

# Are Metal Fans Angrier than Jazz Fans? A Genre-Wise Exploration of the Emotional Language of Music Listeners on Reddit

Vipul Mishra<sup>1</sup>  
vipul-mi@is.naist.jp

Kongmeng Liew<sup>1</sup>  
liew.kongmeng@is.naist.jp

Romain Hennequin<sup>2</sup>  
deezer@research.com

Elena V. Epure<sup>2</sup>  
deezer@research.com

Eiji Aramaki<sup>1</sup>  
aramaki@is.naist.jp

<sup>1</sup>Nara Institute of Science and Technology, Nara, Japan

<sup>2</sup>Deezer Research, Paris-IDF, France

## Abstract

Music forms a big part of our identity and as such, people with a shared preference for certain kinds of music may also share similar traits. In this study, we explore differences in the emotional language of fan communities of different music genres. In focusing on Reddit, we analyze the utterances on online community forums of different music genres using lexicon-based sentiment (emotion) analysis. Upon clustering Subreddit forums, we obtained two clusters: forums discussing genres like Rock, RnB, Country, and Jazz were found to have a higher abundance of positively valenced emotions and a lower amount of negatively valenced emotions. Likewise, Subreddits discussing genres like Metal, Punk, and Rap had a lower amount of positively valenced emotions and a higher abundance of negatively valenced emotion. We observed a high correlation between counts in lyrics of a genre and counts in a fan community for the emotions of anger, disgust, fear, and joy. In sum, we found differences in the emotional language of fan utterances by genre, and these could be partially attributed to the emotions contained in the lyrics.

## 1 Introduction

How music and emotions relate to each other has long been a matter of interest to researchers. Music can induce certain feelings or emotions in the listeners, which can differ from the emotions expressed explicitly in the music. This mood-induction function of music is why music is commonly used for mood regulation or mood management (Saarikallio 2011; Thoma et al. 2012; Papinczak et al. 2015). It has also been found that different styles, or genres of music, are associated with different strategies of mood regulation. For example, Cook et al. (2019) found that genres of electronica and dance music were associated

with use for increasing emotional arousal, while classical music was associated with use for negative mood management. These results suggest that preference for a certain style of music may be associated with a particular set of emotions. Although the link of emotions to various styles of music has been explored in the music psychology literature, these studies have mostly focused on the immediate reactions induced by the music in study participants (Zentner et al. 2008; Sharman and Dingle 2015; Merz et al. 2020). While there have been a few studies that explored the emotions related to specific genre preferences, they have mainly relied on self-reported surveys and questionnaires which can be subjective. Emotionality, or emotional language, associated with preference for music genres has not yet been explored using real-world, naturalistic data. Motivated by this research gap, we explore the emotions expressed in online fan communities of 10 different music genres, to identify the emotion they are strongly associated with. Specifically, we investigate the emotional language associated with 10 Subreddits corresponding to the music genres listed in Table 1.

Reddit<sup>1</sup> is a website having content rating and discussion forums dedicated to different topics and each discussion forum on Reddit is called a Subreddit. For example, r/electronicmusic is a Subreddit for discussing all things related to electronic music.

Additionally, we also run correlation analyses to determine if the prevalence of a type of emotion in the Subreddit is reflective of its prevalence in the music itself. We analyze the track lyrics of each genre in a fashion similar to the comments on the Subreddits to get the prevalence of each type of emotion.

---

<sup>1</sup><https://www.reddit.com/>

Genre Subreddit	Number of comments	Avg. number of tokens	Number of unique users	Lyrics genre	Number of tracks	Avg. number of tokens	Avg. release year
r/Metal	2,069k	30.4	109k	Metal	8.0k	185	2002
r/electronicmusic	858k	26.3	128k	Electronic	5.6k	180	2003
r/punk	728k	26.3	64k	Punk	2.7k	193	1995
r/Jazz	434k	33.0	58k	Jazz	3.3k	173	1991
r/rap	218k	21.4	54k	Rap	7.3k	572	2002
r/country	58k	27.9	12k	Country	6.8k	219	1996
r/blues	48k	26.3	11k	Blues	2.4k	193	1986
r/Rock	44k	25.8	12k	Rock	54.6k	210	1997
r/folk	24k	24.8	8k	Folk	2.8k	214	1996
r/rnb	13k	20.1	4k	RnB	6.6k	332	1993

Table 1: General statistics about Subreddit comments and track lyrics

## 2 Related Work

Literature on Music Psychology is rich with studies measuring the reactions to different genres of music. A study by Cook et al. (2019) found that reactions to music genres like Pop and Rock were associated with anger and revolt while genres like Classical and Jazz were associated with peacefulness and spirituality. Another study also claimed that ‘extreme’ music genres like Metal, punk, etc leads to anger and even aggressive behavior (Zalk et al., 2008). On the other hand, several recent studies claimed no such association exists between intense genres of music and aggression (Merz et al. 2020; Sharman and Dingle 2015; Susino and Schubert 2019). Some of these studies suggested that negative associations with a particular style of music might simply be due to low familiarity with the style of music or even due to holding negative stereotypes about the culture related to that music. Given this ambiguity, we approached this research with the intent to explore and uncover possible relationships in naturalistic data. Furthermore, as we focused on online fan communities, we also assumed that users were familiar with that specific genre when posting on a genre’s Subreddit.

## 3 Methods

### 3.1 Dataset

For all 10 Subreddits, we collected all the comments posted before March 2nd, 2021, using the Pushshift Reddit API (Baumgartner et al., 2020). Table 1 shows the statistical summary about the genre Subreddits. For every Subreddit, we removed the comments that were deleted and were marked ‘[deleted]’, or ‘[removed]’. We also filtered the comments posted by bots using string matching of phrases like ‘I’m a bot’. Finally, the comments for each genre Subreddit were rid of mentions of the particular genre name and its

subgenre. This was done to mitigate bias caused by the genre and subgenre names when computing counts for each emotion. The subgenre names were collected and cleaned manually from Wikipedia.

For analysis with lyrics, we first obtained a list of tracks for each genre using the genre annotations provided with the Million Song Dataset<sup>2</sup> called tagtraum genre annotations. Using the Genius API<sup>3</sup>, we then retrieved the lyrics of tracks using these track names and artist names. The Genius API returned responses for about 160,000 tracks. We filtered the noisy responses from the API using string matching for track title and artist name. We finally got lyrics for a total of about 115,000 tracks. A general statistics of the lyrics collected in this way is given in Table 1.

### 3.2 Sentiment Analysis

We rely primarily on lexicon-based sentiment analysis methods for analyzing the emotionality of each Subreddit, following similar approaches used by Yinger and Springer (2019). Using the emotional categories from the National Research Council of Canada (NRC) Lexicon (Mohammad and Turney, 2013), we group the Subreddits based on how close the emotional language of people is between the Subreddits using the k-means algorithm. Then, we use decision trees to interpret the clustering produced by k-means. Then, we examine if the prevalence of a type of emotion in the comments of a genre Subreddit is reflective of the prevalence of that type of emotion in the music itself.

We used the NRCLex python library<sup>4</sup> to obtain the counts for emotional words for all the Subreddits. The NRC Lexicon consists of words labeled with one or more of the 8 emotion cate-

<sup>2</sup><http://millionsongdataset.com/>

<sup>3</sup><https://docs.genius.com/>

<sup>4</sup><https://pypi.org/project/NRCLex/>

gories: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust. It has been used for a diverse range of tasks, including in the detection of hate speech (Gao and Huang, 2017), and for studying the emotional development of COVID-19 Twitter communities (Marinov et al., 2020), etc. The python library fuses about additional 17,000 wordnet-based synonyms to the NRCLex and holds an emotion dictionary of approximately 27,000 words in total. As counts were larger for Subreddits with more comments, in order to compare the genre Subreddits with subsequent analyses, we normalized the counts of each genre.

### 3.3 Experiments

The Subreddits were clustered using the k-means algorithm according to the normalized emotion counts. The optimal number of clusters was found to be 2 using the elbow method (Thorndike, 1953). We visualize the clustering results using Principal Component Analysis (PCA) in Figure 1. To interpret our results from the k-means clustering, we used the labels predicted by k-means and trained decision trees. Decision trees was chosen due to their ease of interpretation. An example of the decision tree is shown in Figure 3.

To verify whether the results we obtained were not simply due to the a large number of common users between the Subreddits of a cluster, we also clustered Subreddits based on the number of common users. First, about 4000 unique users were randomly sampled from each of the Subreddits. Then, the number of common users within the 4,000 users was calculated for all the Subreddit pairs. We constructed a fully connected graph with the Subreddits as the nodes and the number of common users as the edge weights. We used Gephi’s modularity class function (Bastian et al., 2009) to detect the communities in this graph. The result of the community detection on the common users graph is shown in Figure 2. We also calculated the adjusted rand score to check the similarity of the clustering obtained from the k-means clustering of the normalized emotion counts and the common users graph community detection.

In the second stage of experiments, we examine the correlations between lyrics and Subreddit posts for each emotion category. We obtained normalized counts for the lyrics of each genre in the same way as with the comments. We examined Spearman’s rank-order correlation between

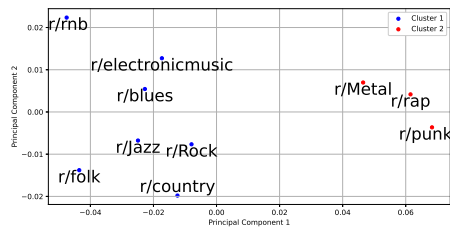


Figure 1: Results of the Subreddits clustering based on the normalized emotion counts. The clustering is visualized using Principal Component Analysis (PCA).

the prevalence of a particular emotion in the comment with that emotion’s prevalence in the lyrics.

## 4 Results

With the k-means clustering of the Subreddits, we got two clear clusters with the first cluster having Subreddits of r/blues, r/country, r/electronicmusic, r/folk, r/Jazz, r/Rock, and r/rnb and the second cluster had Subreddits of r/Metal, r/punk, and r/rap. This clustering is also reflected in the PCA visualization as shown in Figure 1. Similarly, as shown in Figure 2, the common users analysis within the Subreddits revealed two communities. One of the communities contains Subreddits of r/rnb, r/rap, r/Metal, and r/electronicmusic and the other community contains Subreddits of r/blues, r/country, r/folk, and r/Jazz, r/Rock, and r/punk. Subreddits in a community have a higher number of common users amongst themselves compared to Subreddits in other communities. Number of communities detected can differ according to the chosen modularity threshold. However, the modularity threshold was left at a default value of 1.0 which resulted in two detected communities. The adjusted rand score of the clustering obtained from the emotional features and the common users analysis was calculated to be 0.07.

Next, we discuss the analysis with decision trees trained using the cluster labels from k-means. We trained 100 different decision trees, all of which yielded trees of depth 1. The decision trees classified the clusters perfectly using only one of 5 emotions each time i.e a single feature had importance of 1.0 each time. These emotions were anger, disgust, fear, joy, and surprise.

Finally, the correlation analysis with the Subreddit comments and the lyrics revealed a strong positive correlation for anger ( $r=0.818$ ,  $p=0.004$ ), disgust ( $r=0.794$ ,  $p=0.006$ ), and fear( $r=0.697$ ,

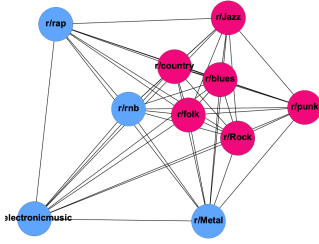


Figure 2: Detection of communities of Subreddits with high number of common users. Force atlas layout is used for arrangement of the nodes. Similar colored nodes belong to the same community.

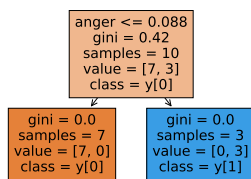


Figure 3: An example of decision tree fit to the clustered Subreddits. This illustrates that the two clusters are clearly distinguished by anger. All Subreddits in cluster 1 have a lower amount of anger words compared to Subreddits in cluster 2.

$p=0.025$ ). A positive correlation was also observed for joy emotion ( $r=0.661$ ,  $p=0.038$ ). No significant correlations were found for the other emotions. The scripts used for the analysis and full results can be found on our GitHub repository<sup>5</sup>.

## 5 Discussion

Our results provide insight on the difference in how fans of different music genres express themselves. The k-means clustering of the emotion counts yielded two clusters of genre Subreddits. The first cluster is characterized by having a relative abundance of positive-valence emotions i.e. joy, anticipation, and surprise. The second cluster is characterized by the higher prevalence of words expressing negative-valence emotions of anger, disgust, and fear. The adjusted rand score between the clustering obtained from the common users analysis and the clustering found through the k-means clustering normalized emotion counts was close to zero. This signifies that emotional difference in the Subreddit clusters is not simply

due to a large number of common users within the Subreddits of the cluster. Rather, the Subreddit clustering hints at a difference in the expressive style of people associated with their music listening habits. Furthermore, we found evidence that a higher degree of high-arousal negative-valence emotions of anger, disgust, and fear expressed in the genre through lyrics is reflected in the fans' emotional expression. These results are in line with previous work by Rubin et al. (2001) that linked fans of metal and rap/hip-hop with trait anger, suggesting that fans of these genres do indeed use more extreme and negative emotional words. However, this does not necessarily disprove research by Merz et al. (2020) that found no correlational or causal link between music genre and psychopathology. One explanation could be due to the specificity of the platform: Subreddit users, in the company of other like-minded users, may express themselves in a manner congruent with the music genre, which may have set the prevailing social norms and standards modeled by the lyrics. Thus, these behaviors may not extend to their daily life beyond the Subreddit. Alternatively, these extreme expressions could function as cathartic release, towards the maintenance of emotional well-being and positive self-regulation Olsen et al. 2020; Sharman and Dingle 2015. This might explain our result, that the prevalence of 'joy' in lyrics of a genre is also reflected with higher amounts of joyful words in the utterances of a genres' Subreddit. However, more research is needed before a definite conclusion can be made, and future research can consider the emotional mechanisms involving genres and fan communities. Our study however has several limitations. We use a single emotion lexicon to quantify the emotions in the comments which can potentially produce biased results. There is a need to validate the results with other emotion lexicons or emotion detection methods. The study also processes comment data only from music listeners on Reddit. Therefore, the findings of this study might also not generalize to the average music listener of a genre. Additionally, we examine the relationship between the emotion expressed by the music and its reflection in the comments using only music with lyrics, without considering the audio component. Therefore, the use of emotion measures computed from the audio could be a direction for future research.

<sup>5</sup><https://github.com/nlp4musa-emotional-language/fan-community-emotion.git>

## 6 References

### References

- Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. [Gephi: An open source software for exploring and manipulating networks](#). *International AAAI Conference on Weblogs and Social Media*.
- Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. [The pushshift reddit dataset](#). *CoRR*, abs/2001.08435.
- Terence Cook, Ashlin R. K. Roy, and Keith M. Welker. 2019. [Music as an emotion regulation strategy: An examination of genres of music and their roles in emotion regulation](#). *Psychology of Music*, 47(1):144–154.
- Lei Gao and Ruihong Huang. 2017. [Detecting online hate speech using context aware models](#). In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, pages 260–266, Varna, Bulgaria. INCOMA Ltd.
- Boris Marinov, Jennifer Spenader, and Tommaso Caselli. 2020. [Topic and emotion development among Dutch COVID-19 Twitter communities in the early pandemic](#). In *Proceedings of the Third Workshop on Computational Modeling of People’s Opinions, Personality, and Emotion’s in Social Media*, pages 87–98, Barcelona, Spain (Online). Association for Computational Linguistics.
- Zachary C Merz, John W Lace, Thatcher R Coleman, and Robert M Roth. 2020. [Challenging the presumptive link between musical preference and aggression](#). *Psychology of Music*, 0(0):0305735620963756.
- Saif Mohammad and Peter Turney. 2013. [Nrc emotion lexicon](#). *National Research Council, Canada*, 2.
- Kirk N. Olsen, Merrick Powell, Aydin Anic, Robert J. Vallerand, and William Forde Thompson. 2020. [Fans of violent music: The role of passion in positive and negative emotional experience](#). *Musicae Scientiae*, 0(0):1029864920951611.
- Zoe E. Papinczak, Genevieve A. Dingle, Stoyan R. Stoyanov, Leanne Hides, and Oksana Zelenko. 2015. [Young people’s uses of music for well-being](#). *Journal of Youth Studies*, 18(9):1119–1134.
- Alan M. Rubin, Daniel V. West, and Wendy S. Mitchell. 2001. [Differences in aggression, attitudes toward women, and distrust as reflected in popular music preferences](#). *Media Psychology*, 3(1):25–42.
- Suvi Saarikallio. 2011. [Music as emotional self-regulation throughout adulthood](#). *Psychology of Music*, 39(3):307–327.
- Leah Sharman and Genevieve A. Dingle. 2015. [Extreme metal music and anger processing](#). *Frontiers in Human Neuroscience*, 9:272.
- Marco Susino and Emery Schubert. 2019. [Cultural stereotyping of emotional responses to music genre](#). *Psychology of Music*, 47(3):342–357.
- Myriam V. Thoma, Stefan Ryf, Changiz Mohiyeddini, Ulrike Ehlert, and Urs M. Nater. 2012. [Emotion regulation through listening to music in everyday situations](#). *Cognition and Emotion*, 26(3):550–560. PMID: 21902567.
- Robert L. Thorndike. 1953. [Who belongs in the family](#). *Psychometrika*, pages 267–276.
- Olivia Swedberg Yinger and D Gregory Springer. 2019. [Using psycholinguistic inquiry to measure emotional response to music: A feasibility study](#). *Psychology of Music*, 47(4):606–614.
- Maarten Zalk, Marc Delsing, Tom Bogt, and Wim Meeus. 2008. [Heavy metal and hip-hop style preferences and externalizing problem behavior: A two-wave longitudinal study](#). *Youth & Society*, 39.
- Marcel Zentner, Didier Grandjean, and Klaus Scherer. 2008. [Emotions evoked by the sound of music: Characterization, classification, and measurement](#). *Emotion (Washington, D.C.)*, 8:494–521.