2015 Pattern Recognition Association of South Africa and Robotics and Mechatronics
International Conference (PRASA-RobMech)
Port Elizabeth, South Africa, November 26-27, 2015

Single-labelled Music Genre Classification Using Content-Based Features

Ritesh Ajoodha, Richard Klein, and Benjamin Rosman

***Abstract*—In this paper we use content-based features to perform automatic classification of music pieces into genres.**

We categorise these features into four groups: features extracted from the Fourier transform's magnitude spectrum, features designed to inform on tempo, pitch-related features, and chordal features.

We perform a novel and thorough exploration of classification performance for different feature representations, including the mean and standard deviation of its distribution, by a histogram of various bin sizes, and using mel-frequency cepstral coefficients.

Finally, the paper uses information gain ranking to present a pruned feature vector used by six off-the-shelf classifiers. Logistic regression achieves the best performance with an 81% accuracy on 10 GTZAN genres.

***Index Terms*—Music genre classification, feature selection, feature representation, MFCC aggregation, area moments, tempo detection, pitch detection, chordal identification, information gain ranking.**

I. Introduction

MUSIC genre, while often being vaguely specified, is perhaps the most common classification scheme used to distinguish music. Although single human responses to genre classification can be biased and stereotypical, there exists a consensus of broad genre definitions across populations worldwide.

Genre classification is one of multiple music classification methods, including mood and artist classification. Although these methods are also similarity-based measures across different music meta-data (e.g. lyrics, artist, timbre), genre offers a culturally authorised prominence on the construction of traditional classes which is more functional for music classification.

Music genre has such a pressing influence on consumers that a listener may prefer one song to another based more on the song's genre than the actual song itself [1] [2]. End-users are more likely to browse music by genre than artist similarity, recommendation, or even music similarity [3]. Therefore, successful music genre classification algorithms will enable users to browse music within genre categories.

Our aim is to explore the space of automatic music genre classification, so as to decrease search-time for music pieces within large databases. This system relies only on the audio signal itself and does not consider any meta-data.

Observing interactions between genre classes through content-based features can unveil cultural associations that exist between these genre classes and is of musicological significance [4].

R. Ajoodha is with the Department of Computer Science and Applied Mathematics, The University of the Witwatersrand, Johannesburg, South Africa. e-mail: ritesh.ajoodha@gmail.com.

R. Klein is with the Department of Computer Science and Applied Mathematics, The University of the Witwatersrand, Johannesburg, South Africa. e-mail: richard.klein@wits.ac.za.

B. Rosman is with the Council for Scientific and Industrial Research as well as the Department of Computer Science and Applied Mathematics, The University of the Witwatersrand, South Africa. e-mail: brosman@csir.co.za.

II. Background

Music Genre Classification is the process of categorising music pieces using traditional and cultural aspects. These traditions and cultures are not precisely defined and so over the years it has become vague as to what characteristics secure music to a particular genre.

Traditional musical aspects are given by four characteristics [5]: melody, harmony, rhythm and sound (timbre, dynamics, and texture) which are hypothesised to contribute considerably to the notion of musical genre. However, standard genre textbook definitions are qualitative, subjective, context dependent, and therefore are difficult to automate.

As a result of the ambiguities that exist between content-based genre definitions the ground truth classification accuracy becomes inescapably bounded as many people may disagree on a particular genre classification of a piece of music.

Composers often do not abide by "genre definitions", which makes us question whether some composers are accepted by currently "defined" music genres. For this reason music genre classification is categorised using human discretion and is therefore prone to errors and subjectivity as many pieces of music sit on boundaries between genres [6].

Successful genre classification makes use of cultural-based rather than content-based feature dissimilarity between genre classes. In this work we do not consider such meta-data, due to lack of availability.

Many genres do not only sound similar but also contain multiple sub-genres which share some similar characteristics. The difficulty of genre classification thereby increases when considering hundreds of other genre types and their respective sub-genres.

Although some authors provide an awareness of genre classification performance bounds imposed by human responses to genre classification [7] [8], further study in experimental research is needed to draw more concise conclusions regarding human responses to genre classification and how this affects ground truth. Humans are biased and subjective in genre classification, which ultimately leads to a lack of consensus in genre labels and thus poor quality of ground truth.

978-1-4673-7450-7/15/$31.00 ©2015 IEEE

To make matters worse, genre definitions evolve over time and continuously give rise to newer structures that have significantly different feature compositions [9]. Therefore, regardless of feature dimensionality, well-built classification procedures are required to classify features successfully with some regard for genre development. This is difficult as the scalability of even the most powerful supervised classification models are unsatisfactory [10].

Benetos and Kotropoulos (2008)	75.0%
Bergstra et al. (2006)	82.5%
Holzapfel and Stylianou (2008)	74.0%
Li et al. (2003)	79.7%
Lidy et al. (2007)	76.8%
Panagakis et al. (2008)	78.2%
Sturm (2013)	83.0%
Tzanetakis and Cook (2002)	61.0%

TABLE I: Noteworthy genre classification algorithms on 10 GTZAN genres.

A review of the literature shows very few capable genre classification systems using the GTZAN dataset. The GTZAN dataset is a collection of 1000 thirty second excepts. The 1000 music excepts are categorised into 10 genres, 100 excepts for each genre. Systems thus far have not adopted automatic genre classification models for media retrieval and recommendation. Successful genre classification includes work by Sturm (2013) [11] who achieved 73-83% on 10 genres; and Bergstra et. al. (2006) [12] who achieved 82.50% on 10 GTZAN genres [13]. Table I shows some noteworthy music genre classification algorithms on 10 GTZAN genres.

III. Feature Analysis

Music comprises of instrument sounds, speech sound, and environmental sounds [14] [15]. In this section we present several features that are hypothesised to be characteristics that can be used to correctly classify musical genre.

These features are organised into four main categories: *Magnitude-based features*, where timbral features that describe loudness, noisiness, compactness, etc. are presented; *Tempo-based features*, where methods that explore rhythmic aspects of the signal are provided; *Pitch-based features*, where algorithms that describe the pitch of music signals are presented; and finally, *Chordal Progression features*, where we explore chroma as a chordal (environmental) distinguishing feature.

Before we present these four families of features, we will firstly introduce four feature representations that will be explored for each feature distribution.

A. Feature Representation

In addition to the **mean**, the following feature representations will be applied to each feature and the best representation for each feature will be used in the final classification.

The Feature Histogram: The feature histogram arranges the feature's local window intensities into bin ranges. The content of each bin is counted and modelled by a frequency histogram. The histogram bin values are normalised and used for classification.

MFCC Aggregation: MFCC representation is a well-known feature representation that takes the first n MFC coefficients (coefficients that make up the short-term power spectrum of sound) as it would a 16khz signal [16] [17]. If the feature contains more than one dimension, then each dimension is assessed independently and n coefficients will be produced per dimension. In this paper we set $n = 4$.

Area Moments: Image moments is a central concept in computer vision and has its root in image processing. Fujinaga (1996) [16] produced 10 such moments for image processing: an image is treated as a 2-dimensional function $f(x, y) = z$, where x and y are indexes of the underlying matrix. The feature values extracted from the audio signal will be treated as a 2-dimensional image and Fujinaga's moments algorithm will be applied to the feature vector.

B. Magnitude-based Features

The magnitude spectrum, obtained from the fast Fourier transform of a signal, houses a family of spectral features which can be used for genre classification. Exploration of the magnitude spectrum has allowed us to identify signal change, noisiness, loudness and many other spectral features that describe aspects of discrete time signals for automatic music genre classification.

Exploring peak-based features, from the local maxima of the frequency domain, creates opportunities to analyse the signal more thoroughly. In this section we explore magnitude-based features for music genre classification.

1) Spectral Slope: The *spectral slope* can be observed when natural audio signals tend to have less energy at high frequencies. Peeters (2004) [18] provides a way to quantify this by applying a linear regression to the *magnitude spectrum* of the signal, which produces a single number indicating the slope of the line-of-best-fit through the spectral data.

2) Compactness: Compactness is a measure of the noisiness of a signal [17] and is calculated by comparing the value of a magnitude spectrum bin with its surrounding values. In many genres (e.g. metal) a random and persistent disturbance that obscures the clarity of sound is desired, which this feature will detect. Figure 3(a) shows the compactness feature values distributed over 10 GTZAN genres.

3) Loudness: Specific loudness is the loudness associated with each critical band of hearing. Total loudness has been used for multi-speaker speech activity detection, automatic speech recognition, instrument recognition and music genre classification.

4) Onset Detection: Onset detection describes information about the initial magnitude of a piece of music [19]. This feature describes the rise in magnitude from zero to some initial value.

5) Peak Detection: Studying the peaks of a signal allows us to account for various principal features that are contained within a signal. For example, peak-based features such as crest factor, peak flux, centroid, and smoothness can help us describe the quality of AC waveform power and detecting vibration. The peak detection algorithm by [20] will be used for extracting peak-based features. Mckay (2005) calculated peaks by detecting local maxima in the frequency bins, and these maximum are calculated within a threshold where the largest maxima within this threshold is considered [20]. These global peaks per threshold are considered without any information about bin location. In our experiments we took a peak threshold of 10. Treating this set of peak values together as a 16khz signal, we then represent these peak values by the centroid, flux, and smoothness features.

6) Spectral Flux: Spectral flux is a content-based feature that measures the rate of change of the magnitude spectrum. This is achieved by comparing every frame of the magnitude spectrum with its previous frame.

7) Spectral Variability: Statistical variability measures dispersion in data, i.e. how closely or spread-out the signal is clustered. We can achieve this by measuring the standard deviation of the magnitude spectrum of the signal.

8) Mel-Frequency Cepstral Coefficients: Mel-frequency cepstral coefficients (MFCCs) are the coefficients that together make up a Mel-frequency cepstrum. The components of MFCC are those from the cepstral representation of the audio signal. In the Mel-frequency cepstrum the frequency bands are equally spaced which favours the human auditory system more than using the cepstrum feature alone, which uses linearly-spaced frequency bands.

9) Spectral Flatness: Spectral flatness is a feature used to calibrate how pure tonal sounds are in comparison to noisy ones. Pure tonal sound refers to resonant structure in a power spectrum, compared to other parts containing white noise.

10) Spectral Rolloff: According to [18], [12], spectral rolloff point is the frequency such that 85% of the signal energy is contained below this frequency. It is correlated with the harmonic/noise cutting frequency [18]. Figure 3(d) shows the spectral rolloff feature values distributed over 10 GTZAN genres.

C. Tempo Detection

Most music retains regular rhythmic formations that creates an impression of tempo. With the purpose of understanding the nature of music to perform genre classification, tempo must be understood and preserved as a feature description. In this section we establish tempo detection schemes for music genre classification. Having already established a method to detect the vitality in a music excerpt by using spectral energy, which is the root mean square (RMS) of the music signal, we present in this section the beat histogram as a crucial feature vector.

1) Energy: Energy is a fundamental descriptor used in speech and audio processing [21]. Energy is measured by calculating the RMS of a discrete-time signal. Figure 3(c) shows the energy feature values distributed over 10 GTZAN genres.

Examining the arithmetic average of the first n windows of a signal (for our experiments we took $n = 100$) and calculating the fraction of these which are below the average, we can calculate the percentage of silence that exists in the signal - as the *fraction of low energy*.

2) The Beat Histogram: The beat histogram is an arrangement of signal strength to yield rhythmic intervals. This is accomplished by measuring the energy of n consecutive windows and computing the fast Fourier transform of the result. This type of feature will produce a very large design matrix and so a simple feature representation is needed. In our experiments the mean feature representation outperformed MFCC and the 20-bin feature histogram.

D. Pitch and Speech Detection

Pitch is a perceived characteristic contained in the frequency of music content. Most music of the same genre exhibit melodies that are just combined notes from a scale set. For example, most notes from an impressionistic piece are taken from whole-tone scales, whereas notes from a jazz pieces of music are taken from pentatonic scales. However, often environmental sounds overtone pitch, disguising available pitch-related elements, which make it difficult to extract pitch computationally. Even human auditory systems can find it difficult to distinguish pitch under these conditions.

In this section we explore pitch and speech related algorithms as an amalgam of these characteristics are hypothesised to describe singing. Together, pitch and speech detection schemes can help us understand gliding, portamento, or even vibrato.

1) Amplitude Modulation: For many musical instruments amplitude periodic modulation is a distinctive quantity. Style introduces characteristic amplitude variation into music tones. It has been observed that changing amplitude envelopes leads to similarity judgments on musical timbre [22]. The energy envelope is useful to extract features measuring amplitude modulation (AM). It has been observed that heuristic strength and frequency of AM can be calculated at two frequency ranges: the first range is between 4 and 8 Hz (where the AM is in conjunction with vibrato) and the second range is between 10 to 40 Hz which correspond to “graininess” or “roughness” of the tone.

2) Zero Crossing Rate: The *zero crossing rate* (ZCR) is the frequency of sign changes that occur along a discrete-time signal. Being a thorough percussive descriptor, this feature has been used in both speech recognition as well as in audio information retrieval. Figure 3(b) shows the strongest beat values distributed over 10 GTZAN genres.

E. Chordal Progressions

Introducing spectral feature extraction to genre detection problems created opportunities to exploit single characteristics of music. Chord structure and progressions have defining traits of music for many years.

1) Chroma: Chroma is defined as a 12 component design matrix where each dimension represents the intensity associated with a particular semitone, regardless of octave [23]. This section implements MFCC-based chroma by extracting MFCCs derived from a 12 component chroma design matrix. Since the components of chroma describe the distribution of semitones in a piece of music, it also informs us how notes are arranged and thus provides information about chordal harmonies. Therefore, modelling chroma indicates if a particular genre displays an attachment or relation to harmonic chordal progressions, as some genres do.

IV. Feature Selection

In the upper part of Table II, we present the features mantained after using the Information gain ranking algorithm. Information gain ranking is a filter method that evaluates the worth of a feature by measuring the information gain with respect to the class. The lower part of Table II lists the eliminated features.

The cut-off point was chosen by considering Figure 1, which shows the results of taking different numbers of features with the highest contributions and using them to classify 10 GTZAN genres. The red-line in Figure 1 shows the cutoff-point taken at 459 features in its respective representation. Figure 1 suggests we could have chosen about 100 features and achieved between 70-75% classification accuracy with minimal performance loss, but we extended this for robustness reasons.

V. Automatic Music Genre Classification

In this section we use the selected features outlined in the upper portion of Table II to perform genre classification on 10 GTZAN genres. Table III yields the results of this experiment: the first column lists the classifiers used; the second column gives us the accuracy for each classifier to correctly identify 10 genres; finally, the third column lists the time to build each classification model. The implementation details of each of the algorithms are outlined in Table IV. Six of-the-shelf classifiers were used: Naïve Bayes; Support Vector Machines; Multilayer Perceptron; Linear Logistic Regression Models; K-Nearest Neighbours; and Random Forests.

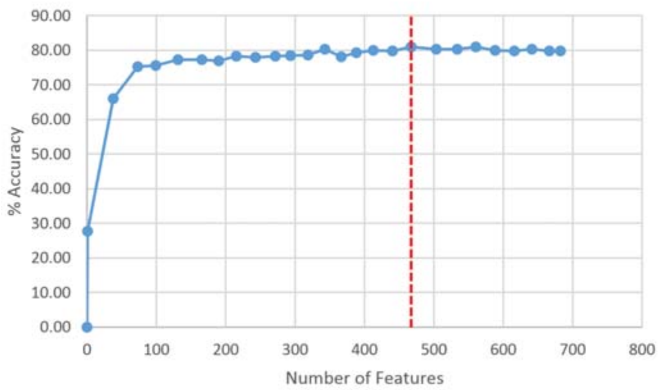


Fig. 1: Classification accuracy vs the number of highest contributing features to classify 10 GTZAN genres.

It is seen that all of the classification algorithms outperform the Naïve Bayes method. The Support Vector machine, Multilayer Perceptron and Random Forests are aligned by their performance with the Multilayer Perception taking the most time to build. The Linear Logistic Regression Model provides the best classification score of 81%. However, with the exception of the Multilayer Perceptron, the Linear Logistic Regression Model takes the longest time to build. Figure 2 shows the confusion matrix for 10 GTZAN genres using Linear Logistic Regression Models with 10-fold cross validation. The particular cluster overlap between rock and country music (and rock and disco) is observed. Although our results are in line with the best performing methods, and we have not exceeded them, we offer a valuable contribution in the form of feature analysis and representation for music genre classification.

Features Maintained	**Rep.**	**Dim. 459**
Spectral Flux	MFCC	4
Spectral Variability	MFCC	4
Compactness	Mean + SD	2
MFCCs	MFCC	52
Peak Centroid	Mean + SD	2
Peak Smoothness	SD	1
Complex Domain Onset Detection	Mean	1
Loudness + Sharpness and Spread	Mean	26
OBSI + Radio	Mean	17
Spectral Decrease	Mean	1
Spectral Flattness	Mean	20
Spectral Slope	Mean	1
Shape Statistic spread	Mean	1
Spectral Centroid	MFCC	4
Spectral Rolloff	SD	1
Spectral Crest	Mean	19
Spectral Variation	Mean	1
Autocorrelation Coefficients	Mean	49
Amplitude Modulation	Mean	8
Zero Crossing + SF	MFCC	8
Envelope Statistic Spread	Mean	1
LPC and LSF	Mean	12
RMS	Mean + SD	2
Fraction of Low Energy	Mean	1
Beat Histogram	SD	171
Strength of Strongest Beat	Mean	1
Temporal Statistic Spread	Mean	1
Chroma	MFCC	48
Features Eliminated	**Rep.**	**Dim. 223**
Peak Flux	20-bin FH	20
Peak Smoothness	Mean	1
Shape Statistic centroid, skewness	Mean	1
Shape Statistic Kurtosis	Mean	2
Strongest Frequency of Centroid	MFCC	4
Spectral Rolloff	Mean	1
Strongest Frequency of FFT	MFCC	4
Envelope Centroid, Skewness and Kurtosis	Mean	4
Beat Histogram	Mean	171
Strongest Beat	Mean + SD	2
Strength of Strongest Beat	SD	1
Fraction of Low Energy	SD	1
Beat Sum	MFCC	4
Relative Difference Function	MFCC	4
Temporal Statistic Centroid	Mean	1
Temporal Statistic Skewness	Mean	1
Temporal Statistic Kurtosis	Mean	1

TABLE II: The features maintained (upper portion) and the eliminated features (lower portion). Column two and three list the feature representation and feature dimension respectively.

VI. Conclusion and Recommendations

Although recent classification accuracy suggests that the performance of learning models for genre classification have become bounded, there is no confirmation to date to suggest these bounds cannot be exceeded. Nonetheless, small changes to existing models are unlikely to produce significantly better classification scores. Therefore, more attention to how feature extraction and classification are performed, or perhaps completely new approaches, are crucial to greatly exceed these bounds.

Classifier	Accuracy	Time to build model
Naïve Bayes	53.2%	0.56 sec
Support vector machines	75.4%	3.82 sec
Multilayer perceptron	75.2%	27.48 sec
Linear logistic regression models	81.00%	25.25 sec
K-nearest neighbours	72.80%	0.01 sec
Random forests	75.7%	18.08 sec

TABLE III: Automatic genre classification using the thinned feature vector.

Classifier	Parameters used
Naïve Bayes	Used a normal distribution for numeric attributes and supervised discretization
Support vector machines	Kernal degree = 3; tolarance of termination criteria = 0.001; epsilon for the loss function = 0.1; did not normalise; used polynomial kernal: $(gamma * u'v + coef0)^{degree}$.
Multilayer perceptron	Number of hidden layers = Number of classes; learning rate = 0.3; training time = 500 epochs; validation threshold = 20.
Linear logistic regression models	Maximum number of iterations for LogitBoost = 500
K-nearest neighbours	Number of neighbours to use = 1; using the absolute error for cross validation; Linear search algorithm
Random forests	Number of trees used = 1000

TABLE IV: Implementation details of each classification algorithm.

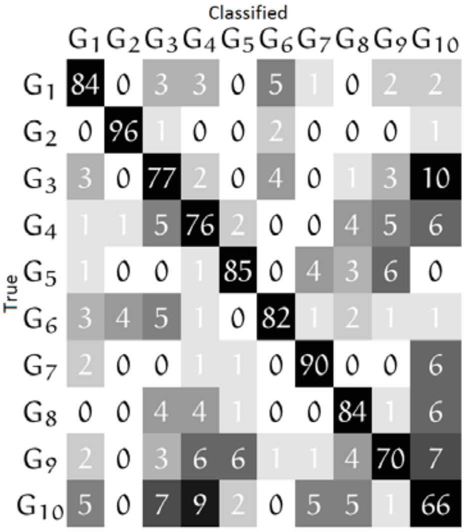


Fig. 2: The confusion matrix for 10 GTZAN genres using linear logistic regression models with 10-fold cross validation. The row and column labels represent genre labels where: G_1 = **Blues**, G_2 = **Classical**, G_3 = **Country**, G_4 = **Disco**, G_5 = **Hiphop**, G_6 = **Jazz**, G_7 = **Metal**, G_8 = **Pop**, G_9 = **Reggae**, and G_{10} = **Rock**.

Erroneous genre labels are often caused by inexperienced respondents and not being exposed to enough of the recording [7], [24], [8]. The reliability of a learning model is purely measured by the quality of its ground truth and so extensive measures must be taken to ensure that the ground truth is well founded and motivated.

Since genre classification is usually performed by humans who observe cultural features (observations of arts and other manifestations of genre cognitively regarded collectively) more than content related features, we should not expect to achieve ground breaking results by classifying genre purely on content-based features. This is evident as the best genre classification algorithms using content-based features only achieve between 75-83% on 10 GTZAN genres.

Incorporating cultural features with structural ones in the feature domain could notably increase current classification rates [25]. Large scale musical structures are present in most music genre types. Understanding the form (cyclic, binary, rondo) of a piece of music can immediately designate a small set of potential genre categories to which the piece could belong. These overall structure-based feature descriptions can be preserved in learning models by using classifiers that exhibit memory[1]. Preserving memory in learning models have been mostly ignored and could hold the key to better understanding chordal progressions and complex melodic structures.

The musicality of a listener is not only required when constructing ground truth, but can also be used to satisfy a particular customer's genre preference. Further empirical research in human responses to genre classification can reveal if certain consumers with different musicality will appreciate music differently. Empirical research should compare and contrast different classification scores for different kinds of customers in terms of age, culture, and musicality. This type of psychological research will enhance our understanding of the possibilities to increase the dependability of ground truth and will also allow us to personalise multiple learning models to cater for groups of individuals' needs rather than forcing a one fits all approach.

[1]Much like how hidden Markov models or recurrent neural networks work.

ACKNOWLEDGMENT

The first author gratefully acknowledges the support of the *NRF Scarce Skills Doctoral Scholarship*, and *the Post-Graduate Merit Award Scholarship* for Doctoral Studies.

REFERENCES

[1] A. C. North and D. J. Hargreaves, "Liking for musical styles," *Musicae Scientiae*, vol. 1, no. 1, pp. 109–128, 1997.

[2] H. G. Tekman and N. Hortacsu, "Aspects of stylistic knowledge: What are different styles like and why do we listen to them?" *Psychology of Music*, vol. 30, no. 1, pp. 28–47, 2002.

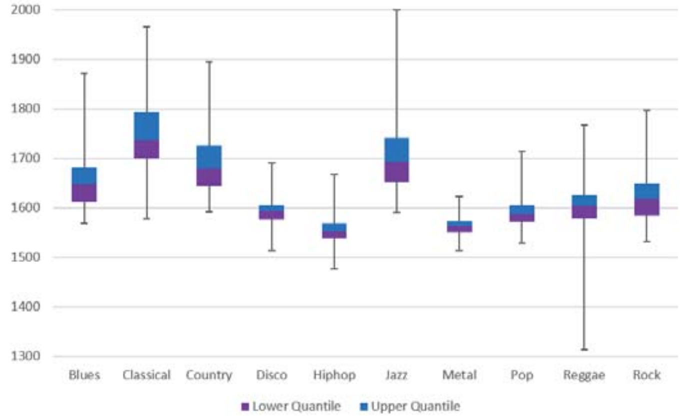
[3] J. H. Lee and J. S. Downie, "Survey of music information needs, uses, and seeking behaviours: Preliminary findings." in *ISMIR*, vol. 2004, 2004, p. 5th.

[4] C. McKay and I. Fujinaga, "Automatic music classification and the importance of instrument identification," in *Proceedings of the Conference on Interdisciplinary Musicology*, 2005.

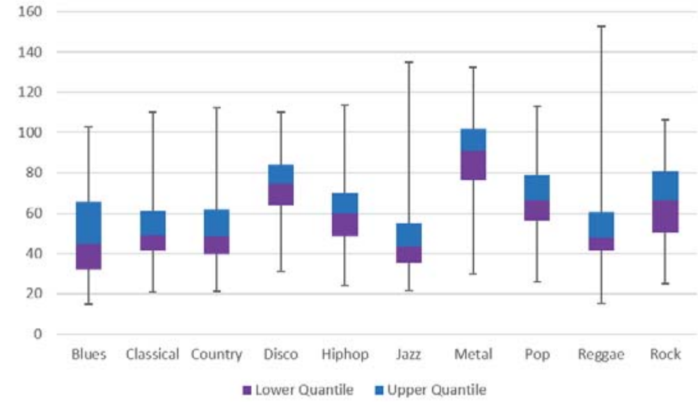
[5] J. D. White, *The analysis of music*. Prentice-Hall Englewood Cliffs, NJ, 1976.

[6] T. Li, M. Ogihara, and Q. Li, "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*. ACM, 2003, pp. 282–289.

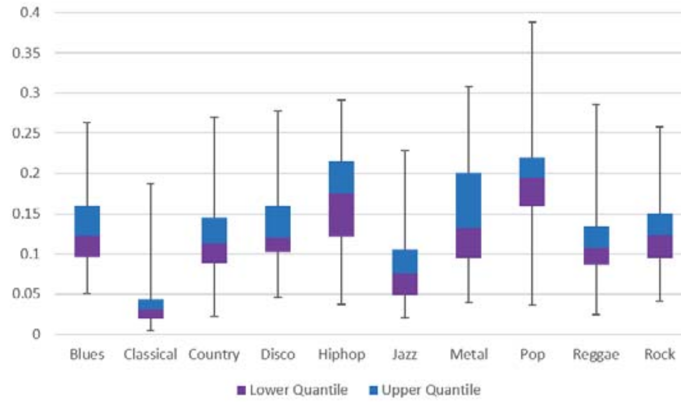
[7] R. O. Gjerdingen and D. Perrott, "Scanning the dial: The rapid recognition of music genres," *Journal of New Music Research*, vol. 37, no. 2, pp. 93–100, 2008.



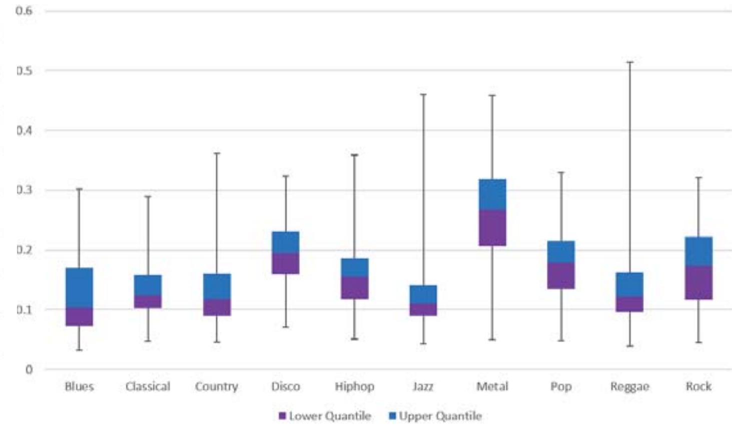
(a) The compactness feature ranges for all 10 GTZAN genres. Compactness distinguishes some genres particularly well from other genre types (e.g. Blues, Classical, Jazz, and Country).



(b) The ZCR feature ranges or all 10 GTZAN genres. Distinguishes metal, disco, and hiphop particularly well from other genres types.



(c) The energy feature ranges for all 10 GTZAN genres. Distinguishes classical, pop, hiphop, metal, and jazz particularly well from other genres types.



(d) The spectral rolloff frequency ranges for all 10 GTZAN genres. Distinguishes metal particularly well from other genres types.

Fig. 3: Different features distinguishing a variety of GTZAN genres.

[8] S. Lippens, J.-P. Martens, and T. De Mulder, "A comparison of human and automatic musical genre classification," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 4. IEEE, 2004, pp. iv–233.

[9] M. Grimaldi, P. Cunningham, and A. Kokaram, "An evaluation of alternative feature selection strategies and ensemble techniques for classifying music," in *Proc. Workshop on Multimedia Discovery and Mining*. Citeseer, 2003.

[10] C. McKay and I. Fujinaga, "Musical genre classification: Is it worth pursuing and how can it be improved?" in *ISMIR*, 2006, pp. 101–106.

[11] B. L. Sturm, "On music genre classification via compressive sampling," in *Multimedia and Expo (ICME), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–6.

[12] J. Bergstra, N. Casagrande, D. Erhan, D. Eck, and B. Kégl, "Aggregate features and adaboost for music classification," *Machine Learning*, vol. 65, no. 2-3, pp. 473–484, 2006.

[13] B. L. Sturm, "The gtzan dataset: Its contents, its faults, their affects on evaluation, and its future use," *arXiv preprint arXiv:1306.1461*, 2013.

[14] E. Wold, T. Blum, D. Keislar, and J. Wheaten, "Content-based classification, search, and retrieval of audio," *MultiMedia, IEEE*, vol. 3, no. 3, pp. 27–36, 1996.

[15] R. Ajoodha, R. Klein, and M. Jakovljevic, "Using statistical models and evolutionary algorithms in algorithmic music composition," in *The Encyclopedia of Information Science and Technology*, 3rd ed., K.-P. Mehdi, Ed. Hershey, Pennsylvania, United States: IGI Global, 2014.

[16] I. Fujinaga, "Adaptive optical music recognition," Ph.D. dissertation, McGill University, 1996.

[17] C. McKay, R. Fiebrink, D. McEnnis, B. Li, and I. Fujinaga, "Ace: A framework for optimizing music classification." in *ISMIR*, 2005, pp. 42–49.

[18] G. Peeters, "A large set of audio features for sound description (similarity and classification) in the cuidado project," 2004.

[19] C. Duxbury, J. P. Bello, M. Davies, M. Sandler *et al.*, "Complex domain onset detection for musical signals," in *Proc. Digital Audio Effects Workshop (DAFx)*, no. 1, 2003, pp. 6–9.

[20] C. McKay, I. Fujinaga, and P. Depalle, "jaudio: A feature extraction library," in *Proceedings of the International Conference on Music Information Retrieval*, 2005, pp. 600–3.

[21] L. Lu, H.-J. Zhang, and H. Jiang, "Content analysis for audio classification and segmentation," *Speech and Audio Processing, IEEE Transactions on*, vol. 10, no. 7, pp. 504–516, 2002.

[22] P. Iverson and C. L. Krumhansl, "Isolating the dynamic attributes of musical timbrea," *The Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 2595–2603, 1993.

[23] D. P. Ellis, "Classifying music audio with timbral and chroma features," in *ISMIR 2007: Proceedings of the 8th International Conference on Music Information Retrieval: September 23-27, 2007, Vienna, Austria*. Austrian Computer Society, 2007, pp. 339–340.

[24] D. Perrot and R. Gjerdigen, "Scanning the dial: An exploration of factors in the identification of musical style," in *Proceedings of the 1999 Society for Music Perception and Cognition*, 1999, p. 88.

[25] C. McKay, "Automatic genre classification of midi recordings," Ph.D. dissertation, McGill University, 2004.