Kumawat, V., & Bangwal, V. (2019). Evolution of different music genres.

Kandel, S., Heer, J., Plaisant, C., Kennedy, J., Van Ham, F., Riche, N. H., ... Buono, P. (2011). Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, *10*(4), 271–288.

Kellogg, J. P. (2013). The urbanization of the" billboard" top album and singles charts: How soundscan changed the game. *MEIEA Journal*, *13*(1), 45–59.

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., ... Turnbull, D. (2010). Music emotion recognition: A state of the art review. In *Proc. ismir* (Vol. 86, pp. 937–952).

Kraft, R., Lu, Q., & Teng, S.-H. (2001, May 1). *Method and apparatus for music summarization and creation of audio summaries.* Google Patents. (US Patent 6,225,546)

Krumhansl, C. L. (1997). An exploratory study of musical emotions and psychophysiology. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *51*(4), 336.

Kuo, F.-F., Chiang, M.-F., Shan, M.-K., & Lee, S.-Y. (2005). Emotion-based music recommendation by association discovery from film music. In *Proceedings of the 13th annual acm international conference on multimedia* (pp. 507–510).

Kwame Harrison, A., & Arthur, C. E. (2011). Reading billboard 1979–89: Exploring rap music's emergence through the music industry's most influential trade publication. *Popular Music and Society*, *34*(3), 309–327.

Lai, J. C.-Y., & Amaladoss, N. (2021). Music in waiting rooms: A literature review. *HERD: Health Environments Research & Design Journal*, 19375867211067542.

Lara, J. A., Lizcano, D., Martínez, M. A., & Pazos, J. (2014). Data preparation for kdd through automatic reasoning based on description logic. *Information systems*, *44*, 54–72.

Laurier, C., Sordo, M., Serra, J., & Herrera, P. (2009). Music mood representations from social tags. In *Ismir* (pp. 381–386).

Lee, J. (n.d.). *Why are spotify songs greyed out? here's how to play them anyway.* Retrieved from `https://whatnerd.com/why-spotify-songs-greyed-out/`

Lems, K. (2001). *Using music in the adult esl classroom.* Citeseer.

Li, S., Mou, C., & Chang, C. (2017). Prediction of genres and emotions by song lyrics.

Li, T., Li, Q., Zhu, S., & Ogihara, M. (2002). A survey on wavelet applications in data mining. *ACM SIGKDD Explorations Newsletter*, *4*(2), 49–68.

Li, T., Ogihara, M., & Tzanetakis, G. (2011). *Music data mining.* CRC Press.

Liao, C., Wang, P. P., & Zhang, Y. (2009). Mining association patterns between music and video clips in professional mtv. In *International conference on multimedia modeling* (pp. 401–412).

Liew, K., Koh, A. H., Brown, C. M., dela Cruz, C., Lee, L. N., Krause, A. E., & Uchida, Y. (2020). *Groovin'to the cultural beat: Preferences for danceable music represent cultural affordances for anger experiences and expressions.* PsyArXiv.

Liu, Y., Bi, J.-W., & Fan, Z.-P. (2017). Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms. *Expert Systems with Applications*, *80*, 323–339.

Malheiro, R., Panda, R., Gomes, P., & Paiva, R. P. (2016). Emotionally-relevant features for classification and regression of music lyrics. *IEEE Transactions on Affective Computing*, *9*(2), 240–254.

Mariscal, G., Marban, O., & Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, *25*(2), 137.

Mathur, M. B., & Reichling, D. B. (2019). Open-source software for mouse-tracking in qualtrics to measure category competition. *Behavior research methods*, *51*(5), 1987–1997.

Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of popular music: Usa 1960–2010. *Royal Society open science*, *2*(5), 150081.

McAuslan, P., & Waung, M. (2018). Billboard hot 100 songs: Self-promoting over the past 20 years. *Psychology of Popular Media Culture*, *7*(2), 171.

McPherson, M., Smith-Lovin, L., & Brashears, M. E. (2006). Social isolation in america: Changes in core discussion networks over two decades. *American sociological review*, *71*(3), 353–375.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, *38*(11), 39–41.

Milliman, R. E. (1982). Using background music to affect the behavior of supermarket shoppers. *Journal of marketing*, *46*(3), 86–91.

Mori, K., & Iwanaga, M. (2014). Pleasure generated by sadness: Effect of sad lyrics on the emotions induced by happy music. *Psychology of Music*, *42*(5), 643–652.

Napier, K., & Shamir, L. (2018). Quantitative sentiment analysis of lyrics in popular music. *Journal of Popular Music Studies*, *30*(4), 161–176.

Nilsson, U. (2008). The anxiety-and pain-reducing effects of music interventions: a systematic review. *AORN journal*, *87*(4), 780–807.

Nishina, Y. (2017). A study of pop songs based on the billboard corpus. *International Journal of Language and Linguistics*, *4*(2), 125–134.

Niven, K. (2015). Can music with prosocial lyrics heal the working world? a field intervention in a call center. *Journal of applied social psychology*, *45*(3), 132–138.

North, A. C., & Hargreaves, D. J. (2006). Music in business environments. *Music and manipulation: On the social uses and social control of music*, 103–25.

North, A. C., Hargreaves, D. J., & Hargreaves, J. J. (2004). Uses of music in everyday life. *Music perception*, *22*(1), 41–77.

Oehler, M., Reuter, C., & Czedik-Eysenberg, I. (2015). Dynamics and low-frequency ratio in popular music recordings since 1965. In *Audio engineering society conference: 57th international conference: The future of audio entertainment technology–cinema, television and the internet.*

Olson, D. L., & Delen, D. (2008). *Advanced data mining techniques*. Springer Science & Business Media.

Pambudi, P., Sarno, R., & Faisal, E. (2018). Searching word definitions in wordnet based on anew emotion labels. In *2018 international seminar on application for technology of information and communication* (pp. 253–256).

Panda, R., Malheiro, R., & Paiva, R. P. (2018). Novel audio features for music emotion recognition. *IEEE Transactions on Affective Computing*, *11*(4), 614–626.

Panda, R., & Paiva, R. P. (2011). Using support vector machines for automatic mood tracking in audio music. In *Audio engineering society convention 130.*

Panda, R., Redinho, H., Gonçalves, C., Malheiro, R., & Paiva, R. P. (2021). How does the spotify api compare to the music emotion recognition state-of-the-art? In *Proceedings of the 18th sound and music computing conference (smc 2021)* (pp. 238–245).

Perone, J. E. (2009). *Mods, rockers, and the music of the british invasion.* ABC-CLIO.

Pettijohn, T. F., & Sacco Jr, D. F. (2009). Tough times, meaningful music, mature performers: Popular billboard songs and performer preferences across social and economic conditions in the usa. *Psychology of Music*, *37*(2), 155–179.

Piatetsky-Shapiro, G. (1990). Knowledge discovery in real databases: A report on the ijcai-89 workshop. *AI magazine*, *11*(4), 68–68.

Prajwala, T. (2015). A comparative study on decision tree and random forest using r tool. *International journal of advanced research in computer and communication engineering*, *4*(1), 196–199.

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Eng. Bull.*, *23*(4), 3–13.

Raskin, R., & Shaw, R. (1988). Narcissism and the use of personal pronouns. *Journal of personality*, *56*(2), 393–404.

Rodriguez, D., & Dolado, J. (2018). *Chapter 2 what is data mining knowledge discovery in databases (kdd).* Retrieved from `https://danrodgar.github.io/DASE/what-is-data-mining-knowledge-discovery-in-databases-kdd.html#ref-FayyadPS1996`

Roßbach, P. (2018). Neural networks vs. random forests–does it always have to be deep learning. *Germany: Frankfurt School of Finance and Management*.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, *76*(5), 805.

Saarikallio, S. (2011). Music as emotional self-regulation throughout adulthood. *Psychology of music*, *39*(3), 307–327.

Sachs, M. E., Damasio, A., & Habibi, A. (2015). The pleasures of sad music: a systematic review. *Frontiers in human neuroscience*, *9*, 404.

Salakka, I., Pitkäniemi, A., Pentikäinen, E., Mikkonen, K., Saari, P., Toiviainen, P., & Särkämö, T. (2021). What makes music memorable? relationships between acoustic musical features and music-evoked emotions and memories in older adults. *PloS one*, *16*(5), e0251692.

Sangnark, S., Lertwatechakul, M., & Benjangkaprasert, C. (2018). Thai music emotion recognition by linear regression. In *Proceedings of the 2018 2nd international conference on automation, control and robots* (pp. 62–66).

Schedl, M. (2019). Genre differences of song lyrics and artist wikis: An analysis of popularity, length, repetitiveness, and readability. In *The world wide web conference* (pp. 3201–3207).

Schellenberg, E. G., Krysciak, A. M., & Campbell, R. J. (2000). Perceiving emotion in melody: Interactive effects of pitch and rhythm. *Music Perception*, *18*(2), 155–171.

Schellenberg, E. G., & von Scheve, C. (2012). Emotional cues in american popular music: Five decades of the top 40. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(3), 196.

Sciandra, M., & Spera, I. C. (2020). A model-based approach to spotify data analysis: a beta glmm. *Journal of Applied Statistics*, 1–16.

Seo, Y.-S., & Huh, J.-H. (2019). Automatic emotion-based music classification for supporting intelligent iot applications. *Electronics*, *8*(2), 164.

Serrà, J., Corral, Á., Boguñá, M., Haro, M., & Arcos, J. L. (2012). Measuring the evolution of contemporary western popular music. *Scientific reports*, *2*(1), 1–6.

Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (kdd, crisp-dm and semma). *International Journal of Innovation and Scientific Research*, *12*(1), 217–222.

Shao, X., Xu, C., Wang, Y., & Kankanhalli, M. S. (2004). Automatic music summarization in compressed domain. In *2004 ieee international conference on acoustics, speech, and signal processing* (Vol. 4, pp. iv–iv).

Sherga Jr, R. M., Wei, D., Benson, N., & Javed, F. (2021). Alternative methods for deriving emotion metrics in the spotify® recommendation algorithm. *SMU Data Science Review*, *5*(3), 3.

Shi, X., Wong, Y. D., Li, M. Z.-F., Palanisamy, C., & Chai, C. (2019). A feature learning approach based on xgboost for driving assessment and risk prediction. *Accident Analysis & Prevention*, *129*, 170–179.

Singh, B. K., Verma, K., & Thoke, A. (2015). Investigations on impact of feature normalization techniques on classifier's performance in breast tumor classification. *International Journal of Computer Applications*, *116*(19).

Skidén, P. (n.d.). *New endpoints: Audio features, recommendations and user taste.* Retrieved from `https://developer.spotify.com/community/news/2016/03/29/audio-features-recommendations-user-taste/`

Soleymani, M., Aljanaki, A., Yang, Y.-H., Caro, M. N., Eyben, F., Markov, K., ... Wiering, F. (2014). Emotional analysis of music: A comparison of methods. In *Proceedings of the 22nd acm international conference on multimedia* (pp. 1161–1164).

Song, Y., Dixon, S., & Pearce, M. T. (2012). Evaluation of musical features for emotion classification. In *Ismir* (pp. 523–528).

Sousou, S. D. (1997). Effects of melody and lyrics on mood and memory. *Perceptual and motor skills*, *85*(1), 31–40.

Spotify. (n.d.). *Web api reference.* Retrieved from `https://developer.spotify.com/documentation/web-api/reference/`

Steffes-Halmer, A. (n.d.). *Selling song rights: Not always good business.* Retrieved from `https://www.dw.com/en/`

selling-song-rights-not-always-good-business/g-57432024#:~:text=The%20musician%20Michael%20Jackson%20famously,first%20gave%20Jackson%20the%20idea.

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of cognitive neuroscience*, *18*(8), 1380–1393.

Stratton, V. N., & Zalanowski, A. H. (1994). Affective impact of music vs. lyrics. *Empirical studies of the arts*, *12*(2), 173–184.

Taruffi, L., & Koelsch, S. (2014). The paradox of music-evoked sadness: an online survey. *PloS one*, *9*(10), e110490.

Thayer, R. E. (1990). *The biopsychology of mood and arousal*. Oxford University Press.

Thompson, C., Parry, J., Phipps, D., & Wolff, T. (n.d.). *spotifyr: R wrapper for the 'spotify' web api (r package version 1.1.0)*. Retrieved from http://github.com/charlie86/spotifyr

Trost, W., Ethofer, T., Zentner, M., & Vuilleumier, P. (2012). Mapping aesthetic musical emotions in the brain. *Cerebral Cortex*, *22*(12), 2769–2783.

Twenge, J. M., & Campbell, W. K. (2003). "isn't it fun to get the respect that we're going to deserve?" narcissism, social rejection, and aggression. *Personality and Social Psychology Bulletin*, *29*(2), 261–272.

Twenge, J. M., & Foster, J. D. (2010). Birth cohort increases in narcissistic personality traits among american college students, 1982–2009. *Social Psychological and Personality Science*, *1*(1), 99–106.

van Balen, J., Burgoyne, J. A., Wiering, F., Veltkamp, R. C., et al. (2013). An analysis of chorus features in popular song. In *Proceedings of the 14th society of music information retrieval conference (ismir)*.

Van den Broek, E. L. (2013). Ubiquitous emotion-aware computing. *Personal and Ubiquitous Computing*, *17*(1), 53–67.

van den Broek, E. L., & Westerink, J. H. (2009). Considerations for emotion-aware consumer products. *Applied ergonomics*, *40*(6), 1055–1064.

van der Zande, M. (2018). *Tune your mood with music: a personalized affective music player* (Unpublished doctoral dissertation). Master thesis) Eindhoven University of Technology.

Van der Zwaag, M. D., Westerink, J. H., & Van den Broek, E. L. (2011). Emotional and psychophysiological responses to tempo, mode, and percussiveness. *Musicae Scientiae*, *15*(2), 250–269.

Vickers, E. (2010). The loudness war: Background, speculation, and recommendations. In *Audio engineering society convention 129.*

Vonderau, P. (2019). The spotify effect: Digital distribution and financial growth. *Television & New Media*, *20*(1), 3–19.

Vuoskoski, J. K., & Eerola, T. (2011). Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences. *Musicae Scientiae*, *15*(2), 159–173.

Vuoskoski, J. K., Thompson, W. F., McIlwain, D., & Eerola, T. (2011). Who enjoys listening to sad music and why? *Music Perception*, *29*(3), 311–317.

Wang, Y., & Ni, X. S. (2019). A xgboost risk model via feature selection and bayesian hyper-parameter optimization. *arXiv preprint arXiv:1901.08433*.

Warner, R. M. (2012). *Applied statistics: From bivariate through multivariate techniques*. Sage Publications.

Watanabe, K., Kawahara, K., Nishida, H., & Okusa, K. (2020). The reality of the loudness war in japan-the case study on japanese popular music. In *Audio engineering society convention 149.*

Webster, G. D., & Weir, C. G. (2005). Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion*, *29*(1), 19–39.

Wong, A. K. C., & Wang, Y. (2003). Pattern discovery: a data driven approach to decision support. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *33*(1), 114–124.

Xiao, L., Tian, A., Li, W., & Zhou, J. (2008). Using statistic model to capture the association between timbre and perceived tempo. In *Ismir* (pp. 659–662).

Xu, L., Zheng, Y., Xu, D., & Xu, L. (2021). Predicting the preference for sad music: the role of gender, personality, and audio features. *IEEE Access*, *9*, 92952–92963.

Yalch, R. F., & Spangenberg, E. R. (2000). The effects of music in a retail setting on real and perceived shopping times. *Journal of business Research*, *49*(2), 139–147.

Zentner, M., Grandjean, D., & Scherer, K. R. (2008). Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion*, *8*(4), 494.