



Using k -Means Clustering to Classify Protest Songs Based on Conceptual and Descriptive Audio Features

Yanru Jiang¹  and Xin Jin^{2,3} 

¹ Department of Communication, University of California Los Angeles, Los Angeles, CA 90095, USA

yanrujiang@g.ucla.edu

² Department of Media and Communication, City University of Hong Kong, Kowloon 999077, Hong Kong SAR

xin.jin@my.cityu.edu.hk

³ Center for Intelligent Media and Communication Research, College of Literature and Journalism, Central South University, Changsha 410017, Hunan, China

Abstract. Protest music is a phenomenal and widely circulated form of protest art in social movements. Previous protest music research has extensively focused on lyrics while ignoring other musical features that also contribute to the role of protest music in social movements. This study fills a gap in previous research by converting 397 unstructured musical pieces into structured music features and proposing a k -means clustering analysis to categorize protest songs based on both high-level conceptual features collected from Spotify and low-level descriptive audio features extracted via Librosa. The Davies–Bouldin index, inertia curve, Silhouette curve, and Silhouette diagram were the main measurements used to compare model performance. An innovative threshold filtering approach (optimizer area) was used to label 128 protest songs. Through a bottom-up folksonomy approach to music classification, this study overcomes the limitations of traditional genre classification by introducing other high-level features (e.g., energy, danceability, instrumentality) and their roles in determining protest music categories.

Keywords: Cultural computing · Music classification · Protest art

1 Introduction

2020 has been “the year of the social movement,” from the Hong Kong Anti-Extradition Bill (Anti-ELAB) movement to Black Lives Matter (BLM). In all of these social movements, music has become a phenomenal and widely circulated form of protest art [1]. Unlike most political songs, which serve the propagandistic function of promoting and solidifying an official state perspective, protest songs challenge dominant ideologies and powerful groups [2]. Previous protest music research has extensively focused on lyrics, discussing their rhetorical effects, while ignoring other musical features—such as melody, harmony, timbre, rhythm, and tempo—that also contribute to the role of protest music in social movements [3].

This study bridges the gap in existing protest music research by converting unstructured musical pieces to structured musical features and proposing a k -means clustering analysis to categorize protest songs based on both high-level conceptual and low-level descriptive audio features extracted using the Spotify API and the Librosa music analysis tool. The Davies–Bouldin (DB) index, inertia curve, Silhouette curve, and Silhouette diagram were the main measurements used to compare model performance of the k -means algorithms. An innovative threshold filtering approach called optimizer area was also employed to generate labels for 128 protest songs based on the clustering analysis.

2 Literature Review

2.1 Protest Music

Mondak [4] defines protest music as a form of political persuasion, as one of its functions is “the attempt to generate outside support for the critical views it expresses” (p. 25). Emphasizing lyrics, Denisoff [5] identified two functions of protest songs—the magnetic and the rhetorical—viewing the music as secondary to the lyrics in realizing these functions. Magnetic songs encourage and recruit non-participants to join the social movement and reinforce existing participants’ commitment through the messages of group solidarity and identity. Rhetorical songs, rather than appealing to group identity and promoting a large movement, describe the dissent, repression, and struggles faced by individuals in a group. In his research on the lyrics of protest music after the 2000s, Cort [6] reflects on the progression of music’s roles in social movements and extends Denisoff’s [5] twofold classification. After advocacy for issues like climate change began to emerge after the millennium, group solidarity and individual adversity were no longer the sole messages delivered by protest songs. Rather, as identified by Cort [6], protest music in the modern world progressively came to include the roles of “challenging a power structure, empowering the listener to overcome adversity, communicating a message to solidify members of a social movement, and encouraging societal change” (p. 5).

As research on the lyrics of protest music began to reach a consistent and comprehensive conclusion regarding the role of protest songs in social movements, Eyerman and Jamison [7] challenged this unilateral emphasis on lyrics for studying protest songs. Responding to Denisoff’s [5] study, which overemphasized lyrics, the authors argued that the music itself also contributes to protest songs’ magnetic and rhetorical functions. Specifically, the structure of protest songs, which are “buil[t] around known and catchy melodies, repeating verses, [and] simple chords” (p. 43), conveys an explicit persuasive effect. Additionally, Eyerman and Jamison [7] focus on a cultural perspective: since protest music’s contextual messages and status as a bearer of musical traditions could be reflected in its familiar tunes and genres, which appeal to certain groups and cultures, the cultural elements of protest music link the past, present, and future of the revolutionary groups. Preexisting African American songs provided new kinds of emancipatory and visionary messages for the U.S. civil rights movement in the late 1950s and early 1960s. White power music—including country, folk, Viking, black metal, and hard rock, which developed in the Western world—is commonly found in neo-fascist movements [8].

At the same time, numerous cognition and psychology studies have investigated the impact of audio features on audiences’ perceptions. Bainbridge et al. [9] found

that both infants and adults demonstrated psychological responses to unfamiliar foreign songs, indicating that audio features alone could affect perception without any additional context. Given that most previous protest music studies have been conducted from lyrics-focused, cultural, generic, and qualitative perspectives, this study selects audio-level music features, such as arousal, melody, rhythm, and timbre, to empirically analyze the following research questions [10, 11]:

RQ1: How should protest songs be classified based on high-level features and low-level descriptors?

RQ2: Which combination of features produces the optimal results for classifying protest songs?

2.2 Music Categorization

Though most previous studies on music classification have focused on genre classification, Scaringella and Zoia [3] claim that most generic approaches remain fuzzy and use loosely defined generic concepts. Possible issues arising from genre classification include the simplification of artists' discourse, unstandardized merging of multiple genres, the need to split one genre into various subgenres, and the overall difficulty in agreeing on a universal taxonomy for music classification. This study challenges the traditional generic approach by incorporating high-level musical features to identify how different sets of features could stimulate different emotional reactions from audiences.

Scaringella and Zoia [3] proposed an innovative solution for classifying music termed "folksonomy," a bottom-up approach wherein users generate classifiers through an interactive mutual agreement. This study recruited two content coders and adopted a folksonomy-based method for labeling clusters of protest songs.

3 Conceptual Framework

Early research on music classification used static clustering methods and has since continually advanced, especially after the introduction of deep learning and neural networks [12]. The primary focuses in this area have two components: (i) feature selection and extraction, and (ii) classifier designs [13]. The majority of music classification research explores genre classification using the GTZAN dataset created by Tzanetakis and Cook [14] which contains 10 genre labels: blues, classical, country, disco, hip hop, jazz, metal, pop, reggae, and rock [15]. Audio feature extraction and music clustering have also shown promise in music tagging, recommendation, and classification with regard to previously undifferentiable music categories, such as traditional music and song emotion recognition [16, 17].

3.1 Audio Feature Extraction

Feature extraction allows researchers to convert unstructured musical pieces into structured data input by using high- and low-level features to capture the differences between pieces of music.

High-Level Features. High-level statistical acoustic features are generally found to have better classifying performance than low-level features [18]. Lidy and Rauber [19] conducted research on psychoacoustic transformation in the context of music genre classification by measuring audiences' psychological responses and connecting those responses to certain high-level audio features of the music. Additionally, high-level features demonstrate better performance in emotion recognition compared with low-level features [19].

Rodà et al. [20] identified valence and arousal as the most frequently researched dimensions among high-level music features. According to Dillman Carpentier and Potter [10], arousal, level of hedonic intensity, and energetic value are commonly studied along with valence (i.e., the negative or positive value of the hedonic level). While the valence–arousal plane has been dominant in studies of musical features, Rodà et al. [20] recommend including other features, such as loudness, roughness, and spectral features, to measure the emotional power of the music.

Low-Level Features. Most of the research on genre classification is conducted on low-level descriptors, mainly focusing on features in music signal processing that simulate human auditory perception, including rhythmic, temporal, cepstral, spectral, and perceptual audio descriptors [21]. Scaringella and Zoia [3] found that low-level descriptors in audio signals could capture melodic and harmonic elements of music more robustly than high-level features.

3.2 Music Classifier Design

With the development of greater computing power and advanced deep learning algorithms, the research focus of music classifier design has shifted from more static, classic models—such as logistic regression, k -nearest neighbors, and support vector machines (SVMs)—to more novel, dynamic approaches, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Tzanetakis and Cook [14] pioneered the genre classification field with their use of a mixed-methods design consisting of k -nearest neighbors and Gaussian mixture models to introduce a comprehensive clustering pipeline. The two authors further standardized the feature categories for music clustering (rhythm, pitch, and temporal structure) for future studies. Li et al. [22] proposed content-based genre clustering, emphasizing the multiclass attributes enabled by the histogram classification approach. Xu et al. [23] applied an SVM for music classification tasks, while Fu et al. [24] challenged the SVM design by combining a naïve Bayes classifier with an SVM to produce an NBSVM algorithm.

Regarding dynamic music classifiers, neural networks have shown success in reflecting the temporal features in a sequence of music timesteps. Vishnupriya and Meenakshi [25] proposed a CNN architecture and implemented it on Mel spectrum and MFCC feature vectors for each song. Grzegorz [26] trained a CNN model for image recognition and conducted transfer learning by reusing the bottom levels of that trained model for genre recognition. Considering music as a sequence of sounds with embedded temporal features, Ghosal and Sarkar [12] designed a model consisting of four layers of CNNs with a final long short-term memory (LSTM) RNN layer to construct a new clustering

augmented learning method for genre classification. With the convolutional layer and LSTM autoencoders in the last layer, the authors were able to capture both spatial and temporal features from the music signal.

Among all classifier designs, k -means clustering is one of the most commonly used and most straightforward algorithms for clustering analysis. Jondya and Iswanto [27] extracted the most relevant audio features through principal components analysis and used k -means clustering to categorize Indonesian traditional music based on 36 low-level descriptors. Kim et al. [17] innovatively applied k -means clustering to a music recommendation system and introduced a dynamic clustering algorithm capable of reflecting changes in users' playlists.

4 Methods

4.1 Data Collection

Several steps were involved in the data collection process. First, I used the Wikipedia protest songs category to generate a list of protest songs from the 19th through 21st centuries. Second, I matched Wikipedia songs with the Spotify music database to collect Spotify music features and 30-s MP3 music demos to be used for Librosa feature extraction in the next step.

Wikipedia List. As there lacked an existing list comprehensively covering global protest songs, I ensured the representativeness and authority of the data by confirming our list synthetically based on qualitative literature on typical protest songs (e.g., *Dear Mr. President*), lists published by authoritative music brands (e.g., RadioX), and the Protest Songs Wikipedia Typology page, which is under the authority of the Library of Congress. I ultimately produced a list of 457 songs with music production information, such as language, release date, genre, and composer.

Spotify. I next used the Spotify API to match songs in the Wikipedia list with the Spotify database according to title, artists, album, and release year. Because the official Spotify Web API is based on the REST (Representational State Transfer) principle, which is not directly compatible with the Python language, Spotipy—a lightweight Python library for the Spotify Web API—was used instead to query music features from the Spotify database. Realizing that the search results in Spotify may not automatically refer to the same songs in the Wikipedia list, I filtered out irrelevant songs when the number of songs in the search results was greater than 50.

Content Coding. Two coders specializing in social movement studies conducted content coding on 457 protest songs by manually verifying whether the search result in the Spotify database matched the music indicated on the Wikipedia list. The MP3 music demos provided by Spotify were used to confirm and select the corresponding song version, while the two coders also used the YouTube video search results to cross-check the accuracy of the Spotify music selection. After a systematic data cleaning procedure, 397 songs that could be matched with the Spotify API were retained for feature extraction.

4.2 Feature Extraction

Feature extraction and selection were performed for both the high-level conceptual features provided by the Spotify API and the low-level descriptive features extracted from the Librosa Python library. The study conducted k -means clustering on high-level and low-level features, respectively, to compare the performance of both sets of features. Considering that differences in feature values may substantially distort the mapping and clustering process, I normalized both high-level and low-level features from 0 to 1.

High-Level Features: Spotify. To operationalize the conceptual-level audio features, the Spotify API was used to extract the technological-level audio features of each protest song. The Echo Nest, a music intelligence and data platform, first introduced audio feature extraction algorithms to the music world; its algorithm has been integrated with the Spotify Audio Feature API since Spotify acquired The Echo Nest in 2014 [28] According to the Object Index in the Spotify API documentation, 10 audio features (acousticness, danceability, energy, instrumentalness, key, liveness, loudness, speechiness, tempo, and valence) are provided by the Spotify API. A detailed description of each audio feature is provided in the Appendix.

For each piece of music, I selected eight features—danceability, energy, speechiness, acousticness, instrumentalness, liveness, valence, and tempo—from the Spotify API for clustering analysis based on their relevance to the music being categorized. Loudness was excluded because its presence would not substantially affect audiences’ emotional perceptions of the protest music. Loudness can fluctuate based on the user’s volume setting, while liveness can also easily vary based on the recording environment during the music production. Therefore, liveness has only a moderate effect on the clustering results (Fig. 1).

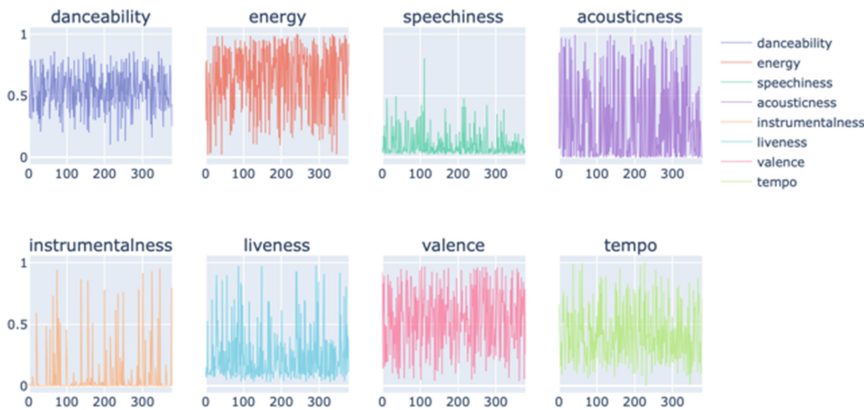


Fig. 1. Distribution of high-level audio features

Low-Level Descriptors: Librosa. Librosa, first introduced by McFee et al. [29], is a commonly used Python package for music and audio signal processing. Unlike the Spotify API, which extracts high-level audio features, Librosa extracts low-level descriptive

features from audio clips, some of which are not readily perceptible to audiences, such as spectral features. Librosa is usually considered comprehensive for music studies because it provides all relevant parameters for audio analysis [29].

In total, 320 music files downloaded using the Spotify API were used for Librosa feature extraction. This number is smaller than the number of songs selected for Spotify feature extraction because 77 songs were excluded due to their lack of MP3 demos. To enable Librosa to process the music files, I converted the 320 MP3 files to WAV format using an online converter tool on a pay-as-you-go basis. Librosa reads WAV files into a one-dimensional NumPy array (denoted as y) and calculates the song duration in seconds based on the sampling rate:

$$\text{Duration Seconds} = \frac{\text{float}(\text{len}(y))}{\text{sampling rate}} \quad (1)$$

Spectral features, defined as “the distribution of energy over a set of frequencies” [29], are the dominant features in Librosa feature extraction and in digital signal processing. Among all spectral features, the Mel spectrogram frequency scale (*melspectrogram*) and its derivatives (*mfcc*, *rms*, *poly_features*) are frequently used to capture the timbral element of music. Specifically, the spectrogram represents the music’s time and frequency information and has been widely used in recently developed music classification models [12].

Besides capturing timbral elements, “pitch class representations (*chroma_stft*, *chroma_cqt*, *chroma_cens*, *tonnetz*) are often used to encode harmony while suppressing variations in octave height, loudness, or timbre” [29]. Additionally, Librosa provides spectral statistic representations for music, including spectral centroid, bandwidth, roll-off, and contrast. Zero-crossing rate, a temporal feature measuring “signal changes from positive to zero to negative or from negative to zero to positive” was also added to the feature selection [3, 30]. One rhythm feature, Fourier tempogram, was also included for the purpose of clustering Librosa audio features.

4.3 Clustering

I used k -means clustering to categorize music computationally, ultimately arriving at 10 clusters for high-level features and seven clusters for low-level descriptors based on the DB index. The clustering results of the perceptual high-level feature set were further verified and adjusted by human coders through a folksonomy classification approach. Six subcategories and three parent categories were ultimately identified for this feature set.

Several metrics, including the inertia curve, Silhouette curve, Silhouette diagram, and DB index, were used to compare and measure model performance. After realizing that the inertia curve, Silhouette curve, and Silhouette diagram could not easily identify the optimal number of music clusters for both datasets in the study, I selected models based on which had the lowest DB index.

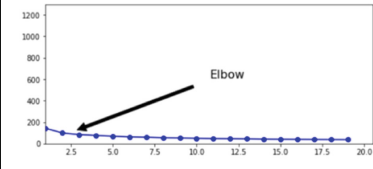
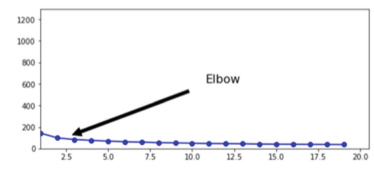
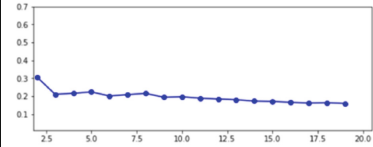
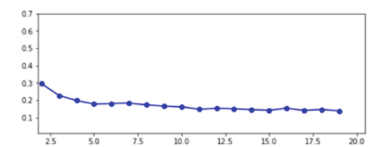
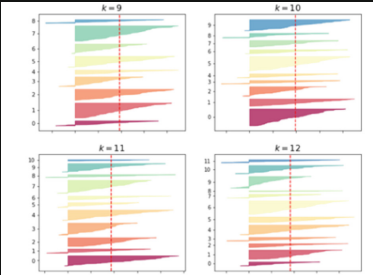
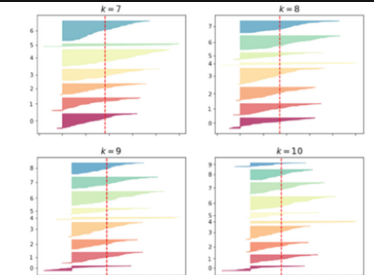
4.4 Music Category Optimization

Once I determined the number of clusters using the DB index, I defined an optimizer area to select protest songs that were more likely to be aligned with the corresponding clusters and filter out the less relevant ones.

In the optimization step, I calculated the mean value of eight features in each cluster and selected them as baselines by adding or subtracting the standard deviation with various multipliers to obtain the optimizer area. Songs with feature values outside this area were excluded. Based on the optimized clusters, whose category attributes are more aggregate and salient, I ultimately identified six categories of protest music by dropping noise clusters and merging similar ones (Table 2; dataset: <https://bit.ly/3vg6n6c>).

$$\text{Optimizer Area} = \text{Mean}_{\text{feature}} \pm \text{Multiplier} \times \text{Standard Deviation}_{\text{feature}} \quad (2)$$

Table 1. Comparison of four measurements used to evaluate clustering model performance

	High-level (Spotify)	Low-level (Librosa)
Davies-Bouldin	The DB index was smallest ($DB = 1.25$) when the number of clusters was set to 10	The DB index was smallest ($DB = 1.48$) when the number of clusters was set to 7
Inertia curve		
Silhouette curve		
Silhouette diagram		

5 Results

5.1 Performance Measurement

For both high-level and low-level features, the inertia curve, Silhouette curve, and Silhouette diagram were not able to produce a distinct pattern in the clustering analysis (see Table 1). The DB index was thus selected for comparing model performance. As the DB index was used to measure similarity across clusters, a lower DB index indicates a better clustering result.

High-Level Features. The DB index was smallest when the number of clusters was set to 10 ($k = 10$) for the high-level feature set. According to the inertia curve and Silhouette curve above, none of these graphs shows a distinct elbow (i.e., turning point) for identifying the best number of clusters for the high-level feature set. The Silhouette diagram indicates the same result as the DB index: protest songs are less likely to be misclassified to the wrong clusters when k equals 10 for high-level features.

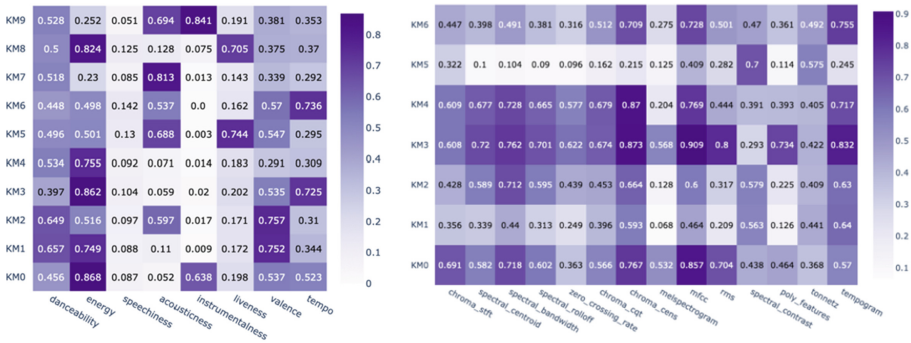


Fig. 2. Heat map of cluster results of high-level (left) and low-level feature set (right)

The heat map (see Fig. 2) reveals the similarities and differences in feature values across the 10 clusters. Specifically, KM4 and KM8 share a similar distribution of music features, and clusters tend to have distinctive values in the energy and valence features.

Using the optimizer area approach, I identified 128 protest songs that were more likely to be aligned with the feature pattern in their corresponding clusters (Fig. 3).

Low-Level Descriptors. Neither the inertia curve nor the Silhouette curve graph below shows a distinct elbow for identifying the best number of clusters for the low-level feature dataset. The Silhouette diagram indicates the same result as the DB index: for low-level features, protest songs are less likely to be misclassified to the wrong clusters when k equals 7.

The heat map for the cluster results of the low-level feature set reveals that KM3 and KM4 have similar distributions of music features, and clusters tend to have distinctive values in the tempogram and MFCC features.

The distribution of the low-level features across the seven clusters reveals that KM0 and KM6 have similar distributions of music features, while most clusters are substantially different from the remaining clusters (Fig. 4).

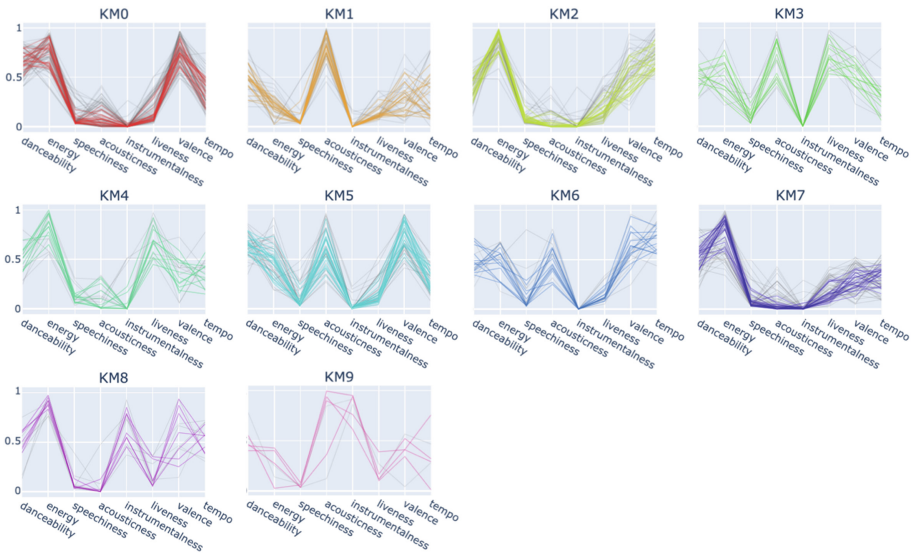


Fig. 3. Distribution of high-level music features by cluster. Gray curves indicate music with feature values outside the optimizer area.

5.2 Clustering and Annotating Results

Through an unsupervised clustering analysis, this study makes a pioneering classification of protest songs into six subcategories within three overarching categories by emphasizing audio feature variation. Category A shows that the salient attributes of rock music are its typically high energy and low acousticness. Category B is associated with a high volume of vocal sounds (high acousticness). Category C, representing the attributes of folk and country music, maintains relatively balanced values of energy and acousticness. Moreover, for each subcategory, some other features were also significant in categorizing the music. For example, high danceability provides songs in subcategory KM0 with strong beats; thereby, our study challenges the traditional generic categorization of protest songs (Table 2).

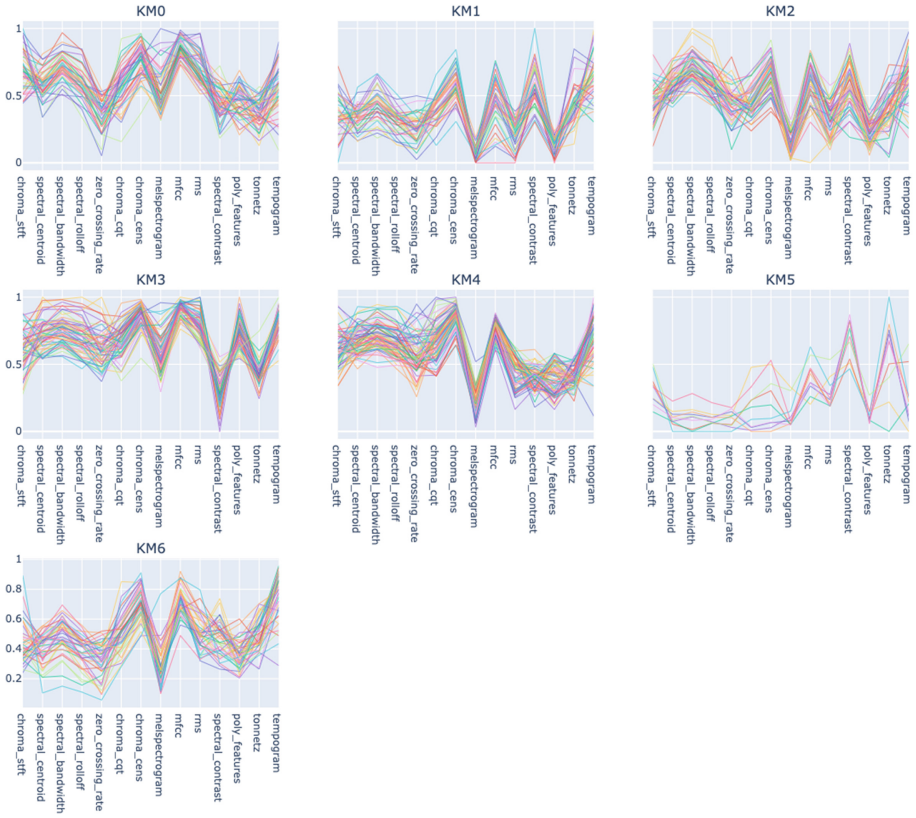


Fig. 4. Distribution of low-level music features by cluster

Table 2. Final music categories with quantities, descriptions, and Spotify links.

Category	N	Description	Representative music
A-km0	21	Dance rock, post-punk, dance music	https://spoti.fi/3pRZUdH
B-km1	21	Highly vocal songs	https://spoti.fi/3bBpgaQ
A-km2	21	Hard rock with high tempo	https://spoti.fi/3dIOWF7
C-km3/5	32	Folk and country music with low tempo	https://spoti.fi/3aRfRl6
C-km6	13	Folk and country music with high tempo	https://spoti.fi/3sqPTWt
A-km7	20	Soft rock	https://bit.ly/3uzmVWk

6 Conclusion

This study fills a gap in previous protest music research, which has historically ignored audio features when classifying music. Specifically, this research converted 397 unstructured musical pieces into structured music features and proposed a k -means clustering

analysis to categorize protest songs based on both high-level conceptual features collected from Spotify and low-level descriptive audio features extracted via Librosa. The DB index was the main metric used to compare model performance. An innovative threshold filtering approach (optimizer area) was used to label 128 protest songs. Through a bottom-up folksonomy approach to music classification, this study overcomes the limitations of traditional genre classification by introducing other high-level features (e.g., energy, danceability, instrumentality) and their roles in determining music clusters.

Due to the lack of perceivability of Librosa features, this study did not manually label the final categorical results for the low-level feature set. Future research could conduct a time-series analysis following the clustering method proposed in this study to investigate how protest music categories shift across centuries.

6.1 Limitations and Future Work

The research has several limitations. First, the Wikipedia typology might not present a comprehensive list of protest songs. Several songs also were dropped due to a lack of matching results in the Spotify database. However, covering almost 400 protest songs from the 19th through 21st centuries for music categorization is already a record-breaking feat in protest music studies [6]. Second, due to the imperceptibility of low-level music descriptors, this study could not manually verify the clustering results produced by the Librosa feature set. In future research, this issue could be overcome by cross-checking with the categorizing results of Spotify features or other state-of-the-art approaches in the music classification field. Lastly, the Librosa feature set contains several audio features that share similar attributes and are highly correlated, such as Mel spectrogram and MFCC, which might lead to collinearity in the feature selection process. It is recommended to verify multicollinearity within spectral and rhythm features before clustering music based on low-level descriptors.

7 Data and Code Availability

Data and code: https://github.com/JoyceJiang73/Protest_Music_Clustering.

References

1. Green Jr., D.F.: Views from the bricks: notes on reading and protest. *CLA J.* **63**, 169–173 (2020). <https://doi.org/10.34042/claj.63.2.0169>
2. Philpott, S.: Of country and country: twang and trauma in Australian Indigenous popular music. *Politik* **23**, 94–98 (2020). <https://doi.org/10.7146/politik.v23i1.120312>
3. Scaringella, N., Zoia, G., Mlynek, D.: Automatic genre classification of music content. *IEEE Signal Process. Mag.* **23**, 133–141 (2006). <https://doi.org/10.1109/MSP.2006.1598089>
4. Mondak, J.J.: Protest music as political persuasion (1988). <https://doi.org/10.1080/03007768808591322>
5. Denisoff, R.S.: Songs of persuasion: a sociological analysis of urban propaganda songs. *J. Am. Folk.* **79**, 581–589 (1966). <https://doi.org/10.2307/538223>
6. Quirk Cort, M.E.: *The Power of Lyrical Protest: Examining the Rhetorical Function of Protest Songs in the 2000s* (2013)

7. Eyerman, R., Jamison, A.: Taking traditions seriously. In: *Music and Social Movements: Mobilizing Traditions in the Twentieth Century*, pp. 26–47. Cambridge University Press (1998). <https://doi.org/10.2307/767983>
8. Eyerman, R.: Music in movement: cultural politics and old and new social movements. *Qual. Sociol.* **25**, 443–458 (2002). <https://doi.org/10.1023/A:1016042215533>
9. Bainbridge, C.M., et al.: Infants relax in response to unfamiliar foreign lullabies. *Nat. Hum. Behav.* **5**, 256–264 (2021). <https://doi.org/10.1038/s41562-020-00963-z>
10. Dillman Carpentier, F.R., Potter, R.F.: Effects of music on physiological arousal: explorations into tempo and genre (2007). <https://doi.org/10.1080/15213260701533045>
11. Rosenthal, R.: Serving the movement: the role(s) of music. *Pop. Music Soc.* **25**, 11–24 (2001)
12. Ghosal, S.S., Sarkar, I.: Novel approach to music genre classification using clustering augmented learning method (CALM). In: *CEUR Workshop Proceedings*, vol. 2600 (2020)
13. Tsai, W.H., Bao, D.F.: Clustering music recordings based on genres. *J. Inf. Sci. Eng.* **26**, 2059–2074 (2010)
14. Tzanetakis, G., Cook, P.: Musical genre classification of audio signals. In: *IEEE Transactions on Speech and Audio Processing*, pp. 293–302. IEEE (2002)
15. Cheng, Y.H., Chang, P.C., Kuo, C.N.: Convolutional neural networks approach for music genre classification. In: *Proceedings of 2020 International Symposium on Computer, Consumer and Control, IS3C 2020*, pp. 399–403 (2020). <https://doi.org/10.1109/IS3C50286.2020.00109>
16. Choi, K., Fazekas, G., Sandler, M., Cho, K.: Convolutional recurrent neural networks for music classification. In: *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing Proceedings*, pp. 2392–2396 (2017). <https://doi.org/10.1109/ICASSP.2017.7952585>
17. Kim, D.M., Kim, K.S., Park, K.H., Lee, J.H., Lee, K.M.: A music recommendation system with a dynamic K-means clustering algorithm. In: *Proceedings of the 6th International Conference on Machine Learning and Applications, ICMLA 2007*, pp. 399–403 (2007). <https://doi.org/10.1109/ICMLA.2007.97>
18. Atmaja, B.T., Akagi, M.: On the differences between song and speech emotion recognition: effect of feature sets, feature types, and classifiers. In: *IEEE Region 10 Annual International Conference Proceedings/TENCON*, November 2020, pp. 968–972 (2020). <https://doi.org/10.1109/TENCON50793.2020.9293852>
19. Lidy, T., Rauber, A.: Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In: *ISMIR 2005, 6th International Conference on Music Information Retrieval*, pp. 34–41 (2005)
20. Rodà, A., Canazza, S., De Poli, G.: Clustering affective qualities of classical music: beyond the valence-arousal plane. *IEEE Trans. Affect. Comput.* **5**, 364–376 (2014). <https://doi.org/10.1109/TAFFC.2014.2343222>
21. Blaszkze, M., Koszewski, D.: Determination of low-level audio descriptors of a musical instrument sound using neural network. In: *Signal Processing: Algorithms, Architectures, Arrangements, and Applications Proceedings, SPA*, September 2020, pp. 138–141 (2020). <https://doi.org/10.23919/spa50552.2020.9241264>
22. Li, T., Ogihara, M., Li, Q.: A comparative study on content-based music genre classification. In: *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pp. 282–289, Toronto, Canada (2003). <https://doi.org/10.1109/ISSPA.2003.1224828>
23. Xu, C., Maddage, N.C., Shao, X., Cao, F., Tian, Q.: Musical genre classification using support vector machines. In: *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 429–432 (2003)
24. Fu, Z., Lu, G., Ting, K.M., Zhang, D.: Learning Naive Bayes classifiers for music classification and retrieval. In: *Proceedings of the International Conference on Pattern Recognition*, pp. 4589–4592 (2010). <https://doi.org/10.1109/ICPR.2010.1121>

25. Vishnupriya, S., Meenakshi, K.: Automatic music genre classification using convolution neural network. In: 2018 International Conference on Computer Communication and Informatics, ICCCI 2018, pp. 4–7 (2018). <https://doi.org/10.1109/ICCCI.2018.8441340>
26. Gwardys, G., Grzywczak, D.: Deep image features in music information retrieval. *Int. J. Electron. Telecommun.* **60**, 321–326 (2014). <https://doi.org/10.2478/eletel-2014-0042>
27. Jondya, A.G., Iswanto, B.H.: Indonesian’s traditional music clustering based on audio features. *Procedia Comput. Sci.* **116**, 174–181 (2017). <https://doi.org/10.1016/j.procs.2017.10.019>
28. Skidén, P.: New Endpoints: Audio Features, Recommendations and User Taste. <https://developer.spotify.com/community/news/2016/03/29/audio-features-recommendations-user-taste/>
29. McFee, B., et al.: librosa: audio and music signal analysis in Python. In: Proceedings of the 14th Python in Science Conference, pp. 18–24 (2015). <https://doi.org/10.25080/majora-7b98e3ed-003>
30. Stürmer, S., Simon, B.: The role of collective identification in social movement participation: a panel study in the context of the German gay movement. *Personal. Soc. Psychol. Bull.* **30**, 263–277 (2004). <https://doi.org/10.1177/0146167203256690>