Similarity Analysis of Modern Genre Music Based on Billboard Hits

AKASH POORANSINGH, (Member, IEEE), AND DYLAN DHORAY, (Member, IEEE)
Department of Electrical and Computer Engineering, The University of the West Indies, St. Augustine Campus, St. Augustine, Trinidad and Tobago
Corresponding author: Akash Pooransingh (akash.pooransingh@sta.uwi.edu)

ABSTRACT Mainstream music can be popularly categorized into specific genres: Rock, Country, Hip Hop, contemporary rhythm and blues (R&B) and Pop. Music of these genres are continually compiled on the Billboard music charts based on popularity. This paper explores the uniqueness of these genres and the possible melding of acoustic characteristics over time. Principal Component Analysis (PCA) are applied to timbral and non-timbral characteristics and compared for each genre. Results show that Hip Hop and Pop maintained a unique distinction over time compared to the other genres while Rock, Country and R&B began to share similar acoustic characteristics within recent times. Further analysis make an attempt to predict the trend of the acoustic nature of genres.

INDEX TERMS Music information retrieval, principal component analysis, acoustic applications.

I. INTRODUCTION
The advent of direct access to digitally stored music and popular streaming services, highlights the importance of content based services for the field. According to [1], the industry has become more competitive in finding innovative ways of sustaining revenue streams. A deeper understanding is important for artists and producers [2], in possibly creating the ideal musical content to maximize the popularity trends of the music consumers. Traditionally, music can be categorized according to genres. Genres are quite diverse and are based on geographical or cultural origin or based on social acceptable classification based on melody, rhythm, timbral and non-timbral characteristics [3]. Mainstream music can be popularly categorized into specific genres such as: Rock, Country, Hip Hop, contemporary rhythm and blues (R&B) and Pop. These music genres are continually compiled on the Billboard music charts based on popularity.

Thus far, most studies on music information retrieval focused on either feature extraction or classification techniques. Reference [3] provided an overview of music information retrieval applications. Here, music features are identified as: (i) Timbral and texture models (temporal features, energy features, spectral shape features (kurtosis, centroid spread, roll off frequency, slope, skewness, MFCCs etc.) and perceptual features), (ii) Melody/harmony as a pitch energy function of music notes (folded and unfolded) and (iii) Rhythm as a periodicity function. Reference [4] noted that timbral features are mostly originated from traditional speech recognition techniques. They are usually calculated for every short-time frame of sound based on the Short Time Fourier Transform (STFT). Here typical timbral features were identified as: Spectral Centroid, Spectral Rolloff, Spectral Flux, Energy, Zero Crossings, Linear Prediction Coefficients, and Mel-Frequency Cepstral Coefficients (MFCCs) and added a proposed the Daubechies Wavelet Coefficient Histograms (DWCH) as a comparable timbral feature. Reference [5] implemented a robust way to extract timbral features. The experimental results show that the employed features have different importance according to the part of the music signal from where the feature vectors were extracted. Reference [6] also used Mel Frequency Cepstral Coefficient (MFCC) as the main feature and reported an accuracy of 76%.

In [7], the impact of frame selection on automatic music genre classification was evaluated in a bag of frames scenario. Here a novel texture selector was used to identify diverse sound textures within each track. The results show that frame selection leads to significant improvement over the single vector baseline on datasets consisting of full-length tracks, regardless of the feature set. Alternative image based features were used in [8]. Here a computer vision technique,
ORB (Oriented FAST and Rotated BRIEF) was used for robust audio identification. The ORB compares the features of the spectogram image query to a database of spectogram images of the songs to good effect. For genre classification, [9] also used spectromer created from the audio signal.

Unsupervised and supervised approaches to classification were identified in [3], included: clustering algorithms, KNN, GMM, LDA, SVM, and ANN approaches. Reference [4] identified K-Nearest Neighbors and Gaussian Mixture models, Support Vector Machines, and Tree-Based Vector Quantization as possible classification techniques. References [6] and [9] provided a Deep Learning approach is used in order to train and classify the system. Here convolution neural network was used for training and classification. Reference [10] applied CNN combined with Recurrent Neural Network (RNN) architecture to implement a music genre classification model. In this study, a model is trained by Mel-Frequency Cepstral Coefficients (MFCC) and CRNN method with the accuracy achieve to 43%.

According to [11], principle component analysis (PCA) provides an approximation of multivariate data into smaller dimensions. Given a multivariate environment, PCA is used ideally for data simplification, dimension reduction outlier detection, classification and prediction. PCA is based upon eigen decomposition upon singular value decomposition of rectangular matrices [12]. Robust PCA has proven very effective in [13] where provided analysis for singing-voice separation from music accompaniment. Reference [14] showed the effective use of PCA for the decorrelation of features over MFCC for music information retrieval. In [15], PCA was used effectively in the automated re-tagging of music track labels with a recall improvement of 7%. The importance of building classification models for predicting popularity metrics using acoustic data was highlighted in [2]. MPEG-7 and Mel-frequency cepstral coefficient (MFCC) features were used to show how popularity metrics of a song can be predicted. However, no work was done thus far on analyzing the popularity and characteristics of each genre. In [16], an attempt to analyze the dependence of genres using an underlying hierarchical taxonomy. The aim here was to identify the relationship of dependence between different genres and provide valuable sources of information for genre classification. Although this was a good approach, the work did not consider the relation of genres. Reference [17] was able to focus on instrument separation to improve music genre classification, and in particular to decrease the misclassification between selected genres in the context of the influence of the specific instrument on selected genres.

No study, thus far, focused on the historical trend of genres and how each influence the other. This paper explores the uniqueness of these genres and the possible melding of acoustic characteristics over time by using PCA on a combination of timbral and non-timbral features. Here, the timbral and non-timbral features are analyzed individually for each genre then compared for notable similarities and differences. PCA was used to reduce the variable dimension of all

considered timbral and non-timbral features and compared for each genre. Finally, the general trend for some features were identified and highlighted as potential triggers for popularity prediction for each genre.

II. ACOUSTICAL METRICS

The acoustical metrics used in this study was split into two groups: non-timbral and timbral characteristics. In [18], it was noted that the MPEG Unified Speech and Audio Coding (USAC) [19] provides the best audio coding quality for both audio and speech signals with MPEG-H 3D providing additional mechanisms for intelligent gap filling. MPEG 7 low-level audio descriptors (LLD) are useful in describing audio and provides seventeen temporal and spectral parameters. The Audio Commons Initiative [20], provide a more practical approach in developing technologies to automatically describe sound and music for retrieval purposes that is easily accessible. This paper focuses on the primary non-timbral low-level audio descriptors as described in [19] and some of the timbral characteristics as described in [21].

A. NON-TIMBRAL

The non-timbral characteristics identified for this paper included the energy, power, pitch and tempo of the songs.

- **Energy** of a sound is the sum of absolute amplitude and was calculated as follows:

  \[ \text{Energy of song} = \sum \text{(Audio data of single channel)}^2 \]

  \[ \text{(1)} \]

  where the audio data represents the amplitude of the wave or loudness of the song.

- **Power** is defined as energy per second and was calculated by:

  \[ \text{Power} = \frac{1}{2\text{(Size of channel)}} \times \frac{\text{Energy of song}}{\text{sampling rate}} \]

  \[ \text{(2)} \]

- **Tempo** of a song is the speed at which it is played. For this application, the bpm was used and calculated by first reading the audio sampling frames and counting the number of beats in each sample. After the beats were calculated, it was then converted to bpm by the following:

  \[ \text{bpm} = \frac{60}{\text{discrete difference of beats}} \]

  \[ \text{(3)} \]

The median value was extracted and used for analysis. The algorithm used for computing the bpm used the Aubio python package and was based on the Spectral difference onset detection created by [22].

- **Pitch** of a song is the highness or lowness of a note being played. The pitch of a musical tone is effectively determined by the repetition rate of the sound [23]. Pitches are measured as a frequency. Like the bpm algorithm, the pitch computation also required the Aubio library...
and the audio file frames were read in the same manner. The algorithm used to calculate the pitch was developed by [24].

B. TIMBRAL

The timbre of music, according to [23], is what allows us to distinguish tones of a person’s voice or instrument. Therefore, the timbre of a song would be what makes it sound different for another song. The list of timbral characteristics fall under psychoacoustic attributes and therefore each characteristic creates a particular sensation when listening to a song. Psychoacoustics is a field which seeks to investigate the relationship between acoustics and psychology i.e. how sound is perceived and interpreted by humans. It should be noted that the score produced from the algorithm was not an exact representation of the characteristics because of the subjectivity associated with human sensations. Reference [25] provides a profile of timbral characteristics. However, [26] provides a deeper insight in describing timbral characteristics. Based on this, eight characteristics, can be identified as: hardness, depth, roughness, sharpness, boominess, brightness, warmth and reverberation.

- **Hardness** of a sound is a mixture of the loudness and harshness. When a sound is harsh it means that the audio core is unbalanced. An audio core is the frequency range (2KHz – 5KHz) which the human ear is most sensitive. So, hardness is a measure of how soothing a sound is i.e. if the balance between loudness and the frequency range is well received by the human ear.
- **Depth** is the front-to-back space, it is a measure of how different frequencies are overlaid to create the spacial effect of near and far sounds. So, a low depth score would mean that all the sounds meshed together and is difficult to discern one from the other while a high depth score would imply that every sound is very distinctive.
- **Roughness** as defined by [27] is a sensation experienced through relatively brisk changes in amplitude modulation, the change referred to in this case is not of fluctuating loudness but of sound quality [26]. The score returned by the roughness extractor is representative of the interaction of the amplitudes which emulate the roughness sensation.
- **Sharpness** as defined in [25] is the sensation associated with the high frequency components of a sound. The score returned from the extractor represents the amount of high frequency energy in comparison to the total energy of the sound.
- **Boominess** is the opposite of sharpness meaning that it is a measure of the amount of low frequency energy compared to the total energy of the sound. Not to be confused with warmth which is a sensation associated with both low frequencies and reverberation while boominess is use of excessive low frequencies. Both boominess and sharpness are associated with a sensation of pain where high sharpness would create a sharp ringing noise while high boominness would create a painful beating noise. So, the score returned for these characteristics are a measure of their respective frequencies relative to the others.
  - **Brightness** as described in [28] of a sound is the amount of mid to high frequency content of the sound. Terms such as “upperpartials” and “harmonics” refer to brightness as well. So, the score returned is an indication of the of the mid to high frequency content used.
  - **Warmth** is the prominence of bass frequencies over higher frequency sounds as well as the time a sound takes to reverberate at bass frequencies compared to higher frequency sounds. The score returned is an indication bass frequency compared to high frequencies.
  - **Reverberation** is a phenomenon that occurs due to a sound persisting after the source has stopped playing it. The persistence is a result of the sound reflecting off surfaces. Additionally, the sound decays as the surface absorbs some of the sound while reflecting it.

C. PRINCIPAL COMPONENT ANALYSIS (PCA) FOR ACOUSTIC ANALYSIS

Analysis of correlations of multiple dimensions of characteristics may be exhaustive and computationally expensive. PCA provides an opportunity for reduced complexity while maintaining original variance information when projecting lower dimensions.

PCA as explained by [11] is an unsupervised learning problem that reduces the dimensionality of a multivariate dataset and aims to preserve as much relevant information from the original dataset whilst undergoing the transformation. The implementation of PCA was adapted from [29], the algorithm comprised of three main steps, standardizing data, projecting data to a two-dimensional space, and visualizing data. The main part of the process is the projection onto a two-dimensional space. This was done in three steps based on [11], computing the covariance matrix, computing the eigen vectors and eigen values of the covariance matrix and sorting the eigen values in descending order. The PCA was applied to both timbral and non-timbral metrics and indices generated in the previous section. Scatter plots were generated to visualize the data.

III. ACOUSTICAL ANALYSIS OF MUSIC GENRES

Music from the Billboard Hot 100 were used, where 70% of the songs were used as the training set and the remaining 30% were used for validation. The top 10 songs for each genre on the Billboard charts on 1 August in five year periods from 1980 to 2020 were used for this analysis. An Intel Core i5-8250 CPU @ 1.60 GHz processor with 8 GB of RAM where the Ubuntu virtual machine was allocated 4 GB of RAM was used for processing. Non-timbral features (energy, power, pitch and tempo) were extracted using the scipy and pydub libraries while the timbral features (hardness, depth, brightness, roughness, warmth, sharpness, boominess and reverberation) were extracted using the timbral models library.
A. NON-TIMBRAL ANALYSIS
Figure 1 shows the variance and mean of each non-timbral characteristic (energy, power, pitch and tempo) across time for all five genres. Across the genres, each non-timbral characteristic followed the same general pattern over time i.e. the characteristics had low variances in 1980 and then it increased over time but coming closer to 2020 the variances generally started to decrease and in the year 2020 the variances were close to matching that of 1980. The same observation was made for the means plot of each characteristic, in 1980 they were generally close in value and over time they dispersed but coming closer to 2020 they began to converge. Additionally, both variance and mean plots for each characteristic generally take the same shape and have proportional changes to each other. This indicated there is a noticeable proportional relationship between variance and mean for non-timbral characteristics. Further observation of Figure 1e and 1f showed that Rock and Country had low variances across time in terms of pitch and their corresponding means were the most stable of the five genres indicating that Country and Rock music
FIGURE 2. Timbral statistical characteristics.

were similar in terms of pitch. From Figures 1a to 1d it was shown that even though there is a steep decline in variance between 2015 and 2020 for power and energy, the mean still increased and when compared to 1980, it was shown that the power and energy of the music genres generally increased over time. In terms of tempo, Figures 1g and 1h showed that there was a relatively high variance between the genres across the years but it reduced to approximately the same range (0-600) in 2020 and the means plot showed that the tempo for all genres over time were approximately stable and within the same range (110 to 130). Considering the variance of the tempo for all genres in 1980 and 2020 were relatively low, it can be inferred that modern songs have the same tempo as the songs in past.

B. TIMBRAL ANALYSIS

This section presents plots of the statistical data (variance and mean) computed from the raw timbral indices and metrics extracted from the audio files. Only the genres and time periods with noticeable trends and patterns were presented here.

With reference to Figures 2a and 2b the graph of variance across time for Country music showed that, generally there was a relatively low variance amongst the timbral characteristics. Because the means plot for each characteristics
were approximately constant from year to year implied that Country music was similar in terms of timbral characteristics across time. Observing Figures 2c and 2d, the same pattern emerges as with the Country music, the variance from year to year was relatively low and the means of the characteristics were approximately the same implying that HipHop music was also similar in terms of timbral characteristics across time. An interesting observation to note is that pre-2005 the timbral characteristics generally varied between 0 to 20 but post-2005 the variance graph took a prominent sinusoidal shape possibly indicating that every 10 years the composition of these songs were reevaluated to suit new generation of listeners. Referencing Figures 2e and 2f even though the mean of the timbral characteristics across genres for the year 1980 was approximately the same, the variance from genre to genre was relatively high indicating that the genres for that time period were distinct. Compared to Figures 2g and 2h the variance for Country, R&B and Rock in 2020 are low and fall within the range 0 to 15. Considering that the means of the characteristics in 2020 were more stable than in 1980 implied these genres are more similar in terms of timbral characteristics in 2020 than 1980.

Comparing the non-timbral variance across time to the timbral variance across time, it was clear that there exists an inverse relationship between them because the genres seemed to become more distinct over time in terms of non-timbral characteristics as opposed to them becoming less distinct over time in terms of timbral characteristics.

C. PCA FOR ACOUSTIC ANALYSIS
Scatter plots of PC1 vs PC2 which were the result of the PCA done in the methodology using the raw data extracted from the audio files. The specific scatter plots shown here correspond to the genres and time periods presented in the previous subsection.

Referencing the PCA scatter plots shown in Figures 3a and 3b, it was also shown that in 1980 the points were generally compact whereas in 2020 the points were more dispersed across the plot. This further justified the point made that in terms of non-timbral characteristics the genres were more distinct in 2020 than in 1980. A specific example of this can be seen upon observation of the points for HipHop and R&B in Figure 3a compared to Figure 3b. A shift from being relatively compact in 1980 to dispersed in 2020 was shown. The pattern observed from these PCA scatter plots were also reflective of the variance shown in Figures 1a to 1g, as mentioned in the timbral analysis, the dispersion of the points reflect the relatively high variance and by extension distinctiveness while the compactness reflects the relatively low variance and by extension similarity. It was inferred that based on the variance of genres over time, they were generally similar in 1980 and distinct in 2020, to further support this, Figures 3c and 3d showed the PCA scatter plots of HipHop and Country over time respectively. These genres were chosen to compare the observations made about these genres in term of timbral characteristics. It can be seen in both figures that in the year 1980 the points were compacted into a relatively small section of the graph and for 2020 the points were relatively more dispersed across the plot. From this behavior, the inference was further justified that in terms of non-timbral characteristics the genres became more distinct over time and the behavior was opposite to that of the timbral characteristics.

The inferences made from the analysis of the statistical data can be further supported by the PCA scatter plots shown in Figures 3e to 3h. Comparing Figures 3g to 2a, looking at years 1980, 2005 and 2015 when the variance of the timbral characteristics were relatively high, the points on 3g were very spaced out compared to the years 2000, 2010 and 2020 where the variance was relatively low. Further observation of the remaining years show that the points were approximately compacted the same way as the years with low variance except for some outliers. Additionally, since the means for the timbral characteristics were approximately constant, confirming that Country songs were generally similar to each other over the years. Looking at Figures 3h and 2c the same pattern with the variance and mean shown for Country music emerged for HipHop. The points representing years with relatively high variance (2010 and 2020) had a wider spread across the plot compared to years with lower variance (2005 and 2015). The remaining years were relatively close such that the inference of HipHop music being similar in terms of timbral characteristics over the years was justified. Comparing Figures 3e to 3f and Figures 2f and 3f to Figures 2g and 2h respectively. The points for the year 1980 across the genres were quite spread out in comparison to the points for 2020. This spread for the year 1980 was also reflective of the variance plot shown in Figure 2e further indicating that the genres were distinct in 1980. Now comparing Figures 3f and 2g, the points representing the genres Country, R&B and Rock are more compact compared to those of 1980. Looking at Figures 2g and 2h, these genres had variances within the same range (3 to 15) as well as approximately stable means for each characteristic thus implying that these genres were less distinct from each other in 2020. Therefore, in terms of timbral characteristics Country, R&B and Rock were less distinct in 2020 than they were in 1980.

D. CORRELATION ANALYSIS
Another important aspect of the similarity analysis involves correlating the characteristics to each other, hence the use of correlation matrices.

Heat maps of correlation matrices that were generated from the raw timbral data extracted from the audio files. The colour scale to the right of the map are an indication of how close the score is to 1. The heat maps shown here correspond to the genres and time periods chosen in the statistical data subsection. The colour scale on the right are an indication of how close a value is to 1.

From the correlation matrix shown in Figure 4a, the pairs of boominess & brightness and brightness & warmth have a relatively high negative correlation. Looking at the means
plot in Figure 2b this relationship was seen by the subtle changes observed. This indicates that over the years when country songs were being composed these characteristics were adjusted inversely to each other. On the other extreme, boominess & warmth, hardness & sharpness, roughness & sharpness, and reverb & warmth have high positive correlation to each other which is also supported by the means plot in figure 16 as the characteristic pairs vary proportionally to each other. Observing the correlation matrix for HipHop music over the years in Figure 4b. The same positive and negative characteristic pairs from the Country genre tend more towards neutrality in value indicating that the strength of the correlations decreased for HipHop music. With respect to Figure 4c the characteristics pairs of boominess & depth, brightness & hardness, brightness & roughness, brightness & sharpness and hardness & sharpness have high positive correlations across genres in 1980. The pairs boominess & roughness, brightness & warmth, depth & roughness,
hardness & warmth, sharpness & warmth have high negative correlations. The remaining pairs are more neutral. Comparing the correlations of Figures 4c to 4d there is a large increase of neutrality amongst the characteristic pairs with brightness & sharpness and roughness & sharpness being the only high positive correlations and sharpness & warmth pairs being the only high negatively correlated pairs. These correlations are verified by observing the means plot for 1980 and 2020. It can therefore be said that the correlations between the timbral characteristics have become weak over the years and are more neutral in relationship i.e. they were more independent of each other in 2020 than they were in 1980. Finally, the correlations between the timbral characteristics and the principal components indicate how each component was influenced by each characteristic i.e. how the points are placed on the scatter plot. For example, in Figure 4a, it was shown that PC1 was mostly negatively correlated to the characteristics and PC2 was positively correlated. So, points further to the right on the x axis correspond to lower values of timbral characteristics while points higher on the y axis correspond to higher values of timbral characteristics.

**E. TIMBRAL VERSUS NON-TIMBRAL ANALYSIS**

The PCA scatter plots presented here are the timbral PC1 against the non-timbral PC1. As with the previous PCA scatter plots, the data used to generate it was the raw metrics and indices. The genres and time periods shown here correspond with those chosen in the timbral and non-timbral subsections. For this analysis PCA scatter plots of timbral PC1 vs non-timbral PC1 were shown. As mentioned in the previous sub-section, it was inferred that the timbral and non-timbral characteristics were inversely related in terms of distinction. This meant that for timbral characteristics, the genres became less distinct over time or at least, they were less distinct in 2020 than in 1980. But, in terms of non-timbral characteristics, the genres were more distinct in 2020 than in 1980. So, observing Figure 6a, it was shown that for Country music in 1980 the points were relatively close to each other along the x-axis and more dispersed across the y-axis. The points representing the year 2020 were more dispersed along the x-axis and closer together along the y-axis. This matched the
variance plots shown for 1980 Country music in Figure 2e for the timbral characteristics, there was a relatively high variance while in the non-timbral variance plot there was relatively low variance of the characteristics. Then in the year 2020 the timbral characteristics had a lower variance amongst them but the non-timbral characteristics had a higher
variance. The same observation was made for the HipHop genre in Figure 6b. In Figure 6c, the points representing the genres Country and HipHop were relatively close to each other along the x-axis which is reflective of the low variance they had in the non-timbral variance plot. Compared to the y-axis the points were more dispersed which also reflected the variance of the timbral characteristics shown in Figure 2e. Figure 6d was essentially the transpose of Figure 6c, the points for Country and HipHop were dispersed across the x-axis which reflected the variance shown in Figures 1a to 1g and closer along the y-axis, which was reflective of the variance plot shown in Figure 2g of the non-timbral and timbral characteristics.

1) PREDICTION ANALYSIS

Figures 7a and 7b show the regression model generated from the raw timbral and non-timbral data using the song database as the training set and one country song from 2021 to test how accurate the regression model was. The figures were shown to illustrate how music producers and artists can trend the general behavior of the timbral and non-timbral characteristics and compare new music samples to the general trend. Predictions for future years can be implied by the general trends of these graphs. In this example, it was observed that data has relatively high variation along the y-axis for each year. Given this behavior, the regression model took a path that passed approximately through the middle of the set of points. It can also be seen that for the non-timbral characteristic (i.e tempo) the regression model was closer to the actual value compared to the regression model for the timbral characteristic. Furthermore, the direction that the regression model was headed towards was opposite to the real data point, indicating further inaccuracies from the model. The reason for this difference can be attributed to the score returned for the timbral characteristics are also predicted using a regression model. It can then be inferred that using predicted values as input for a prediction algorithm may not result in accurate results.

IV. CONCLUSION

In this paper we have successfully analyzed specific genres of Billboard music over time (1980 to 2020): Rock, Country, Hip Hop, contemporary rhythm and blues (R&B) and Pop using a traditional music information retrieval approach. Our method used PCA to aggregate timbral and non-timbral characteristics and compared each genre for independence and similarity over time. This paper found that over time, Hip Hop and Pop maintained a unique distinction compared to the other genres while Rock, Country and R&B began to share similar acoustic characteristics within recent times. Further analysis make an attempt to predict the trend of the acoustic nature of genres that may be useful for artist and producers. For the future, a similar analysis can be conducted to include a wider range of acoustic features (timbral and non-timbral), as well as an improvement on prediction methodology with the use of different classification systems including deep neural networks.

REFERENCES

[1] K. Naveed, C. Watanabe, and P. Neittaanmäki, “Co-evolution between streaming and live music leads a way to the sustainable growth of music industry—Lessons from the U.S. Experiences,” Technol. Soc., vol. 50, pp. 1–19, Aug. 2017.
[2] J. Lee and J.-S. Lee, “Music popularity: Metrics, characteristics, and audio-based prediction,” IEEE Trans. Multimedia, vol. 20, no. 11, pp. 3173–3182, Nov. 2018.
[3] N. Scaringella, G. Zoa, and D. Mlynek, “Automatic genre classification of music content: A survey,” IEEE Signal Process. Mag., vol. 23, no. 2, pp. 133–141, Mar. 2006.
[4] T. Li, M. Ogihara, and Q. Li, “A comparative study on content-based music genre classification,” in Proc. 26th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR), 2003, pp. 282–289.
[5] C. N. Silla, Jr., A. L. Koerich, and C. A. A. Kaestner, “Feature selection in automatic music genre classification,” in Proc. 10th IEEE Int. Symp. Multimedia, Dec. 2008, pp. 39–44.
[6] S. Vishnupriya and K. Meenakshi, “Automatic music genre classification using convolution neural network,” in Proc. Int. Conf. Comput. Commun. Informat. (ICCCI), Jan. 2018, pp. 1–4.
[7] J. H. Foleis and T. F. Tavares, “Texture selection for automatic music genre classification,” in Appl. Soft Comput., vol. 89, Apr. 2020, Art. no. 106127.
[8] D. Williams, A. Pooransingh, and J. Saitoo, “Efficient music identification using ORB descriptors of the spectrogram image,” EURASIP J. Audio, Speech, Music Process., vol. 2017, no. 1, pp. 1–17, Dec. 2017.
[9] R. L. Aguiar, Y. M. G. Costa, and C. N. Silla, “Exploring data augmentation to improve music genre classification with ConvNets,” in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2018, pp. 1–8.
[10] Y-H. Cheng, P-C. Chang, D-M. Nguyen, and C-N. Kuo, “Automatic music genre classification based on CRNN,” Eng. Lett., vol. 29, no. 1, 2020.
[11] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” Chemometrics Intell. Lab. Syst., vol. 2, nos. 1–3, pp. 37–52, 1987.
[12] H. Abdi and L. J. Williams, “Principal component analysis,” Wiley Interdiscip. Rev., Comput. Statist., vol. 2, no. 4, pp. 433–459, 2010.

[13] P.-S. Huang, S. D. Chen, P. Smaragdis, and M. Hasegawa-Johnson, “Singing-voice separation from monaural recordings using robust principal component analysis,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Mar. 2012, pp. 57–60.

[14] Y. Vaizman, B. McFee, and G. Lanckriet, “Codebook-based audio feature representation for music information retrieval,” IEEE/ACM Trans. Audio, Speech, Language Process., vol. 22, no. 10, pp. 1483–1493, Oct. 2014.

[15] Y.-H. Yang, D. Bogdanov, P. Herrera, and M. Sordo, “Music retagging using label propagation and robust principal component analysis,” in Proc. 21st Int. Conf. Companion World Wide Web (WWW Companion), 2012, pp. 869–876.

[16] T. Li and M. Ogihara, “Music genre classification with taxonomy,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), vol. 5, Mar. 2005, p. v-197.

[17] A. Rosner and B. Kostek, “Automatic music genre classification based on musical instrument track separation,” J. Intell. Inf. Syst., vol. 50, no. 2, pp. 363–384, Apr. 2018.

[18] S. R. Quackenbush and J. Herre, “MPEG standards for compressed representation of immersive audio,” Proc. IEEE, vol. 109, no. 9, pp. 1578–1589, Sep. 2021.

[19] Information Technology—MPEG Audio Technologies—Part 1: MPEG Surround, Standard ISO-IEC23003-1, 2007.

[20] F. Font and X. Serra, “The audio commons initiative,” in Proc. Int. Soc. Music Inf. Retr. Conf. (ISMIR), 2015.

[21] A. Pearce, S. Safavi, T. Brookes, R. Mason, W. Wang, and M. Plumbley, “D5.6 second prototype of timbral characterisation tools for semantically annotating non-musical content,” Audio Commons, Ecosyst. Creative Reuse Audio Content, Inst. Sound Recording, Univ. Surrey, Tech. Rep., 2017.

[22] J. Foote and S. Uchihashi, “The beat spectrum: A new approach to rhythm analysis,” in Proc. IEEE Int. Conf. Multimedia Expo (ICME), Aug. 2001, p. 224.

[23] R. D. Patterson, E. Gaudrain, and T. C. Walters, “The perception of family and register in musical tones,” in Music Perception. New York, NY, USA: Springer, 2010, pp. 13–50.

[24] A. de Cheveigné and H. Kawahara, “YIN, a fundamental frequency estimator for speech and music,” J. Acoust. Soc. Amer., vol. 111, no. 4, pp. 1917–1930, 2002.

[25] T. H. Pedersen, “The semantic space of sounds,” Delta, Denmark, Tech. Rep., 2008.

[26] P. N. Vassilakakis and K. Fritz, “SRA: A web-based research tool for spectral and roughness analysis of sound signals,” in Proc. 4th Sound Music Comput. (SMC) Conf., 2007, pp. 319–325.

[27] E. Zwicker and H. Fastl, Psychoacoustics: Facts and Models, vol. 22. Berlin, Germany: Springer, 2013.

[28] C. Saitis and S. Weinzierl, “The semantics of timbre,” in Timbre: Acoustics, Perception, and Cognition. Cham, Switzerland: Springer, 2019, pp. 119–149.

[29] M. Galarnyk, “PCA using Python (scikit-learn),” Towards Data Sci., Canada, Tech. Rep., 2017.