A Survey of Style Identification Approaches in Music Information Retrieval

Santosh Pakhare[#1], Mrs M. A. Potey[*2]

[#]*Department of Computer Engineering Savitribai Phule Pune University*
DYPCOE, Akurdi, Pune, India

Abstract: *One of the problems to solve in Music Information Retrieval (MIR) is the modelization of music style. The system could be trained to identify the main features that would characterize music genres or style so as to look for that kind of music over large musical corpus. So in this paper multimodal approach, pattern recognition approach and co-updating approach is been studied for identifying the style from different genre of the music. Considering the intuitive feelings of similarity from the listeners perspective, the focus on features that are computed using similarity metrics for melodies, harmonies, and audio signals for style identification. A multimodal approach mostly considered support vector machine as a binary classifier to determine if two songs or music played by the same artist given their similarity metrics in the three aspects and also discussed the experimental methodologies of the two different approaches.*

Keywords — *Gaussian mixture models, melodic contour, music similarity, n-grams, style.*

I. Introduction

Music is various types of information, and may evoke in the listener different moods or emotions. Such type of information is retrieved (known as Music Information Retrieval) through some existing systems such as machine learning algorithms (MLs) and pattern recognition (PR) techniques which process of classification and indexing in large digital music corpus. Hence managing ever growing multimedia database on day by day basic is very time consuming when done completely manually, this is why automatic system should be required for minimize the job. In [Huron, 2000] Huron finds out that since the mail features of music are social and psychological, the most useful characterization would be based on four types of information: *genre*, *emotion*, *style*, and *similarity* [1].

Music style identification is one of the key problems in MIR. It is very difficult for machine to classify and identify exact pieces of abstract information because of the multiplex secular relationships present in musical features, including pitch, rhythm and dynamics, however, humans can easily identify different styles and genres [2]. The human brain is having extraordinary capability to process musical information very efficiently. For the task of computer-based music analysis, here we have studied some approaches which help to the researcher for identifying the style from the musical clip.

The rest of the paper is structured as follows. In Section II we have discussed some style identification challenges. Section III describes work related to music retrieval based on style. Section IV describes different musical phrase representation. Section V enlisted different musical style recognition approaches. Later in the section, we have done some data-table discussion of three different approaches. In section VII, is briefly reviewed conclusion.

II. STYLE IDENTIFICATION CHALLENGES

1. According to Ching-Hua et al. [3] music creation is dynamically changing. Music pieces or songs created by the same artist may not always sound the same, although they may sound more similar to each other than the other artists work. Songs by the same artist may share certain components but not necessarily contain all of them, depending on how versatile the artist is.

2. Roger B. et al. [4] stated that it is very difficult for computers to classify and analyze the pieces of non-realistic information because of the composite temporal relationships present in musical features, including pitch, rhythm and dynamics.

3. Tao Li et al. [14] point out that similarity between artists reflects personal tastes and suggest that different features have to be combined together so as to achieve reasonable results in similar artist discovery.

Therefore, in order to understand the nature of musical data, all related features should be considered (i.e., Melody, Harmony and Acoustic signals) while each music piece should be regarded as an individual case. So identifying the style three different approaches are enlisted i.e., Pattern Recognition Approach, Multimodal Approach and Co-updating Approach.

III. RELATED WORK

There is not much work is done for identifying the style from instrumental music. Pampalk *et al.* [1] use self-organizing maps (SOMs) to solve the problem of managing digital libraries according to different sound features of musical clips, in such a way that similar clips are group together, and perform a content-based classification of the sounds. In [2] a system is developed by using neural networks and support vector machines able to classify fragmented samples into a given list of sources or artists. A neural system to identify different music types from sound inputs is described in [3]. For genre classification [4] this paper helps to understand different features and classification techniques.

Stamatatos and Widmer [11] focused on stylistic performance measures and analysis technique to obtain an group of simple classifiers that work together to recognize the most likely music artist of a piece given a set of piano corpus. Input data has been taken from a computer-monitored piano.

The most popular classification task for musical data is genre classification. Genre classification involves classifying a given music example into one of several pre-defined classes/genres, such as classical, pop, rock, jazz. Most genre classification systems take audio recordings as input. Lower level acoustic features such as spectral centroid, zero-crossing rates, mel-frequency cepstral coefficients [3] and features related to psychoacoustics [4] are commonly used. Commonly used classifiers include support vector machine [5], genetic algorithms [6], hidden Markov models [7]. However, identifying an artist's signature style is a different task from genre classification. For example, an artist can be labelled with multiple genres while even artists in the same genre may have very different styles.

IV.MUSICAL PHRASE REPRESENTATION

A. ***Melody:*** "Melody is a tune which is a part of music you hum, sing, whistle or play." Melodies are very identifiable and sometime sing-able. However, just the series of pitches doesn't make a melody [3]."

Fig. 1 Representation of melody phrase [11]

B. ***Harmony:*** "In music, there are simultaneous pitches (tones, notes), or chords called harmony". It involves construction of chords and progressions [11].

Fig. 2 Representation of harmony phrase [11]

Harmony is represented by set of "vertical" lines which easy to distinguish from melodic line and "horizontal" aspect [3].
Harmony consists of set of 5 lines and 4 spaces on which music notes are written.

C. ***Acoustic Signals:*** In addition to symbolic data such as melody and harmony, acoustic data extracted from CD or mp3 recordings provide abundant information about which instruments were used and how the notes were played.

.
Fig. 3 Audio signal waveform [12]

The three most distinct acoustic features are as follows.
Volume: This feature analyse the loudness of the audio clip, which represent the amplitude of the signals. Whereas it is also called as energy or intensity of the audio signals.
Pitch: Vibration rate of audio signals has been analysed by pitch with help of the fundamental frequency, or the reciprocal of the fundamental period of voiced audio signals.
Timbre: Meaningful content (such as a vowel in English) of audio signals has been represented by waveform within a fundamental period of voice signals.

V.MUSICAL STYLE RECOGNITION APPROACHES

In this section three different approaches for style identification are enlisted i.e., Pattern recognition approach, Multimodal approach and co-updating approaches. Where in the pattern recognition approach music style is identified from melody representation only whereas in the multi-modal approach the style is identified from melody, harmony and acoustic features.

A. *PATTERN RECOGNITION APPROACH*

Pattern recognition is a branch of machine learning that focuses on the recognition of patterns and identify the similar pattern from large corpus. Pattern recognition approach is trained on labeled data called supervised learning, but when there lack of labeled data are available then we have to consider some other algorithms which discovers previously unknown patterns called unsupervised learning.

Musical Input:
Pedro J. et al. [8] aim is to develop a framework for experimenting on musical style automatic recognition from symbolic representation of melodies (digital scores) by using shallow statistical features. His framework used all the main stages of a pattern

recognition system which is feature extraction, feature selection and classification [8].

Shallow Statistical Descriptors:
A descriptive scheme is nothing but descriptive statistics which summarize from melody features such as Intervals, durations, pitches, silences, harmonicity, rhythm, etc. So considering all this features is called as shallow structure description [8]. This initial model is made up of 28 descriptors summarized in Table I.

Table I.
Musical Descriptors [8]

Category	Descriptors
Overall	Number of notes Number of significance silences Number of non-significant silences
Pitch	Pitch range Average pitch Dev. Pitch
Note duration	Note duration range Avg. note duration Dev. Silence duration
Silence duration	Silence duration range Avg. silence duration Dec. silence duration
Inter Onset Interval	IOI range Avg. IOI Dev. IOI
Pitch Interval	Internal range Avg. Internal Dev. Internal
Non-diatonic notes	Num. non-diatonic notes Avg. non-diatonic degrees Dev. Non-diatonic degrees
Syncopation	Number of syncope
Normality	Pitch distribution normality Note duration distribution Normality Silence duration distribution Normality IOI distribution. Normality Interval distribution normality Non-diatonic degree distribution normality

Classification Techniques:
In this section three different classifier are enlisted i.e., Naive Bayes(NB), K-Nearest Neighbor (K-NN), and Self-organizing maps (SOMs) where NB, K-NN, and SOMs are widely used in pattern recognition approach. Three of them are fully supervised methods: the Bayesian, and KNN [22]. The other one is an unsupervised learning neural network, the SOM [13].

1. Naïve Bayes Classifier:
The Bayesian classifier is parametric and, when applied to a two-class problem, computes a discriminant function

$$g(x) = log\frac{P(X)|\omega_1}{P\ (X)|\omega_2} + log\frac{\pi_1}{\pi_2} \qquad (2)$$

For a test sample X, where $P(X)|\omega_1$ is the conditional probability density function for class i and π_i are the priors of each class. For each statistical descriptor Gaussian probability density functions for each style are considered. Hence means and variances are calculated separately. The classifier assigns a sample to ω_1 if $g(x) > 0$ and to ω_1 otherwise. The decision boundaries, where $g(x) = 0$, are in general hyper quadrics in the feature space [8].

2. KNN Classifier:

The k-NN classifier uses a Euclidean metrics approach to calculate the distance between the training sample and the test samples. For each test samples style label is assigned by author [11] and on that basic KNN (the k-neighbourhood) did the classification.

3. Self-organizing map :
This classifier is a type of artificial neural network (ANN) which is trained using unsupervised learning to produce low dimensional and discretized representation of the supplied input called a map.

B. ***MULTIMODAL APPROACH***
Music is a unique type of multimedia data [5]. It not only exists in many different data formats, but it also conveys numerous musical ideas via hierarchical instrumentations and/or sounds. For example, symbolic data such as melodies and harmonies indicate the fundamental components in music compositions while acoustic data present another layer of information created by performances or sound effects. So combination of melodies, harmonies and acoustic signals as a input to the system which helps to identify the style and gets the exact match to the supplied input.

Musical Input:
Ching-Hua et al. [5] analyse the musical data in three aspects: melodies, harmonies, and audio signals. Melodic similarity is represented by the cosine distance between two pitch class distributions and pitch interval distributions, as well as cosine distance between melodic contours in phrases. Harmonic similarity is computed by comparing chord profiles of two songs. A chord profile consists of n-gram chord patterns weighted by the pattern's durations. Acoustic similarity is produced by computing the Mahalanobis distance between the feature vectors extracted from two audio recordings, as well as by comparing the Gaussian mixture models built for the two recordings using Monte Carlo sampling.

Classification Technique:
Ching-Hua et al. [5] has used support vector machine as a binary classifier task to examine the effectiveness of the similarity metrics in melody, harmony, and audio signals for artist differentiation. An input instance for the classifier consists of similarity values between two songs and the classifier aims to report

whether the given two songs are from the same artist(s).

C. *CO-UPDATING APPROACH*

This approach [14] trained the system from both labelled and unlabelled data which helps to minimize dissimilarity on untagged data. A co-updating approach is an iterative Expectation-Maximization (EM)-type procedure. It works as follows. First from tagged data weak classifiers f_1 and f_2 are trained. For each iteration, with the help of expectation step uses the current classifiers to identify the labels of untagged data and the maximization step re-configure the classifiers using the tagged samples and a random collection of untagged samples on which the classifiers agree. This procedure is end up once some termination criterion is going too met. The intuition behind this approach used by author [14] is that stochastically select the untagged samples on which the two component classifiers agree and confident, and then use them along with the tagged samples to train/update the classifiers.

VI. EXPERIMENTAL METHODOLOGIES

This section focuses on the experimental methodologies used by Ching-Hua et al. [5], Pedro J. et al. [8] and Tao Li et al .[14] are discussed.

In the multi-modal approach [5] the main study focuses on different features that are calculated using similarity metrics for melodies, harmonies and audio signals for style identification. Melodies similarity is represented by the cosine distance between the two pitch class distributions and pitch interval distribution, as well as cosine distance between melodic contours in phrases. Harmonic similarity is computed by comparing chord profiles of two songs. A chord profile consists of n-gram chord patterns weighted by the pattern's durations. Acoustic similarity is produced by computing the Mahalanobis distance between the feature vectors extracted from two audio recordings, as well as by comparing the Gaussian mixture models built for the two recordings using Monte Carlo sampling. Experiments are conducted using songs of well-known pop/rock bands from 6 albums. Here the two sets of experiments are conducted.

1. Positive Instances
2. Negative Instances

Positive instances are created by comparing two songs from the same artist using these similarity metrics, while negative cases are produced by two songs from different artists. For this experiment Ching-Hua at al. [5] used support vector machine as the classifier. Classification results are generated using 10-fold cross validation whereas the classifier has given 85 % accuracy for identifying the style.

In the pattern recognition approach author [8] has conducted the experiments for automatic music style recognition from symbolic representation of melodies (digital scores) by using shallow structural features. The two well defined music styles like jazz and classical have been chosen as workbench for these experiments. The author [8] has used three different classifiers for these experiments i.e., Bayes classifier, K-NN classifier and SOM classifier. A bayes and K-NN performed comparatively well.

Tao Li et al. [14] studies the problem of identifying "similar" artists using both lyrics and acoustic data. The approach for using a small set of tagged samples for the seed tagging to build classifiers that improve themselves using untagged data is presented. This approach is tested by author [14] on a data set of 43 artists and 56 albums using artist similarity provided by All Music Guide [15]. Whereas results of this approach as mentioned in the table I.

VI. DATA-TABLE DISCUSSION

After the literature survey the summary of several research works on musical style identification is mentioned in following table.

TABLE II
Summary of several research works on musical style identification

Author	Data set	Features	Processing Technique	Class-ifiers used	Accuracy
Ching-Hua-Chuaa [5]	Last.fm And CD	Melody, Harmony and Acoustic	GMM and Similarity Metrics	SVM	85%
Pedro J. et al.[8]	110 MIDI files	Melody	Shallow Statistical descriptors	KNN, Bayesian and SOM	KNN=93% SOM=56% Bayes=79%
Tao Li et al.[14]	All Music Guide	Timbral and DWCH	Co-updating Approach	SVM	0.78%

VII. CONCLUSION

In this paper musical style recognition approaches are discussed i.e. Pattern recognition approach, Multimodal approach and Co-updating approach with their comparative study. In the multi-modal approach a binary classification task is used to examine the effectiveness of the similarity metrics in melody, harmony, and audio signals for artist differentiation. An input instance for the classifier consists of similarity values between two songs, and the classifier aims to report whether the given two songs are from the same artist(s).

ACKNOWELGEMENT

We express our thanks to publishers, researchers for making their resource available & teachers for their guidance. We also thank the college authority for providing the required infrastructure and support. Last but not the least we would like to extend a heartfelt gratitude to friends and family members for their support.

REFERENCES

[1] Shi-Huang Chen and Shih-Hao Chen. "*Content-based music genre classification using timbral feature vectors and support vector machine*." In Proceedings of the 2Nd International Conference on Interaction Sciences: Information Technology, Culture and Human, ICIS '09, pages 1095{1101, New York, NY, USA, 2009.

[2] Ching-Hua Chuan. "*Pop-rock musical style as defined by two-chord patterns at segmentation points in the melody and lyrics.*" In Multimedia, 2009. ISM '09. 11th IEEE International Symposium on, pages 448{452, Dec 2009.

[3] Ching-Hua Chuan. "*Harmonic style-based song retrieval using n-gram.*" In Proceedings of the 1st ACM International Conference on Multimedia Retrieval, ICMR '11, pages 17:1{17:7, New York, NY, USA, 2011. ACM.

[4] Ching-Hua Chuan. "*Audio classification and retrieval using wavelets and Gaussian mixture models".* volume 4, pages 1{20, Hershey, PA, USA, January 2013.

[5] Ching-Hua Chuan. "*A multimodal approach to song-level style identification in pop/rock using similarity metrics".* In Proceedings of the 2013 12th International Conference on Machine Learning and Applications - Volume 02, ICMLA '13, pages 321{324, Washington, DC, USA, 2013. IEEE Computer Society.

[6] Maxime Crochemore, Christos Makris, Wojciech Rytter, Athanasios Tsakalidis, and Kostas Tsichlas. Approximate "*string matching with gaps*". 2002.

[7] Roger B. Dannenberg, Belinda Thom, and David Watson. "*A machine learning approach to musical style recognition.*" In Proc. International Computer Music Conference, pages 344{347, 1997.

[8] P.P. de Leon and J. Inesta. "*Pattern recognition approach for music style identification using shallow statistical descriptors."* volume 37, pages 248{257, March 2007.

[9] J. Foote. "*An overview of audio information retrieval.*" In Multimedia Systems, volume 1, pages pp. 307{328, 1999.

[10] Asif Ghias, Jonathan Logan, David Chamberlin, and Brian C. Smith. *"Query by humming: Musical information retrieval in an audio database.*" In Proceedings of the Third ACM International Conference on Multimedia, MULTIMEDIA '95, pages 231{236, New York, NY, USA, 1995. ACM.

[11] GeorgeV. Karkavitsas and GeorgeA. Tsihrintzis. "*Automatic music genre classification using hybrid genetic algorithms".* In GeorgeA. Tsihrintzis, Maria Virvou, Lakhmic. Jain, and RobertJ. Howlett, editors, Intelligent Interactive Multimedia Systems and Services, volume 11 of Smart Innovation, Systems and Technologies, pages 323{335, 2011.

[12] Kunio Kashino, Kazuhiro Nakadai, Tomoyoshi Kinoshita, and Hidehiko Tanaka. "*Organization of hierarchical perceptual sounds: Music scene analysis with autonomous processing modules and a quantitative information integration mechanism." In* Proc. International Joint Conf. on Arti_cial Intelligence, pages 158{164, 1995.

[13] T Kohonen. Self-organizingmaps.

[14] Tao Li and Mitsunori Ogihara. "*Music artist style identification by semi-supervised learning from both lyrics and content."* In Proceedings of the 12th Annual ACM International Conference on Multimedia, MULTIMEDIA '04, pages 364{367, New York, NY, USA, 2004. ACM.

[15] H. B. Lincoln. "*Some criteria and techniques for developing computerized thematic indices".* In Computers and the Humanities, volume 2, pages pp. 71{93. Springer, 1967.

[16] Michael Mandel, Michael I. M, and Daniel Ellis. "*Song-level features and support vector machines for music classification."* 2005.

[17] Matthias Mauch, Simon Dixon, Christopher Harte, and Queen Mary. *Discovering chord idioms through beatles and real book songs.* 2007.

[18] Martin Mckinney and Jeroen Breebaart. *"Features for audio and music classification. In Proceedings of the International Symposium on Music Information Retrieval,*" pages 151{158, 2003.

[19] Mitsunori Ogihara and Tao Li. Ismir 2008 session 5d mir methods n-gram chord profiles for composer style representation.

[20] Elias Pampalk, Simon Dixon, and Gerhard Widmer. Exploring music collections by browsing different views. volume 28, pages 49{62, Cambridge, MA, USA, June 2004.

[21] Yuting Qi, J.W. Paisley, and L. Carin. Music analysis using hidden markov mixture models. volume 55, pages 5209{5224, Nov 2007.

[22] P. E. Hart R. O. Duda and D. G. Stork. *Pattern classification*. Hagen Soltau, T. Schultz, M. Westphal, and Alex Waibel. Recognition of music types. In Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on, volume 2, pages 1137{1140 vol.2, May 1998.

[23] J. Kangas J. Laaksonen T. Kohonen, J. Hynninen. Som pak: *The self-organizing map program package.*

[24] B. Whitman, G. Flake, and S. Lawrence. *Artist detection in music with minnowmatch.* In Neural Networks for Signal Processing XI, 2001. Proceedings of the 2001 IEEE Signal Processing Society Workshop, pages 559{568, 2001}