[2009]

RESEARCH REPORT

MULTIEXPERT SYSTEM FOR AUTOMATIC MUSIC GENRE CLASSIFICATION

Aliaksandr Paradzinets

Hadi Harb

Liming Chen

[June 2009]

MULTIEXPERT SYSTEM FOR AUTOMATIC MUSIC GENRE CLASSIFICATION

***Aliaksandr Paradzinets**, Hadi Harb, Liming Chen*

Ecole Centrale de Lyon

Departement MathInfo

aliaksandr.paradzinets@ec-lyon.fr, liming.chen@ec-lyon.fr, hadi.harb@ec-lyon.fr

ABSTRACT

Automatic classification of music pieces by genre is one of the crucial tasks in music categorization for intelligent navigation. In this work we present a multiExpert genre classification system based on acoustic, musical and timbre features. A novel rhythmic characteristic, 2D beat histogram is used as high-level musical feature. Timbre features are extracted by multiple-f0 detection algorithm. The multiExpert classifier is composed from three individual experts: acoustic expert, rhythmic expert and timbre analysis expert. Each of these experts produces a probability of a song to belong to a genre. The output of the multiExpert classifier is a neuron network combination of individual classifiers. It was shown in this work a 12% increase of classification rate for the multiExpert system in comparison to the best individual classifier.

Keywords – Music genre classification, acoustic features, rhythmic features, timbre features, multi expert system.

1. INTRODUCTION

Digital music distribution and consumption is gaining in popularity. A significant amount of music is nowadays sold digitally via networks, such as the Internet or mobile phone networks. In the digital era, consumers have access to millions of songs. Also, artists and producers are able to produce more songs and distribute them instantly. Due to this democratization of access to digital music, efficient categorization systems and intelligent search engines become crucial.

Music genres are categories used by music editors and distributors. These categories facilitate the navigation into a physical music collection. Genres are naturally used to categorize digital music collections. To date, genres are manually associated by music editors, distributors and aggregators. Associating genres automatically to music titles is becoming important for several reasons. First, in the digital era the music unit is the title in contrary to the album in the physical era, meaning a clear increase in the number of units to be categorized. Second, an automatic music genre classification system generates categories that are independent from the subjective categories given manually.

The present work is dedicated to automatic music genre classification.

Several systems have been proposed in the literature for automatic genre classification. In their majority, these systems are an adaptation of a general audio classifier to the task of music genre classification. They use a signal-only analysis.

In this paper we hypothesize that a signal analysis approach for music genre classification has its limitations. First, due to the complexity of the task and second due the inappropriateness of the signal features generally used. We propose therefore to use different characteristics of music titles: timbre-related, rhythm-related, and artist name-related ones. Multi-expert classifier architecture is proposed in order to be used with different characteristics.

Another important contribution of this work resides in a reference database of 37480 seconds of signal from 822 different artists classified manually in 6 musical genres. The database is publicly available for research purposes.

The purpose of this work is to analyze the effect of the combination of several characteristics and classifiers capturing different aspects of musical information in the context of automatic genre classification.

Experimental results on the reference database have showed that the combination of different types of characteristics dramatically improves classification results.

Genre classification of music is a complicated problem of automatic signal classification. Many recent works have tackled this problem.

[1] uses frequency centroid, spectral flux, zero crossing rate, cepstral characteristics as well as characteristics of musical rhythm and other aspects. Proposed features are classified by GMM classifier. For six musical genres: Classic, Country, Disco, Hip Hop, Jazz and Rock the results in terms of average classification precision were of 62% for 30-second segments.

[2] proposes to use peak valleys of spectral characteristics coupled with GMM for genre classification. For three genres of music: Pop, Jazz and Rock a precision of classification of 81% was reported for 10-second segments.

A classical approach of genre classification, based on cepstral features (MFCC and delta MFCC) and GMM classifier was used in [3]. The author reports a precision of 92% in classification of entire songs (musical titles) of the following genres: Blues, Easy Listening, Classic, Opera, Techno and Indy Rock.

An original approach of temporal structure modeling of musical signals using neural networks is introduced in [4]. This approach was evaluated in genre classification of four genres: Rock, Pop, Techno and Classic. Authors provide an average precision of 70% for 4-second segments.

Recent approaches presented in literature use spectral features such as MFCC, ZCR etc. together with support vector machines [5] and AdaBoost methods [6]. They reported nearly 72% and 78% correspondingly on *Magnatune* database at MIREX2005 Genre Classification Contest.

Direct comparison of algorithms found in literature is a complicated question since these algorithms were evaluated on different databases with different genres. For example, similar approaches used correspondingly in [1] and [3] have produced precision rates rather different as 62% and 92%. In addition, length of segments put to the classification also has an influence on results. It is very probable, for example, that one song could be correctly classified by major vote over all segments of the song when only 30% of its duration is classified correctly.

2. DATABASE

One important difficulty to overcome in the development of an automatic music genre classification system is the constitution of a reference database. A reference database that is sufficiently representative of the real world situation in order to draw reliable conclusions on system architecture, features, classifiers etc. The database must also reflect the real needs in real world applications, especially in the definition of genres and the variability of music excerpts for each genre.

We have chosen six music genres for the reference database. The genres were chosen as being the six genres we generally found on several online music stores. The selected list of genres includes: Rock (Pop), Rap (HipHop, R&B), Jazz (Blues), Classic, Dance (Disco, Electro, House), Metal (Hard Rock, Heavy Metal)

Each of these “general” genres consists of several sub-genres which have more precise definition. For example, the Rap genre consist of such sub-genres as Rap, HipHop, R&B, Soul etc… each sub-genre corresponds to a specificity which means that two songs of the given sub-genre are closer at least from musical edition’s point of view than two songs from different sub-genres. Unfortunately, detailed genre taxonomy can be defined in multiple ways [7] which is a limit for the definition of a universal musical genres taxonomy. Hence, we propose to choose from each “general” genre a well defined sub-genre which represents the main genre. The choice of sub-genres lies on the most representative sub-genre in the meaning of number of songs associated to it by a musical distributor, for instance fnacmusic.

For each representative sub-genre we have selected the list of artists associated to it on the music distributor store. This list was then used to capture music from webradios [www.shoutcast.com]. The musical segments were captured as 20-seconds records starting from the 50th second of the play and saved as PCM 8KHz 16bit Mono files. In total the reference database consists of 1873 titles from 822 artists which make 37480 seconds in total (see Table 1).

It is crucial to note an important variability of musical titles in this reference database owing to an important number of artists. Notice that a genre classification system applied on a reference database of limited variability, in terms of artists, may capture the specific style of the artists instead of capturing the general genre characteristics. As far as we know, this is the first reference database where the attribution of genres to each title is not made in subjective manner by one person but takes into account the musical distribution attribution. Also, in comparison with other databases like magnatune, the current reference database is better balanced in the meaning of representation of classes (~1000 classic vs. ~70 for jazz in the case of magnatune).

Table 1. ECL database details.

	Titles	**Artists**	**Duration**
Classic	214	113	4280
Dance	335	226	6700
Jazz	305	104	6100
Metal	324	105	6480
Rap	311	152	6220
Rock	384	122	7680
Total	**1873**	**822**	**37480**

3. CHARACTERISTICS

Different kinds of features are needed in order to discriminate between musical genres. We suppose that timbre-related acoustic features, such as MFCC features, are not sufficient for musical genre classification. Some musical genres are for example better described by specific beat patterns such as classical or Dance genres, while other genres are better described by some cultural characteristics such as Pop Rock. We propose in this work to combine acoustic features with beat-related features and artist name related features.

3.1. Acoustic features

In a previous work [8] we have proposed the use of the statistical distribution of the audio spectrum to build feature vectors in what we call the Piecewise Gaussian Modeling (PGM) features. PGM features constitute an interesting alternative for the MFCC features. They offer a condensed view for the audio signal without losing important information necessary for audio classification [9]. In this paper we propose to use PGM features for genre classification. We introduce also perceptually motivated PGM which are PGM features enhanced by modeling of human auditory filter.

The PGM is inspired by several aspects of psychoacoustics. It is fair to suppose that the perception of a stimulus is strongly related to its correlation with the past memory. The context effect is a known effect in the human speech recognition domain. We model the context effect by an auditory memory and we suppose that the classification of a stimulus at time instant *t* is based on the status of the memory at time instant *t*. The auditory memory is supposed to be a Gaussian distribution of the spectrum in the past time window, called the Integration Time Window (ITW). A concatenation of mean and variance values defines the feature vector of corresponding ITW and called PGM features. We can summarize the PGM feature extraction by the following:

1) Computation of the spectral coefficient vectors with 30 ms Hamming window and 20 ms overlap.
2) Grouping spectral coefficient vectors in ITW windows, estimating the Gaussian parameters of each ITW window. The Gaussian parameters are the mean and variance vectors and constitute 2 vectors of 20 elements each.
3) Normalising the mean values by their respective maximum and normalising the variance values by their respective maximum.

The mean and variance characteristics of basic PGM are issued from basic frequency spectrum of a signal. No information about modeling of human auditory filter is taken into account. It may be valuable to include a human auditory model in PGM characteristics. We hypothesize, however, that including basic perceptual aspects, such as a simplified auditory filter model, in the PGM features cannot considerably improve music genre classification results. In fact, music genre classification probably requires a high level analysis: an analysis by rule on high level characteristics such as rhythm, instruments involved etc. Hence, low level characteristics like spectrum-based features have a limit in their capability of musical genres discrimination.

To compute the perceptually motivated PGM, after calculation of frequency spectrum we apply critical bands filter (Bark scale [11]), equal loudness contour and specific loudness sensation [12].

Specific Loudness Sensation is a measure that can be obtained as follows:

$$L = \begin{cases} 2^{\frac{1}{10}(E-40)} & if\ E > 40\, phon \\ (\frac{1}{40}E)^{2.642} & otherwise \end{cases} \qquad (1)$$

To implement the Equal Loudness Contour we have summarized it by a single curve of weighs corresponding to different critical bands. Thus the value of perceptual loudness at 2KHz is half of that at 200Hz. The application of Equal Loudness Contour issues then energy values in Phon.

3.2. Beat-related characteristics

Rhythmical features are features obtained from beat or onset detection in musical signal [13][14]. We have developed a rhythmical description which we call 2D Beat Histogram. Beat detection in musical signal results normally in a curve which can be then compared by application of a threshold to detect instants of beats or onsets and beat period can be subsequently calculated. The beat histogram is a histogram of values of beat periods. It contains rhythmical information more rich than just the number of beats per minute (BPM) [15]. The 2D beat histogram is a combination of beat histograms for different values of threshold applied in beat detection.

Our approach in beat detection bases on a special version of Continuous Wavelet Transform – Variable Resolution Transform as spectral representation of the signal [17]. VRT as the basic time-frequency transformation was chosen due to flexible possibility of time/frequency resolution adjustment in desired way. Unlike the FFT, which provides uniform time resolution, the VRT provides high time resolution and low frequency resolution for high frequencies and low time resolution with high frequency resolution for low frequencies. In that respect it is similar to the human ear which exhibits similar time-frequency resolution characteristics [16].

An example of a musical excerpt representation after application of this VR-transform is depicted on Figure 1.

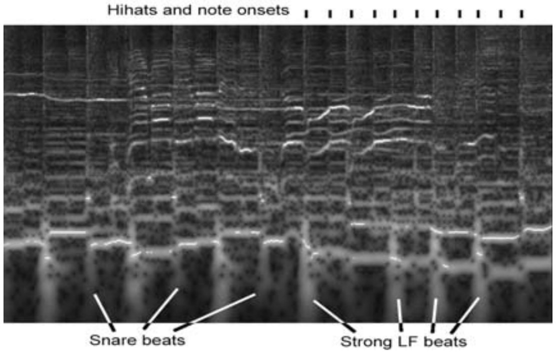

Figure 1. VRT representation of musical excerpt.

Since the information about beats and onsets is assumed to be concentrated in vertical constituent of VRT spectrogram, an image treatment technique can be applied to mark out all fragments in this “spectral image” connected with beats and onsets.

Usage of image treatment technique has been described in literature by few works. In [18] the authors apply edge enhancement filter on the Fast Fourier Transform (FFT) image in preprocessing phase. In the current work, preliminary experiments with VRT spectrum showed good results with the use of Sobel X operator [19].

Subsequently, the enhanced spectrogram $W^{*}(t,scale)$ is treated by calculating a beat curve in the following way. A small 5-sample window together with preceding large 100-sample window is moved across the enhanced spectrogram. The value of the beat curve in each time moment is the number of points in the small window with values higher than a threshold which is obtained from the average value of points in the large window. Numerous beat curves may be computed separately by dividing the spectrum into bands. For the general question of beat detection the only one beat curve is used.

The probable beats are situated in beat curve's peaks. However, the definition of final beat threshold for the beat curve is problematic. Adaptive and none-adaptive algorithms for peak detection may be unstable. Many weak beats can be missed while some false beats can be detected.

Thus, we propose to build a 2D form of beat histogram with a beats period on the X axis and with amplitude (beat detection threshold) on the Y axis (Figure 2). It is hence possible to avoid the disadvantage of recording conditions dependency (e.g. volume) and peak detection method. The range of threshold variation is taken from 1 to the found maximum-1. Thus, the beat strength is taken relatively and the volume dependency is avoided

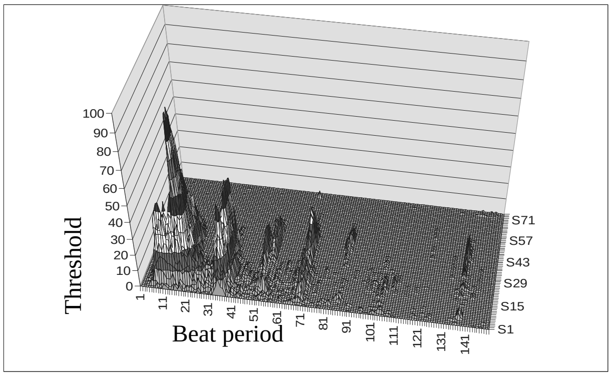

Figure 2. A 2-D beat histogram.

To prove the independency from recording conditions we have carried out an experiment. For this purpose a musical composition has been filtered with treble and bass cut filters. The resulting histograms of beats still had the same forms and peaks.

The described rhythmical image representation foresees a resemblance measure of two musical compositions in the meaning of rhythm. As the 2D beat histogram is not affected neither by volume of music nor by conditions of recording (e.g. frequency pass band), it can be used directly in a distance measure.

Possible ways to calculate a distance between two histograms is to compute a sum of their common part or to calculate the difference between bins with slight variations on axes. For two normalized histograms we define it as:

$$Dist_{H1,H2} = \sum_{x=1,y=1}^{N,M} \frac{1}{2}\left(\min_{R}\left(\left|H1_{x,y} - H2_{(x,y)+R}\right|\right) + \min_{R}\left(\left|H1_{(x,y)+R} - H2_{x,y}\right|\right)\right) \qquad (5)$$

where

H1, H2 – beat histograms to compare

N, M – beat histogram size

R – tolerance area of the following form (Figure 3)

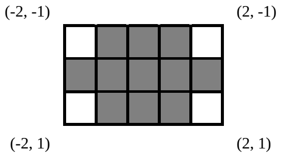

Figure 3. Tolerance area in histogram distance calculations.

3.3. Timbre characteristics

The third characteristic we extract from a musical piece is the timbre histogram. In general, “voiced” instruments differ from each other also by their timbre – profile of their partials. In our work we collect all detected notes with relative amplitude of their harmonics (this information is extracted using the algorithm we proposed in [17]). Further, relative amplitudes of harmonics are reduced to 3-4 bits and attached together in order to form an integer number. Histogram of these numbers is then computed.

Unlike the classical approach which consists of describing the timbre globally, our timbre features are first defined based on separate notes isolated by the multiple-f_0 estimation algorithm and then they are summarized in a histogram. Comparing such histograms gives a similarity measurement, which is supposed to be somehow instrument-related.

4. CLASSIFICATION SYSTEM

The problem of genre classification is naturally a complex one. Several characteristics are needed in order to distinguish between musical genres. Several approaches exist for incorporating the different characteristics into a classification system. E.g. [21] describes a probabilistic approach to the combination of music similarity features.

In this work we choose to build a classifier as a combination of several individual classifiers, so-called experts. That is we build a multiExpert system. Each of the individual experts is specific to one type of characteristics. Hence an acoustic expert uses the acoustic features, a rhythmic expert uses the beat-related features, and a timbre expert uses the timbre features. Each of these experts produces a probability of belonging of a song to a genre. In previous works by our team a weighted sum of individual classifiers was used for producing resulting probabilities [22]. However, a contribution of each classifier is not necessarily linear. Hence in this work we used a method of combining based on MLP as a generic non-linear classifier with the following architecture. Normalized outputs of individual classifiers form the input for the MLP. Six outputs of the MLP are probabilities that a song belongs to six genres. The MLP is trained on the testing set where target vectors for training are formed with one member value equal to 1 for the genre in question and 0-values for all the rest. The MLP which was used has one hidden layer with 24 neurons. Further increase of the number of neurons in the hidden layer leads to augmenting of the network complexity and requires more training data. It can therefore result in overtraining with bad generalization.

4.1. Acoustic expert

The expert of classification by acoustic analysis is one or several MLPs have the PGM or perceptually motivated PGM features as input characteristics. Each MLP of this expert is trained independently on the whole learning dataset or on a part of it.

4.2. Rhythmical expert

The expert of classification by rhythmical analysis is a basic k-NN classifier based on the 2D beat histogram and the rhythmical similarity measure. The rhythmical distances between musical files to be classified and musical files from the test set are calculated. The probability of belonging of the file in question to a class (genre) is proportional to the number of files of the same class returned in the top 15. Hence, this is a 15-NN classifier.

4.3. Timbre expert

Timbre expert is a k-NN classifier fed by timbre histograms (§3.3) as feature vectors. Minkosvki-form distance is used in place of a distance measure.

5. EXPERIMENTATION

We have carried out a series of experiments to analyze the effect of different characteristics in order to evaluate the multiExpert architecture.

5.1. PGM vs. perceptually motivated PMG

The aim of this experiment is to compare the effect of human auditory filter modeling in the context of PGM characteristics. In this experiment 400 seconds of signal per genre were used as learning dataset. The rest of data which makes 16330 seconds was classified. We should mention that only 10 seconds per title where used for learning purposes. In fact, there were no changes observed when increasing this duration to 20 seconds.

One of MLPs was trained using basic PGM characteristics and the other – using PGM characteristics after application of human auditory modeling.

The result of classification is as follows.

Table 2. Perceptually motivated PGM vs. basic PGM classification results (average rate 43.0% vs. 40.6%)

	C	D	J	M	Ra	Ro
C	**42 / 47**	2 / 3	6 / 16	1 / 0	3 / 4	7 / 11
D	10 / 4	**41 / 49**	9 / 13	4 / 14	18 / 23	10 / 16
J	25 / 16	6 / 5	**32 / 31**	3 / 4	10 / 5	23 / 19
M	4 / 6	20 / 15	11 / 11	**69 / 50**	17 / 20	16 / 13
Ra	5 / 9	16 / 16	18 / 9	11 / 24	**41 / 43**	11 / 17
Ro	14 / 18	15 / 12	24 / 20	12 / 8	11 / 5	**33 / 24**

Confusion matrixes for the PGM and perceptually motivated PGM characteristics show the following conclusions. First, the use of human auditory modeling yield an improvement of classification rates. In average this improvement is about 3%. Second, this improvement is not of large significance, which comes from the fact that the classification of music by genres bases mainly on high level analysis instead of low level characteristics. We can expect a limit of classification precision which cannot be overcame using only low level acoustic-based analysis. This fortifies our hypothesis of combination of experts where each expert does different analysis of the classification problem.

As a result of this experiment we chose perceptually motivated PGM characteristics instead of basic PGM as acoustic features in our classification system.

5.2. Mono expert PGM-MLP vs multiExperts PGM-MLP

The idea here is to split the training dataset into sub-ensembles several hundred seconds long. For each sub-ensemble of the training dataset a PGM-MLP classifier is trained. During the phase of classification every expert (PGM-MLP classifier) gives the probabilities of belonging of the characteristic vector to different classes. The sum of the probabilities is used as the output of multiExpert PGM-MLP classifier. The aim of this method of classification is to simplify the training procedure of the neural networks while improving the classification precision [22]. For the neural network it is known that if a classification problem is complex, the training time can be long at the same time with high probability to converge to a local minimum.

We apply here this method to simplify the training process while a dataset of the system of genre classification can reach thousands songs.

The same datasets as in pervious test were used in this experiment. However, 4 MLP neural networks were trained on 4 sub-ensembles of learning dataset. A simple solution of estimation of relative weights for multi-MLP combination was used where the weights are equal and static with the sum equal to 1.

Table 3. multi MGI-MLP classification results (average 49.3%)

	C	D	J	M	Ra	Ro
C	**53**	5	12	2	5	10
D	3	**40**	7	7	11	8
J	23	4	**38**	2	6	21
M	7	24	15	**75**	16	19
Ra	2	16	12	7	**55**	7
Ro	12	11	16	7	7	**35**

As it can be seen from the result table, the rate of classification is improved in comparison to basic training of a single MLP. The improvement is about 5% in average with considerable simplification of the training process. These results confirm our results obtained in [22].

5.3. Genre classification using beat histogram

In this experimentation we apply a rhythmical analysis classifier to carry out the genre classification. The datasets here are the same as in previous experiments. Every musical title in the testing set is compared to all titles in the learning set by the distance between 2D beat histograms described in 3.2. We proceed then with classic 15-KNN classification. As it cat be seen from the confusion matrix (Table 4), the classification results for musical genres which have specific rhythmical structure are clearly superior to others. Precision of classification for such genres as Classic, Dance and Rap are 82.8%, 76.8% and 75.7% correspondingly.

Table 4. Beat expert classification results (average 71.4%)

	C	D	J	M	Ra	Ro
C	**82.8**	0.5	4.4	5.4	1.0	5.9
D	0.3	**76.8**	2.1	4.6	7.9	8.2
J	5.5	3.8	**68.9**	1.4	10.0	10.4
M	8.3	3.5	8.0	**69.7**	2.2	8.3
Ra	0.0	6.0	10.7	3.3	**75.7**	4.3
Ro	7.4	3.6	15.3	8.2	11.2	**54.4**

The same algorithm was applied to classification of *ISMIR2004 magnatune* database where there were 729 files in learning and 729 files in testing set within 6 genres: CLASSICAL, ELECTRONIC, JAZZ_BLUES, METAL_PUNK, ROCK_POP, WORLD. The average classification rate was similar to the average classification rate on our database (Table 5).

Table 5. Beat expert on *Magnatune* (average 54.6%, raw accuracy 68.1%)

	C	E	J	M	Ro	W
C	**89.7**	0.6	0.6	0.3	3.1	5.6
E	6.1	**56.8**	1.7	3.1	13.1	19.2
J	11.5	1.9	**30.8**	0	3.8	51.9
M	4.4	3.3	0	**46.7**	30.0	15.6
Ro	10.8	7.4	1.0	13.3	**52.7**	14.8
W	23.0	9.0	0	0.4	16.8	**50.8**

5.4. Genre classification by timbre classifier

This experiment concerns another musical similarity metric – the timbre distance. A k-NN classifier based on this distance has performed with lower but still correct results. Table 6 describes the classification results obtained for the ECL database.

Table 6. Timbre analysis expert (average 42.4%)

	C	D	J	M	Ra	Ro
C	**49.3**	3	13.8	17.7	4.4	11.8
D	7.9	**19.8**	5.8	30.8	7.6	28.0
J	8.0	11.4	**31.5**	14.9	9.0	25.3
M	5.1	3.2	1.6	**85.7**	0.3	4.1
Ra	5.3	12.0	15.0	23.0	**28.3**	16.3
Ro	12.3	7.9	6.0	30.6	3.6	**39.6**

In the case of *Magnatune* database the raw classification accuracy obtained was 52.2% which makes 380 songs from 729 to be classified exactly. The normalized mean accuracy was 39.6%. The whole confusion matrix is given in Table 7.

Table 7. Timbre analysis expert on *Magnatune* dataset (average 39.6, raw 52.2%)

	C	D	J	M	Ra	Ro
C	**75.6**	1.9	0	2.5	5.0	15.0
D	16.6	**17.0**	0.4	14.0	15.3	36.7
J	19.2	0	**5.8**	3.8	13.5	57.7
M	10.0	2.2	0	**58.9**	18.9	10.0
Ra	23.2	2.0	0.5	24.1	**30.5**	19.7
Ro	27.5	9.0	0	5.3	7.4	**49.6**

5.5. Committee of experts

As explained in the section on architecture of classifiers, our multi-expert system aims at fusing single music feature based-experts into a global one for music genre classification. As the goal of this thesis work is the use of music features in a complementary way to purely acoustic features, we have also included a pure acoustic feature based classifier as one of our single classifiers as a baseline music genre classifier. Recall also that the fusion strategy is a Multi-Layer Perceptron having one hidden layer which synthesizes a global classification result from the outputs of single classifiers. Table 8 shows the result obtained for combinations of all experts on the Magnatune dataset.

Table 8. All experts combined by MLP on *Magnatune* dataset, (normalized mean accuracy 66.9%, raw accuracy 74.2%)

	C	D	J	M	Ra	Ro
C	**88.7**	0.6	0	0.6	1.2	8.9
D	3.5	**58.8**	9.6	3.5	7.9	16.7
J	7.7	3.8	**57.7**	0	11.5	19.2
M	0	8.9	0	**66.7**	22.2	2.2
Ra	1	11.8	2	10.9	**64.7**	9.8
Ro	13.9	8.2	2.5	0.8	9.83	**64.6**

As it can be seen from the table, normalized mean accuracy rate is up to 66,9%. Thus the combination of experts that uses music features in addition to purely acoustic feature based PGM-MLP expert, brings a significant improvement of classification precision as compared to the normalized mean accuracy rate of 49,6% achieved by the single PGM-MLP expert, rhythmic or timbre expert. The best improvements were achieved on Electronic and World genres, changing from an accuracy rate of 33,6% and 35,2% for single PGM-MLP to 58,8% and 64,6% for multi-expert system, respectively. These two music genres presumably need more music features for a better discrimination due to their varieties.

With the ECL Music genres dataset, which has a higher artist variability and generally better defined genres, our multi-expert system achieves even better results. Table 9 gives the classification results by our Multi-expert system. Indeed, our multi-expert system displays a normalized mean classification accuracy rate up to 80.9% as compared to 49.3% achieved by the single PGM-MLP expert.

Table 9. All ECL_genres database classification results (normalized mean and raw accuracy 80.9%)

	C	D	J	M	Ra	Ro
C	**91.6**	0	2	0	0	6.2
D	0	**69.2**	1	3.2	6.5	19.7
J	1.1	8	**71.2**	0	3.4	16
M	0	1	0	**89.2**	2.1	7.5
Ra	0	4.6	2.3	4.6	**80.2**	8.1
Ro	1.7	4.4	2.6	5.3	1.7	**83.9**

Experimental results and final comparison can be summarized by the following figures. Figure 4 depicts the comparison of classification accuracies of separate classifiers and their mult-expert combination for both databases.

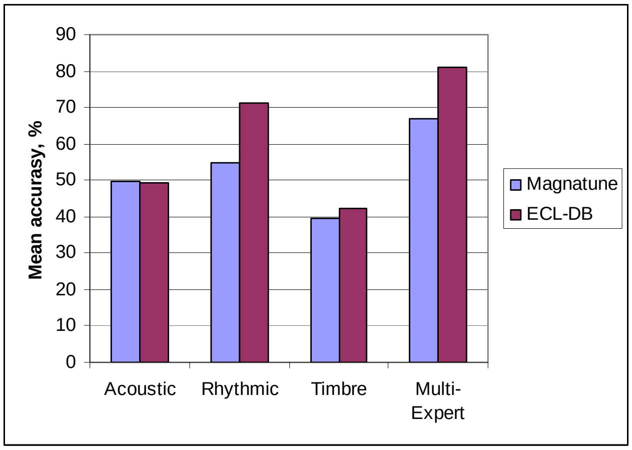

Figure 4. Comparison of classification results issued by different classifiers and their multi-expert combination for both databases.

All classifiers behave quite similarly in the case of both databases except the rhythmic classifier which performed better on the ECL database. In both cases there is a significant increase of classification rates with combined experts.

The figures (Figures 5) shows the performances of separate classifiers and their combinations on the ECL database according to genre.

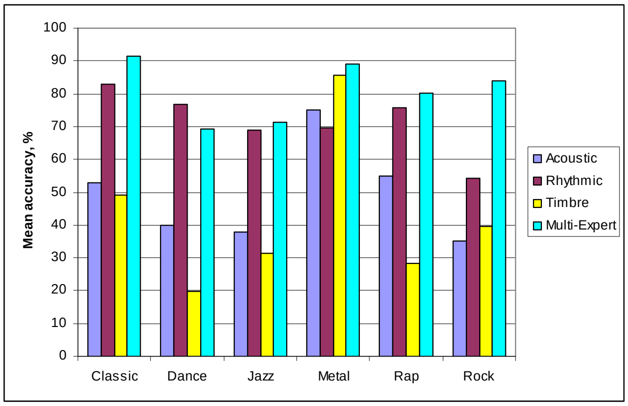

Figure 5. Performance of separate classifiers and their combination according to genre in the case of ECL database.

Generally a superiority of Multi-Expert classification results is observed for the majority of classes except just two cases – Classic of Magnatune database and Dance of ECL database. It can be explained by a tendency of the Multi-Expert configuration to average per-class accuracies together with a high Dance-to-Rock confusion in the latter case.

6. CONCLUSION

In this paper we have proposed a genre classification system which is based on an expert committee architecture where each individual expert uses its own specific characteristics. We have utilized such experts as rhythmical characteristics expert, acoustic characteristics expert and timbre analysis expert. It was shown that the expert with the highest performance is the rhythmical expert while the lowest classification rate was obtained from timbre expert.

In the case of the acoustic expert, the application of human auditory filter modeling before Piecewise Gaussian Modeling brings a slight improvement of classification rates in comparison to the basic PGM. More significant improvement of performance is achieved using Multi PGM-MLP. However, these improvements are not statistically significant.

The rhythmical expert uses the 2D beat histogram combined with a basic similarity measure between histograms and gives satisfactory results. However, the rhythmical expert finds its limits in classification of such genres as Jazz, Metal, Rock (Pop Rock).

Rock and in reality Pop Rock genre proves to be the most difficult genre from the point of view of signal analysis. The expert, which analyzes a number of Internet sites with the name of the artist and the name of musical genre gives the best results for such genres as Rock.

The combination of three experts, acoustic, rhythmic and timbre significantly raises the classification performance. Classification rate passes form 54.6% for the best individual classifier to 66.7% for the combination of all three experts in the case of *Magnatune* dataset.

An advantage of the proposed system is that it is highly extendable. It can for example incorporate other experts which are known in literature in order to give higher classification rates in comparison to single isolated classifiers.

We can conclude that the problem of music genre classification is a problem which requires the collaboration of multiple types of classifiers and characteristics. Acoustic analysis finds quickly its limits and must be supplemented by characteristics of cultural order.

7. REFERENCES

[1]. Tzanetakis G., Essl G., Cook P., Automatic Musical Genre Classification of Audio Signals, ISMIR01, 2001

[2]. Jiang Dan-ning, Lu Lie, Zhang Hong-Jiang, Cai Lian-Hong, Tao Jian-Hua, MUSIC TYPE CLASSIFICATION BY SPECTRAL CONTRAST FEATURES, Proc. of IEEE International Conference on Multimedia and Expo (ICME02), Lausanne Switzerland, August, 2002

[3]. Pye D., Content-based methods for the management of digital music, Proceedings of IEEE International Conference on, Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Volume: 4,2000 Page(s): 2437-2440 vol.4

[4]. Soltau Hagen, Schultz Tanja, Westphal Martin. Recognition of Music Types. Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, Seattle, WA. Piscataway, NJ 1998

[5]. Michael Mandel, Dan Ellis, Song-Level Features And Support Vector Machines For Music Classification, ISMIR2006

[6]. Bergstra J., Casagrande N., Erhan D., Eck D., Kegl B., Aggregate Features and AdaBoost for Music Classification, Machine Learning, 2006

[7]. Pachet F., Cazaly D., A Taxonomy of Musical Genres, Proceedings of Content-Based Multimedia Information Access Conference (RIAO) Paris, France 2000

[8]. Hadi Harb, Liming Chen, Voice-Based Gender Identification in Multimedia Applications, Journal of Intelligent Information Systems JIIS, special issue on Multimedia Applications, 24:2, 179-198, 2005

[9]. Hadi Harb, Classification d'un signal sonore en vue d'une indexation par le contenu des documents multimédias, PhD Thesis, Ecole Centrale de Lyon, December 2003

[10]. Harb H., Chen L. (2003) HIGHLIGHTS DETECTION IN SPORTS VIDEOS BASED ON AUDIO ANALYSIS, Proceedings of the Third International Workshop on Content-Based Multimedia Indexing CBMI03, September 22 - 24, IRISA, Rennes, France, pp 223-229

[11]. Zwicker E., Fasl H., Psychoacoustics, Facts and Models, vol 22 of Springer Series of Information Sciences, Springer, 2nd updated edition 1999

[12]. Bladon R., Modeling the judgment of vowel quality differences, Journal of the Acoustical Society of America, 69:1414-1422, 1981

[13]. Foote, J., M. Cooper, and U. Nam., Audio retrieval by rhythmic similarity, In Proceedings of the International Conference on Music Information Retrieval, 2002

[14]. Paulus, J., and A. Klapuri., Measuring the similarity of rhythmic patterns, In Proceedings of the International Conference on Music Information Retrieval, 2002

[15]. Tzanetakis G., Essl G., Cook P., Human Perception and Computer Extraction of Musical Beat Strength, Conference on Digital Audio Effects (DAFx-02), 2002

[16]. Tzanetakis G., Essl G., Cook P., *Audio Analysis using the Discrete Wavelet Transform*, Proc. WSES Int. Conf. Acoustics and Music: Theory 2001 and Applications (AMTA 2001) Skiathos, Greece

[17]. Paradzinets A., Harb H., Chen L., *Use of Continuous Wavelet-like Transform in Automated Music Transcription*, EUSIPCO2006

[18]. Nava G.P., Tanaka H., (2004) *Finding music beats and tempo by using an image processing technique*, ICITA2004

[19]. Sobel L., *An isotropic image gradient operator*, Machine Vision for Three-Dimensional Scenes, 376-379, Academic Press, 1990

[20]. Knees P., Pampalk E., Widmer G., *Automatic Classification of Musical Artists based on Web-Data* ÖGAI Journal, vol. 24, no.1, pp. 16-25, 2005.

[21]. Flexer A., Gouyon F., Dixon S., Widmer G., *Probabilistic Combination of Features for Music Classification,* ISMIR2006

[22]. Harb H., Chen L., Auloge J-Y., *Mixture of experts for audio classification: an application to male female classification and musical genre recognition,* In the Proceedings of the IEEE International Conference on Multimedia and Expo, ICME 2004