MSc. IT & Cognition
University of Copenhagen
Cognitive Science 2

Detecting Emotion with Spotify: A Multimodal Approach

By
Flora Haahr Kringelbach <bps775@alumni.ku.dk>
Georgios Petkakis <ghw792@alumni.ku.dk>
Lukasz Zajaczkowski <nxv526@alumni.ku.dk>

Link to project repository:
`https://github.com/giorgospetkakis/multimodal-mer-spotify`

June 3rd, 2020

June 3rd 2020
Flora Haahr Kringelbach, Georgios Petkakis,
and Lukasz Zajaczkowski

Abstract

Music Emotion Retrieval has been an active field of research for decades. Recently Spotify released an online API that provides automatically generated emotion-related features for tracks in their music library. We used the MoodyLyrics4Q dataset to test the ability of these features to correctly classify emotions as defined by Russell's circumplex model of affect. We also tested classifiers trained only on external lyrics data. We hypothesized that supplementing the audio features collected from Spotify with textual information from the tracks' lyrics would improve our ability to recognize the emotion the tracks evoked. We proved our hypothesis, with our fusion models outperforming both unimodal approaches.

NB: we will denote our different sections in the paper using our initials.

Introduction [GP]

As the amount of online content increases, so does the need for personalized user recommendations. Basic recommender systems may consider a user's content history or what 'similar users' enjoy. Music presents a unique opportunity to extend these systems to consider the user's emotional response. Music Emotion Recognition (MER) has been a growing academic field for decades and has recently also attracted the attention of the music industry. In an attempt to harness this body of research, streaming services like Spotify have already created technologies that automatically recognize the emotion content of music based on its audio features [25]. Spotify recently released some of its emotion recognition data through an online API, giving us the opportunity to test them. Recent studies have shown that both audio and textual data can affect music emotion classification performance [30] [41] [37] [11] [51]. We hypothesize that supplementing the audio features collected from Spotify with textual information from the tracks' lyrics will improve our ability to recognize the emotion they evoke. We will consider important psychological models of affect, developments in emotion recognition from audio and textual features as well as previous unimodal and multimodal studies. We will then attempt to create a multimodal machine learning classifier that outperforms models built on each of its constituent modalities.

Emotion [FHK]

Emotion vs. mood [FHK]

It is necessary to specify the definitions used when working within the topic of mood or emotion. There is general consensus within the fields of cognitive science and psychology that the two describe similar affective states, but differ in how they manifest. Some argue that emotion is a brief, intense state with synchronized psychological and physical responses to an internal or external event, while mood is a diffuse, longer state marked by a change in subjective feeling, often without a clear cause [7]. Others have argued that mood and emotion are at the opposing ends of a spectrum. The way an affective state manifests determines its placement on this spectrum [4]. Due to this lack of a clear-cut definition, we will not be distinguishing between the concepts of mood and emotion within this paper.

Theories of emotion classification [FHK]

We can divide models of emotion classification into two dominant groups: discrete and dimensional. Discrete models suggest that human emotion fits into a limited number of discrete categories (e.g. anger and fear). Dimensional models propose that emotions are continuous and can be mapped based on underlying dimensions (e.g. arousal, valence, and dominance) "in combination with cognitive processes" [20]. Basic emotion theory is among the most used discrete emotion models [20] [12]. This theory divides affective states into "families" of related states which share characteristics that separate them from states belonging to other families [13]. Ekman states that the strongest evidence for distinct emotions comes from studies on universal facial expressions for anger, disgust, enjoyment, fear, and sadness.

The dimensional theory includes models that map emotions onto a multidimensional vector space. Some of the most used models are those by Thayer and Russell. Thayer proposed a two-dimensional model in which two types of moods determine affect: energetic and tense [52]. According to the model, these two moods interplay in different ways depending on arousal levels and arousing factors. For example, at low to moderate arousal levels, an arousing factor such as stress will lead to both high energy and tension. However, high levels of arousal will result in higher tension, but lowered energy. Within the framework of this model, valence is a result of varying levels of energy and tension [12].

In comparison, Russell's circumplex model of affect considers valence and arousal as two dimensions ranging from negative to positive valence and low to high arousal [45]. Russell's model maps emotions as points onto the arousal/valence plane. Simplified versions of Russell's model divide the plane into four quadrants. Each quadrant's extremal point serves as a label for all points in that same quadrant (see figure 1). In the present study, we will be relying on the simplified version of the model.

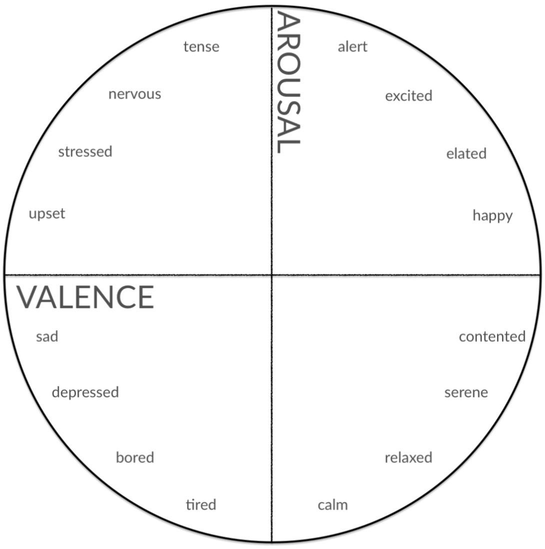

Figure 1: Russell's circumplex model of affect. In the simplified model, high arousal and positive valence correspond to the label "excitement" or "happiness", low arousal and positive valence to "calmness", low arousal and negative valence to "sadness", and high arousal and negative valence to "anger".

Perceived vs. induced emotion [FHK]

Another important element of studying emotion in music is the distinction between perceived and induced emotion. Perceived emotion denotes a listener's perception of the emotion expressed in music (e.g. through the lyrical content) [18]. Induced emotion describes the listener's emotional response to a piece of music. Studies have shown that listeners can perceive emotions such as joy, anger, and sadness in music from other cultures based on acoustic features such as tempo [2] [3]. People can also recognize emotions in music from cultures they have never been exposed to before [17]. This indicates that the ability to perceive emotion in music is universal. While there appears to be an advantage in perceiving emotion within cultures than across them, people are still able to recognize basic emotions such as "anger, fear, happiness, humor, peacefulness, and sadness" in music regardless of cultural familiarity [29]. The concept of universal recognition of basic emotions in music is further supported by evidence that communication of basic emotion in both music and general vocal expression reaches accuracy levels well above chance both within and across cultures [26].

While evidence indicates a certain level of universality for perceived emotion in music, the study of induced emotion is complicated by the idiosyncrasies of the listeners. These include differences in their personal experiences, personality types, descriptions of individual emotional experiences, mental illnesses, etc. Schoen and Gatewood found that, while emotions induced in listeners follow certain general patterns, "most music will call up secondary or related feelings dependent largely on the individual differences of the listeners' mood and experience" [47]. Eerola et al. have shown that extraversion tends to correlate positively not just with the intensity of emotional responses and experience of happiness, but also with the experience of sadness and tenderness [54]. They also found that openness to experience has a positive relationship with the emotional intensity experienced by sad and tender musical pieces. Findings such as these emphasize the complicating factors of individual effects in the study of induced emotion.

In this study, we based the majority of our modelling on perceived emotion in the form of lyrics and audio features. However, given how induced and perceived emotion manifest, there is no way to confirm whether our mood labels, which were user-generated, are representative of perceived, induced, or both types of emotion.

Related work

[GP]Early studies of emotion in music recorded the perceived emotion of the listener. Hevner and Schoen studied the emotional response to music by having participants select the most appropriate adjectives to describe musical pieces played to them [21] [47]. These studies produced clusters of related adjectives for different musical pieces. Farnsworth later refined these adjective clusters [14]. Hu analyzed the cooccurence of the adjectives from these psychological studies with crowdsourced tags from Last.fm, one of the largest online music services at the time [23]. Hu concluded that future research should extend the existing psychological models produced by Hevner, Russell and others with the user-generated data on sites like Last.fm and AllMusic.

[LZ]Other researchers also recognized that a unimodal approach to analyzing a complex medium such as music is insufficient. Spectral-based methods were not robust enough at identifying higher level semantic or stylistic features [40]. This resulted in researchers combining audio and textual features [32] [39] [38] [55] [31]. Studies that focus on identifying emotions in lyrics fall under the wider category of studies utilizing NLP techniques to detect mood, sentiment, or stance. Early studies in this category were binary classification problems [42]. Such methods were adopted in the musical domain to categorize songs by genre, artist, or mood. Later, lexica of emotionally charged words (such as ANEW or NRC) proved to be useful in identifying the polarity of songs [24] [28], and purely lexical methods such as Bag of Words and tf-idf were supplemented by Part-of-Speech tags and additional statistical information like the number of function words and the vocabulary size [24] [30]. Hirjee and Brown further enriched their set of features by exclusively analyzing rhymes in hip hop music [22]. More recently, Fell and Sporleder pointed out the importance of meta-features such as song structure and the 'orientation of the song towards the world' [15]. Both of these were later utilized in combination with word embeddings by Giammusso et al. [19] in predicting the emotion of playlists.

[LZ]Modern word embeddings, first described by Bengio et al. in 2003 [5], are the most active area of research in mood and sentiment detection based on textual information. They are currently considered the state of the art in this domain. In a word embedding language model, words are represented as high-dimensional vectors that contain syntactic and semantic information. For example, synonyms would be closer to each other in the vector space than two words that are completely unrelated. Embeddings are trained on large corpora using models such as 'skip-gram' or 'continuous-bag-of-words' to capture the semantics and role of the word relative to its context. Word embeddings are often used side by side with statistical methods described earlier because they are able to capture more abstract features of the lyrics that traditional methods would ignore. This makes word embeddings a powerful tool for predicting emotion as described in dimensional models such as Russell's.

[GP]Studies investigating these models in relation to audio music data first worked with signal processing [36] or regression models [16] [56] [49]. In 2015, Plewa and Kostek used Self-Organizing Maps, a type of neural network, to automatically predict levels of arousal and valence in music [44]. While methods like fuzzy classifiers had already achieved high prediction accuracy [57], machine learning and specifically artificial neural networks have dominated the field in recent years. Multiple studies have utilized the power of large datasets and faster processing speeds to create predictive models of acoustic [35] [34] [33] [27] [1] and lyrical [19] [10] features. There has also been an increase in studies using a multimodal approach [30] [41] [37] [11] [51]. Among the most popular prediction methods in these studies are Support Vector Machines (SVMs) and deep neural networks.

[GP]Spotify data has also been studied, albeit less frequently. For example, some researchers generated playlists based on emotion [50]. Others examined which kinds of music a specific subset of Spotify users tends to listen to [43]. Another study used Spotify data to predict the popularity of songs and concluded that energy and valence were among the most important predictors of popularity [48]. Sangnark et al. mapped the 'valence' and 'energy' values from the Spotify data onto the valence/arousal axes of Russell's model [46]. In their study, the quadrant that a song was mapped to determined its label. With academic interest in Spotify's API increasing and multimodal approaches to predicting emotion in music becoming more prevalent, it is important to test the validity and usability of Spotify data in a multimodal setting.

Method

Data

[LZ]The emotion labels we use in our study come from the MoodyLyrics4Q dataset. Following Russell's circumplex model, Çano chose to label each song in this dataset with a 'sad', 'happy', 'relaxed', or 'angry' tag. To do this, he retrieved 150 mood terms from relevant research papers and extracted the ones that fell into one of the four quadrants of Russell's model. He then used word embeddings to measure inter- and intra- cluster similarity and chose tags based on those results. Next, he queried the Last.fm API[1] to retrieve the tags of songs present in the Million Song Dataset [6]. Tracks with his pre-selected tags were placed into one of the four quadrants based on strict criteria[2] [8].

[GP]We collected our audio data from Spotify's Web API[3]. The data for each track consist of 13 distinct features: duration, key, mode, time signature, acousticness, danceability, energy, instrumentalness, liveness, loudness, speechiness, valence, and tempo. Some of these features, like the key and mode, are conventional musical attributes. Others, like 'acousticness' and 'instrumentalness', are values derived through Spotify's proprietary machine learning platform [25]. Jehan et al. engineered these features to capture a "fairly consistent" representation of the emotional response to a given musical track [25]. This makes the features particularly attractive for the present study. We standardized all continuous features in this data.

[LZ]We used the Lyrics Genius API[4] to collect song lyrics. In addition to the lyrics, the API returns basic information on song structure (tags for choruses, verses, and bridges). Although we were able to collect audio features for 1797 songs, we could only retrieve the lyrics for 1745 of them. This reduced the final number of tracks considered in our analyses. We used the crawl840 dataset[5], a set of 300-dimensional GloVe word embeddings pre-trained on web crawl data. We chose these based on Çano's finding that crawl840 outperforms other similar word embedding datasets, including the one trained on the same data as the MoodyLyrics4Q[6] dataset [9]. We also extracted several linguistic features from the lyrics as indicated by [19]. These include the relative frequency of adjectives, punctuation marks, and verbs in the past, present and future tenses, the sentiment polarity and subjectivity, line duplications, repeated sounds, and the presence of the title of the song in its lyrics for each song.

Models

Experiment 1[LZ]

In the first experiment, we focused on the ability of the Spotify audio features to correctly classify the songs in our dataset. We split this experiment into two parts. In the first part, we considered all the features in the dataset. In the second part, we excluded 'energy' and 'valence', as we thought they were too closely aligned with Russell's diagram.

We chose NuSVC and Logistic Regression as classifiers as they are frequently used in other studies. After experimenting with other algorithms, we also found that ExtraTrees produced good results. Even though our study takes into consideration a larger number of songs than similar studies, the dataset was still too small to sufficiently train an artificial neural network.

[1] `http://last.fm/`

[2] A song was placed into quadrant Qx if it had at least 4 Qx tags and no tags from other quadrants, 6-8 Qx tags and 1 tag from other quadrants, 9-13 Qx tags and 2 tags from other quadrants, 14 or more Qx tags and 3 tags from other quadrants

[3] `https://developer.spotify.com/documentation/web-api/`

[4] `https://docs.genius.com/`

[5] `https://nlp.stanford.edu/projects/glove/`

[6] `http://softeng.polito.it/erion/`

Experiment 2[LZ]

In our second experiment, we focused on generating predictions based only on the lyrics data. Each song was represented by a tf-idf weighted average of the word embeddings for each word in its lyrics. Tf-idf is the ratio of a word’s frequency within the song divided by its frequency in all other songs. This makes it possible to identify words characteristic of specific songs and their respective mood classes.
To accurately replicate the results in [19], we trained models with our own data as well as data from their project, which is openly available.

Experiment 3[GP]

We tested our hypothesis by combining the data from our previous experiments into a single multimodal dataset. We first tried early fusion. We compared the accuracy of our three selected classifiers on a concatenation of our audio and text features. To compare the crawl840 word embeddings with the pre-trained embeddings in [19], we repeated the experiment by replacing our embeddings and statistical features with theirs.
Finally, we tested a late fusion model. We split the prediction phase into three stages. First, we fed the textual and audio features into separate Logistic Regression classifiers. Those classifiers generated probabilities for each class, which we then used as input for a final model. We compare the results achieved by ExtraTrees, NuSVM, and Logistic Regression in both early and late fusion models.

Results[joint]

Figures 2 and 3 show different visualizations of the Spotify data on a two-dimensional plane based on Russell’s model. We noticed significant improvement in the definition of the clusters of both ’angry’ and ’happy’ tracks when using the principal components instead of the data’s ’energy’ and ’valence’. There is also less overlap between the ‘sad’ and ‘relaxed’ clusters when compared to the ’energy’/’valence’ graph.

Experiment 1

Table 1 includes the results of the first experiment. We achieved the highest accuracy when using all the features and the Extra Trees algorithm. The ’energy’ and ’valence’ values did not improve the accuracy by much. This led us to investigate the relative contribution of each feature to the variance of the data. Figure 4 shows the cumulative variance of the top 10 principal components. We excluded the musical key from this visualization since its high variance is unrelated to emotion. Figure 5 shows the correlation of each feature with other features. Notable pairs include ’acousticness’ with ’energy’ and ’loudness’, ’valence’ with ’danceability’, and ’energy’ with the ‘angry’ and ‘relaxed’ targets.

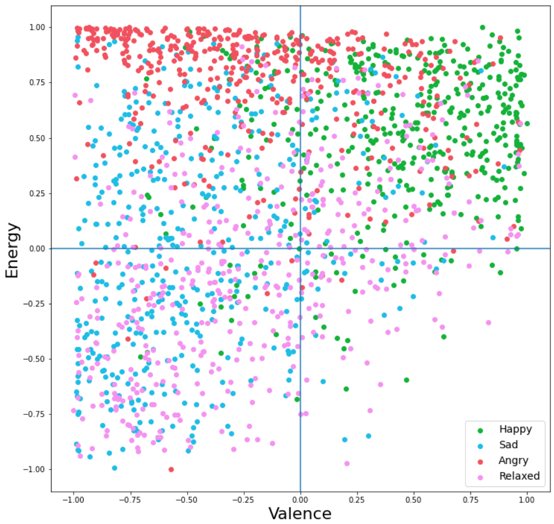

Figure 2: A visualization of ’energy’ and ’valence’ from the Spotify data. The was scaled data to fit in the -1 to 1 range on both the x- and y-axes.

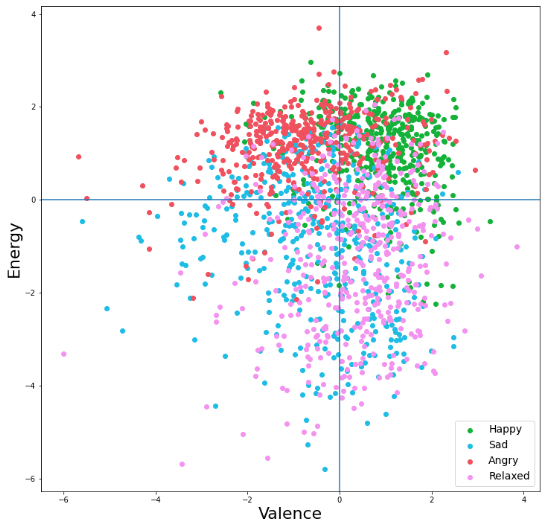

Figure 3: A visualization of the two most principal components of the Spotify data after removing the key

	All features	**Without valence/energy**
Extra Trees	65.3%	61.7%
Logistic Regression	61.9%	57.9%
NuSVC	61.9%	57.0%

Table 1: Results from experiment 1.

Experiment 2

Table 2 shows the results of the classification task using only lyrics data. We noticed a large difference be-

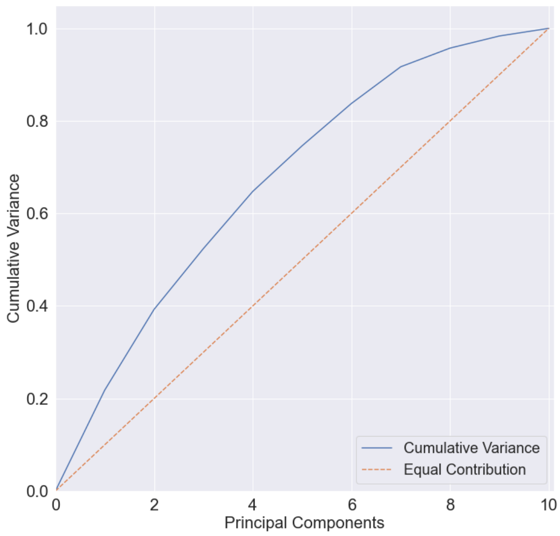

Figure 4: The relative contribution of each principal component to the variance. The orange line represents equal contribution among all components.

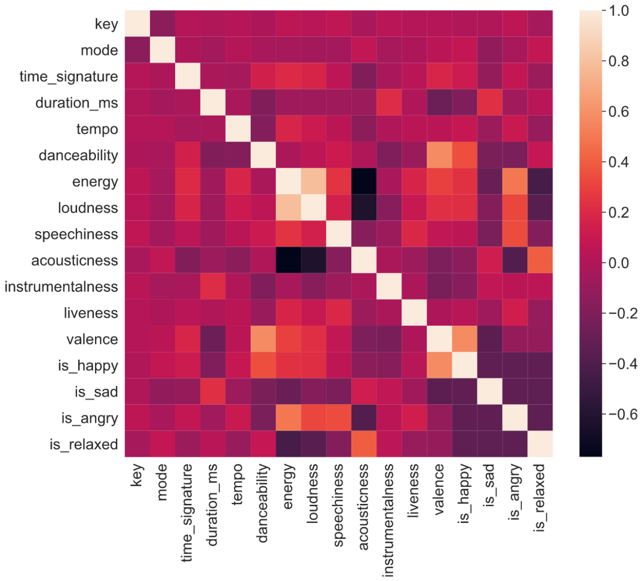

Figure 5: Correlation matrix for the Spotify audio features.

tween our results and the results of Giammusso et al. [19]. This happened even though we extracted the same data and used the same modeling techniques. By using their word embeddings, we were able to replicate their results, raising our accuracy from 45.3% to 61.3%.

Experiment 3

	a)	b)	c)	d)
Extra Trees	45.1%	44.0%	53.6%	58.2%
Logistic Regression	45.3%	40.9%	59.8%	61.3%
NuSVC	42.2%	45.7%	60.1%	60.3%

Table 2: Results from experiment 2. Classification using only lyrics data **a)** is GloVe word embeddings and manually mined features; **b)** is only GloVe word embeddings; **c)** is Giammusso, et al.'s word embeddings; and **d)** is Giammusso, et al.'s word embeddings and mined features.

	Early fusion	Late fusion
Extra Trees	69.0%	68.3%
Logistic Regression	68.8%	70.3%
NuSVC	64.5%	69.0%

Table 3: Results from experiment 3. This shows a comparison of early and late fusion.

The best early fusion classifier was the Extra Trees classifier, with an average accuracy of 69.0%. This is a significant improvement over the performance in the first experiment. The late fusion model outperformed all previous models with an average accuracy of 70.4%. Results varied between 65% and 76% depending on the testing and training split. Table 3 shows the accuracy of the chosen algorithms.

Discussion [joint]

The late fusion experiment proves our hypothesis that Spotify data performs better on classification tasks when supplemented with textual features. Despite its low dimensionality, Spotify data contains a high amount of information and is easily interpretable. This makes it a useful source of information for analyzing mood induced by music, and this is amplified in a multimodal setting. Still, the extraction process and exact specification of Spotify data is proprietary. This makes it hard to verify the quality of the data.
Well-chosen word embeddings significantly improved accuracy. They provided a level of abstraction that audio features could not capture. However, general-purpose word embeddings underperformed relative to those trained for this specific task. This shows that metrics of the robustness of word embeddings are not limited to the size and the quality of their training corpus. Future research should focus on creating embeddings that capture more context-specific information. Designing algorithms that capture more granular textual information may achieve this. Such embeddings could make other textual features redundant, as these would already be represented in high dimensional vector space. Sentence embeddings achieve this by capturing more abstract features such as tone and style.

As discussed earlier in this paper, we could only rely on classification based on perceived emotion. However, we need a better understanding of induced emotion to design MER systems that take individual users into account. Eerola and Vuoskoski found that listeners with high levels of empathy and openness to experience enjoy sad music more than other people. This happens due to the induction of feelings such as "nostalgia, peacefulness, and wonder," not just sadness [53]. This indicates that recommendation of music based on emotion cannot rely only on perceived emotion. Future studies should use datasets with labels for both perceived and induced emotion. Like Giaummusso et al., we found that discriminating between 'sad' and 'relaxed' songs was difficult both for machine classifiers and the humans labelling the data [19]. Brain scanning technology could provide more reliable data on induced emotion, which could help make this distinction.

MER algorithms could also benefit from capturing the temporal aspect of music. Most pieces are made up of multiple segments which may express different emotions, both through text and audio features. Algorithms that analyze the interplay between those segments, and how a song starts, develops, and ultimately resolves, could prove useful for the development of more refined multimodal models.

References

[1] S. Amiriparian, M. Gerczuk, E. Coutinho, A. Baird, S. Ottl, M. Milling, and B. Schuller, "Emotion and themes recognition in music utilising convolutional and recurrent neural networks", 2019.

[2] L.-L. Balkwill and W. Thompson, "A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues", eng, *Music Perception*, vol. 17, no. 1, pp. 43–64, 1999, ISSN: 0730-7829. [Online]. Available: `http://search.proquest.com/docview/1300609128/`.

[3] L. Balkwill, W. F. Thompson, and R. Matsunaga, "Recognition of emotion in japanese, western, and hindustani music by japanese listeners", *Japanese Psychological Research*, vol. 46, no. 4, pp. 337–349, 2004, ISSN: 0021-5368.

[4] C. Beedie, P. Terry, and A. Lane, "Distinctions between emotion and mood", eng, *Cognition and Emotion*, vol. 19, no. 6, pp. 847–878, 2005, ISSN: 0269-9931. [Online]. Available: `http://www.tandfonline.com/doi/abs/10.1080/02699930541000057`.

[5] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model.", eng, *Journal of Machine Learning Research*, vol. 3, no. 6, pp. 1137–1155, 2003, ISSN: 1532-4435. [Online]. Available: `http://search.proquest.com/docview/27916616/`.

[6] T. Bertin-Mahieux, D. P. Ellis, B. Whitman, and P. Lamere, "The million song dataset", 2011.

[7] J. C. Borod, *The Neuropsychology of Emotion.* eng, ser. Series in Affective Science Ser. 2000, ISBN: 9780198027409.

[8] E. Çano, "Text-based sentiment analysis and music emotion recognition", eng, *arXiv.org*, 2018, ISSN: 2331-8422. [Online]. Available: `http://search.proquest.com/docview/2117278212/`.

[9] E. Çano and M. Morisio, "Quality of word embeddings on sentiment analysis tasks", in *International Conference on Applications of Natural Language to Information Systems*, Springer, 2017, pp. 332–338.

[10] H. Corona and M. P. O'Mahony, "An exploration of mood classification in the million songs dataset", 2015, ISSN: 2518-3672.

[11] R. Delbouys, R. Hennequin, F. Piccoli, J. Royo-Letelier, and M. Moussallam, "Music mood detection based on audio and lyrics with deep neural net", eng, *arXiv.org*, 2018, ISSN: 2331-8422. [Online]. Available: `http://search.proquest.com/docview/2110114602/`.

[12] T. Eerola and J. K. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music", eng, *Psychology of Music*, vol. 39, no. 1, pp. 18–49, 2011, ISSN: 0305-7356.

[13] P. Ekman, "An argument for basic emotions", eng, *Cognition and Emotion*, vol. 6, no. 3-4, pp. 169–200, 1992, ISSN: 0269-9931. [Online]. Available: `http://www.tandfonline.com/doi/abs/10.1080/02699939208411068`.

[14] P. Farnsworth, "A study of the hevner adjective list", eng, *The Journal of Aesthetics and Art Criticism*, vol. 13, p. 97, 1954, ISSN: 0021-8529. [Online]. Available: `http://search.proquest.com/docview/740804876/`.

[15] M. Fell and C. Sporleder, "Lyrics-based analysis and classification of music", in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 620–631.

[16] A. Friberg, E. Schoonderwaldt, P. N. Juslin, and R. Bresin, "Automatic real-time extraction of musical expression", eng, 2002.

[17] T. Fritz, S. Jentschke, N. Gosselin, D. Sammler, I. Peretz, R. Turner, A. Friederici, and S. Koelsch, "Universal recognition of three basic emotions in music", eng, *Current Biology*, vol. 19, no. 7, pp. 573–576, 2009, ISSN: 0960-9822. [Online]. Available: `http://search.proquest.com/docview/20102742/`.

[18] A. Gabrielsson, "Emotion perceived and emotion felt: Same or different?", eng, *Musicae Scientiae*, vol. 5, no. 1_suppl, pp. 123–147, 2001, ISSN: 1029-8649.

[19] S. Giammusso, M. Guerriero, P. Lisena, E. Palumbo, and y. =. 2. Troncy Raphaël, "Predicting the emotion of playlists using track lyrics",

[20] S. Hamann, “Mapping discrete and dimensional emotions onto the brain: Controversies and consensus”, eng, *Trends in Cognitive Sciences*, vol. 16, no. 9, pp. 458–466, 2012, ISSN: 1364-6613.

[21] K. Hevner, “The affective character of the major and minor modes in music.”, eng, *American Journal of Psychology*, vol. 47, 1935, ISSN: 0002-9556. [Online]. Available: `http://search.proquest.com/docview/1289721973/`.

[22] H. Hirjee and D. Brown, “Using automated rhyme detection to characterize rhyming style in rap music”, eng, *Empirical Musicology Review*, vol. 5, no. 4, pp. 121–145, 2011, ISSN: 1559-5749. [Online]. Available: `https://doaj.org/article/daf65bcc29124bba9d26603c6d4e1d3c`.

[23] X. Hu, “Music and mood: Where theory and reality meet”, 2010.

[24] Y. Hu, X. Chen, and D. Yang, *Lyric-based song emotion detection with affective lexicon and fuzzy clustering method*, 2009.

[25] T. Jehan and N. Montecchio, *Automatic prediction of acoustic attributes from an audio signal*, US Patent 10,089,578, Oct. 2018.

[26] P. N. Juslin and P. Laukka, “Communication of emotions in vocal expression and music performance: Different channels, same code?”, eng, *Psychological Bulletin*, vol. 129, no. 5, pp. 770–814, 2003, ISSN: 0033-2909.

[27] P. Keelawat, N. Thammasan, B. Kijsirikul, and M. Numao, “Subject-independent emotion recognition during music listening based on eeg using deep convolutional neural networks”, eng, in *2019 IEEE 15th International Colloquium on Signal Processing Its Applications (CSPA)*, IEEE, 2019, pp. 21–26, ISBN: 9781538675632.

[28] S. Kim, H. Kim, T. Weninger, and J. Han, “Authorship classification: A syntactic tree mining approach”, eng, in *Proceedings of the ACM SIGKDD Workshop on useful patterns*, ser. UP ’10, ACM, 2010, pp. 65–73, ISBN: 9781450302166.

[29] P. Laukka, T. Eerola, N. S. Thingujam, T. Yamasaki, and G. Beller, “Universal and culture-specific factors in the recognition and performance of musical affect expressions”, eng, *Emotion*, vol. 13, no. 3, pp. 434–449, 2013, ISSN: 1528-3542.

[30] C. Laurier, J. Grivolla, and P. Herrera, “Multimodal music mood classification using audio and lyrics”, eng, in *2008 Seventh International Conference on Machine Learning and Applications*, IEEE, 2008, pp. 688–693, ISBN: 9780769534954.

[31] C. M. Lee, S. S. Narayanan, and R. Pieraccini, “Combining acoustic and language information for emotion recognition”, in *Seventh International Conference on Spoken Language Processing*, 2002.

[32] T. Li and M. Ogihara, “Music artist style identification by semi-supervised learning from both lyrics and content”, eng, in *Proceedings of the 12th annual ACM international conference on multimedia*, ser. MULTIMEDIA ’04, ACM, 2004, pp. 364–367, ISBN: 1581138938.

[33] H. Liu, Y. Fang, and Q. Huang, “Music emotion recognition using a variant of recurrent neural network”, in *2018 International Conference on Mathematics, Modeling, Simulation and Statistics Application (MMSSA 2018)*, Atlantis Press, 2019.

[34] T. Liu, L. Han, L. Ma, and D. Guo, “Audio-based deep music emotion recognition”, in *AIP Conference Proceedings*, AIP Publishing LLC, vol. 1967, 2018, p. 040 021.

[35] X. Liu, Q. Chen, X. Wu, Y. Liu, and Y. Liu, “Cnn based music emotion classification”, eng, *arXiv.org*, 2017, ISSN: 2331-8422. [Online]. Available: `http://search.proquest.com/docview/2074461998/`.

[36] L. Lu, D. Liu, and H.-J. Zhang, “Automatic mood detection and tracking of music audio signals”, eng, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 5–18, 2006, ISSN: 1558-7916.

[37] R. Malheiro, R. Panda, P. Gomes, and R. Paiva, “Bi-modal music emotion recognition: Novel lyrical features and dataset”, 9th International Workshop on Music and Machine Learning–MML’2016–in . . ., 2016.

[38] R. Mayer and A. Rauber, “Musical genre classification by ensembles of audio and lyrics features”, in *Proceedings of International Conference on Music Information Retrieval*, 2011, pp. 675–680.

[39] R. Neumayer and A. Rauber, “Integration of text and audio features for genre classification in music information retrieval”, in *European Conference on Information Retrieval*, Springer, 2007, pp. 724–727.

[40] F. Pachet and J.-J. Aucouturier, “Improving timbre similarity: How high is the sky?”, *Journal of negative results in speech and audio sciences*, vol. 1, no. 1, pp. 1–13, 2004.

[41] R. Panda, R. Malheiro, B. Rocha, A. Oliveira, and R. P. Paiva, “Multi-modal music emotion recognition: A new dataset, methodology and comparative analysis”, in *International Symposium on Computer Music Multidisciplinary Research*, 2013.

[42] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up? sentiment classification using machine learning techniques”, 2002.

[43] F. Pinarbaşi, “Demystifying musical preferences at turkish music market through audio features of spotify charts”, *TURKISH JOURNAL OF MARKETING*, vol. 4, no. 3, pp. 264–279, 2019, ISSN: 2458-9748.

[44] M. Plewa and B. Kostek, "Music mood visualization using self-organizing maps", eng, *Archives of Acoustics*, vol. 40, no. 4, pp. 513–525, 2015, ISSN: 01375075. [Online]. Available: `http://search.proquest.com/docview/1860866483/`.

[45] J. A. Russell, "A circumplex model of affect", eng, *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980, ISSN: 0022-3514.

[46] S. Sangnark, M. Lertwatechakul, and C. Benjangkaprasert, "Thai music emotion recognition by linear regression", in *Proceedings of the 2018 2nd International Conference on Automation, Control and Robots*, 2018, pp. 62–66.

[47] M. Schoen, *The effects of music : a series of essays*, eng, ser. International library of psychology, philosophy and scientific method. London: Kegan Paul, Trench, Trubner Co., 1927.

[48] M. Sciandra and I. C. Spera, "A model based approach to spotify data analysis: A beta glmm", *d/SEAS Working Paper Forthcoming*, 2020.

[49] M. Soleymani, M. N. Caro, E. M. Schmidt, C.-Y. Sha, and Y.-H. Yang, "1000 songs for emotional analysis of music", in *Proceedings of the 2nd ACM international workshop on Crowdsourcing for multimedia*, 2013, pp. 1–6.

[50] G. Subramaniam, J. Verma, N. Chandrasekhar, N. K. C, and K. George, "Generating playlists on the basis of emotion", eng, in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, 2018, pp. 366–373, ISBN: 9781538692769.

[51] K. R. Tan, M. L. Villarino, and C. Maderazo, "Automatic music mood recognition using russell's twodimensional valence-arousal space from audio and lyrical data as classified using svm and naïve bayes", *IOP Conference Series: Materials Science and Engineering*, vol. 482, no. 1, pp. 1–6, 2019, ISSN: IOP Conference Series: Materials Science and Engineering.

[52] R. E. Thayer, *The biopsychology of mood and arousal*, eng. Place of publication not identified: Oxford University Press Incorporated, 1990, ISBN: 9786610440825.

[53] J. K. Vuoskoski and W. F. Thompson, "Who enjoys listening to sad music and why?", eng, *Music Perception*, vol. 29, no. 3, pp. 311–317, 2012, ISSN: 07307829.

[54] J. K. Vuoskoski and T. Eerola, "Measuring music-induced emotion: A comparison of emotion models, personality biases, and intensity of experiences", eng, *Musicae Scientiae*, vol. 15, no. 2, pp. 159–173, 2011, ISSN: 1029-8649.

[55] D. Yang and W.-S. Lee, *Disambiguating music emotion using software agents*, 2004.

[56] Y.-H. Yang, Y.-C. Lin, Y.-F. Su, and H. Chen, "A regression approach to music emotion recognition", eng, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 448–457, 2008, ISSN: 1558-7916.

[57] Y.-H. Yang, C.-C. Liu, and H. Chen, "Music emotion classification: A fuzzy approach", eng, in *Proceedings of the 14th ACM international conference on multimedia*, ser. MM '06, ACM, 2006, pp. 81–84, ISBN: 1595934472.