SMU Data Science Review

Volume 5 | Number 3 Article 3

2021

Alternative Methods for Deriving Emotion Metrics in the Spotify® Recommendation Algorithm

Ronald M. Sherga Jr.
Southern Methodist University, msherga@smu.edu

David Wei
Southern Methodist University, davidwei@smu.edu

Neil Benson
Southern Methodist University, ndbenson@smu.edu

Faizan Javed
Southern Methodist University, fjaved@smu.edu

Alternative Methods for Deriving Emotion Metrics in the Spotify® Recommendation Algorithm

David Wei[1], Neil Benson[1], Matt Sherga[1], Faizan Javed[1]
[1] Master of Science in Data Science, Southern Methodist University,
Dallas, TX 75275 USA
{davidwei, ndbenson, msherga, fjaved}@smu.edu

Abstract. Spotify's® recommendation algorithm tailors music offerings to create a unique listening experience for each user. Though what this recommender does is highly impressive, there is always room for improvement given that these techniques are not fully prescient. This study posits that in addition to creating certain features based on audio analysis, incorporating new features derived from album art color as well as lyrical sentiment analysis may provide additional value to the end user. This team did not find that a significant difference existed between color valence and Spotify® valence; however, all other comparisons resulted in statistically significant difference of means using paired t-tests. Due to the failure in finding a significant difference between color valence and Spotify® valence, this team is of the opinion that if a relationship between the two is found after additional exploration, they could be used in conjunction with one another to recommend music more accurately to listeners. Alternatively, there may be value in the statistical difference of the other variables, whereby further research may demonstrate a purpose in leveraging the differences.

1 Introduction

In 2020, the Recording Industry Association of America (RIAA) found that music streaming platforms account for 83% of today's digital music industry revenue [1]. The shift to digital consumption of music has also brought with it the technology to recommend music and enrich the listener's experience, including recommendations based on mood or emotion of the music. In a "traditional" sense, the methods for defining and modeling these emotional measures are derived from mechanical analysis of the music itself (i.e., the audio signal) [2]. Existing models consider volume, tempo, note, timbre, key, pitch, and modulation features including acoustic frequency and modulation frequency [3][2] to supply the recommendation algorithms with attributes that ultimately provide listeners personalized listening suggestions. These attributes, *valence* and *arousal*, can be used to describe the mood of the music.

Valence is a measure that explains the overall positivity or negativity of an emotion, while arousal is a measure that portrays the energy of the emotion. Valence ranges from positive to negative, and arousal ranges from passive to active.

Many streaming services, including Spotify®, describe their music moods using these features based on prior research that emotions can categorized and codified into a two-dimensional space known as Russel's Circumplex Model of Affect [4] in which emotions are mapped by their degree of valence and arousal. In the circumplex model noted, arousal occupies the vertical axis and valence the horizontal.

It is thought these two attributes for most streaming services are derived from somewhat sophisticated modeling techniques applied to audio signals, as is the status quo; it is unknown whether an album's cover art or the song's lyrics are instrumental in measuring valence and arousal.

Given how lyrics carry the semantic blueprints of a song, and visualizing colors can evoke certain emotions within people, augmenting music recommendation engines with measures derived from color and lyrics seems like a logical next step.

Spotify's® valence and energy measures will be used as baselines for comparison to this study's work of extracting valence and arousal measures from album art color and lyrical sentiment. This study posits that in addition to recommendation engine's "traditional" emotion features based on audio signal analysis, incorporating new features derived from album art color as well as lyrical sentiment analysis could provide additional value to the streaming listener and enhance recommendation algorithms.

To extract valence and arousal from album art color and lyrical sentiment, this study employed a variety of techniques.

The *Color Thief* API was used under MIT License to identify the dominant colors from the album cover art image [5]. These colors are in three dimensions; red, green, and blue or RGB as the primary colors. Prior research between color and emotions has been based on the primary triad of red, yellow, and blue or RYB. Conversion from RGB to RYB was performed [6].

Each color was then rotated and transformed to map on a two-dimensional space, or the same as Russel's Circumplex Model of Affect [4], providing valence and arousal from color. The transformations resulted in a loss of brightness or shade/tint within the color spectrum, leaving hue and light/dark variations occupying the same two-dimensional space. This loss meant that all perfectly neutral colors, black to white, were mapped to the same axis at the origin, and that light red and dark red were measurably the same, for example. Neutral colors denote neutral valence and arousal, and dark shades are mapped to the same emotions as light shades.

To extract lyrical sentiment, lyrics were downloaded and assessed through various natural language processing techniques including a mixed-hybrid approach of lexicon and machine learning techniques applied to the corpus of lyric data. Once the entire corpus of lyrics was labeled using a combination of lexical sentiment tools, the lyrics were then transformed into a TF-IDF (term frequency-inverse document frequency) sparse document-term matrix by finding the frequency of each word in the lyrics according to how often each appeared in all songs.

Because of the size of the document-term matrix, Latent Semantic Analysis (LSA) was performed to reduce the overall dimensionality of the matrix to increase performance by converting the TF-IDF matrix into a topic model resulting in a smaller and denser matrix.

The four new features extracted, color valence, color arousal, lyric valence, and lyric arousal were compared to the baseline, Spotify's® arousal and energy measures, and

to each other, to determine their statistical similarity. This study has shown that extracting valence and arousal from an album's cover art color and a song's lyrical sentiment resulted in statistically dissimilar measures for three of the four features, valence and arousal, from color and lyrics. Color valence was not shown to be significantly different base on the paired t-tests employed.

For this study's measures that were found to be statistically dissimilar to those of the industry standard, more research is suggested to improve the derived values by adjusting the sentiment analysis process or color mapping. Alternatively, these may be used as independent features that no longer represent valence and arousal, but unnamed features that warrant further exploration. More research would be needed to determine if these variables as independent can be defined or improved upon. Considering color valence was found to have a mean that was not statistically different than Spotify's®, this team suggests reinforcing the baseline value with color's since they are found to be in the same domain. The bimodal distribution of color valence is of some concern, so further exploration may be needed in order to confidently integrate into the recommendation algorithm.

2 Related Work

Emotions, and analogously moods, are intrinsically hard to define, assess, describe, or discern. To many, they are subjective. "Similar to the spectrum of color, emotions seem to lack the discrete borders that would clearly differentiate one emotion from another" (Posner, Russell, & Peterson, 2005, p.6). This obscured ambiguity led Russell in 1980 to codify emotional states in what is known as the circumplex model of affect, which is a two-dimensional circumplex model that proposes to map emotions to varying degrees of activation of valence and arousal [4, 7].

2.1 Music and Emotion

An emotion can be described as a linear combination of both states; valence being a pleasure-displeasure continuum and arousal being a level of alertness. For instance, fear is mapped as being moderately negative in valence; but high in active arousal; similarly, sleepiness is its contrarian counterpart, with passive arousal and moderately positive valence. This provides an index for which to map emotions into a two-dimensional space [7]. The circumplex model of affect is an ideal framework for the mapping of mood to music.

According to Agarwal and Om [8], "Music is the art of 'language of emotions'" (p. 1). Curiously, the connection between music and mood has long been studied [9] and is the subject of countless pieces of research in the field of knowledge.

The current attribution of mood, sentiment, or emotions to music is generally described using the feature valence, the x-axis of the Circumplex Model of Affect, as is the case with Spotify® [10]. Spotify® defines valence as:

> "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)" (Spotify®, 2021, para. 126).

Additionally, Spotify® depicts the arousal as music "energy," the y-axis of the Circumplex Model of Affect and defines it as:

> "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." (Spotify®, 2021, para. 113).

Spotify® and others use more contemporary methods for attributing mood to music which are somewhat sophisticated, and employ various modeling techniques applied to audio signals and run the gamut in style, technique, and accuracy; Agarwal et al. [8] developed an efficient supervised framework for music mood recognition via audio signal processing and Support Vector Regression. Similarly, Chapaneri et al. [11] showed that valence and arousal, the two emotions mapped in the Russell Circumplex Model, can be moderately estimated by applying a Deep Gaussian process to variable-length segments of the audio songs. Much of the current body of work aims to predict music mood utilizing audio signal processing automatically. This can be construed as a traditional, very "mechanical" approach and limiting to possibilities of music exploration and consumption.

In yet another example, diverging from the status quo, Tan et al. [12] showed that through a combination of audio signal processing and supplemental lyric analysis using Naive Bayes and Support-Vector Machine classifiers, high accuracy audio classification for valence and arousal could be achieved. Tan, et al.'s prior research [12], among other bodies of work such as MusicMood: predicting the mood of music from lyrics using machine learning (2014), Lyric-based music mood recognition (2015), and Music mood classification using intro and refrain parts of lyrics (2013), is a marked divergence from the norm [3][2] by fortifying mood music classification with lyrical analysis and appears to be the trend moving forward.

2.2 Lyrics and Emotion

In one particular study, when asked 141 music listeners about their judgments of emotional expression in music, 29% of them described the lyrics as a primary factor in expressing emotion in music [15]. In a separate study comparing audio-based emotion detection to that of text-based lyric methods, it was found that the text-based methods outperformed the audio features and were much more effective at predicting emotion in music [16]. It is no surprise then, that current efforts are underway to augment audio signals with lyrical data to produce greater accuracy in the classification of music emotion and musical genres in the topic of Music Information Retrieval [17]. Given how lyrics carry the semantic blueprints of a song, lyrical sentiment analysis has become a heavily researched area in recent years due to increased consumption of social media and subscription platforms.

Many of the current methodologies in approaching Natural Language Processing and lyrical emotion involve utilizing sentiment analysis and traditional machine learning models. At a very simplistic and basic level, Hu et al. [18] proposed using user input tags that described the emotions found in the songs as "ground truths" to feed into a clustering algorithm that derived three non-topical mood clusters from the associations of the user labeled tags. Taking a more mixed approach, Laurier et al. [19] demonstrated that an overall classification accuracy by normalizing the modalities across both audio and lyrics and then combining both features into the same feature space using SVM increased overall mean performance accuracy of 80.7% as compared to its counterparts of lyric and music features alone of only 61.3%. Similarly, a separate research team [14] also approached automatic lyric mood classification by utilizing the ANEW word dictionary and K-Means clustering on non-matching word level features to first create initial label features as training data and then combining these features with keyword frequencies using TF-IDF and Naive Bayes probabilistic classifier to reach a relatively high accuracy of 90.5%. Deviating from classification modeling approaches, one research team evaluated sentiment analysis comparing regression and classification methods [20]. Though the data used by this team was not domain-specific to lyrics and music, the team observed that classification accuracies became less meaningful as the spectrum of sentiments increased beyond a 2-class scenario and that regression algorithms resulted in much better predictions as it modeled fine-grained sentiment distributions with far less error.

It was found that in most cases regardless of modeling techniques, a common approach of using the NLTK preprocessing framework for tokenization, stemming, and normalization along with Naive Bayes or SVM was found successful among most research utilizing similar extraction and modeling methodologies [12][13][20]. These approaches resulted in arguably equally successful predictions of the emotional mood of the song using lyrics though nearly all research concluded that future studies be done with larger sample sizes of songs. Interestingly, Yang et al. [20] found that this hypothesized approach of increasing overall performance by increasing sample size had diminishing returns as adding more songs does not significantly increase the number of new words used for training after 10,000 songs.

2.3 Color and Emotion

Given that the recommender is already analyzing the auditory component of music to identify its emotional valence [10], an additional feature informing valence and arousal is proposed by this study: color. It is generally accepted that visualizing colors can evoke certain emotions within people. According to Solli & Lenz [21], the emotions experienced tend to be universal, unlike other visuals like faces or objects, as those tend to vary much more from person to person. They studied whether emotions are evoked in the same way with multicolored images versus single or paired colors. The researchers concluded that a "color emotion" can be perceived the same way with multicolored images, in general. The study intended to support the usage of an image retrieval method via color emotions.

This study intends to extract the dominant color(s) from album art and map them to valence and arousal which is the same emotional indicators used by Spotify®. The

"Color Thief" API was used for the extraction step, as it provided the best expected output when sampled by this team [5]. In exploring the connection between music and color, a motivating factor is a common exercise of tailoring specific store atmospheres to encourage customer behavior. According to Cheng et al., "environmental psychologists have determined that factors of an environment can work together [synergistically] to influence persons in the environment." They further assert that consumer activities can be motivated by a mixture of related influencers [22]. Their study tested for an association between physical versus cyber store fronts, in relation to how they can leverage music and color to influence customers. Their findings were that "both music and colour reveal significant effects on respondents' emotional responses" (Cheng et al., 2009, p.1) and that they work together to boost the effect on customer's emotions. This same logic is part of this study's rationale: by including the visual component of color, the valence and arousal features (music-derived) may be improved for Spotify's® recommender system.

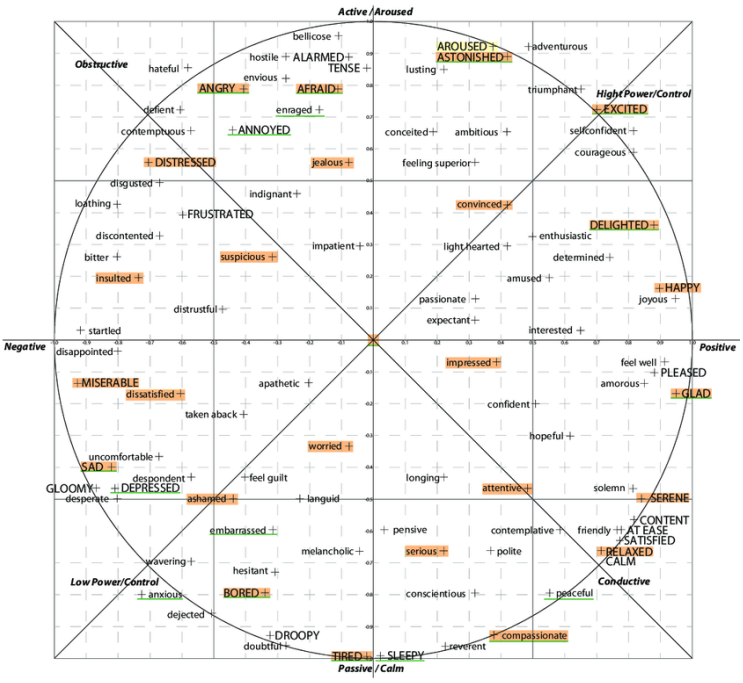


Fig. 1. Russell's Circumplex Model of Affect

This study mapped color to an emotional plot via a combination of Itten's color wheel and Russell's Circumplex Model (see Figure 1) [7]. This study's color guide is based on Itten's color wheel, which was created by a Swiss art theorist, Johannes Itten. It is one of the most common color relationship representations used today [23]. Color was transformed using a coordinate map called Russell's Circumplex model ("RCM") to provide valence and arousal. The RCM was developed by James Russell, Jonathan Posner, and Bradley Peterson to demonstrate that emotions are determined by an

interaction between "two fundamental neurophysiological systems […] valence [and] arousal." (Posner, Russell, & Peterson, 2005, p.2). This is the same type of emotions that are determined in this study's music and lyrical sentiment analysis.

To take the first step from color to valence, the Itten's color wheel can be transformed and combined with the RCM to provide two-dimensional coordinates (x = valence, y = arousal) of the associated emotion. This transformation was developed by Fagerberg et al. as part of an effort to implement an emotional recognition & expression technology that could be built into SMS (text messages) called "eMoto" [24]. The project intended to translate users' physical gestures into colors representing their associated emotions, to then be added to messages. This was achieved through participants' providing their preferred colors for association with their messages to communicate context to the recipient. Their study used physical feedback to allow emotional navigation throughout the RCM's two axes which output the associated colors. In contrast, this study used color as the input with the valence and arousal to be extracted for statistical comparison to the music measures.

This paper aims to establish if there is a statistically significant difference among music valence and arousal as stated by Spotify® API, and the respective variables derived from 1) color sourced from album cover art and 2) sentiment analysis of individual song lyrics.

3 Data

3.1 RGB (Color) Dictionary

There is a large, though finite number of RGB colors available (256^3); a three-dimensional library was created and be used as the reference for subsequent RGB to RYB conversion and dimension reduction methods. This color dictionary is represented as a three-dimensional cube that was rotated and transformed to reduce the dimensions down to a two-dimensional plane. Each color was mapped to a corresponding space in two dimensions, and scaled to map to RCM.

3.2 Album Art & Song Lyrics

The primary data source was the Spotify® API, which provides a rich library of audio features including music metadata about any artist, album, or track, including proprietary measures of valence and energy (arousal). The album art itself is provided as an album level feature. Additionally, to get song lyrics, the Genius® API was used which provides a host of lyrics and annotations across a wide range of genres.

At the lowest level, every song name and artist name was randomly searched using a single letter wildcard and extracted from the Spotify® API, which was the referenced as the search criteria in the Genius API. In total, this study includes 3,564 randomly selected songs, each representing a unique album. The filter used to narrow down the scope of this study includes only songs with album art, and lyrics containing only

English words. Interestingly, it was found that after 3,564 unique song ids, there was diminishing returns in the search criteria using Spotify® as the majority of the songs included both classical and non-English lyrics.

3.3 Sentiment Lexicon Knowledge Base

Instead of building a custom linguistic knowledge base of sentiment headwords from scratch, a series of existing toolkits and knowledge bases were used to provide pre-labeled sentiment scores. The VADER (Valence Aware Dictionary and sEntiment Reasoner) was used as the primary labeler for valence values consisting of a rule-based model of evaluating corpus sentiment. SentiWordNet was also used with VADER as an additional sentiment scorer that extends the WordNet database with roughly 100,000 auto-generated synsets to extract word level senses.

Lastly, an expanded version of the ANEW (Affective Norms of English Words) dictionary known as the NRC VADER lexicon was used to provide both labeled arousal and valence values. This sentiment dictionary, compiled in 2018 by Mohammad et al. [25] is by far, the largest manually created lexicon with roughly 20,000 English words and their measured sentiment scores.

4 Methods

4.1 Color Mapping:

4.1.1 Extract Dominant Album Art Color via Color Thief

The Color Thief API [5] allows an image to be loaded and analyzed, returning a color palette of six dominant color values in the form of RGB. Seen below is a custom image analyzed this way where first non-neutral dominant color is plotted in matplotlib.

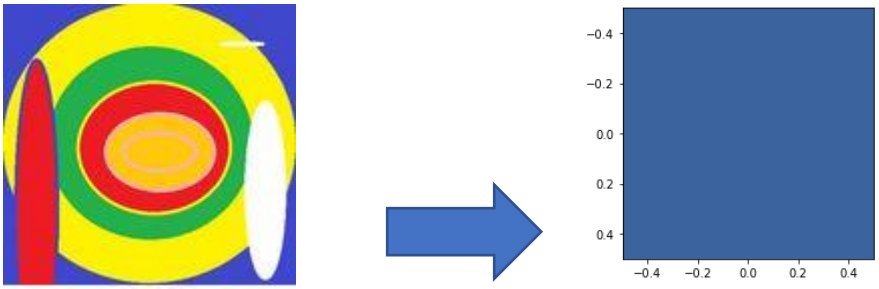


Fig. 2. Dominant Color Extraction by Color Thief, with Plotted Outcome

Because the world is mostly portrayed in neutral or muddy colors nearest to the center of the color space, with splashes of color here and there, this study aimed to find the first predominant, non-neutral color in the color palette.

In order to do so, each of the six extracted dominant colors measured distance from center by finding the difference between each of the three colors, Red, Green, and Blue, and taking the average, where distance from center of 0 is perfectly neutral. The distance from center/neutral is stated as:

$$\text{avg}(\text{abs}(R-B),\ \text{abs}(B-G),\ \text{abs}(G-R)). \quad (\mathbf{1})$$

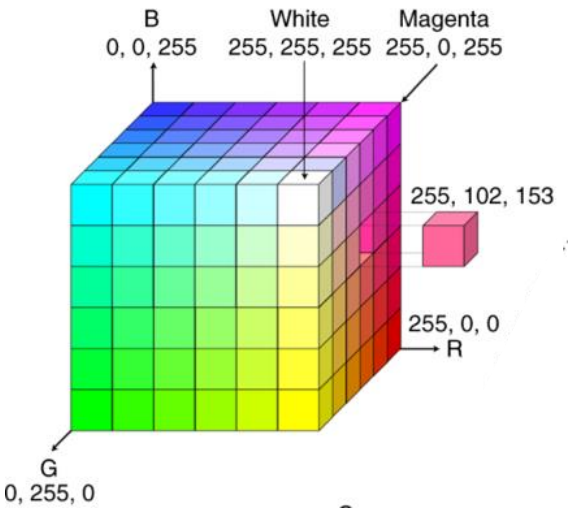
It was observed that approximately $1/11^{th}$ (or 23/256) of the color space near the center was considered neutral or indiscernibly neutral by this study. This study took the first color that was non-neutral, or center from distance was greater than or equal to 23. This extracted non-neutral color was then used to measure valence and arousal related to the album artwork.

In the instances where all six colors were considered neutral, then the most predominant neutral color of the palette was taken to measure valence and arousal.

While there are limitations to this methodology, including noted instances where a non-neutral color is clearly not dominant, but is taken as the color to analyze valence and arousal, this paper recognizes these shortcomings and notes these cases as potential outliers that could benefit from human review.

4.1.2 Reduce RYB Dimensions to Two Dimensions by Converting to HSV (Hue, Saturation, Value) Color Space

To reduce the number of dimensions of the RGB color cube (see Figure 3) from three to two, this was accomplished by converting to HSV color space. This effectively puts shade/tint/brightness on the same central Z-axis so that all neutral colors, black to white, were aligned following [0, 0, n] (see Figure 4).



Woolf, M.S., Dignan, L.M., Scott, A.T. et al. Digital postprocessing and image segmentation for objective analysis of colorimetric reactions. Nat Protoc 16, 218–238 (2021).

Fig. 3. Because color is mapped in three dimensions as noted in Nature [26], RGB dimensions were reduced to two in order to map Itten's Original Color Wheel [23], and subsequently transformed to map to Russel's Circumplex Model [4].

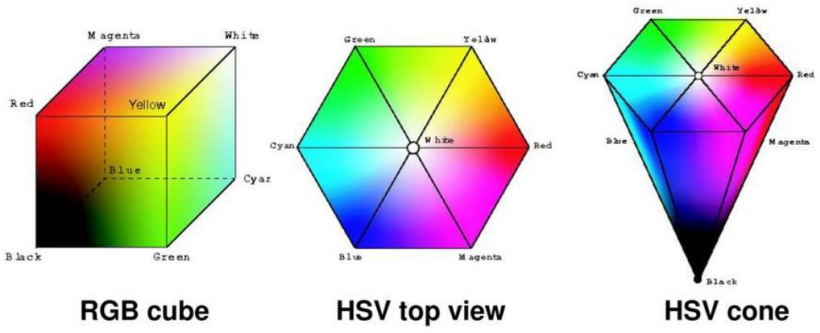


Fig. 4. A representation of the HSV color space projected down into two-dimensions. Neutral colors from black to white are placed on the same axis perpendicular to the plane of this paper/monitor [27].

4.1.3 Data Loss with Dimension Reduction

While reducing the number of dimensions, data and color mappings were lost. This study focused on intentionally losing data specific to brightness (shade/tint) and was accounted for by aligning black and white coordinates to the same neutral axis.

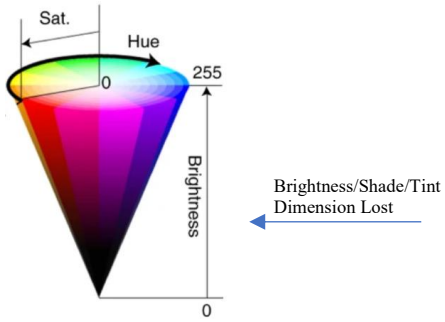


Fig. 5. [25] The reduction resulted in an intended loss related to brightness (or shade/tint) from black to white as a result of rotating the original color cube so all neutral colors were aligned on the same axis (z). The vertical dimension noted above. This means that albums with perfectly neutral predominant colors such as blacks or grays or whites, default to neutral mappings of valence and arousal in a two-dimensional space. All neutral colors are considered the same (neutral) in the realm of reduced dimensions. Neutrality is a state of emotion that was assessed.

In instances where an album art's predominant color is perfectly neutral then the song was scored as neutral; or having valence of or near 0 and arousal of or near 0. All other non-neutral colors were mapped to valence and arousal within the two-dimensional plane of Russell's Circumplex Model.

4.1.4 Convert RGB Dictionary to RYB

Given that Itten's Original Color Wheel [23] references subtractive primary colors of Red (R), Yellow (Y), and Blue (B) or RYB, and current technology references colors of light, or additive primary colors of Red (R), Green (G), and Blue (B) or RGB, this research converted the color library's RGB colors to RYB using trilinear interpolation as described in Paint Inspired Color Compositing by Gossett et al [6].

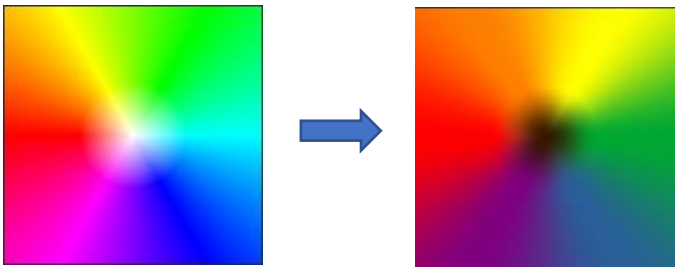


Fig. 6. Converting RGB to RYB

It should be noted that during the conversion from RGB to HSV in two-dimensions and subsequently to RYB, not all colors have a direct mapping such as magenta or cyan, or colors from the vertical dimension. Blues are less vivid, purple is more mundane, and the red, yellow, orange spectrum is much broader. This means that there were cases of extracted color from the album art that didn't map to this paper's RYB valence and arousal space. In order to mitigate these instances, this study used K-nearest Centroid clustering to classify an RGB value in its closest RYB equivalent.

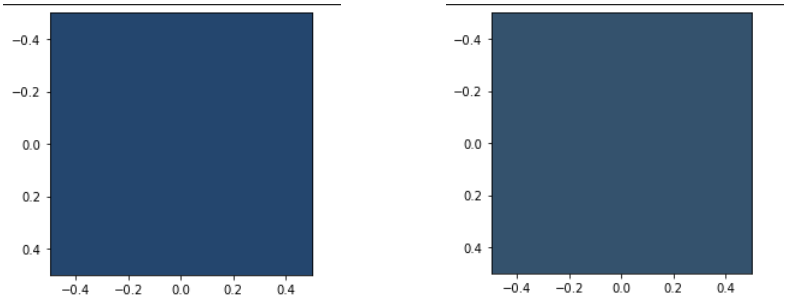


Fig. 7. On the left, the extracted predominant non-neutral color from an album. It was ***not*** found in the color dictionary, however on the right is its closest match from the RYB color dictionary based on Nearest Centroid clustering. In this case, the valence and arousal values were used from the color on the right.

4.1.5 Rotate to Match Ståhl, et al [24].

In order to map to Ståhl, et al.'s [24] color wheel, it was necessary to rotate the hue 300 degrees counterclockwise. By proxy, this study respects that of Itten's Original Color Wheel [23].

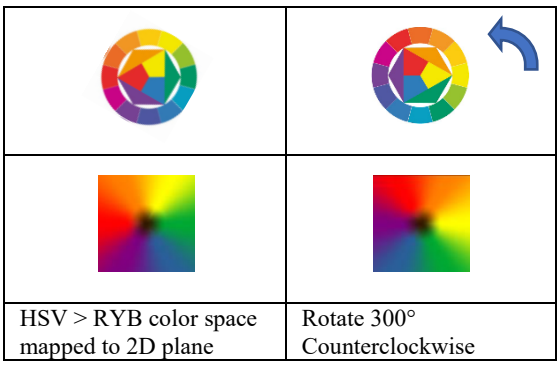


Fig. 8. In previous works, most notably Ståhl, et al. [24], Itten's Original Color [23] was transformed to map color to the emotions of Russel's Circumplex Model with low valence and arousal colors in the purple/blue space, and high valence and arousal colors being represented in the red/orange space. This paper aims to utilize Ståhl's previous mapping by performing and applying the following two transformations to the two-dimensional color data; rotate 300° counterclockwise.

By mapping the three-dimensional RGB colors to a valence/arousal dictionary in two-dimensions, an RGB value from an album could then be mapped to valence and arousal by simply getting RGB from the color mapping created. Each song's ID was then mapped to a color-based valence and arousal value to compare against the baseline and lyrical sentiment analysis measures.

4.2 Semantic Lyric Sentiment

4.2.1 Lyric Preprocessing

The lyrics were extracted from the Genius API based on the reference song found in Spotify® API. These lyrics are assumed to be accurate as Genius allows artists to verify the lyrics directly. However there have been cases found where an editorial's annotation may differ in interpretation though the original lyrics itself were sound. As a result, there was no major changes required for the lyrics as the extracted version of the lyrics required only preprocessing. The song parts that provide structure to the song which separates the sections (ex. 'Intro', 'Verse', 'Chorus', etc.) however was removed from the song as these break points were not intrinsic to the song lyrics themselves. A general language filter using the langdetect [28] Python package ported from Google's language-detection library was used as a final filter to filter out any remaining non-English words existing in songs that the initial Spotify® left in.

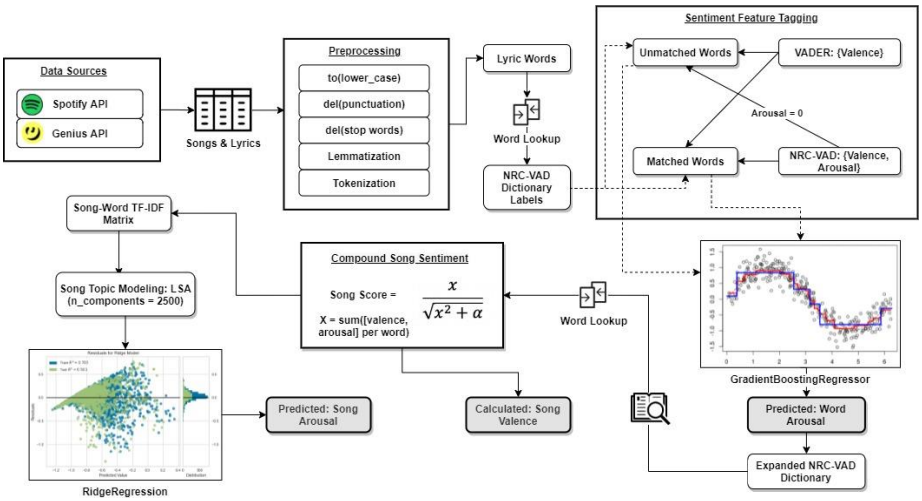


Fig. 9. Lyrical Sentiment Analysis Pipeline

Each song's lyrics were then normalized by removing any punctuation and stop words. Lemmatization was also performed as this applied a low-level standardization to the lyrics to reduce the amount of word variations found in each song. No stemming was performed as the original form of the word was important since past and future tense had differing sentiments for certain words and word phrases.

Word-level features were then processed to extract the number of syllables in a word, the number of senses found using WordNet, as well as the part-of-speech using NLTK's POS-tagger. Prior research had found that only POS tags including "nouns, verbs, adjectives, and adverbs can present an emotional meaning" (Albornoz, et al. 2010, p. 5) in words, thus word filtering based on only these tags were processed with the final feature set of words consisting of: <word, number of synsets, syllables, POS tag, valence, and arousal>.

4.2.2 Predicting Word Arousal

A hybrid approach (see Figure 9) of both Sentiment Tools and Machine Learning was applied to the corpus of lyric data. Multiple lexical knowledge-based tools were used, namely the VADER and NRC-VAD models, theorizing that overlap would exist at a word level. This was further proved using a two-sample t-test for all 8,194 song lyrics, finding that there were no significant differences between the measured valence and arousal values (p-value = 0.127) using either approach. However, in application, a proposed method of ensembling the sentiment scores from VADER and NRC-VAD of a given word did not work (Figure 10) since VADER measured 87% the NRC-VAD word corpus as neutral (0 overall sentiment) though the distributions of the two were very similar.

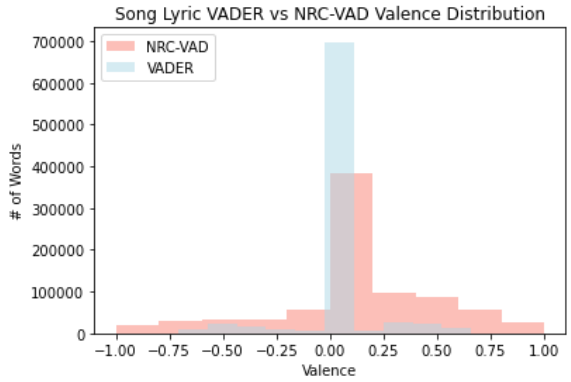


Fig. 10. VADER vs NRC-VAD Lyric Distributions. NRC-VAD's labeled sentiment have a more normal distribution of Valence values though in effect, applying sentiment scores to otherwise neutral words as shown by VADER's distribution resulting in higher, more inflated valence scores.

This hybrid approach was done to first utilize pre-existing labeled sentiment scores which prior research had shown success in predicting overall text sentiment [34]. However, more importantly, since emotion is defined as a two-dimensional plane consisting of both valence and arousal per the Russel's Circumplex Model [4], this revealed a major drawback with many of the existing popular text sentiment tools since most out-of-the-box sentiment analyzers conveyed emotion as a one-dimensional value of valence alone.

To address this unrepresented dimension of emotion in song lyrics, a proposed method of using the NRC-VAD labeled sentiment which contains manually labeled arousal values was used as a primary feature in addition to other word features extracted during preprocessing. These features were then used to predict lyric word arousal values that existed outside the NRC-VAD's labeled dictionary. It was found that using a Gradient Boosting Regression model provided the best results which provided the overall lowest R2, mean-squared average (MAE), and mean-square error (MSE) fit among the competing Support Vector (SVR), Random Forest, Ridge and baseline Linear Regression models. Ultimately, using the R2 metric as a measure of overall fit, it was found that a multiple regression approach to using word properties as predictive features resulted in a nonlinear relationship between the properties of a word and its arousal sentiment, with high variance found in the residuals. This indicated that measuring word arousal on the properties of a word alone to measure arousal was not sufficient.

Table 1. Predicting Word Arousal using Word Features

Word Prediction Models	R^2	MAE	MSE	RMSE
Gradient Boosting Regression (GBR)	0.203	0.244	0.093	0.306
Support Vector Regression (SVR)	0.194	0.709	0.800	0.894
Ridge Regression	0.1027	0.262	0.105	0.324
Linear Regression	0.1026	0.262	0.105	0.324
Random Forest	0.023	0.267	0.115	0.338

Despite a lower-than-expected measure of fit on the measured word arousal values, it did provide usefulness at providing data points on an otherwise unknown variable. By using the predicted word arousal values found between the NRC-VAD dictionary and song corpus, a small portion of lyrics were then used to build an expanded version of the NRC-VAD labeled dictionary to account for a wider range of lyrics. By expanding the original NRC-VAD dictionary of 20,006 unique words to a much larger dictionary of 51,886 unique words, this provided a much larger corpus of words that would've otherwise been excluded. Using the expanded sentiment dictionary, an overall sentiment compound score was calculated for each song. A normalizing function defined below was applied to maximize the overall polarity of a song' sentiment between the range of the Russel's Circumplex two-dimensional plane of emotion [-1,1].

$$Compound\ (Valence, Arousal)\ Score = \frac{x}{\sqrt{x^2 + \alpha}} \quad (\mathbf{2})$$

4.2.3 Topic Modeling and Song Arousal Predictions

Once the entire corpus of lyric sentiments was labeled using both a dictionary and predicted approach, the lyrics were then transformed into a TF-IDF (term frequency-inverse document frequency) sparse document-term matrix. This conversion from words to word vectors provided the frequency of each word in an individual song based on how often it appeared in the entire corpus of songs and its lyrics. Considering the dataset comprises a total of 8,194 songs, the size of this matrix resulted into a tall and fat matrix structure with 43,647 words as column vectors. As a result, latent semantic analysis (LSA) was proposed to address the issues of potential collinearity existing in the sparse TF-IDF matrix by utilizing an unsupervised latent semantic analysis (LSA) model to reduce the overall dimensionality of the matrix into a smaller, denser topic-based vector space. It was found that using 2500 components or topics ($1/8^{th}$ the size of NRC-VAD word base) explained 77% of the variance with the first topic holding the most model weight.

Table 2. Predicting Song Arousal using Topic Modeling

Song Prediction Models	R^2	MSE	MAE	RMSE
Expanded Dictionary LSA	0.531	0.142	0.041	0.202
Expanded Dictionary TF-IDF	0.521	0.143	0.042	0.204
Non-expanded Dictionary TF-IDF	0.499	0.212	0.081	0.284
Non-expanded Dictionary LSA	0.490	0.214	0.082	0.287

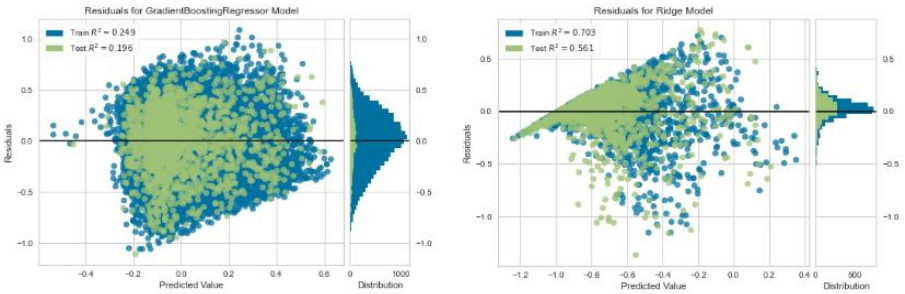


Fig. 11. Word Feature (left) vs Topic Model (right) Arousal Predictions.

Variations of the LSA topic based model and TF-IDF model (Table 2) were then fed into separate Ridge regression models to predict on a continuous [-1, 1] scale the final song arousal values. The model results showed that by converting a word feature model into word vector space, this approach nearly tripled (R^2 = 0.531) the original model's fit (R^2 = 0.203), explaining half of the overall variance found in the predicted outcomes along with lower overall error as well. Though there were insignificant differences between that of the TF-IDF model and the LSA model, by comparing model run times, the topic-based model took 10 times faster to run on a fraction of the TF-IDF feature set, making it the preferred approach. It should be noted that though promising, by increasing the number of features included in the model in both the LSA and TF-IDF model approaches, there is the possibility that the dramatic increase in overall fit could be due to the model fitting the subspace of values by decreasing the sum of squared estimate errors for every increase in estimated coefficient. In other words, by simply throwing more predictors (words) at the model, a word matrix-based approach may be suspect to some level of overfit. Ultimately, the expanded dictionary LSA model with final predicted arousal values along with VADER's compounded valence values were then used for final comparison back to color and Spotify's® sentiment scores.

5 Results

This experiment derived arousal and valence (N = 3,564) from two sources: dominant color in album cover art, and sentiment analysis of lyrics. Both features from each

source were compared to those provided by the Spotify® API, individually using the pairwise t-test as well as collectively using One-Way Repeated Measures Analysis of Variance (rANOVA). These techniques were chosen because of the lack of independence across the variables, as each measurement from each source corresponds to the same respective song. For example, "Across the Universe" by The Beatles from the album Let It Be, would have three arousal and three valence measurements: Spotify®, dominant album art color, and lyrical sentiment.

Before proceeding with the testing of the data, an EDA was performed to check that it meets the assumptions: continuous dependent variable, observations independent of one another, dependent variable normally distributed, and no outliers.

Firstly, valence and arousal are being measured on a continuous scale.

Secondly, the selections of tracks and albums were done randomly to ensure independence of measurements within each dependent variable.

Thirdly, each variable was plotted in a histogram to check for normal distribution.

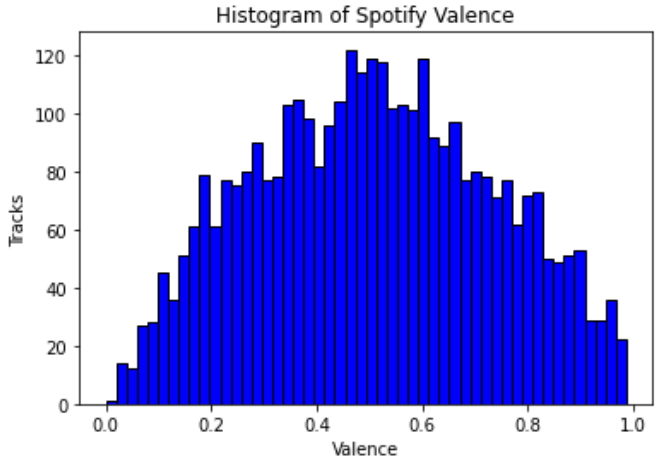


Fig. 12. Spotify® Valence Histogram

In Figure 12, you can see that Spotify's® valence is normally distributed.

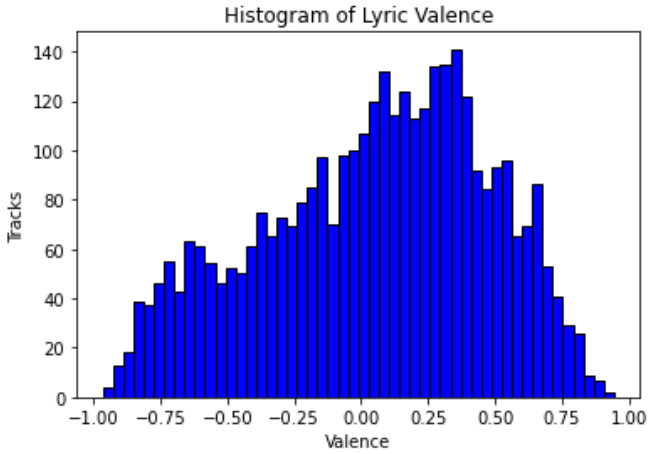


Fig. 13. Lyric Valence Histogram

Lyric Valence shows a slight skewness to the left, but this team did not deem it significant enough to require transformation.

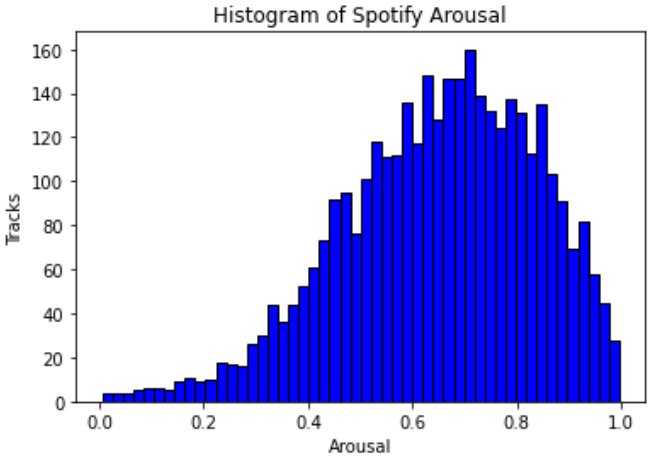


Fig. 14. Spotify® Arousal Histogram

The above figure shows Spotify's® arousal values, which are clearly left skewed. In order to normalize the data, it was squared before used in the test. All data is scaled to -1:1 before testing.

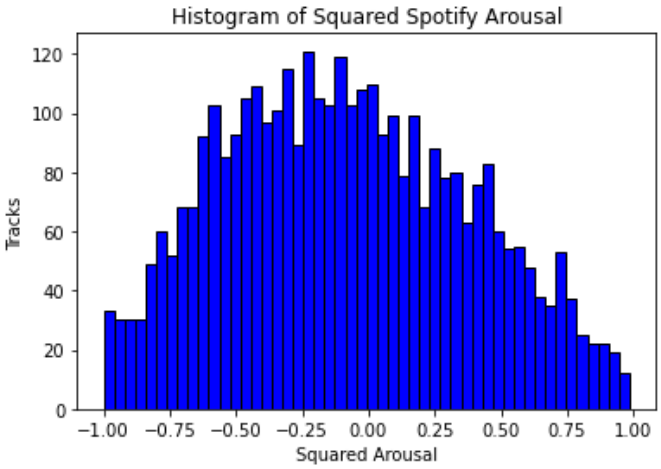


Fig. 15. Squared Spotify® Arousal Histogram

The resulting histogram is much more normal.

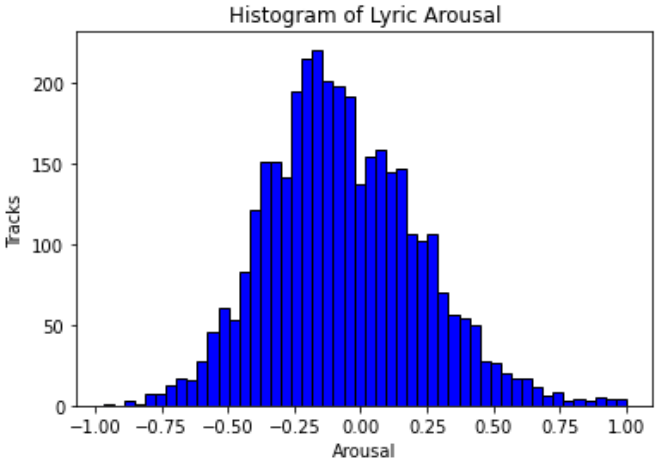


Fig. 16. Lyric Arousal Histogram

The Lyric Arousal is mostly normally distributed.

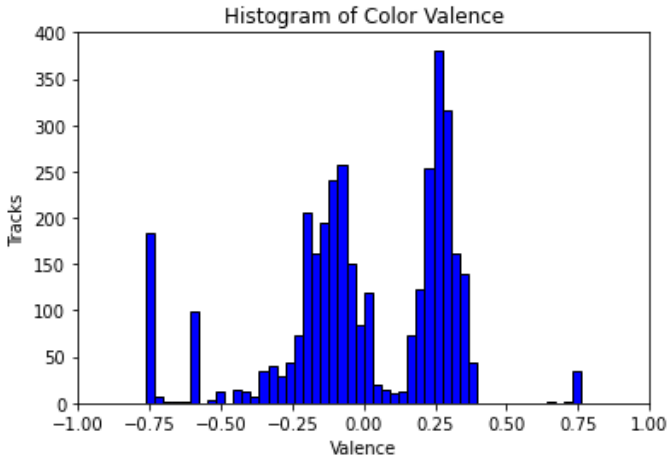


Fig. 17. Color Valence Histogram

The color valence is a bimodal distribution, but based on the number of samples to be tested (n=3564), this team decided the ANOVA and t-test to be robust to the normality assumption and thusly proceeded with the data as-is.

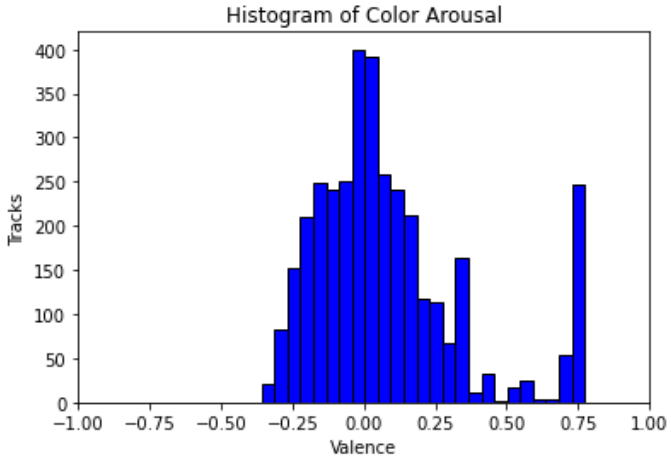


Fig. 18. Color Arousal Histogram

The color arousal distribution would be mostly normal, however there is a large concentration of values on the higher end. This presents a similar issue to a bimodal distribution, however again this team proceeded with the belief that the tests will be robust to these normality violations.

Lastly, boxplots were created to check for outliers. Skew was already noted in some variables so outliers were expected in the boxplots.

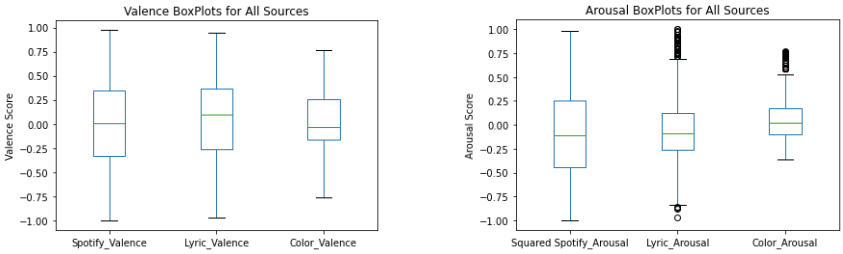


Fig. 19. Valence and Arousal Boxplots for all Sources

In order to diagnose the influence of these outliers, to determine if it is necessary to remove them, linear model residuals for each valence and arousal were plotted against the Spotify® baseline variables using R. A Cook's distance of .1 was used as a threshold for determining if a point was influential enough to warrant removal. None of the variables contained a point that exceeded this threshold.

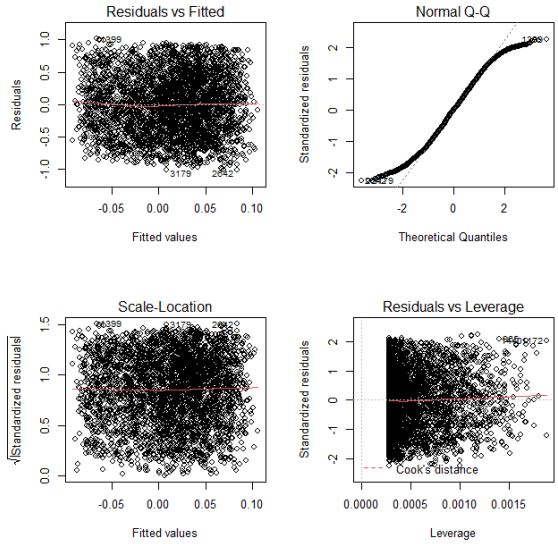


Fig. 20. Lyric Valence Residuals

As you can see, the residuals and squared residuals look to be sufficiently random across the zero line with a random scatter. The QQ plot shows excellent adherence to the baseline for most of the points. Lastly, it's clear that there aren't any noticeable points with high influence, although there is data with high leverage.

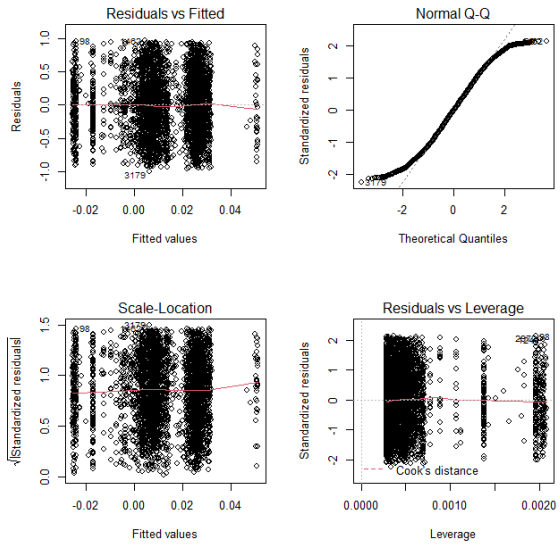


Fig. 21. Color Valence Residuals

As you can see, there are striations within the residuals, and this can be attributed to the method by which color space was used to transform RGB into valence. Because Nearest Centroid was required due to an incomplete dictionary, the continuous data had to be placed into its nearest color. Aside from these striations however, there are no outliers that have significant influence, and although the data is bimodal, the residuals follow a very linear path on the QQ plot. Notice that the fitted residuals plots are lacking in much presence on the higher end, which might inform improvements for future study.

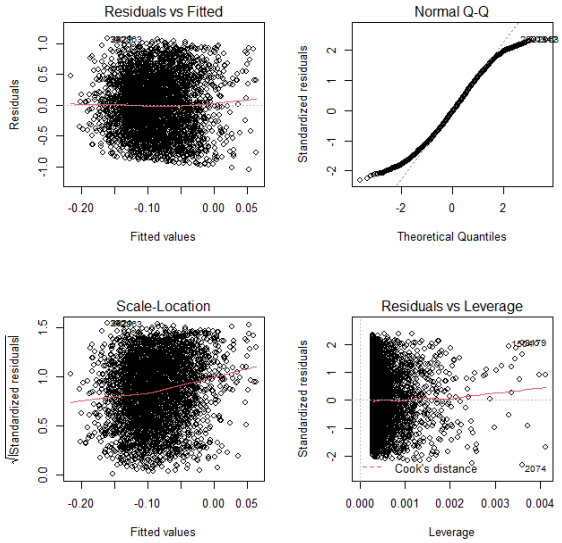


Fig. 22. Lyric Arousal Residuals

Lyric arousal looks excellent, with a mostly linear baseline around which the residuals are randomly scattered. The QQ plot also shows sufficient adherence to the baseline, and there were no points that violated Cook's distance threshold.

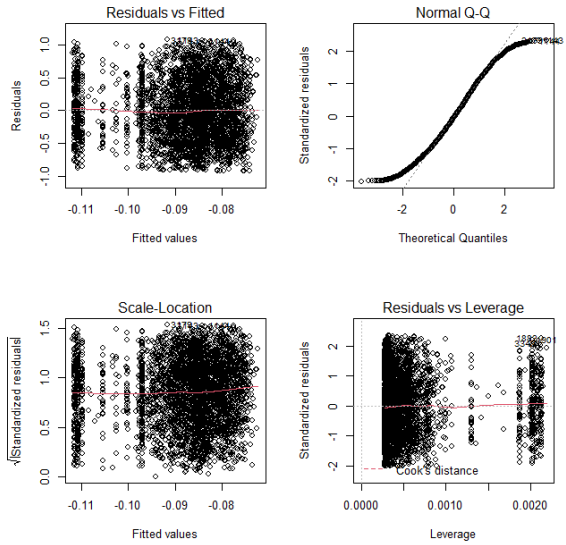


Fig. 23. Color Arousal Residuals

Lastly, here are the residuals for color arousal. Again, the striations can be attributed to the need for rounding certain data. Because of the valence and arousal coming from the same plane of measurement, it also is interesting that the data is tending to bunch up on the right. However, the baseline is again very linear, the QQ plot is mostly linear, and the points with leverage are not exerting influence on the regression.

Aside from the bimodal distribution and the unique behavior of the color measurements, this team agrees the assumptions are sufficiently met, and proceeds with repeated-measures testing.

rANOVA:

First, Repeated Measures ANOVA was performed on all three sources in the data for each variable compared to the baseline Spotify® measurements. For arousal and valence both, there was found to be a statistically significant difference within subjects (songs). The arousal test yielded an F-Score of 260.8, with a p value of < .0001. The valence test had an F-Score of 13.19, with a p value of < .0001. These results indicate that the pairwise t-tests should indicate a statistically significant difference between at least two of the sources for both arousal and valence.

Valence

- **Null Hypothesis 1** (H_{v0}):

  $\mu(\text{Spotify® Valence}) = \mu(\text{Lyric Valence}) = \mu(\text{Color Valence})$

  - **Alternative Hypothesis** (H_{va}):

  At least one Valence Mean is statistically significantly different from the rest.

Arousal

- **Null Hypothesis 2** (H_{a0}):

  $\mu(\text{Spotify® Arousal}) = \mu(\text{Lyric Arousal}) = \mu(\text{Color Arousal})$

  - **Alternative Hypothesis** (H_{aa}):

  At least one Arousal Mean is statistically significantly different from the rest.

Table 3. ANOVA – Arousal and Valence per Color, Music and Lyrical Features

Music Feature	Sum of Squares	Df	F	PR(>F)
Valence	3.93	2	13.19	<.0001
Arousal	61.64	2	260.80	<.0001

Pairwise Comparisons:

Valence

- **Null Hypothesis 3** (H_{v0}):

  $\mu(\text{Spotify® Valence}) = \mu(\text{Lyric Valence})$

  - **Alternative Hypothesis** (H_{va}):

  The mean valence of the baseline is not equal to the mean valence of lyrics.

- **Null Hypothesis 4** (H_{v0}):

  $\mu(\text{Spotify® Valence}) = \mu(\text{Color Valence})$

  - **Alternative Hypothesis** (H_{va}):

  The mean valence of the baseline is not equal to the mean valence of color.

Arousal

- **Null Hypothesis 5** (H_{a0}):

  $\mu(\text{Spotify® Arousal}) = \mu(\text{Lyric Arousal})$

  - **Alternative Hypothesis** (H_{aa}):

  The mean arousal of the baseline is not equal to the mean arousal of lyrics.

- **Null Hypothesis 6** (H_{a0}):

  $\mu(\text{Spotify® Arousal}) = \mu(\text{Color Arousal})$

  - **Alternative Hypothesis** (H_{aa}):

  The mean arousal of the baseline is not equal to the mean arousal of color.

In the pairwise comparisons, Color Arousal was tested against Squared Spotify® Arousal. There was a major statistically significant difference between them (t-stat = -19.02, 95% CI [-.19, -.15], p = <.0001). However, Color Valence was not found to be statistically significantly different from Spotify® valence (t-stat = 1.38, 95% CI [-.01, .03], p = .168).

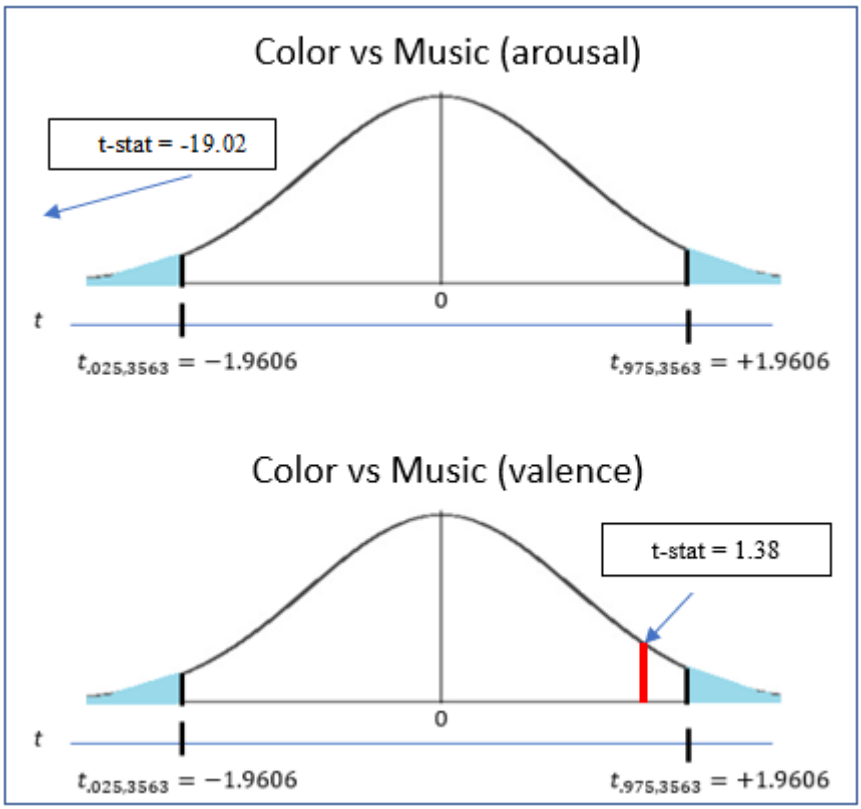


Fig. 24. Student's t Curves for Color vs Music

Lastly, lyric arousal and Spotify® arousal were compared, resulting in a statistically significant difference, though not extreme (t-stat = -2.29, 95% CI [-.04, -0.0], p = .022). Valence between the two sources also showed a statistically significant difference, but more extreme (t-stat = -3.36, 95% CI [-.05, -.01], p = .0008).

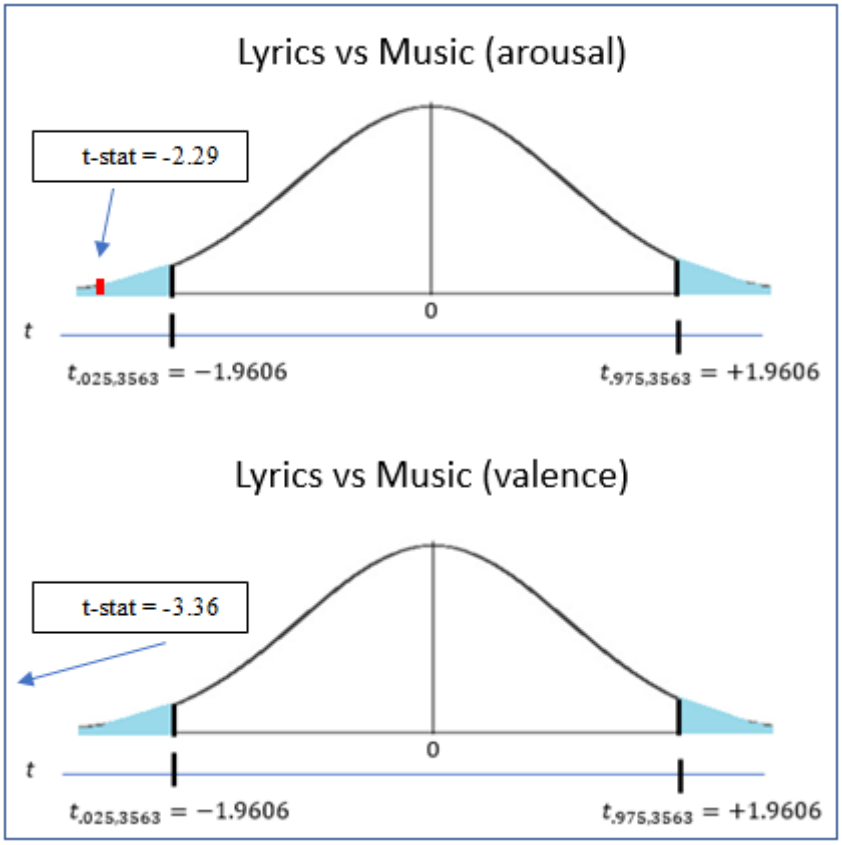


Fig. 25. Student's t Curves for Lyrics vs Music

Table 4. Paired t-tests – Arousal/Valence of Color and Lyrics against the Spotify® Baseline

Test Against Baseline	t-stat	95%CI	P-Value	Outcome
Color Arousal	-19.02	[-.19, -.15]	<.0001	Reject
Color Valence	1.38	[-.01, .03]	.168	Fail to Reject
Lyric Arousal	-2.29	[-.04, <0]	.022	Reject
Lyric Valence	-3.36	[-.05, -.01]	.0008	Reject

6 Discussion

This paper aimed to establish if there was a statistically significant difference among music valence and arousal as stated by Spotify® API, and the respective variables derived from 1) color sourced from album cover art and 2) sentiment analysis of individual song lyrics.

Lyrics' metrics were found not to match Spotify® for both variables, which suggests either the methodology needs improvement, or that these features may be useful together in a different way as unidentified features. The same could be said about color, except that one variable was like Spotify's whereas the other was not. This makes it harder to determine if it is the methodology of transformation, collection, or bias that is more in need of improvement, as valence and arousal exist on the same plane of the circumplex. It is fascinating however that of all the features, color valence ended up being the only one that gave us results we were hoping for. This may suggest that music's effect on the level of positivity is either stronger or more consistent than on excitement, or simply that album cover artists use color theory and are decently good at what they do. With a revised approach, this could be a valuable source to include in creating the recommendation algorithm's valence.

This study encountered a few limitations in the methodology and data collection. One difficulty encountered was the dominant color extraction. Aside from songs that have no associated album art, due to the nearly limitless variation of images that can be found on album art, generalizing the most dominant color across all covers assumes that certain colors are more dominant than others. In other words, non-neutral colors were favored even though they potentially might only be a small portion of an album's cover art. Along the same lines, the color transformation was a novel approach, and wasn't able to capture all 256^3 possible variations that could be found within the RGB color space. In fact, there is data loss in the conversion that moves from the RGB to RYB color space which lets it map to two dimensions. More work is clearly necessary in aligning the more common RGB with RYB and the planar model. A side effect of the limitations in the color transformation was the bimodal distribution of valence. This is assumed to be a result of how the colors must be relocated to the new space, but may very well be attributable to what was deemed the dominant color in the album art or that albums skew heavy in particular color palettes and there could be a bias due to a general human preference of colors used in these images.

The sentiment analysis was hindered by a few factors: English-only words were used, slang varies the meaning of words across songs, and with the rapid evolution of language hastened by social media, the variable interpretations should change over time. Fortunately, there was a pre-existing approach for collecting valence of text, but it was not designed for music, so there may be lost meaning since language changes context across media. Arousal was not a readily derived feature, so its creation was more ad hoc than valence's. With the novelty of approach and application of this sentiment analysis, it would need further exploration to provide more confident values.

Spotify® was used as the baseline by which to compare the variables created in this study which prevents more widespread use of the methods performed here, as their validity is based solely on the assumption that the baseline is the gold standard. Additionally, due to the mystery of how Spotify® creates their variables, it is entirely

possible that sentiment analysis of lyrics and album art are already being used in their internal analyses.

Lastly, although randomly selected, only a single track was assessed per album, so expanding to include more tracks from an album may yield results counter to what this study has gathered. Similarly, the body of available music that was accessed through the API doesn't contain everything, so possibly niche markets may not have been available for including in the dataset.

6.1 Ethical Concerns

Misclassification of emotional measures within music platforms could indicate to recommendation systems the wrong emotions. This could slowly lead to moving the user away from intended emotions, or in some cases to emotional extremes.

Emotional misclassification could be a result classifying only on a singular emotion, as is the case with this study's album art methodology for mapping color to emotion, namely clustering to nearest representation from RGB to RYB, for example. Similarly, in the instance of lyrical emotional measures, moving users through emotional spaces based on similarity vectors between words.

In the opposite direction of moving users to extremes, there exists the possibility that emotional recommendations would drive users into a silo like state or keep them in the same emotional space; an emotional echo chamber based on lyrics or album art color.

Moving users to extreme emotions could have lasting, real world effects on the users as well as the world around them. The same could be observed for users who get stuck in emotional echo chambers. There are a lot of psychological underpinnings when it comes to mapping and manipulating emotions based on lyrics or album color and listening patterns.

Beyond emotional extremes and echo chambers, tracking or monitoring an emotional state could be used to target individuals in harmful and invasive ways. Users could be targeted for advertising or by government agency or intervention such as when Facebook and Cambridge Analytica used psychographic targeting in the 2016 US Presidential Election and similarly Brexit in the UK. While there are plenty of companies that would not allow the exploitation of mental health instability of their users, there are plenty of unethical companies that would.

Platforms could also use emotional manipulation to drive specific outcomes, either with advertisers within the platform, revenue driving metrics within the platform, or outside the platform with real world consequences. Real world outcomes that could be derived from irrational emotions.

Beyond advertising and platform manipulation, governments could subpoena users' emotional maps to target and influence individuals in ways that might violate their human or civil rights, or worse death or imprisonment. Similar to concerns regarding governments seizing records in DNA databases and targeting genetic markers, governments could use emotional maps to target users of specific emotional markers of their perceived enemies.

6.1.1 Future Study

With respect to color and generalizing the most dominant non-neutral color from album art, this study's methodology ignored value in assessing neutral colors and their respective neutral emotions where neutral colors were the dominant, but a non-neutral color was present. Non-neutrals were given preference. Future work could focus on fine-tuning parameters to determine if a mostly, or overwhelmingly neutral album, with a splash of color, maps effectively to neutral emotions.

With respect to mapping the three-dimensional RGB color cube to two dimensions, colors were subsequently lost. Additionally, the conversion between RGB and RYB resulted in further color loss. Future works could find a way to map all RGB colors into a two-dimensional plane. Additionally, mapping RGB (instead of RYB) colors to emotions might open a whole new world for mapping color to emotions. Combining the two, would result in evaluating the emotions of all ~16.8m colors of the color cube.

With respect to the bimodal distribution of valence found in color extraction from album art, future works could evaluate the why behind these distributions. Are these colors and their respective emotions really that popular in the visualization of album art? Is this purely a coincidence regarding this paper's sampling method?

With respect to lyrical lexicons, given the lack of arousal measurements provided by the majority of sentiment analyzers, a novel approach of predicting lyric word arousal was done by utilizing regression models to predict words outside the range of pre-labeled sentiment. Features for these regression-based models relied mainly on the properties of a word which naturalistically, are limited in nature. Results from this method indicated less-observable features beyond word properties alone played a much stronger role in predicting arousal. Though it was found that by moving from a regression-based approach to a topic-model increased overall predictive ability at a song level, a neural network or deep learning model could theoretically further improve word level predictions by utilizing nonlinear activation layers to better fit the model's variance.

Additionally, by observing the strong increase in the model's performance to predict arousal by transforming into vector space, it is speculated that models utilizing a graph-based method, such as measuring the probabilistic distribution between words, would also increase arousal predictions.

7 Conclusion

In today's competitive music streaming industry, the ability for listeners to consistently discover new content is one of the leading factors of success. A crucial step in the discovery process depends on how music is broken into organizable ingredients that create distinguishable features users could then apply as part of their overall experience. In addition, driving forward fresh ways to consistently curate new music allows for complex recommendation systems to further expand their abilities in producing a more personal experience for its consumer base.

This paper explores going beyond traditional means of extracting feature elements that are exclusive to characteristics of audio features by evaluating the significance of other properties inherent to almost all forms of music; the emotion that is evoked from a song's album art and lyrics. By extracting mood sentiment from the dominant color

of album art, there is evidence suggesting similarity to the sentiment measures Spotify defines. This finding opens new avenues for both music exploration and recommendation features utilizing a common property of music found throughout the majority of songs. Though in contrast it was found that the song's lyrics had measurably different statistical representations of the same baseline measurement of mood, there is room for further study on the meaning behind such discrepancy, which may reveal truths to the differences in how emotion is conveyed through various means. Ultimately, the research done on both the lyrics and the album color offer promising results for new approaches to music curation, recommendation and interaction based on the intrinsic emotional value hidden in acoustic evaluations of music.

References

1. Friedlander, J. P. (2020). Year-End 2020 RIAA Revenue Statistics. riaa.com. https://www.riaa.com/wp-content/uploads/2021/02/2020-Year-End-Music-Industry-Revenue-Report.pdf.

2. Maezawa, A., & Okuno, H. (2015). Bayesian Audio-to-Score Alignment Based on Joint Inference of Timbre, Volume, Tempo, and Note Onset Timings. Computer Music Journal, 39(1), 74–87. https://doi.org/10.1162/COMJ_a_00286

3. Jia-Min Ren, Ming-Ju Wu, & Jang, J. (2015). Automatic Music Mood Classification Based on Timbre and Modulation Features. IEEE Transactions on Affective Computing, 6(3), 236–246. https://doi.org/10.1109/TAFFC.2015.2427836

4. Russell J A 1980 A circumplex model of affect J. of Personality and Soc. Psych.39(6) pp 1161-78 doi:10.1037/h0077714

5. Dhakar , L. (n.d.). color-thief Version (2.3.2). MIT License.

6. Gossett, N., & Chen, B. (n.d.). Paint Inspired Color Compositing. Minneapolis/St. Paul ; University of Minnesota at Twin Cities.

7. Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and Psychopathology, 17(3), 715–734. https://doi.org/10.1017/S0954579405050340

8. Agarwal, G., & Om, H. (2021). An efficient supervised framework for music mood recognition using autoencoder-based optimised support vector regression model. IET Signal Processing, 15(2), 98–121. https://doi.org/10.1049/sil2.12015

9. Alsted, J. H. (1664). Templum musicum, or, The musical synopsis of the learned and famous Johannes-Henricus-Alstedius being a compendium of the rudiments both of the mathematical and practical part of musick, of which subject not any book is extant in our English tongue / faithfully translated out of Latin by John Birchensha.

10. Spotify® AB. (2021). Web API Reference. Spotify® for Developers.https://developer.spotify.com/documentation/web-api/reference/.

11. Chapaneri, S., & Jayaswal, D. (2020). Deep Gaussian processes for music mood estimation and retrieval with locally aggregated acoustic Fisher vector. Sadhana (Bangalore), 45(1). https://doi.org/10.1007/s12046-020-1313-8

12. Tan, K., Villarino, M., & Maderazo, C. (2019). Automatic music mood recognition using Russell's two dimensional valence-arousal space from audio and lyrical data as classified using SVM and Naïve Bayes. IOP Conference Series. Materials Science and Engineering, 482(1), 12019–. https://doi.org/10.1088/1757-899X/482/1/012019

13. Raschka S 2014 MusicMood: predicting the mood of music from lyrics using machine learning (Michigan: Michigan State University) Preprint arXiv:1611.00138

14. Ascalon E I V and Cabredo R 2015 Lyric-based music mood recognition DSLU Res. Congress 2015 De La Salle University, Manila, Philippines

15. Oh S, Hahn M and Kim J 2013 Music mood classification using intro and refrain parts of lyrics 2013 Int. Conf. on Inform. Sci. and Appl. (ICISA) pp 1-3 doi:10.1109/ICISA.2013.6579495

16. Juslin, P., & Laukka, P. (2010). (tech.). Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening. Journal of New Music Research. Retrieved from https://www.tandfonline.com/doi/pdf/10.1080/09298210420003178 13?needAccess=true

17. Xia, Y., Wang, L., Wong, K.-F., & Xu, M. (2008, June 15). *Lyric-based Song Sentiment Classification with Sentiment Vector Space Model.* ResearchGate. https://www.researchgate.net/publication/220874069_Lyric-based_Song_Sentiment_Classification_with_Sentiment_Vector_Space_Model.

18. Hu, X., Bay, M., & Downie, J. (2007). (rep.). CREATING A SIMPLIFIED MUSIC MOOD CLASSIFICATION GROUND-TRUTH SET. ISMIR . Retrieved from https://ismir2007.ismir.net/proceedings/ISMIR2007_p309_hu.pdf

19. Laurier, C., Grivolla, J., & Herrera, P. (2008). (tech.). Multimodal Music Mood Classification Using Audio and Lyrics. IEEE. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.182.426&rep=rep1&type=pdf

20. D. Yang and W. Lee, "Music Emotion Identification from Lyrics," 2009 11th IEEE International Symposium on Multimedia, 2009, pp. 624-629, doi: 10.1109/ISM.2009.123.

21. Solli, M., & Lenz, R. (2011). Color emotions for multi-colored images. Color Research and Application, 36(3), 210–221. https://doi.org/10.1002/col.20604

22. Cheng, F.-F., Wu, C.-S., & Yen, D. C. (2009). The effect of online store atmosphere on consumer's emotional responses - an experimental study of music and colour. Behaviour & Information Technology, 28(4), 323–334. https://doi.org/10.1080/0144929070177057

23. Itten J (1971) Kunst der Farbe. Otto Maier Verlag, Ravensburg, Germany

24. Fagerberg, P., Ståhl, A., & Höök, K. (2004). eMoto: emotionally engaging interaction. Personal and Ubiquitous Computing, 8(5), 377–381. https://doi.org/10.1007/s00779-004-0301-z

25. Mohammad, S. M. (2015, July). Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. Saif M. Mohammad. Retrieved September 24, 2021, from https://saifmohammad.com/WebDocs/acl2018-VAD.pdf.

26. Woolf, M.S., Dignan, L.M., Scott, A.T. et al. Digital postprocessing and image segmentation for objective analysis of colorimetric reactions. Nat Protoc 16, 218–238 (2021). https://doi.org/10.1038/s41596-020-00413-0

27. Ahmad, M. (n.d.). Color Space. HSV is a projection of the RGB space. Digital Publisher: slideplayer.com

28. Shuyo, N. (2015, December 1). language-detection/ProjectHome.md at WIKI · Shuyo/language-detection. GitHub. Retrieved September 25, 2021, from https://github.com/shuyo/language-detection/blob/wiki/ProjectHome.md.

29. Hutto , C. J., & Gilbert, E. (2015, January). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. https://www.researchgate.net/publication/275828927_VADER_A_Parsimonious_Rule-based_Model_for_Sentiment_Analysis_of_Social_Media_Text.

30. Albornoz, J., Plaza, L., & Gervás, P. (2010). Improving Emotional Intensity Classification using Word Sense Disambiguation.

31. *Listening is everything*. Spotify. (n.d.). https://www.spotify.com/us/about-us/contact/.

32. *Company Info*. Spotify. (n.d.). https://newsroom.spotify.com/company-info/.