International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:3, 2008

Automatic Musical Genre Classification Using Divergence and Average Information Measures

Hassan Ezzaidi, and Jean Rouat

***Abstract*—**Recently many research has been conducted to retrieve pertinent parameters and adequate models for automatic music genre classification. In this paper, two measures based upon information theory concepts are investigated for mapping the features space to decision space. A Gaussian Mixture Model (GMM) is used as a baseline and reference system. Various strategies are proposed for training and testing sessions with matched or mismatched conditions, long training and long testing, long training and short testing. For all experiments, the file sections used for testing are never been used during training. With matched conditions all examined measures yield the best and similar scores (almost 100%). With mismatched conditions, the proposed measures yield better scores than the GMM baseline system, especially for the short testing case. It is also observed that the average discrimination information measure is most appropriate for music category classifications and on the other hand the divergence measure is more suitable for music subcategory classifications.

***Keywords*—**Audio feature, information measures, music genre.

I. Introduction

MUSICAL Genre is widely used to categorize and label extremely vast world of music. This task can be achieved by human experts for the music industry or by consumers themselves. As a result, many different taxonomies are used classify the same musical genres. This is related to the fact, that many different descriptors and semantic ambiguities exist determine genre classification [5]. For example, names categories associated to each genre are not always similar coherent including the hierarchical structure (subcategories) itself. The manufacturing of new instruments and the realization new albums continue to accentuate this tendency. In future, genre taxonomy will remain in an elastic and dynamical structure. As argued by Aucouturier et Pachet [5], genre may be used in intentional or extensional concept. For each concept, genre taxonomy is related to different interpretations at descriptor level or at the semantical level. The authors describe three approaches to establish musical genre classification: manual classification (projection of human or expert knowledge), *prescriptive* approach that relies on supervised learning using signal processing techniques (classify genre as they are found) and finally emergent classification approaches that are based similarity measures to automatically produce the hierarchical genre structure.

We propose (for musical genre classification) to investigate new classification technique based on the information theory measures. The *prescriptive* approach (classify genre as they are found) is adopted in this work. The interest of the proposed technique resides primarily on the simplicity of its mathematical formalism and on its potential to be implemented for real or differed time applications. It requires little memory capacity to store the reference prototypes (2 parameters), not much computing time and remains very flexible over the testing/training duration. The parameters can be recursively estimated when the time duration of musical piece is long enough. Experiments are carried out according to different strategies as matched and mismatched conditions, long or short testing with either long or short training. Results are compared to a Gaussian Mixture Model (GMM) recognizer system.

Hassan Ezzaidi is with Ermetis, Université du Québec à Chicoutimi, Chicoutimi, Qu´ebec, Canada, G7H 2B1 (e-mail: hezzaidi@uqac.uquebec.ca).
Jean Rouat is with NECOTIS, GEGI, Université de Sherbrooke, Sherbrooke, Québec, Canada, J1K 2R1(e-mail: Jean.Rouat@ieee.org).

II. Related Work

Several features and models were proposed and experimented for genre classification. Generally, parameters can be divided into three feature families. The *first family* represents the timbral texture of audio signal and usually comprises:
Spectral Centroid [8], Spectral Rollof [8], Zero crossing Rate [8] [9], Vector of the fast Fourier transform samples, Spectral Flux [8] [9], Mel Frequency Cepstral [7] [8] [9], Linear Prediction Coefficients and Linear Prediction Coefficient [8]. The *second family* represents the rhythmic content features as proposed by [9]. Theses features are based on detecting the periodicities of the signal. For the extraction of theses features, a discrete wavelet transform, envelope extraction, autocorrelation function and finally the peak detection are elaborated to built a beat histogram. The *third family* is based on pitch content features [9] [10]. The pitch features are based on a pitch histogram obtained from multiple pitch detections. Five features are extracted from the pitch histogram and used for musical genre classification. *Other features* have been proposed and experimented for automatic genre classification such as like the audio low-level descriptor in the context of MPEG-7 standard, Root mean square, periodicity rate, various order central moments [2].

A. Classification Techniques

The majority of proposed works in automatic musical genre

classification, use as classifiers the Gaussian Mixture models (GMM). Support vector machine is investigated in Toa et al. [10] upon various features and compared to GMM and Linear Discriminant Analysis (LDA). Fisher's linear discriminant analysis was proposed by Toa et al. [10] for genre classification. The radial basis function network (RBF) was examined by [12] using various combinations of initialization methods. A feedforward neural network (FFNN) is also proposed cas classifiers for five subgenres of classical music [6]. The K-Nearest Neighbor is non-parametric classifier that was proposed in various work [11] [9].

III. Proposed Classification Measures

A. Information Theory Measures

Let $\{m_R(i)\}_{1\le i\le M}$ be a sequence of M independent parameter vectors related to the source information noted as R, extracted from an acoustical signal. All vectors are p dimensional, assumed to be independent and distributed like a Gaussian function. Therefore, they are characterized in the parametric form only by 2 parameters: a mean vector noted $\overline{m_R}$ and a covariance matrix noted Σ_R as:

$$\overline{m_R} = \frac{1}{M}\Sigma_{i=1}^{i=M} m_R(i)$$

$$\Sigma_R = \frac{1}{M}\Sigma_{i=1}^{i=M}\left(m_R(i)-\overline{m_R}\right)^T\left(m_R(i)-\overline{m_R}\right)$$

where $(\,)^T$ is the transpose.

Similarly, a sequence of N vectors $\{m_T(i)\}_{1\le i\le N}$ corresponds to a target information source to be classified and that obeys to the same properties as the reference source. Hence, the target source can be represented by 2 parameters: the mean vector $\overline{m_T}$ and the covariance matrix noted Σ_T.

A discrimination information in the Bayes classifier sense, for class W_R versus class W_T, can be measured by the logarithm of likelihood ratio as defined in [3]:

$$\mu_{R,T} = \ln\left\{\frac{p_R(x)}{p_T(x)}\right\} \quad (1)$$

where $p_R(x)$ and $p_R(x)$ correspond to the probability densities for the reference and target classes, respectively. The averaged information for class W_R versus class W_T is the expectation of $\mu_{R;T}$ and is defined as:

$$I_{R,T} = \int_x p_R(x)\ln\left\{\frac{p_R(x)}{p_T(x)}\right\}dx \quad (2)$$

If the distribution of each class is assumed to be Gaussian and multivariate, I_{RT} can be expressed in another form according to mean vector and covariance matrix:

$$I_{R,T} = 0.5\left(\ln\left(\frac{|\Sigma_T|}{|\Sigma_R|}\right) + tr\left(\Sigma_R\left(\Delta_m\Sigma^{-1}{}_{RT}\right)\right) + tr\left(\Sigma_T^{-1}\Delta_m\Delta^T{}_m\right)\right) \quad (3)$$

where:

$\Delta_m = \overline{m_R} - \overline{m_T}$, $\Delta\Sigma_{TR}^{-1} = \Sigma_T^{-1} - \Sigma_R^{-1}$, $\Delta\Sigma_{RT}^{-1} = \Sigma_R^{-1} - \Sigma_T^{-1}$ and $\Delta\Sigma_{RT} = \Sigma_R - \Sigma_T$.

A divergence measure or *total average information discrimination* is defined as the sum of the average information discrimination $I_{R,T}$ and $J_{R,T}$ and can be expressed as:

$$J_{R,T} = \frac{tr\left[\Delta\Sigma_{RT}\Delta\Sigma^{-1}{}_{TR}\right]}{2} + \frac{tr\left[\Delta\Sigma^{-1}{}_{RT}\Delta_m\Delta_m^T\right]}{2} \quad (4)$$

This divergence provide a dissimilarity measure between the normally distributed classes W_R and W_T. The information iscrimination $I_{R;T}$ and divergence $J_{R;T}$ are tested in the content of automatic musical genre classification. The development and details for equations 3 and 4 can be found in [3].

The motivation for theses measures is based on the fact that the musical signal is generally characterized by rhythmicity and regularity which cover a long temporal period.

IV. Features Extraction and the Reference Measures

A. Music DataBase

The RWC Music Database [4] is a copyright-cleared music database and it is the world's first large-scale music database compiled specifically for research purposes. It is composed of 100 musical pieces: 73 pieces originally composed and arranged and 27 pieces come from the public-domain.

Among the many characteristics of the RWC database, this includes the following: 91.6 hours recording, performance of about 150 instrument bodies, 3 variations for each instrument, variations in instrument manufacturers and musicians, different manufacturer/different musician, wide variety of sounds. The music database (RWC) is divided into main categories and subcategories of genres.

B. Features Vector Extractions

Each musical piece is first downsampled from 44.4Khz to 16Khz. Then the musical signal is divided into frames of 1024 samples with 50% overlap. It is assumed that the musical signal is more stable and quasi-stationary than the speech signal where coarticulation is dominant. For each frame, a Hamming window is applied without pre-emphasis. Then, 29 averaged Spectral energies are obtained from a bank of 29 Mel triangular filters followed by a discrete cosine transform, yielding 12 Mel frequency Cesptrum Coefficients. Cepstral mean normalization is not used because it removes important genre attributes that characterize the piece style (see [1]).

Since one uses a classifier based on time averaged measures, we assume that the influence of delta and delta-delta MFCC coefficients is not of major importance.

C. The Reference System

For comparison purposes, Gaussian Mixture Models (GMM) of the MFCC is used and 16 mixtures and diagonal covariance matrices are estimated via the Expectation Maximization (EM) algorithm.

V. Experiments

A. Prescriptive Taxonomies

While exploring the RWC database in its hierarchical structural for genre classification, we found that the number of musical pieces corresponding to each subcategory is not equal. This suggests the use of the database with two procedures:

1) All subcategories of which the number of musical pieces is not three are rejected. In this case, the number of categories and subcategories are reduced to 9 and 29, respectively.
2) The number of musical piece is not important, thus, all the existent categories and subcategories are used. Then, the number of categories and subcategories are respectively 12 and 40.

B. Experiments

Two strategies are investigated the testing:

1) Strategy 1: long training and long testing. A half musical piece was taken for the training session and the rest (second half piece) was used for the testing session.
2) Strategy 2: long training and short testing.

In this case, the same scenario as strategy 1 is kept for the training session but, for the testing session only one minute is extracted from second half musical piece. Matched and mismatched conditions are also reported for each strategy. Matched conditions refer to experiments where each musical piece is used in training (first half peace) and testing sessions (second half peace). Mismatched conditions refer to experiments where the musical test piece was never used or presented during the training session.

TABLE I
Score Recognition (in%) for Matched Conditions; Long Train/Test

		I_{ij}	J_{ij}
Categories	$\mu(T; R)$	98	99
(12)	$\mu(R; T)$	98	99
Subcategories	$\mu(T; R)$	96	97
(40)	$\mu(R; T)$	97	97

TABLE II
Score Recognition (in%) for Matched Conditions; Long Train/Test

		I_{ij}	J_{ij}
Categories	$\mu(T; R)$	98	99
(11)	$\mu(R; T)$	98	99
Subcategories	$\mu(T; R)$	96	98
(29)	$\mu(R; T)$	97	98

TABLE III
Score Recognition (in%) for Mismatched Conditions; Long Train/Test

		I_{ij}	J_{ij}
Categories	$\mu(T; R)$	75	70
(12)	$\mu(R; T)$	70	70
Subcategories	$\mu(T; R)$	58	62
(40)	$\mu(R; T)$	60	62

TABLE IV
Score Recognition (in%) for Mismatched Conditions; Long Train/Test

		I_{ij}	J_{ij}
Categories	$\mu(T; R)$	70	64
(11)	$\mu(R; T)$	64	64
Subcategories	$\mu(T; R)$	49	55
(29)	$\mu(R; T)$	51	55

VI. Recognition Criterion

During the training session, mean vectors and covariance matrices are estimated and stored as prototype reference to characterize each musical genre. The indice R is used to design the reference. Similarly, during the testing session, the measures between the test file and all reference prototypes of musical genre are evaluated. The reference prototype style with the minimal distance to the test is assigned to the recognized style.

VII. Results and Discussion

All results presented here are based on supervised learning techniques. Particularly, the average information is tested in her asymmetrical form $I_{i,j}$ and $I_{j,i}$ because the divergence distance is originally symmetrical. As mentioned above, we are interested in carrying out an automatic classification where the hierarchical structure is pre-established. Two structures were defined for the RWC database, each comprising two levels with a different node number on each level. Based on the prescriptive approach and the proposed measures, we are interested in automatically reproduce the proposed taxonomy of the RWC musical database.

For each category and subcategory, results are reported on tables I o VIII. Table IX illustrates the recognition scores of the GMM reference system for each strategy and specific experimental conditions. Tables I to VIII report the scores for the two information theoretical measures based upon information theory concept, for all strategies and experimental conditions. Each measure $I_{i,j}$, $J_{i,j}$ and GMM models, is tested and evaluated for the high hierarchical level where the classification is carried out based on the category label, and on a lower level where the classification is carried out based on the subcategory label. For all experiments the best performance (for every measures and training/testing conditions) is observed for the matched conditions. With the $I_{i,j}$ measure, we obtain recognition scores from 95% to 98% for genre classification by categories and from 88% to 97% when genre classification was addressed by subcategories. With the $J_{i,j}$ measure, we obtain recognition scores from

97% to 99% for genre classification by categories and from 92% to 97% when genre classification was addressed by subcategories. Scores for these different measures and strategies remain similar with the matched conditions. It is also observed that the long and short testing does not have many influence on the scores. The good performance obtained with the proposed measures confirms that previous theoretical assumptions are probably verified. For mismatched conditions, where tests were never seen during training sessions, the scores drop significantly. The recognition scores varied from 55% to 73% for the genre classification by categories, and from 40% to 58% by subcategories. The information theoretical measures seem to be more interesting and yield the best scores in comparison to the other measures for all experimental strategies. Particularly, it is observed that the discrimination measure $I_{i,j}$ yields a better score when the classification is addressed by category. When the problem is addressed to subcategories, the best score is obtained with the divergence measure $J_{i,j}$ However, the decrease in performance with mismatched conditions can partially be explain by the fact that human experts themselves do not always agree on the category or sub-category for a specific musical. Furthermore, the auditory system is able to perceive subtle features (likes tremolo, hangs of rhythm,etc..) that are not encoded in the MFCC parameters. The reference classification system yields the same recognition rates in matched conditions comparatively to the proposed measures, except with short testing utterances where the score drops significantly to 30%. With mismatched condition, the recognition scores of categories are similar for all measures, but with the recognition scores of subcategories the best performance of the reference system is globally 10% lower than $I_{i,j}$ and $J_{i,j}$

VIII. Conclusion

Automatic musical genre taxonomy has been realized. Matched/mismatched and long/short testing strategies have been studied. The best results were observed in all matching conditions and yielded score until 99% and 97% recognition for categories and subcategories, respectively. Worst results were observed in all mismatched conditions and decreased to about 73% and 58% for categories and subcategories, respectively. A Gaussian Mixture Model is used as a reference system. With mismatched conditions, the proposed measures yield better scores than the reference system especially for the short testing case. Particularly, it is observed that the averaged discrimination information measure is most appropriate for musical categories classification and on the other hand the divergence measure is most suitable for music subcategories classification. We plan to adapt this technique for unsupervised musical genre classification in the same conditions. We will to focus our research on the automatic generation of new categories and subcategories.

TABLE V
Score Recognition (in%) for Matched Conditions; Long Train and Short est

		I_{ij}	J_{ij}
Categories	$\mu(T; R)$	96	97
(12)	$\mu(R; T)$	95	97
Subcategories	$\mu(T; R)$	89	93
(40)	$\mu(R; T)$	93	93

TABLE VI
Score Recognition (in%) for Matched Conditions; Long Train/ and Short Test

		I_{ij}	J_{ij}
Categories	$\mu(T; R)$	96	97
(11)	$\mu(R; T)$	96	97
Subcategories	$\mu(T; R)$	88	92
(29)	$\mu(R; T)$	93	92

TABLE VII
Score Recognition (in%) for Mismatched Conditions; Long Train/ and Short Test

		I_{ij}	J_{ij}
Categories	$\mu(T; R)$	73	68
(12)	$\mu(R; T)$	65	68
Subcategories	$\mu(T; R)$	53	58
(40)	$\mu(R; T)$	53	58

TABLE VIII
Score Recognition (in%) for Mismatched Conditions; Long Train/ and Short Test

		I_{ij}	J_{ij}
Categories	$\mu(T; R)$	67	60
(11)	$\mu(R; T)$	55	60
Subcategories	$\mu(T; R)$	42	48
(29)	$\mu(R; T)$	40	48

TABLE IX
GMM Score Recognition (in%) for Matched and Mismatched Conditions

	Time	matched	mismatched
Categories 11	long	100	73
	Short	81	60
Categories 12	Long	100	75
	short	81	61
SubCategories 34	Long	100	44
	short	80	38
SubCategories 40	Long	95	50
	short	77	42

References

[1] H. Ezzaidi and J. Rouat, Speech, music and songs discrimination in the context of handsets variability, In *proceedings of ICSLP 2002*, 16-20 September 2002.

[2] FT. Lambrou, P. Kudumakis, M. Sandler, R. Speller and A. Linney, Classification of Audio Signals using Statistical Features on Time and Wavelet Transform Domains, In *IEEE ICASSP 98*, May 1998, Seattle, USA.

[3] J. Tou and R. Gonzalez. Pattern recognition principles, Addison-Wesley Publishinig Company, Reading, Massachusetts , 1974.

[4] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka,RWC Music Database: Music Genre Database and Musical Instrument Sound Database, In *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, pp. 229-230, October 2003.

[5] Aucouturier, J.J and Pachet, F. Musical Genre: a Survey. In *Journal of New Music Research*, Vol. 32, No 1, pp.83- 93, 2003.

[6] Ricardo Malheiro, Rui P. Paiva, Ant ˆ Unio Mendes, Teresa Mendes, Amilcar Cardoso, A Prototype for Classification of Classical Music using Neural Networks, In *Proc. of the 8th IASTED International Conference on Artificial Intelligence and Soft Computing*, pp. 294-299, ASC'2004, Marbella, Spain, September-2004.

[7] H. Soltau, T. Schultz, M. Westphal, and A. Waibel. Recognition of Musical Types. In *Proceedings International Conference on Acoustics, Speech and Signal Processing (ICASSP).*, May 1998, vol. 2, pp. 1137˜n1140.

[8] Tzanetakis, G., Essl, G., and Cook, P., Automatic Musical Genre Classification of Audio Signals, In *Proceedings of the 2001 International Symposium on Music Information Retrieval*, 2001.

[9] Tzanetakis, G., and P. R. Cook, Musical Genre Classification of Audio Signals, In *IEEE Transactions on Speech and Audio*, July, 2002.

[10] Li, Tao and Tzanetakis, George, Factors in Automatic Musical Genre Classification of Audio Signals, In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY October 2003.

[11] Tao Li, Mitsunori Ogihara, and Qi Li. A Comparative Study on Content-Based Music Genre Classification, In *Proceedings of Annual ACM Conference on Research and Development in Information Retrieval*, (SIGIR 2003),Pages 282-289.

[12] Douglas Turnbull, Charles Elkan, Fast Recognition of Musical Genres Using RBF Networks, In *IEEE Trans. Knowl. Data Eng*, 17(4): 580-584 (2005).