FEATURE ANALYSIS FOR MUSICAL GENRE CLASSIFICATION TASK

Sefik Emre Eskimez
University of Rochester
eeskimez@ur.rochester.edu

ABSTRACT

In this project automatic genre classification problem will be studied and a new system for music genre will be developed. This paper focuses on impact of features on music genre classification task. The following genres are used in this project: Jazz, Metal, Hip-Hop, Classical and Disco, Reggae, Country, Pop. It is found that 3 features out of 26 performs better in genre classification task.

1. INTRODUCTION

Music genres are human made labels, basically divides music into categories. The members of the same category usually carry the same characteristics. Those categories mainly depend on the instruments that are used in the music piece, harmonics, and rhythmic content of the piece. The most popular music genres are rock, pop, metal, classic, jazz, blues and electronic.

There are vast amount of digital music available online, most of them are classified according to their genre. This classification allows end users to find songs easier according to their favorite genre. The current media on the Internet are manually labeled (manual labor). With increasing number of music content every day, automatic genre classification becomes non-trivial task.

1.1 Features

According to [1], the features that can be used in genre classification systems are timbre, melody, harmony, rhythm, and spatial location.

Timbre can be defined as a feature that makes two sounds with the same pitch and loudness different from each other. For characterizing Timbre there are different features that can be summarized as follows [1-2],

- Temporal features: Zero-crossing rate and linear prediction coefficients.
- Energy features: Root mean square energy of signal frame, energy of harmonic component of power spectrum, and energy of the noisy part of power spectrum.
- Spectral shape features: Centroids, spread, skewness, kurtosis, slope, roll-off frequency, variation, and Mel-frequency Cepstral coefficients (MFCCs).
- Perceptual features: Relative specific loudness, sharpness, and spread.

Melody is a memorable block of pitch series. Harmony can be defined as supporting notes that forms chords. If Harmony is the vertical element in music then Melody is said to be the horizontal element [1]. [3] proposed a genre classification method using timbral texture (which is based on STFT), rhythmic content (which is based on wavelet transform (WT)) and pitch content.

Rhythm can be defined as all the temporal aspects of musical piece. There are lots of applications, which use rhythm descriptors, such as beat tracking, tempo induction. These descriptors can be extracted from autocorrelation functions, which tell the periodicity in the perceivable tempo.

1.2 Classifiers

For genre classification there are different approaches on classifier. There are two types of machine learning approaches: unsupervised learning and supervised learning.

Unsupervised learning uses the data to classify based on objective similarity measures. Similarity measures can be chosen as Euclidian distance, cosine distance, distance of Gaussian Mixture Model (GMM), and distance of Hidden Markov Models (HMM). Some examples of clustering algorithm are K-means, Agglomerative Hierarchical Clustering and Self-Organizing Map (SOM).

Supervised learning algorithms uses manually labeled data to train their systems and classify unknown data with information on the system. Some of supervised classifiers are K-nearest neighbor (KNN), Gaussian Mixture Models (GMM), Hidden Markov Model (HMM), Linear Discriminant Analysis (LDA), Support Vector Machines (SVM) and Artificial Neural Networks (ANN).

K-nearest neighbor (KNN) is a nonparametric classification technique. Classifier will use the information on k nearest feature vectors and give label to current feature vector. [3] and [4] explores the use of KNNs on genre classification.

Feature vectors are considered as Gaussian distributions in Gaussian Mixture Models (GMM). Each distribution parameters are usually estimated using the iterative expectation minimization algorithm. [3] models musical genres as GMMs and using maximum likelihood classifier. [5] also models genres as GMMs but uses tree-like approach. It can make rough classification first, and then a finer classification follows. Another approach is to use maximal classification binary tree [6]. The tree's root contains all the information and divides into two leaves by using single Gaussian classifier.

Support Vector Machines (SVM) can be used for classification or regression. They are based on non-linear kernel functions. [7-8] uses SVM for genre classification task. SVM with Kullback-Lleiber divergence-based kernel is used on [9]. [10] uses mixture of SVM experts to increase accuracy. Multi-layer SVM is used on [11] as genre classifier.

Linear Discriminant Analysis (LDA) is the task to find the most suitable linear transformation so that classes can be easily separated from each other. The classification then can be done in transformed space according to some measure. An example measure can be Euclidian distance. The feature vector dimensionality is reduced using LDA on [6].

Hidden Markov Model (HMM) is statistical model that consists of two kinds of states, observable and hidden. It requires time set of observations. [7]and [12] uses HMM for genre classification.

[13] measures performances of different classifiers, namely SVM and LAD and reports the results.

Non-negative matrix multiplication is used for timbre representation, which then used to form a spectrogram on [14]. With sets of spectrogram given classification of genre is done using GMM.

The next section will cover the feature that will be used in project and briefly describe the feature extraction method. Section 3 explains the classifier that will be used in project. Section 4 will contain the results and discussion parts and Section 5 will be the conclusion part.

2. FEATURE VECTOR

In this project multiple features are used to form feature vector, which will be used in classification. Input music is chopped into frames and for most of the cases features are calculated per frame.

The features are in three categories. First category is the Timbre characteristics of input music segment. The features chosen for timbre representations are listed below

First 5 **MFFCs** are good enough to represent timbre. DC component is not included on this list. Because MFFCs are calculated per frame and we cannot use them all, means and squares of these five coefficients are used. They are also the first five elements in feature vector.

Spectral shapes of Spectrogram are used as for next features. Because the calculations are frame-wise, the means and variances of those variables are used. Feature vector contains means and variances of **Spectral Centroids, Spectral Roll off** and **Spectral flux.**

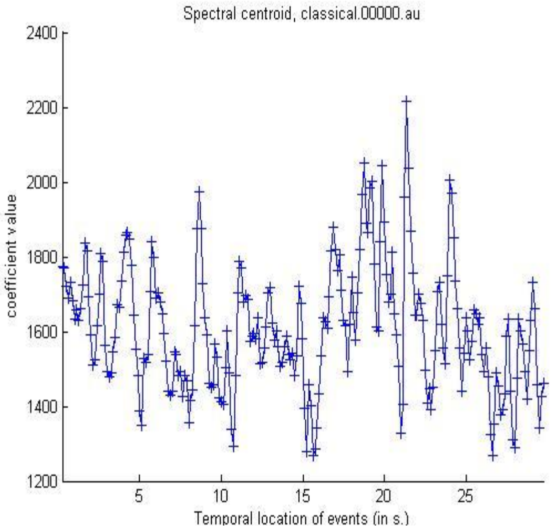

Figure 1 Spectral Centroids of classical music sample.

Spectral Centroids can be defined as center of gravity of STFT. Figure 1 shows an example Spectral Centroids of a sample. It is represented as,

$$C_t = \frac{\sum_{n=1}^{N} M_t[n] * n}{\sum_{n=1}^{N} M_t[n]} \tag{1}$$

where M_t represents magnitude of STFT for time frame t. Spectral rolloff (R_t) is the frequency point, which %85 of the magnitudes are distributed below that point. Figure 2 illustrates Spectral rolloff of the same sample. Mathematically;

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 * \sum_{n=1}^{N} M_t[n] \tag{2}$$

Spectral flux indicates the spectral change. Mathematically it is squared difference between two consecutive spatial elements,

$$F_t = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[n])^2 \tag{3}$$

where N_t represents normalized magnitude of STFT for time frame t. Figure 3 illustrates Spectral flux of classical sample.

Next feature in feature vector is **zero cross rates** of the frames. This feature determines the noisiness of the signal. Figure 4 shows the zero-crossing rates for classical.00000.au sample. It can be shown as,

$$Z_t = \frac{1}{2} \sum_{n=1}^{N} |(sign(x[n]) - sign(x[n-1])^2| \tag{4}$$

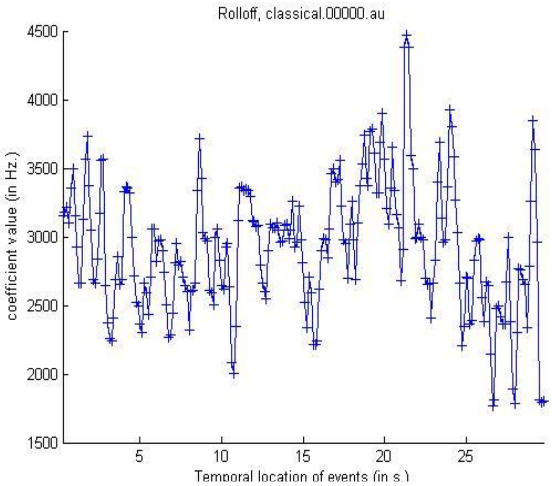

Figure 2 Spectral Rolloff of classical music sample.

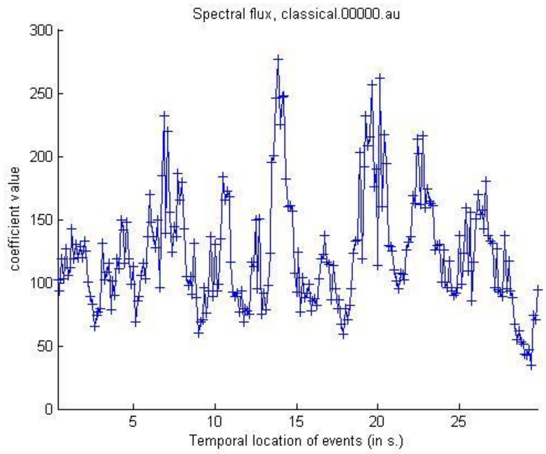

Figure 3 Spectral flux of classical.00000.au sample.

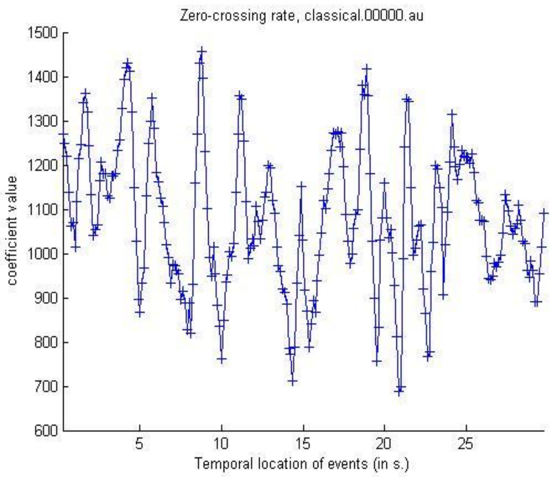

Figure 4 Zero-crossing rate of classical music sample.

where x[n] is the time domain signal. Another feature is the **low-energy**, which is the percentage of energies whose RMS values are lesser than the average RMS values of musical piece's frames.

Second category of features represents beat characteristics of musical piece. Beat spectrum is derived from similarity matrix. D represents the (dis)similarity between i^{th} and j^{th} frames of input audio,

$$D(i,j) = \frac{v_i . v_j}{\|v_i\| \|v_j\|} \tag{5}$$

Similarity matrix S is formed from all frame combinations of D. Beat spectrum is formed from autocorrelation of S,

$$B(k,l) = \sum_{i,j} S(i,j)S(i+k, j+l). \tag{6}$$

Because *B* is symmetric, one can get a one dimensional array by summing rows of *B*, resulting in *B(l)*. Fig 5 shows example beat spectrum.

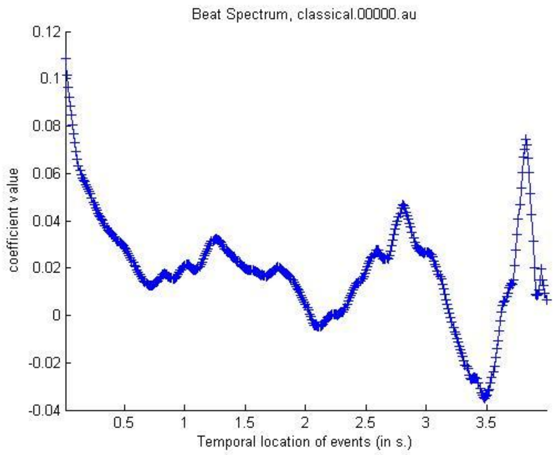

Figure 5 Beat spectrum of classical music sample.

Beat spectrum is calculated for each sample and means, variances and sum of the beat spectrum are used in feature vector for representing beat characteristics.

Third category of features contains pitch characteristics of a musical piece. Only dominant pitch for the whole musical piece is extracted and used. For extracting pitch information MIR toolbox [15-16] is used. The implementation details of this function are unknown, but any pitch detector can be a good choice. This particular toolbox is chosen because of execution time considerations. Most of the algorithms in this project are similar or exactly the same with this courses assignments, therefore existing codes can be used. But they will result in poor execution time. For all feature extraction purposes in this project, MIR toolbox is used.

As an extra feature to test on, tempo feature is added to feature vector. Next section describes KNN classifier.

3. KNN CLASSIFIER

K-Nearest Neighbor (KNN) classifiers use a distance metric to calculate the distance between the input feature vectors and trained feature vectors. It chooses k lowest distance feature vectors and assigns the label of majority of those vectors.

Distance measure used in this project is Euclidian distance. It can be shown as,

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + \cdots + (p_N - q_N)^2} \quad (7)$$

where N is the length of feature vector. MATLAB built-in implementation of KNN classifier is used in this project.

4. DATABASE

Musical Analysis, Retrieval and Synthesis for Audio Signals (MARSYAS) genre collection database is used in this project [3, 17]. The samples are 22050 Hz Mono 16-bit .wav files. Each of them is 30 seconds long. First set contains five genres, and those are Jazz, Metal, Hip-Hop, Classical and Disco. Second set contains Metal, Country, Pop, Classical, Reggae. There are 100 samples for each genre.

5. CLASSIFICATION

Feature vectors for all training and test data are extracted. 20 songs from each genre class are picked as test data and rest of them are treated as train data.KNN classifier, with Euclidian distance measure, does classification. Accuracy is calculated according to ground truth data.

There are two sets of genres inspected in this project to validate the inspected features are not just successful for a certain database. Table 1 shows 3-nearest neighborhood musical genre classification results for first set where Table 2 shows results for second set. Selecting k = 1 yields better result than 3-nearest neighbor result.

Accuracy	
Total	**43.80%**
Jazz	**33.33%**
Metal	**42.85%**
Hip-Hop	**52.38%**
Classical	**57.14%**
Dance	**33.33%**

Table 1 Accuracy values for KNN classification results for first set (3-Nearest Neighbors)

The accuracy values are relatively low compared to [3], where total precision is reported as 59 percent. To improve classification results, next section inspects each feature's impact on classification results.

Accuracy	
Total	**%51.42**
Metal	**%57.14**
Country	**%47.61**
Pop	**%66.66**
Classical	**%57.14**
Reggae	**%28.57**

Table 2 Accuracy values for KNN classification results for second set (3-Nearest Neighbor)

6. FEATURE'S IMPACTS ON CLASSIFICATION RESULTS

To measure impact of a particular feature on classification, only one feature is used for classification and total accuracy results are stored. Table 3 shows resulting accuracies for corresponding feature.

Feature	1st Set Accuracy	2nd Set Accuracy
Mean of MFCC(2)	**30.47%**	**27.61%**
Mean of MFCC(3)	**33.33%**	**38.09%**
Mean of MFCC(4)	**26.66%**	**23.80%**
Mean of MFCC(5)	**32.38%**	**34.28%**
Mean of MFCC(6)	**34.28%**	**33.33%**
Variance of MFCC(2)	**40.00%**	**40.95%**
Variance of MFCC(3)	**40.00%**	**36.19%**
Variance of MFCC(4)	**34.28%**	**31.42%**
Variance of MFCC(5)	**31.42%**	**33.33%**
Variance of MFCC(6)	**38.09%**	**34.28%**
First Dominant Tempo	**35.23%**	**23.80%**
Second Dominant Tempo	**37.14%**	**27.61%**
Mean of Beat Spectrum	**32.38%**	**22.85%**
Variance of Beat Spectrum	**40.00%**	**35.23%**
Sum of Beat Spectrum	**32.38%**	**22.85%**
RMS value	**35.23%**	**41.90%**
Mean of Spectral Centroids	**44.76%**	**34.28%**

Variance of Spectral Centroids	28.57%	39.04%
Mean of Spectral Rolloff	42.85%	49.52%
Variance of Spectral Rolloff	37.14%	38.09%
Mean of Spectral Flux	44.76%	46.66%
Variance of Spectral Flux	46.66%	48.57%
Mean of Zero-crossing rate	39.04%	33.33%
Variance of Zero-crossing rate	29.52%	33.33%
Low Energy	23.80%	25.71%
Dominant Pitch	35.23%	34.28%

Table 3 Feature's success measure on genre classification.

6.1 Feature Reduction

Some features are not contributing to classification; instead they are reducing the accuracy of results. Therefore a threshold of %38 is applied to accuracy results of features. The features whose results past the threshold are added to new feature list. %38 threshold value is found empirically for the first set, and %40 threshold used for the second set.

New training and test datasets are formed, which are just cropped versions of first datasets. These new datasets only contain features, which are successful at genre classification.

The results of new classification are shown on Table 4 and Table 5, for first and second sets respectively.

	Accuracy	Precision
Total	62.85%	
Jazz	52.38%	57.89%
Metal	47.61%	58.82%
Hip-Hop	61.90%	59.09%
Classical	80.95%	80.95%
Dance	71.42%	57.69%

Table 4 Accuracy values for KNN classification results for first set (3-Nearest Neighbor)

	Accuracy	Precision
Total	60.00%	
Metal	57.14%	52.17%
Country	52.38%	50.00%
Pop	61.90%	65.00%
Classical	85.71%	85.71%
Reggae	42.85%	47.36%

Table 5 Accuracy values for KNN classification results for second set (3-Nearest Neighbors)

7. EVALUATION

The most important features for the genre classification are variance of second, third and sixth MFCCs, variance of beat spectrum, means of Spectral Centroids, Spectral Rolloff, Spectral Flux and Zero-crossing rate, and variance of Spectral Flux. Those features have above %38 accuracy and they are found by experimenting. Any additional feature from current list will decrease the overall accuracy of genre classification.

Final accuracy of genre classification (%62.8) is a decent value compared to other results [3, 18].Compared to Table 1, accuracy is increased %43, after feature reduction. Precision for each class is above %55.

From the results of this project it can be said that classical music can be identified well, compared to other genres (%80 precision).

Genre classification is an ill-defined problem and highly depends on training dataset. Also taxonomies (labels) does not have a standard, it can vary on globally. Having globally standardized genres will be a good start for improved genre classification systems.

8. CONCLUSION

This report briefly explains the genre and the features and classifiers used in genre classification. A brief literature review is presented within the content limit of this report.

Composition of feature vector is explained, along with feature extraction. KNN classifier is explained. Implementation results are shown for different number of nearest neighbors.

Additional feature reduction step is implemented to increase accuracy and precision of genre classifier. It is found that some of the proposed features are not as effective as other features and lowers the overall genre classification quality.

The final results are listed and evaluated. It is concluded that accuracy and precision of proposed genre classifier are decent compared to literature.

It is found that the most important features for genre classification task are as follows (which are marked green on Table 3),

- Second MFCC,
- Spectral Rolloff,
- Spectral Flux.

9. FUTURE WORK

Additional features can be tested to see their impact on genre classification. With decent amount of quality features, it is possible to achieve higher accuracy values.

Different classifiers can be tested and feature reduction step can be arranged according to new classifiers. Classifier's impact on genre classification can be explored.

Working on bigger datasets and more genre classes can reveal more interesting properties of features. Maybe advanced feature reduction can be implemented on such discovery.

10. REFERENCES

[1] Scaringella, N.; Zoia, G.; Mlynek, D., "Automatic genre classification of music content: a survey," *Signal Processing Magazine, IEEE*, Vol. 23, No.2, pp.133-141, 2006.

[2] G. Peeters, “A large set of audio features for sound description (similarity and classification) in the CUIDADO project,” CUIDADO I.S.T. Project Rep., 2004.

[3] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” IEEE Trans. Speech Audio Processing, vol. 10, no. 5, pp. 293–302, July 2002.

[4] E. Pampalk, A. Flexer, and G. Widmer, “Improvements of audio based music similarity and genre classification,” in Proc. 6th Int. Symp. Music InformationRetrieval, London, UK, pp. 628–633, 2005.

[5] J.J. Burred and A. Lerch, “A hierarchical approach to automatic musical genre classification,” in Proc. 6th Int. Conf. Digital Audio Effects (DAFx), London, UK, 2003.

[6] K. West and S. Cox, “Finding an optimal segmentation for audio genre classification,” in Proc. 6th Int. Symp. Music Information Retrieval, London, UK, pp. 680–685, 2005.

[7] N. Scaringella and G. Zoia, “On the modeling of time information for automatic genre recognition systems in audio signals,” in Proc. 6th Int. Symp. Music Information Retrieval, London, UK, pp. 666–671, 2005.

[8] T. Lidy and A. Rauber, “Evaluation of feature extractors and psycho-acoustic transformations for music genre classification,” in Proc. 6th Int. Symp. Music Information Retrieval, London, UK, pp. 34–41, 2005.

[9] M. Mandel and D. Ellis, “Song-level features and support vector machines for music classification,” in Proc. 6th Int. Symp. Music Information Retrieval, London, UK, pp. 594–599, 2005.

[10] N. Scaringella and D. Mlynek, “A mixture of support vector machines for audio classification,” Music Information Retrieval Evaluation Exchange (MIREX), 2005.

[11] ChangshengXu; Maddage, M.C.; Xi Shao; Fang Cao; Qi Tian, "Musical genre classification using support vector machines," *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.5, 429-32 vol.5, 2003.

[12] H. Soltau, T. Schultz, M. Westphal, and A. Waibel, “Recognition of music types,” in Proc. IEEE Int. Conf. Acoustics, Speech Signal Processing (ICASSP),Seattle, WA, USA, vol. II, pp. 1137–1140, 1998.

[13] Tao Li; Tzanetakis, G., "Factors in automatic musical genre classification of audio signals," *Workshop on Applications of Signal Processing to Audio and Acoustics, IEEE.* pp.143,146, 19-22 Oct. 2003.

[14] Holzapfel, A.; Stylianou, Y., "Musical Genre Classification Using Nonnegative Matrix Factorization-Based Features," , *IEEE Transactions on Audio, Speech, and Language Processing*, vol.16, no.2, pp.424,434, Feb. 2008.

[15] Olivier Lartillot, Petri Toiviainen, “A MATLAB Toolbox for Musical Feature Extraction From Audio”, International Conference on Digital Audio Effects, Bordeaux, 2007.

[16] Olivier Lartillot, Petri Toiviainen, TuomasEerola, “A Matlab Toolbox for Music Information Retrieval”, in C. Preisach, H. Burkhardt, L. Schmidt-Thieme, R. Decker (Eds.), Data Analysis, Machine Learning and Applications, Studies in Classification, Data Analysis, and Knowledge Organization, Springer-Verlag, 2008.

[17] http://marsyas.info/

[18] http://www.music-ir.org/mirex/wiki/2008:Audio_Genre_Classification_Results