Audio Engineering Society

Convention Paper

Presented at the 121st Convention
2006 October 5–8 San Francisco, CA, USA

Music Genre Categorization in Humans and Machines

Enric Guaus[1,2], Perfecto Herrera[1,2]

[1] Sonology Department, High Music School of Catalonia, Barcelona
{enric.guaus, perfecto.herrera@esmuc.net}

[2] Music Technology Group, Pompeu Fabra University, Barcelona
{eguaus, pherrera@iua.upf.edu}

ABSTRACT

Music Genre Classification is one of the most active tasks in Music Information Retrieval (MIR). Many successful approaches can be found in literature. Most of them are based on Machine Learning algorithms applied to different audio features automatically computed for a specific database. But there is no computational model that explains how musical features are combined in order to yield genre decision in humans. In this work we present a listening experiment where audio has been altered in order to preserve some properties of music (rhythm, harmony, etc) but at the same time degrading other ones. Results are compared with a series of state-of-the-art genre classifiers based on these musical properties and we draw some lessons from that comparison.

1. INTRODUCTION

In the last few years, the distribution of digital music is changing the habits of music consumers in our society. It is not so strange that personal music collections use more than 20Gb in Hard Disks. This is a large amount of data and, obviously, it has to be organized under some criteria that produce better results in search, browse and retrieval processes.

Traditionally, genre has been considered the most important descriptor used in CD stores and online music dealers. In CD stores, it is used to guide costumers into a specific album from a specific author who is located in a specific shelf. In online shops, genre is only one of the multiple criteria in the search engine. In many cases, genre descriptor has been also applied as the main classification criteria in personal CD collections but using Genre in web-stores and remote databases is done according to criteria that are not homogeneous, yielding sometimes inconsistent results (i.e., the same artist being classified in different classes by different music providers). Audio files are often organized according to an “Author – Album – Track” tree and genre is only included as metadata attached to each audio file (iTunes, Winamp, etc.). Whatever the use of the genre descriptor is, it helps users to find their preferred music. The problem arises when manual labeling has to be applied to such amount of data [1] [2]. Although there is no a ground-truth in musical genre taxonomies [3], automatic genre classification systems can help both musicologists and personal users in this task.

Many efforts have been done in automatic musical genre classification in the last years. Soltau achieves an accuracy of 86% of correct classification using 4 genres in [4]. Tzanetakis achieves a 61% of accuracy over 10 genres in [5]. Kosina obtains an accuracy of 88% for the 3 genres with only 3 seconds audio excerpts in [6]. Li show classification rates of 79% over 10 genres in [7]. McKay achieves an accuracy of 42% over 38 musical genres in [8]. Finally, Pampalk gets an accuracy of 84% over 6 genres in the MIREX Contest as reported in [9].

Most of the proposed algorithms in literature usually rely on timbre and rhythmic features that do not cover the whole range of musical facets, nor the whole range of conceptual abstractness that seem to be used when humans perform this task.

Contrastingly, research on how humans categorize music genres is still in its infancy. There is no computational model that explains how musical features are attended, selected and weighted in order to yield genre decision. We also lack of a model that explains how new categories are created and integrated into our musical knowledge structures. There are some studies which try to study the musical organization in humans and how different facets of music affect genre classification [10] [11] [12] [13].

The aim of our work is to improve our knowledge about the importance of different musical facets and features on genre decisions. We present a series of listening experiments where audio has been altered in order to preserve some properties of music (rhythm, timbre) but at the same time degrading other ones. It was expected that genres with a characteristic timbre provide good classification results when users deal with rhythm modified audio excerpts, and vice versa. We also want to study whether the different levels of distortion affect the classification o not.

Results of listening experiments will be compared with the output of some “state of the art” machine learning classifiers with a similar structure than the listening experiment. Conclusions of this comparison should be the starting point to set up a new generation of automatic genre classifiers based on musical (and other) aspects that define musical genre.

The paper is organized as follows: In Sec. 2 the listening experiments, results and related conclusions are presented. The design of state-of-the-art automatic classification, results and local conclusions are shown in Sec. 3. Finally, overall conclusions and future work are discussed in Sec. 4.

2. LISTENING EXPERIMENTS

2.1. Database and Ontology

Our experiment uses music from 6 genres (Alternative, Classical, Electronic, Jazz, Pop, Rock) taken from the database proposed by Rentfrow et Al. in [14]. This database, also called STOMP (Short Test of Music Preferences) is made up of 14 music genres (alternative, blues, classical, country, electronica/dance, folk, heavy metal, rap/hiphop, jazz, pop, religious, rock, soul/funk, soundtrack) according to musicological criteria (a group of experts were asked), commercial criteria (taxonomies in online music stores were consulted) and also the familiarity of participants with the proposed genres. A list of 10 songs for each of these genres is proposed, assuming that they are clear prototypes for each one of the genres. In our experiment, we discarded some of these genres to avoid possible confusions to participants due to several reasons (e.g. religious). Furthermore, we intentionally mix genres with widely accepted boundaries (classical, jazz, electronic) with some other with more debatable limits (alternative, pop, rock).

2.2. Data Preparation

We selected 5 seconds-long audio excerpts. According to the main goal of this work, some rhythm and timbre modifications have been applied to the audio, in order to create excerpts where the timbre or rhythm information of the music has been somehow degraded.

On one hand, rhythmic modifications are designed to preserve timbre avoiding the participant to extract any temporal information from the audio excerpt. This modification is based on the segmentation of the original audio into short frames. These frames are shuffled to create a new audio segment with the same length as the original, but preserving the average global spectral envelope (i.e., the timbre information). The length of the segmented frames varied among 3 values: 125ms, 250ms and 500ms. It was expected that genres with a particular or personal timbre yielded good classification results when users deal with these audio excerpts. We also wanted to study whether the different

		Timbre															Rhythm															None				
		1/3rd. Oct.					1/3rd. Oct.					1/3rd. Oct.					125ms					250ms					500ms									
Block	Presenting	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e
Alternative	Alternative	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Classic	1					1					1					1					1					1					1				
	Electronic		1					1					1					1					1					1					1			
	Jazz			1					1					1					1					1					1					1		
	Pop				1					1					1					1					1					1					1	
	Rock					1					1					1					1					1					1					1
Classic	Alternative					1					1					1					1					1					1					1
	Classic	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Electronic	1					1					1					1					1					1					1				
	Jazz		1					1					1					1					1					1					1			
	Pop			1					1					1					1					1					1					1		
	Rock				1					1					1					1					1					1					1	
Electronic	Alternative				1					1					1					1					1					1					1	
	Classic					1					1					1					1					1					1					1
	Electronic	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Jazz	1					1					1					1					1					1					1				
	Pop		1					1					1					1					1					1					1			
	Rock			1					1					1					1					1					1					1		
Jazz	Alternative			1					1					1					1					1					1					1		
	Classic				1					1					1					1					1					1					1	
	Electronic					1					1					1					1					1					1					1
	Jazz	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Pop	1					1					1					1					1					1					1				
	Rock		1					1					1					1					1					1					1			
Pop	Alternative		1					1					1					1					1					1					1			
	Classic			1					1					1					1					1					1					1		
	Electronic				1					1					1					1					1					1					1	
	Jazz					1					1					1					1					1					1					1
	Pop	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Rock	1					1					1					1					1					1					1				
Rock	Alternative	1					1					1					1					1					1					1				
	Classic		1					1					1					1					1					1					1			
	Electronic			1					1					1					1					1					1					1		
	Jazz				1					1					1					1					1					1					1	
	Pop					1					1					1					1					1					1					1
	Rock	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 1: Details of the presented audio excerpts to the participants: The experiment is divided in 6 blocks (corresponding to 6 musical genres). A total of 70 audio excerpts are presented in each block. 35 excerpts belong to the musical genre that defines the block. 15 excerpts have timbre distortion (3 different levels), 15 excerpts have rhythmic distortion (3 different levels) and 5 excerpts do not have any musical distortion.

levels of distortion affect the classification o not. On the other hand, timbre modifications are designed to preserve rhythm while avoiding the participant to easily extract any timbre information from the audio excerpt. This modification is based on the filtering of the input signal into frequency bands. The energy for each the log-scale band is used to modulate Gaussian noise centered in that specific frequency band. The energies are computed for each frame, then, this process is similar to basic vocoding. Three different filter banks have been applied (3rd octave, 6th octave and 12th octave) to study the discrimination power affected by this parameter in the classification results (See http://www.iua.upf.edu/~eguaus/aes121 for mp3 examples). It was expected that genres with a particular or personal rhythm would yield good classification results when users had to deal with these audio excerpts.

In summary, we will use excerpts from 6 different genres, sometimes “distorted” with either timbre or rhythm alterations, and in some cases “clean” (i.e., with no alteration). We have 3 levels for each modification (125ms, 250ms or 500ms for the rhythmic modification; 3rd. octave band, 6th. octave band or 12th. octave band for timbre modification) and no further option for the unmodified condition. The task presented to the subjects is a dichotomic decision (yes/no) task where a genre label is presented in the screen, a 5 seconds excerpt is played and they have to decide if it belongs to the genre which label was presented in the screen. In order to keep balanced the proportion of answers, half of the excerpts belonged to the targeted genre and a half of "fillers" was used, according to the schema depicted in Table 1, which provides an overview of the trials, events and blocks that were used.

2.3. Participants

In the experiment participated 42 music students from the High Music School of Catalonia (ESMUC - www.esmuc.net), 27 males and 15 females. The age of participants was between 18 and 43 years old (Mean=25.43; Standard Dev=5.64). All of them were students of the first two years in different areas: Early Music (6 students: 14.3%), Classical (25 students: 59.5%) and Jazz (11 students: 26.2%).

Students spent an average of 2.2 hours (Standard Dev=1.4) everyday listening to music. The associated activities in this period of time are usually traveling or doing homework. Rehearsals and instrument training are excluded from these statistics.

Participants were also asked to define the familiarization degree from 1 (I've never heard about this kind of music) to 5 (I'm an expert in this type of music) for all the selected genres: Alternative (Mean=2.16; Standard Dev=1.09), Classical (Mean=3.83; Standard Dev=1.12), Electronic (Mean=2.20, Standard Dev=0.93), Jazz (Mean=3.47; Standard Dev=0.94), Pop (Mean=3.03, Standard Dev=0.85) and Rock (Mean=3.04; Standard Dev=0.96).

We were also interested in detecting how participants classified their CD collection. The proposed options were Geographical Criterion (0.0%), Cover Color (0.0%), Alphabetical Order (22.6%), Genre (47.2%), Chronologic (13.2%) and Other (17%) which includes "No Order", "Recently Bought" or Favorites on the top". This test shows how important are genre labels in classification process for CD collections.

Finally, we also propose to repeat the experience proposed by Uitdenbogerd in [12]. Participants have been asked to categorize music into exactly 7 categories. Results are shown in Figure 1. Genres only proposed once are not shown in this table (Ambient, Bossa-Nova, Songwriter, Flamenco, etc.). A note of

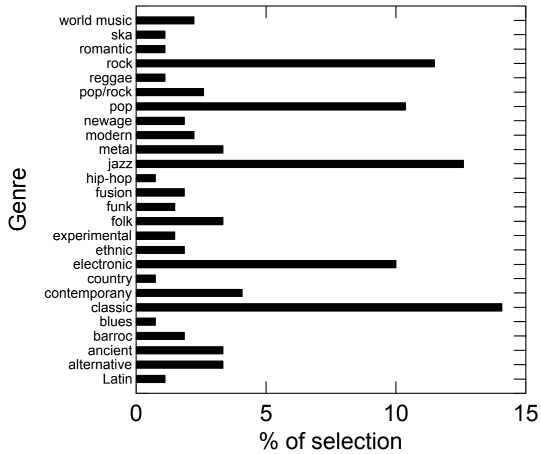

Figure 1 : Percentage of proposed musical genres by participants when they were asked to categorize music into exactly 7 categories

caution has to be raised here as this question was asked after the instructions for the experiment had been presented and therefore the proposed genres might be influenced by those used in it.

2.4. Procedure

The instructions of the experiment informed about its goals and its general structure (one block per genre, expected binary responses "Yes" or "No" for all the audio excerpts, time information relevant but not crucial, etc.). Then, a training block was presented, divided into three parts: 1) Participants were asked to familiarize with the response keys used during the entire test. No distinction between left-handed and right-handed people was applied. 2) Participants were invited to listen to some audio excerpts for each one of the musical genre in order to adjust the genre boundaries. 3) Participants were invited to listen to some rhythmic and timbre modifications of the original excerpts in order to familiarize with the modifications used. At this point, the survey began and the first block (Alternative) started until the last one (Rock) finished. Participants could only relax for a short period between blocks. The presentation of different audio excerpts inside a block was randomized according to the Table 1. The overall required time for completing the experiment was about 30 minutes.

2.5. Results

Figure 2 shows the percentage of correct classified instances for different genres and modifications and Figure 3 shows the corresponding averaged response times. Table 2 show numerical results.

Observing genres individually, alternative music yields similar results for all kind of distortions. The number of correct classifications is a slightly higher for rhythm distortion as well as the response time is slightly lower. The most clear pattern is that the timbre distortion in all cases yields near-chance responses (e.g. around 50%).

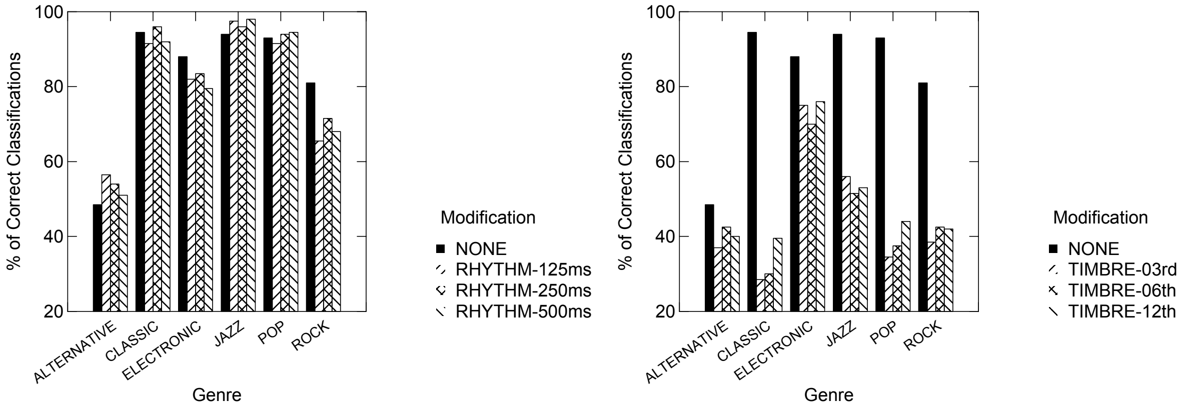

Figure 2: Percentage of correct classifications for different genres and modifications

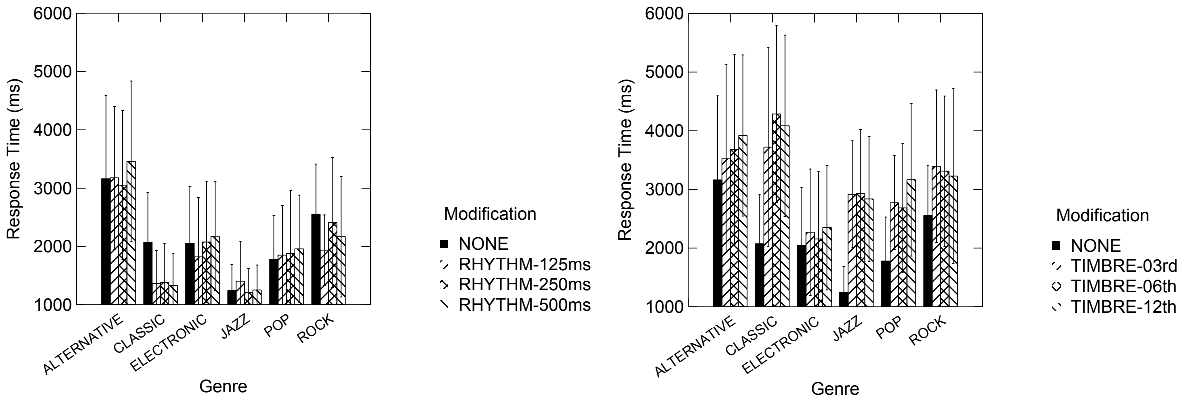

Figure 3: Mean and Standard Deviation of the response time for different genres and modifications

		Timbre			Rhythm			Orig
		1/3rd. Oct	1/6th. Oct	1/12th. Oct	125ms	250ms	500ms	
Alternative	Hits (max=5)	1.85	2.125	2	2.825	2.7	2.55	2.425
	Resp. Time (ms)	3522.4	3682.3	3918.1	3177.2	3049.7	3458.6	3161.0
Classic	Hits (max=5)	1.425	1.5	1.975	4.575	4.8	4.6	4.725
	Resp. Time (ms)	3722.1	4284.6	4083.9	1362.3	1385.7	1328.8	2073.0
Electronica	Hits (max=5)	3.75	3.5	3.8	4.1	4.175	3.975	4.4
	Resp. Time (ms)	2270.0	2156.1	2349.5	1822.5	2077.4	2176.7	2050.6
Jazz	Hits (max=5)	2.8	2.575	2.65	4.875	4.8	4.9	4.7
	Resp. Time (ms)	2915.3	2930.7	2838.3	1404.6	1206.5	1253.4	1242.8
Pop	Hits (max=5)	1.725	1.875	2.2	4.575	4.7	4.725	4.65
	Resp. Time (ms)	2773.2	2684.5	3164.4	1851.4	1881.6	1958.5	1780.4
Rock	Hits (max=5)	1.925	2.125	2.1	3.275	3.575	3.4	4.05
	Resp. Time (ms)	3394.3	3311.3	3228.2	1936.1	2411.1	2167.7	2553.5

Table 2: Numerical results for listening experiments

This could mean that the amount of distortion we created was too radical and that the resulting music did not preserve a minimum of information to identify most of the genres on basis of timbre, not even using the lowest degree of distortion (the only exception would be for “electronic”).

Looking at the specific genres, classical and jazz music show good classification results and low response times for rhythmic distortion, but the opposite tendency is observed for timbre distortion. A possible interpretation for this is that these two musical genres are clearly defined by particular timbres. In contrast, electronic music is the only one that presents good classification results and low response times with timbre distortion. This musical genre is equally defined by rhythm and timbre because of the similar results with both types of distortions. Pop music also presents better results under timbre identification and, finally, rock music is in between pop and alternative. The conclusion is that, according to the selected taxonomy, “alternative” music has no clear difference in rhythm or timbre with pop and rock. Maybe “alternative” music is an artificial genre without musical fundament, or maybe the difference is in another musical component like the harmony or the lyrics.

	# Hits				Response Time			
	Modification		**Degree**		**Modification**		**Degree**	
	F	p	F	p	F	p	F	p
Alternative	6.567	**0.014**	0.508	0.604	4.274	0.053	4.831	**0.014**
Classic	169.319	**0.000**	4.115	**0.020**	45.349	**0.000**	2.216	0.127
Electronic	5.124	**0.029**	0.451	0.639	4.532	**0.040**	3.067	0.053
Jazz	121.268	**0.000**	1.450	0.241	190.42	**0.000**	0.506	0.605
Pop	156.465	**0.000**	4.166	**0.019**	37.805	**0.000**	1.028	0.366
Rock	21.644	**0.000**	3.243	**0.044**	28.046	**0.000**	1.599	0.213

Table 3: Results of the ANOVA test for distortion analysis

	# Hits		**Response Time**	
	F	p	F	p
1/3rd. Octave	20.718	**0.000**	4.723	**0.001**
1/6th. Octave	11.412	**0.000**	7.023	**0.008**
1/12th. Octave	13.356	**0.000**	6.585	**0.008**
125ms	33.394	**0.000**	27.903	**0.000**
250ms	38.516	**0.000**	18.551	**0.000**
500ms	42.236	**0.000**	32.371	**0.000**

Table 4: Results of the ANOVA test for Genre analysis

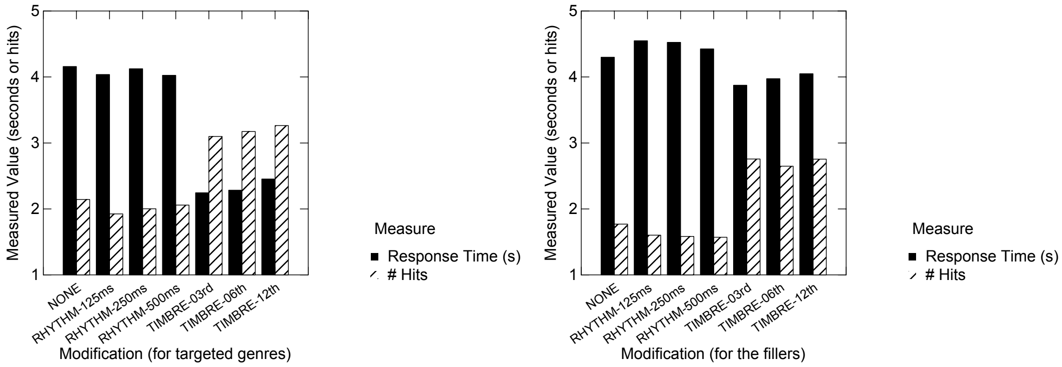

Figure 4: Results for genre identification as a function of the presented distortion when (a) the presented audio excerpt is the one targeted in the block (left panel), and (b) the presented audio excerpt belongs to the non-targeted genres (i.e., presentation of “fillers”) (right panel).

One-way Analysis of variance (ANOVA) is used to test the null-hypothesis within each genre block, assuming that the sampled population is normally distributed. ANOVA is computed on Response Time and on Number of Hits. First, we test whether the distortion degree for both modifications had real influence in classification results. The null-hypothesis is defined as follows:

H_0= Presented distortions, in one genre, do not influence classification results

Results show that, concerning modifications, we have to reject the null-hypothesis and accept that different modifications do affect classification results. As shown in Table 3, the distortion degree in the Number of Correct Classifications shows a significant effect on Classical, Pop, and Rock, while the distortion degree for the Response Time shows a significant effect on Classical, Electronic, Jazz, Pop and Rock ($p < 0.05$). This different pattern can be associated to the high variability of responses in those cases where classification is not clear, i.e. timbre distortions in classical music or confusions between alternative and rock. Whatever the result, the null-hypothesis can not be rejected for this case. Now, we test whether the distortions have the same influence in different genres. The null-hypothesis is defined as:

H_0= Presented genres are equally affected for each specific distortion

Results in Table 4 show how the null-hypothesis can be rejected with a high level of confidence.

Finally, results for overall detection independent of the genre are shown in Figure 4. Roughly speaking, rhythm modifications provide better classification results and lower response times than timbre ones. Furthermore, it is easier to generate a negative response ("it does not belong to the target genre") than a positive one.

Genre	Train	Test
Classical	320	318
Electronic	115	114
Jazz	26	26
Metal	29	29
Pop	6	5
Punk	16	16
Rock	95	96
World	122	123

Table 5: Details for the Magnatune Database

Collecting all the information provided above, we can conclude that, according to the configuration of this experiment, the easiest musical genre classification for humans is to detect when a specific timbre does not belong to classical music, and the more difficult is to detect whether a given rhythm belongs to (again) classical music.

3. AUTOMATIC GENRE CLASSIFICATION

In this section, we will study the behavior of an automatic classification system for musical genre. The study is not focused on the performance by itself. Results will be compared with the results obtained under different conditions in the listening experiments.

3.1. Description

Most of the genre classification studies found in the literature have a similar structure. First of all, the audio database has to be collected and labeled according to a specific taxonomy [15] [3]. Then, some audio descriptors are automatically computed. These features provide certain information related to timbre, rhythm or melody from the original audio data. Other semantic descriptors can be computed [16] but we will focus only on those aspects of music studied in the listening experiments. Sometimes, it is necessary to compute some statistics of these data in order to include temporal information. At this point, the classifier needs to be trained. Although some classifiers use unsupervised methods to classify musical genres [17], most of them use manually annotated labels from the audio database. Finally, the evaluation is normally performed by cross-validation of the training data or by the split of the audio database into two or three subsets (typically 66% for training and 33% for testing or 70% for training, 15% for validation and 15% for testing).

3.1.1. Databases

Two databases have been included in this experiment:

- Magnatune repository (http://www.magnatune.com). This database is used to verify that descriptors and classification schemes we propose are not so far than those used in the Genre Classification contest organized in the context of the International Symposium on Music Information Retrieval - ISMIR 2004 (http://ismir2004.ismir.net). The details of this database are shown in Table 5.

- STOMP Database used in the Listening Experiments and described above.

3.1.2. Descriptors

The audio descriptors that we have used for automatic classification are divided in two main groups: timbre and rhythmic.

- Our timbre description is defined by a compact set of 39 descriptors which include: Zero Crossing Rate (1); Spectral Centroid (1), Spectral Flatness (1), MFCC (12), derivative (12) and acceleration (12). Means and variances of these descriptors are computed for the whole song.

- Our rhythmic description is also defined by a compact set of 39 descriptors which include: Zero Crossing Rate (1), Spectral Centroid (1), Spectral Flatness (1), MFCC (12), derivative (12) and acceleration of data in Rhythm domain (see [18] for more information about these descriptors) Means and variances of these descriptors are computed for the whole song.

All the classification experiments have been made using WEKA [19]. After some initial tests using Support Vector Machines (SMO), Naive Bayes, Nearest Neighbors (kNN) and Decision Trees (J48), the classification algorithm finally used is Support Vector Machine with different exponential parameters (See http://www.iua.upf.es/~eguaus/aes121 for more information).

3.2. Results

Results for the automatic classification experiments are shown in Figure 5. According to the musical aspects discussed in Sec. 2, the used descriptors are grouped in timbre, rhythm and both.

Results are shown independently for these three configurations. Although different train-set and test-set were provided for the Magnatune database, 10-fold cross validation method has been used for all the cases, and then, results are more comparable. Results for Magnatune database show accuracies up to 80% in classification using both timbre and rhythm descriptors. Roughly speaking, these results are comparable to the results obtained by Pampalk in the 2004 genre evaluation contest (see http://ismir2004.ismir.net/genre_contest/results.htm).

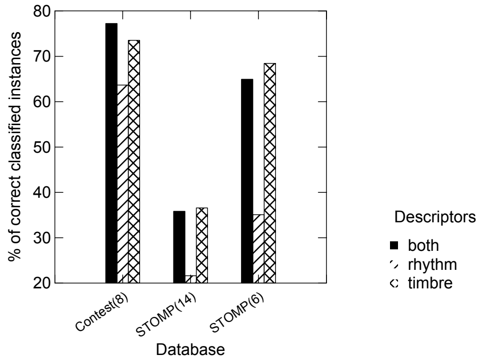

Figure 5: Comparison of automatic classification results for three databases (Magnatune, STOMP and reduced STOMP used in Listening Experiments) for three different sets of descriptors

MAGNATUNE								
classified as --->	a	b	c	d	e	f	g	h
(a) classical	311	0	0	0	0	0	0	9
(b) electronic	3	87	0	0	0	0	15	10
(c) jazz	2	0	12	0	0	0	6	6
(d) metal	0	2	0	6	0	0	21	0
(e) pop	0	1	0	0	0	0	5	0
(f) punk	0	0	0	0	0	7	9	0
(g) rock	8	12	0	2	0	0	67	6
(h) world	28	11	0	0	0	0	10	73

Table 6: Confusion matrix for the automatic classification of the Magnatune database

RENTFROW														
classified as --->	a	b	c	d	e	f	g	h	i	j	k	l	m	n
(a) alternative	5	0	0	1	0	0	0	0	0	1	0	1	2	0
(b) blues	1	1	0	0	0	4	0	0	0	1	0	0	0	2
(c) classical	0	1	8	0	0	0	0	0	0	0	0	0	0	0
(d) country	2	1	0	2	0	2	1	0	0	0	1	0	1	0
(e) electronica	1	1	0	0	3	0	2	1	1	0	0	1	0	0
(f) folk	0	5	0	2	0	1	0	0	0	0	0	0	0	2
(g) funk	0	0	0	0	1	1	4	1	0	0	0	2	1	0
(h) heavymetal	2	0	0	0	0	0	0	4	2	0	0	0	2	0
(i) hip-hop	1	0	0	1	0	0	0	0	5	0	0	1	0	2
(j) jazz	1	2	0	0	0	2	0	0	1	4	0	0	0	0
(k) pop	0	0	0	0	0	0	1	1	1	0	3	1	0	2
(l) religious	1	1	0	2	0	0	0	0	0	0	1	3	0	0
(m) rock	3	0	0	1	0	0	1	1	0	0	0	1	2	0
(n) soul	0	2	0	1	0	1	0	0	3	0	0	0	0	3

Table 7: Confusion matrix for the automatic classification of the STOMP database

RENTFROW (Reduced)						
classified as --->	a	b	c	d	e	f
(a) alternative	6	0	0	1	1	2
(b) classical	0	9	0	0	0	0
(c) electronica	1	0	6	2	1	0
(d) jazz	1	1	1	6	1	0
(e) pop	1	0	1	0	5	2
(f) rock	1	0	1	0	2	5

Table 8: Confusion matrix for automatic classification of reduced STOMP database

Timbre related descriptors provide more classification power than rhythm ones with differences about 15% but the inclusion of rhythm information improves results in all cases.

Applying the same classification conditions to the STOMP database, accuracies decrease because of the high number of musical genres in the taxonomy (14) and the size of the database (10 songs/genre). But the pattern of results depending on the timbre and rhythm facets is similar to that obtained using the Magnatune database.

Finally, accuracies near 70% are obtained with the reduced version of STOMP. Here again, the contribution of timbre is more important than rhythm in the classification process. Furthermore, timbre classification can yield better results than those obtained with both timbre and rhythm descriptors.

Tables 6, 7 and 8 show confusion matrices for the classifications done using the three databases and including timbre and rhythmic descriptors. Note how classical music is correctly classified, pop and rock have some kind of confusion between them, and jazz, electronic and alternative music are worse classified. We will compare these results with the results of the listening experiments in the next section.

Whatever the absolute value is, this value depends on the selected genres to classify (not only the number of genres). In Table 9 there are some classification results using a simple k-NN classifier for two different possible selections of genres from 2 to 14.

4. CONCLUSIONS

A listening experiment for musical genre classification has been presented. Results for a state-of-the-art automatic genre classification algorithm are also shown. Assuming that both experiments are not identical, results are quite similar in such a way that timbre features of music provide more genre discrimination power than rhythm. Even so, genres like electronica require some rhythmic information for better

Genres	#	%
alternative blues	2	84.21
alternative blues classical	3	78.57
alternative blues classical country	4	50.00
alternative blues classical country electronica	5	61.70
alternative blues classical country electronica folk	6	47.36
alternative blues classical country electronica folk funk	7	47.76
alternative blues classical country electronica folk funk heavymetal	8	42.85
alternative blues classical country electronica folk funk heavymetal hiphop	9	45.97
alternative blues classical country electronica folk funk heavymetal hiphop jazz	10	43.29
alternative blues classical country electronica folk funk heavymetal hiphop jazz pop	11	40.56
alternative blues classical country electronica folk funk heavymetal hiphop jazz pop religious	12	35.96
alternative blues classical country electronica folk funk heavymetal hiphop jazz pop religious rock	13	30.89
alternative blues classical country electronica folk funk heavymetal hiphop jazz pop religious rock soul	14	29.32
soul rock	2	84.21
soul rock religious	3	66.66
soul rock religious pop	4	33.33
soul rock religious pop jazz	5	41.30
soul rock religious pop jazz hiphop	6	44.64
soul rock religious pop jazz hiphop heavymetal	7	39.39
soul rock religious pop jazz hiphop heavymetal funk	8	34.21
soul rock religious pop jazz hiphop heavymetal funk folk	9	32.55
soul rock religious pop jazz hiphop heavymetal funk folk electronica	10	34.73
soul rock religious pop jazz hiphop heavymetal funk folk electronica country	11	28.57
soul rock religious pop jazz hiphop heavymetal funk folk electronica country classical	12	32.45
soul rock religious pop jazz hiphop heavymetal funk folk electronica country classical blues	13	30.08
soul rock religious pop jazz hiphop heavymetal funk folk electronica country classical blues alternative	14	29.32

Table 9: Comparison of classification results for different groups of genres

discrimination results. Furthermore, for humans, it seems easier to identify music that does not belong to a given genre than identifying that it belongs to a specific genre. These results show that automatic classification could be based on expert systems for a specific genre instead of global systems. The selected taxonomy also affects directly to classification results, as shown in the listening experiments tests: confusions between alternative and rock music appear as well as for an automatic classifier when different subgroups of taxonomies provide different results in genre classification.

For the listening experiments, the two proposed distortions provide differences in the classification results depending on the musical genre. Results show how the distortion degree has not a direct relationship with the discrimination power. The response time for non distortion audio excerpts can be a measure of how assimilated and musically defined the musical genres are. Alternative music provides response times higher than 3 seconds while classical or jazz music provide response times between 1.5 and 2 seconds. Maybe "Alternative" label was created under some commercial criteria while the jazz music can be defined exclusively by musical properties.

Results of this listening experiment need to be extended to non musician participants. The inclusion of other facets of music like harmony or tonal information as well as other high level semantic descriptors is also crucial for a full characterization of musical genre discrimination in humans. This will help to design new musical genre classifiers that need to deal with new and previously unseen music styles and could be included as a part of music recommenders.

5. ACKNOWLEDGEMENTS

This work was partially funded by the EU-FP6.IST-507142 SIMAC project and also supported by the High Music School Music of Catalonia (ESMUC).

6. REFERENCES

[1] Kemp, C; "*Towards a Holistic Interpretation of Musical Genre Classification"*; Graduate Dissertation, University of Jyväskylä; 2004

[2] Lopez Cano, R; "*Favor de no tocar el género. Géneros, estilo y competencia en la semiótica musical cognitiva actual*"; Actas del VII Congreso de la Sib; 2002

[3] Pachet, F.; "*A Taxonomy of Musical Genres*"; Proc. RIAO; 2000.

[4] Soltau,H., Schultz,T., Westphal, M. and Waibel, A; "*Recognition of music types*"; Proc. ICASSP; 1998

[5] Tzanetakis, G., Essl, G., and Cook, P.: "*Automatic Musical Genre Classification of Audio Signals*"; Proc. ISMIR; 2001.

[6] Kosina, K; "*Music Genre Recognition*"; Fachhochschule Hagenberg, Medientechnik und -design; 2002

[7] Li, T., Ogihara, M., and Li, Q.; "A comparative study on content-based music genre classification"; Proc. SIGIR; 2003.

[8] McKay, C. and Fujinaga, I.; "*Automatic genre classification using large high-level musical feature sets*". Proc. ISMIR; 2004.

[9] Pampalk, E; "*Computational Models of Music Similarity and their Application to Music Information Retrieval*"; PhD Thesis: 2006

[10] Perrot, D. and Gjerdingen, R. O.; "*Scanning the dial: An exploration of factors in the identification of musical style*". Proc. ICMPC; 1999

[11] Heittola, T.; "*Automatic Classification of Music Signals*"; Master of Science Thesis; 2003

[12] Uitdenbogerd, A.; "*A study of automatic music classification methodology*"; Proc. ISMIR 2004

[13] Hannah, W.P.; "*Automated Music Genre Classification Based on Analyses of Web-Based Documents and Listeners' Organizational Schemes*"; Master of Science dissertation; 2005.

[14] Rentfrow, P.J. and Gosling, S.D.; "*The Do Re Mi's of Everyday Life: The Structure and Personality Correlates of Music Preferences*"; Journal of Personality and Social Psychology, Vol. 84, No. 6; 2003.

[15] Aucouturier, J.J and Pachet, F.; "*Representing Musical Genre: A State of the Art*"; Journal of New Music Research, 32(1), 2003.

[16] Herrera, P. Bello, J. Widmer, G. Sandler, M. Celma, O. Vignoli, F. Pampalk, E. Cano, P. Pauws and S. Serra, X; "*SIMAC: Semantic interaction with music audio contents*"; Proc European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies; 2005

[17] Shao, X., Xu, C. and Kankanhalli, M.; "*Unsupervised classification of musical genre using hidden Markov model*"; Proc. ICME; 2004

[18] Guaus, E. Herrera, P; "*The Rhythm Transform: Towards A Generic Rhythm Description*"; Proc. ICMC; 2005

[19] Witten, I.H. and Frank, E.; "*Data Mining: Practical machine learning tools and techniques*"; Morgan Kaufmann, San Francisco, 2005.