MUSIC AND GEOGRAPHY: CONTENT DESCRIPTION OF MUSICAL AUDIO FROM DIFFERENT PARTS OF THE WORLD

Emilia Gómez, Martín Haro, Perfecto Herrera
Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain
{emilia.gomez,martin.haro,perfecto.herrera}@upf.edu

ABSTRACT

This paper analyses how audio features related to different musical facets can be useful for the comparative analysis and classification of music from diverse parts of the world. The music collection under study gathers around 6,000 pieces, including traditional music from different geographical zones and countries, as well as a varied set of Western musical styles. We achieve promising results when trying to automatically distinguish music from Western and non-Western traditions. A 86.68% of accuracy is obtained using only 23 audio features, which are representative of distinct musical facets (timbre, tonality, rhythm), indicating their complementarity for music description. We also analyze the relative performance of the different facets and the capability of various descriptors to identify certain types of music. We finally present some results on the relationship between geographical location and musical features in terms of extracted descriptors. All the reported outcomes demonstrate that automatic description of audio signals together with data mining techniques provide means to characterize huge music collections from different traditions, complementing ethnomusicological manual analysis and providing a link between music and geography.

1. INTRODUCTION

Most of existing Music Information Retrieval (MIR) technologies and systems focus on mainstream popular music from the so-called "Western tradition". The term *Western* is generally employed to denote most of the cultures of European origin and most of their descendants. The unavailability of scores for most musical traditions makes necessary to work with audio recordings, and some recent works have studied if the available techniques and descriptors for audio content description are suitable when analyzing music from different traditions [12].

We provide in [3] an initial contribution in this direction, with the goal of analyzing the descriptive power of tonal features to discriminate Western vs non-Western music material. These tonal features are derived from chroma representations, computed using an interval resolution of 10 bins per semitone and representative of the employed tuning system and gamut. We found that tonal descriptors were able to distinguish these two classes with an 80% accuracy using different classifiers and an independent set for testing. The music collection was made of 1,500 pieces from different areas of the world. In a similar way, Liu et al. have recently performed a study on the classification, by means of Support Vector Machines, of a music collection of 1,300 pieces containing Western classical music, Chinese and Japanese traditional music, Indian classical music and Arabic and African folk music [7]. The best result (84.06%) was obtained using timbre features, and the results for standard chroma features was very low. This might indicate that one semitone resolution is not accurate enough to represent non-equal tempered scales and gamuts found in various cultures.

The goals of this paper can be summarized as follows: first, to analyze the contribution of the different facets of music description (timbre, rhythm, tonality) for the automatic classification of Western vs non-Western music; second, to evaluate the validity of the different features to characterize certain types of music; and third, to investigate the relationship between extracted descriptors and geographical location of the analyzed pieces (latitude and longitude). In order to do that, we have gathered a music collection covering traditional music from different geographical zones and countries as well as a varied set of Western musical styles. Up to our knowledge, the relationship between geography and extracted descriptors has not been addressed in any previous existing piece of literature, and the present study provides an attempt in this direction.

2. METHODOLOGY

2.1 Music collection

For this study, we gathered a music collection comprising 5,905 pieces from different musical traditions and styles. They were manually divided into Western and non-Western categories and labelled according to the musical genre and geographical location (area and country).

For non-Western music, we gathered a total of 3,185 audio recordings distributed by geographical region, as

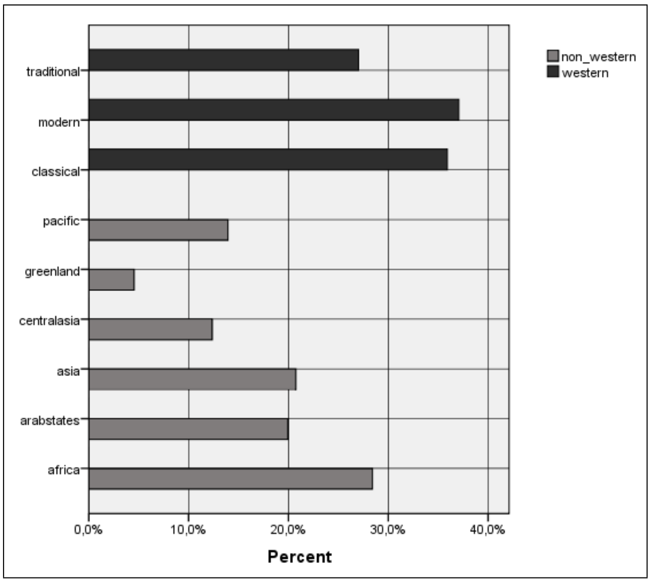

Figure 1. Distribution of the music collection.

defined by UNESCO [1] . They were distributed among the different countries and labelled according to the country of origin and geographical region. We defined the categories *Pacific, Greenland, Central Asia, Asia, Arab States* and *Africa*. These samples contain representative recordings of traditional music from different countries, discarding those having some Western influence (e.g. equal-tempered instruments). They were extracted from CD collections used for ethnomusicological studies (field recordings and compilations of traditional music).

We also considered 2,720 recordings from Western music assigned by UNESCO to the region *Europe and North America*. A set of this data was gathered from commercial CDs and is scattered across different musical genres (alternative, blues, classical, country, disco, electronica, folk, funk, hip-hop, jazz, metal, pop, reggae, rock and soul). A subset of the ”Western” collection that was chosen has been widely used within the MIR community [6, 10, 11]. We also added a collection of traditional music from Western countries (Europe and American folk). This data was labelled according to country of origin and musical genre.

Figure 1 shows the class distribution of the music collection under study. For Western music, we have distinguished between three main classes: classical, traditional music and a general class called *modern* that groups the remaining musical genres. For non-Western material, we have grouped the different countries into the mentioned categories. As it can be seen in the figure, classes are not equally distributed. One reason for that is the variability of pieces available to our analysis, which made it very hard to find the same number of excerpts for all the considered countries (e.g. we only found around 10 pieces for countries such as Vanuatu, Oman, Zimbabwe or Tanzania while the number of pieces for traditional music of European countries had to be restricted to 90 excerpts per country). On the other hand, geographical regions differ on the number of countries and musical traditions. For instance, there were few recordings from Greenland compared to the different styles present in Asia (including Indian music for instance). We will minimize the impact that this might have in the classification problem by balancing the distribution of Western vs non-Western material.

For this study we analyzed the first 30 seconds of each musical piece, and we discarded few non representative parts containing silences or ambiguous introductions (the music on these introductions was not related to the overall content of the piece).

2.2 Feature extraction

A main goal of this study is to provide a multi-faceted description of the music collection and compare the relative performance of different musical facets (tonal, timbre and rhythm) for comparative analysis of music from around the world. In order to do that, extracted audio features are related to these different facets:

Tonality: tonal features are related to the pitch class distribution of a piece, its pitch range or tessitura and the employed scale and tuning system. The features in this group include the *tuning frequency*, which estimates the frequency used to tune a musical piece if we consider an equal-tempered scale. This feature is expected to be close to zero for pieces tuned in this temperament. High-resolution pitch class distributions are also obtained as the Harmonic Pitch Class Profile (HPCP), computed with a resolution of 10 bins per semitone and averaged for the analyzed segment. We also obtain a ”transposed” version of the HPCP that we call the THPCP, by ring shifting the HPCP vector according to the position of the maximum value. Some tonal features are then derived from them (*equal-tempered deviation, non-tempered energy ratio* and *diatonic strength*). We finally consider a dissonance measure and a descriptor called *octave centroid*, which is obtained from a multi-octave fundamental frequency representation and corresponds to the geometry centre of the played pitches. We compute this description on a frame basis and then obtain the average and variance for the considered segment. This set of features was used in a previous study [3].

Timbre: we gather here a standard set of timbre features including loudness, spectral flux, spectral flatness, roughness, MFCCs and energy computation in bark bands. These features are computed in a frame basis, and we then obtain statistical measures such as maximum and minimum value, mean and variance. Timbre features are obtained as explained in [9].

Rhythm: in terms of rhythmic features, we consider different attributes such as the estimated global tempo for the analyze excerpt as well as some features obtained from Inter-Onset Interval (IOI) histograms (peak positions and values) and onset rate (number of onsets per second). The algorithm for rhythmic feature computation is based on the system described in [2].

[1] http://portal.unesco.org/geography

Drum: this group is composed by a set of song-level percussion descriptors computed from the output of a transcription system that detects drum kit events (i.e. bass drum, snare drum and hi-hat) [5]. Other instruments sounding like them are probably detected and considered as being them. These song-level descriptors include: the ratio between the number of detected events per instrument and the total number of onsets (e.g. bass drum/total), the ratio between the number of instances among instruments (e.g. bass drum/hi-hat), the number of detected events per minute (e.g. hi-hat/min) and the peak values of the histogram of the inter-instrument intervals.

2.3 Classification algorithms

We have approached several classification methods but, for the sake of summarization, we only present the results obtained for Support Vector Machines (SVM), considered as one of the best-performing learning algorithms currently available. We have employed the data mining software RapidMiner [8] [2] , which implements SVM using LibSVM [1].

We have used a grid search facility available in RapidMiner to find the following optimal values for the kernel function: linear ($u' \cdot v$), polynomial ($(\gamma \cdot u' \cdot v + coef_0)^{degree}$) and radial basis function ($e^{-\gamma \cdot |u-v|^2}$). $coef_0$ has been set to its default value ($coef_0 = 0$) and a grid search has been run to find the optimal values of kernel type, γ, $degree$ and C, which corresponds to the cost parameter that controls the trade off between allowing training errors and forcing rigid margins. A soft margin then permits some misclassifications. Increasing the value of C increases the cost of misclassifying points and forces the creation of a more accurate model that may not generalize well [3] . We have adopted an evaluation procedure based on 10-fold cross-validation over equally-distributed classes.

3. RESULTS

3.1 Distribution of features

In order to have a preliminary idea of the usefulness of the different features for comparative analysis, we provide here some analysis of the feature distributions. Figure 2 shows the distribution of the tuning descriptor for Western and non-Western music. As expected, the distribution of tuning deviation with respect to 440 Hz is centered on 0 cents for Western music and equal distributed between -50 and 50 cents for non-Western pieces. We also find some differences in other tonal descriptors such as equal-tempered deviation, representing the deviation from an equal-tempered scale, which also appears to be lower for Western than for non-Western music (see Figure 3).

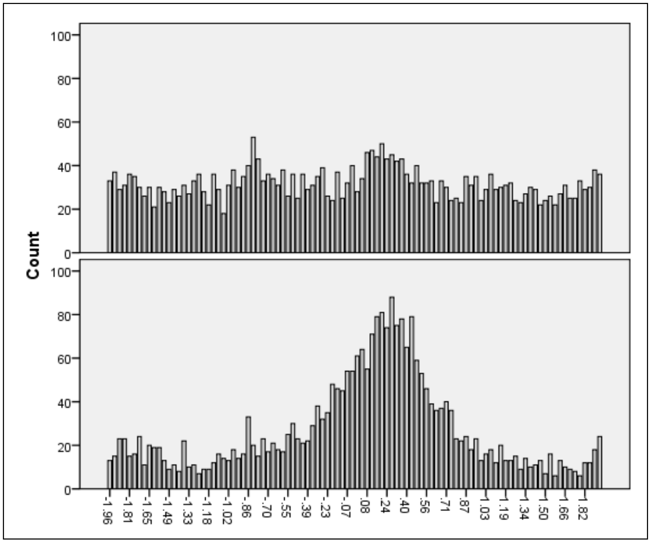

Figure 2. Distribution of tuning frequency (normalized value) for non-Western (top) and Western music (bottom).

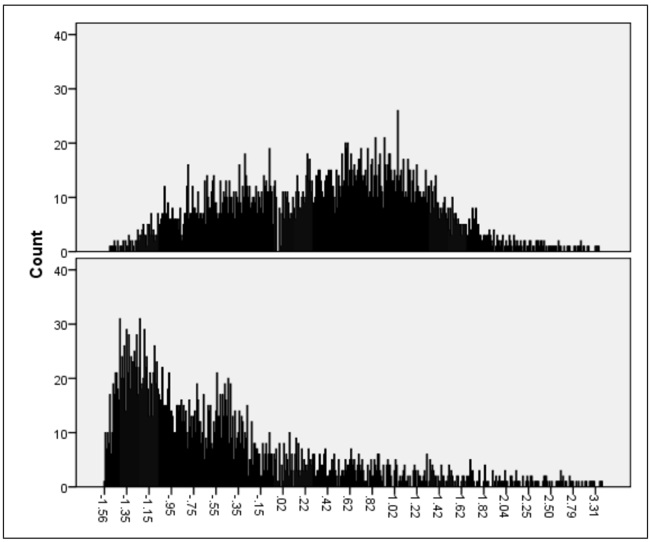

Figure 3. Distribution of equal tempered deviation (normalized value) for non-Western (top) and Western music (bottom).

We can also analyze the feature distribution for the different geographical areas and musical genres. One example is shown in Figure 4, where we represent the distribution of the *drum/total* descriptor, for the different geographical areas and musical styles. This descriptor represents the presence of drum sounds (or other instruments with similar sound) in the analyzed piece. As expected, we observe that the values for this feature are high for the class *modern* (including musical genres such as jazz, pop and rock) and African music (with a significant presence of percussive instruments), and are low for classical music.

3.2 Western vs non-Western classification

Our goal here is to have a classifier that automatically assigns the label ”Western” or ”non-Western” to any audio file that is input and analyzed with the mentioned features. We are aware of the limitations of the concept of Western as opposed to non-Western, as this is a first step towards the definition and formalization of stylistic features proper to different kinds of music.

For the Western vs non-Western categories, the achieved

[2] http://rapid-i.com

[3] http://www.dtreg.com/svm.htm

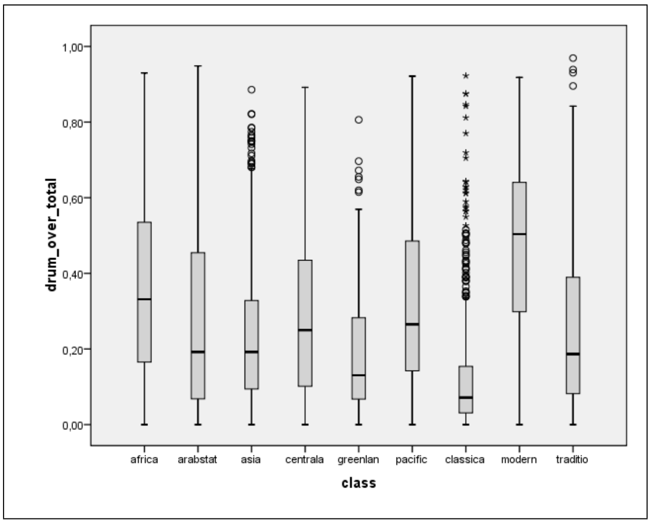

Figure 4. Distribution of the *drum/total* descriptor for the different geographical areas and musical genres.

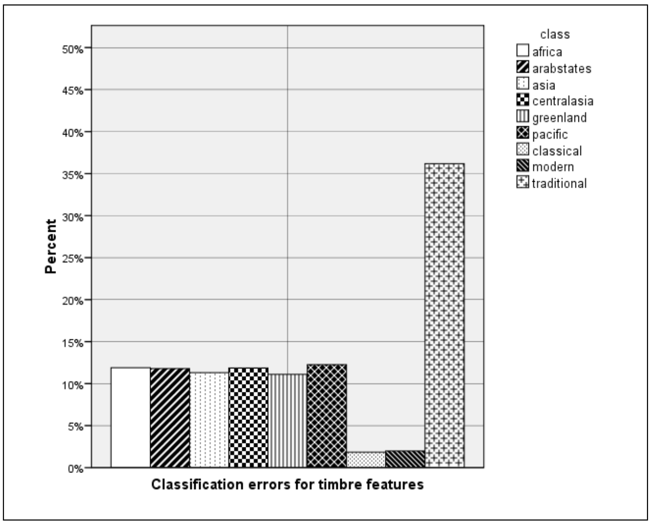

Figure 5. Classification errors (%) for timbre descriptors.

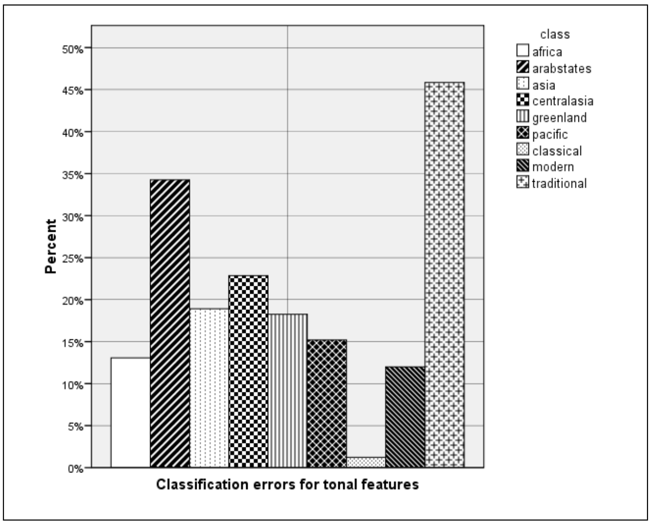

Figure 6. Classification errors (%) for tonal descriptors.

classification results for the different feature sets are summarized in Table 1. The second row indicates the accuracy of timbre features after applying an attribute evaluation method for feature selection, correlation-based feature selection (CFS) [4]. This algorithm selects a near-optimal subset of features that have minimal correlation between them, and maximal correlation with the to-be-predicted classes. This procedure was performed 10 times by means of a 10-fold cross-validation procedure, and only the timbre features that were selected more than 8 times were considered. They include descriptors based on spectral MFCCs, Bark-band energy, spectral flux and roughness.

The last row indicates the accuracy after applying the same feature selection method to the whole feature set. The feature-selection procedure was performed 10 times as well by means of a 10-fold cross-validation procedure, and only the features that were selected 10 times were considered. The selected features include a combination of tonal (tuning frequency, deviations from equal tempered scale and relative intensity of the fourth and fifth degree of a diatonic scale), timbre (features derived from MFCCs, energy in bark bands and spectral flux) and drum features (number of detected hit-hat and bass-drum per minute).

As a general conclusion, we observe that the highest classification accuracy, 88.53%, is obtained using timbre features. Nevertheless, the number of features for this set is very high (176 descriptors). Using a feature selection procedure, the set can be reduced to 25 timbre features with a 83.36% of accuracy. As the non-Western collection contains many field recordings, we think that timbre descriptors may be related to recording quality instead of musical properties of the pieces under study. In this regard, we observe that 81.23% of accuracy is obtained by using 41 tonal features, and, using the feature selection procedure described above, 23 features from the different sets yield to a global accuracy of 86.88%. This descriptor set should be considered robust to recording quality.

We can also see that the performance for rhythmic and drum features is very low, indicating that these descriptor sets are incomplete to discriminate Western from non-Western material. This was expected because Western music includes pieces without drums (e.g. classical and traditional music) and without a steady rhythm. Looking at the F-measure for Western and non-Western classes, we do not find significant differences for timbre, tonal and drum features. Nevertheless, we observe that the value of F-measure for Western music is more than 10% lower than for non-Western music when using rhythmic features. In general, the low performance of rhythmic descriptors may suggest that the implemented features only represent periodicity and tempo of music with a steady rhythm, but can be insufficient to capture more subtle rhythmical aspects of classical and traditional music.

3.3 Classification accuracy for different musical genres and traditions

Figures 5 to 8 present the percentage of classification errors per each class for the different feature sets. We observe

Set	Nb	Kernel function parameters			
		(type, degree, cost, gamma)	Accuracy (%)	F-measure W	F-measure non-W
Timbre	176	polynomial, 3, C=8.87, gamma=0.4	88.53	0.8856	0.8850
Timbre (CFS)	25	polynomial, 3, C=8.87, gamma=0.4	83.36	0.8345	0.8327
Tonal	41	linear, C= 2.14	81.23	0.8152	0.8095
Rhythm	23	linear, C= 2.14	62.02	0.5520	0.6704
Drum	17	radial basis function, C= 0.0, gamma= 0.4	69.83	0.7036	0.6929
Selection (CFS)	23	radial basis function, C= 0.0, gamma= 0.4	86.88	0.8600	0.8765

Table 1. Accuracy using SVM classifier and a grid search procedure.

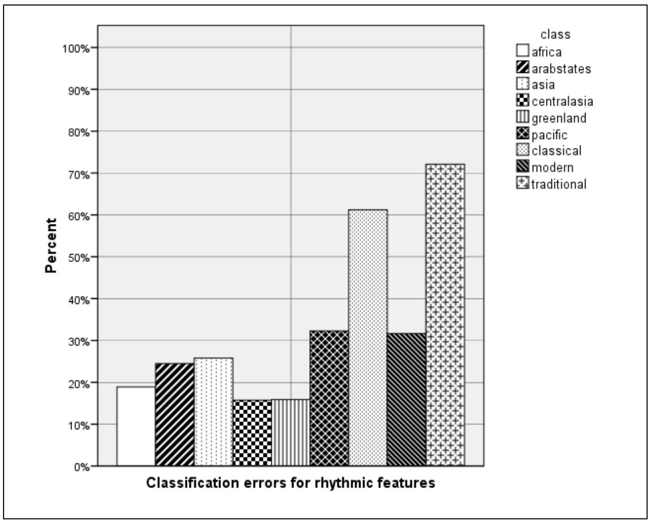

Figure 7. Classification errors (%) for rhythmic descriptors.

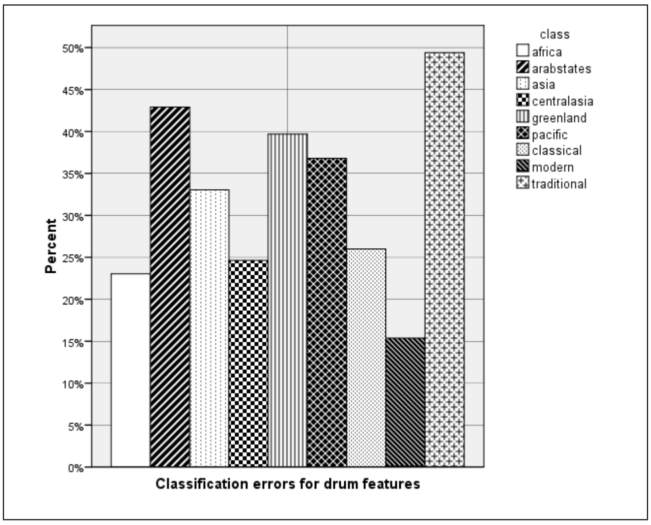

Figure 8. Classification errors (%) for drum descriptors.

that *traditional* music is the most misclassified Western class for all the feature sets. The reason for that may be that traditional pieces are closer to non-Western material with respect, for instance, to instrumentation (e.g. a small number of instruments, similar recording conditions) or tonality (e.g. high degree of ornamentation or scales with deviations from equal tuning).

On the other hand, we observe that more than 60% of the *classical* excerpts are not correctly classified when using rhythmic descriptors, and only 25% when using drum descriptor. We can think that these feature sets may consider Western music as having a constant rhythm and with a high presence of drum sounds, and that classical pieces differ from this assumption. We also observe that *arabic* music is sometimes labeled as Western when using tonal features, as found in [3].

3.4 Geographical location and feature distance

We can also analyze the correspondence between audio features and geographical location by studying the geographical distribution of feature values. In order to do that, we have computed a set of statistics over the audio features for the considered countries. Figure 9 shows an example of the geographical distribution of the equal-tempered deviation feature. We observe that low values are found in Europe, United States and Australia, while higher values are found in African and Asian music.

We have then studied the correlation of audio features averages and the average latitude and longitude coordinates for each country. In the Pearson correlation computation, the Bonferroni method was used to adjust the observed significance level for the fact that multiple comparisons were made. The following average descriptors showed a low (i.e., $0.3 < |r| < 0.45$) but significant correlation with the geographical coordinates:

Latitude: 3rd peak of the Inter-Onset Interval histogram, transposed chroma features (2nd, 3rd, 5th, 7th and 9th equal-tempered positions), chroma features (7th, 10th and 12th equal-tempered positions), equal tempered deviation and non-tempered energy ratio.

Longitude: 4th peak of the Inter-Onset Interval histogram and number of onsets per second.

From the previous list, it is worth to note that latitude is mostly associated to tonal features, while longitude is more associated to rhythmic descriptors. We can also build a regression model for latitude using only the above-mentioned 13 significant descriptors, yielding a correlation value of 0.59. Regarding longitude, the correlation is 0.31. These initial observations need to be carefully re-assessed in the context of theories that might relate geographical and climate variables to constraints on musical instrument construction or on music-centred social activities.

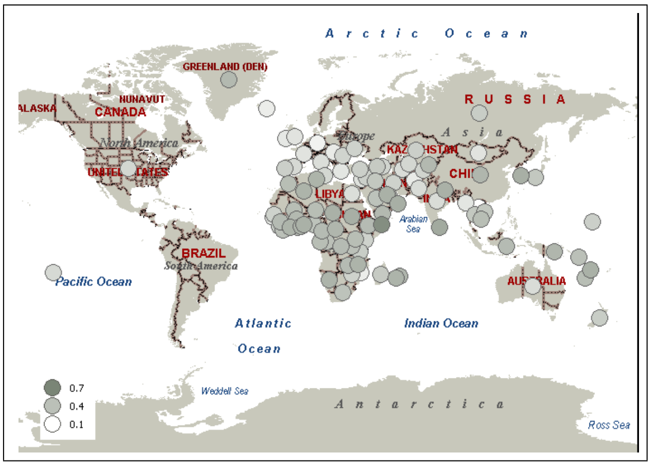

Figure 9. Geographical distribution of the equal tempered deviation descriptor (normalized value).

4. CONCLUSIONS AND FUTURE WORK

We have presented an empirical approach to the comparative analysis of music audio recordings based on tonal, timbre and rhythmic features using a music collection from various parts of the world. We tried to automatically distinguish music from Western and non-Western traditions by means of automatic audio feature extraction and classification. An accuracy of 86.68% was obtained for a music collection of around 6,000 pieces, using only 23 features from different musical facets. This confirms that timbre and rhythmic descriptors complement high-resolution tonal features for the characterization of music from various cultures. Furthermore, each feature set helped to discriminate certain types of music (e.g. drum features were suitable to identify pieces from the *modern* category).

As a future work, we would like to extent this analysis to more specific music collections and complement our current description from an ethnomusicology perspective. In this regard, we will attempt the clustering of the pieces in terms of musical culture. Ideally, we should then be able to define and formalize stylistic features proper to different traditions, and approach genres not just geographically but as a set of traits. We think that this will help to refine our descriptors and similarity measures accordingly. We also plan to complement the current collection with music from the UNESCO region *Latin America and the Caribbean* and explore influences and ”frontier music” with this procedure.

As a final consideration, we conclude that existing MIR techniques are of great interest for the comparative study of all existing music traditions in the world, and audio description tools have a great potential to assist in ethnomusicological research. We hope that the present work contributes to the understanding of our musical heritage by means of computational modeling.

5. ACKNOWLEDGEMENTS

The authors are grateful to Georges Tzanetakis for sharing his genre collection. They also want to thank their colleagues at the Music Technology Group (UPF) and the anonymous reviewers for their helpful comments and suggestions. This research has been partially funded by the EU-IP project PHAROS6 (IST-2006-045035) and the e-Content Plus project VARIAZIONI7 (ECP-2005-CULT-038264).

6. REFERENCES

[1] Chih-Chung, C. and Chih-Jen, L. “LIBSVM : a library for support vector machines”, 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm

[2] Dixon, S. ”Automatic extraction of tempo and beat from expressive performances”. *Journal of New Music Research*, 30, pp. 39-58, 2001.

[3] Gomez, E. and Herrera, P. “Comparative analysis of music recordings from Western and non-Western traditions by automatic tonal feature extraction”, *Journal of Empirical Musicology*, 3(3), pp. 140-156, 2008.

[4] Hall, M.A. “Correlation-based feature selection for discrete and numeric class machine learning”, *7th International Conference on Machine Learning*, pp. 359-366, San Francisco, CA, USA, 2000.

[5] Haro, M. and Herrera, P. “From low-level to song-level percussion descriptors of polyphonic music”, *International Conference on Music Information Retrieval*, Kobe, Japan, 2009.

[6] Holzapfel, A. and Stylianou, Y. “A statistical approach to musical genre classification using non-negative matrix factorization”, *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2, pp. 15-20, Honolulu, Hawai, USA, 2007.

[7] Liu, Y., Xiang, Q., Wang, Y. and Cai, L. “Cultural style based music classification of audio signals” *IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei Taiwan, April 2009.

[8] Mierswa, I.,Wurst, M., Klinkenberg, R., Scholz, M. and Euler, T. “YALE: Rapid prototyping for complex data mining tasks”, *12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (KDD-06), pp. 935-940, Philadelphia, USA, 2006.

[9] Peeters, G. “A large set of audio features for sound description (similarity and classification) in the cuidado project”, Technical report, CUIDADO I.S.T. project, IRCAM, 2004.

[10] Rentfrow, P. J. and Godsling, S. D. “The do re mi’s of everyday life: The structure and personality correlates of music preferences”, *Journal of Personality and Social Psychology*, 84(6), pp. 1236-1256, 2003.

[11] Tzanetakis, G., Essl, G. and Cook, P.. “Automatic musical genre classification of audio signals”, *International Symposium on Music Information Retrieval*, Bloomington, Indiana, USA, 2001.

[12] Tzanetakis, G., Kapur, A., Schloss, W. and Wright, M. “Computational ethnomusicology”, *Journal of Interdisciplinary Music Studies*, 1(2), pp.1-24, 2007.