Artist filtering for non-western music classification

Anna Kruspe Hanna Lukashevich Jakob Abeßer

Fraunhofer IDMT
Ehrenbergstr. 31
Ilmenau, Germany
Correspondence should be addressed to Anna Kruspe (kpe@idmt.fhg.de)

ABSTRACT

The "album effect" is a known phenomenon in musical artist and genre recognition. Classification results are often better when songs from the same album are used in the training and evaluation data sets. Supposedly, this effect is caused by the production conditions of the album, e.g. recording quality, mixing and equalization preferences, effects etc. This behavior does not represent a real world scenario, though, and should therefore be avoided when evaluating the performance of a classifier.
The related "artist effect" also affects the results of genre recognition. It is caused by the appearance of the same artists in the training and evaluation data sets. Artist filters have been proposed previously to remove this influence. We perform three different experiments to characterize the "artist effect" somewhat better. First, we test its influence on the classification of musical pieces into their regions of origin. We then repeat this experiment using only specific sets of features (timbre, rhythm, and tonality). Finally, we perform a finer genre recognition with genres from four different world regions. The influence of the aforementioned effect is evaluated for all experiments.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information search and retrieval; H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing; I.5.4 [**Pattern Recognition**]: Applications; J.5 [**Computer Applications**]: Arts and Humanities

General Terms

Performance, Algorithms

Keywords

Genre recognition, Album effect, Artist filters, Non-Western music

1. INTRODUCTION

The topic of *Music Information Retrieval (MIR)* has experienced a strong development in recent years. The purpose of this field of study is the automatic extraction of information from and about music. One major application scenario are recommendation systems, which offer music to their users that is similar to their known favorites. Retrieval systems, which allow users to find music with specific properties or metadata, are also widely used [5]. The recognition of musical genres such as classical music, rock, jazz, or funk plays an important role here since users tend to group music into these relatively simple categories and expect to find new music using these descriptions [23]. Supposedly, most people also define their taste in music by the genres they listen to. The task of genre recognition is problematic due to a variety of reasons. The poor definition of genre labels and the unclear distinction between them are two of them. One musical piece nowadays hardly ever belongs to a single genre, but posesses a range of influences [17]. Discovering distinct properties of each genre is another unsolved problem.
Two aspects of musical genre classification are further explored in this paper. The first one is the application of MIR techniques to non-western music. This field of application has only started to receive attention in the past few years. The classification of musical pieces into their cultural areas of origin is a new and interesting topic that requires the expansion and broadening of existing technologies. For this purpose, Tzanetakis et al. coined the term *Computational Ethnomusicology* [24].
The second aspect concerns the influence of the data sets on the classification results. The album effect is a known phenomenon for classification tasks. When doing genre classification, a related effect appears additionally: Rather than classifying musical pieces by their genre, the classifier gains its information from the performing artists. In [20], a so-called "artist filter" is applied to the training and evaluation data to mediate this problem. This filter ensures that all the musical pieces by one artist are either in the training or in the evaluation data set.
In this paper, we run experiments to characterize the influence of erroneous artist classification on genre recognition. We do this on a database containing both western and non-western music.

All music classification tasks follow rougly the same procedure. First of all, a music database is compiled. It is then annotated with the labels used for classification afterwards. Features are extracted from the audio data for all pieces to gain a more meaningful representation of their

musical and auditory properties and to remove redundancy and irrelevance. An automatic feature selection algorithm is usually applied to these features to select the most salient ones. Then, a training algorithm generates models for all annotated classes. Unknown musical pieces can be classified based on these models, which on the other hand allows to evaluate the models. In this publication, we follow the described procedure. Details will be explained in the upcoming sections.
The remainder of this paper is organized as follows. In Sec. 2, we give a brief overview over related work. The various experiments and their conditions are described in Sec. 3. We present our results in Sec. 5 and discuss them in Sec. 6. Finally, we draw conclusions and give an outlook on possible future research in Sec. 7.

2. STATE OF THE ART

2.1 Non-western genre recognition

An overview on the classification of non-western music into its regions of origin has been given in [13]. In summary, the results are comparable to those for the classification of western music. However, no standardized data set such as the one created for the *MIREX*[1] contest for western music is widely used. Therefore, the results are not directly comparable. Notable achievements were made in [17], [11], [10], [16], [15], and [22].
In [17], a new classification system is proposed. It breaks the labeling of the musical pieces down into the three domains timbre, rhythm, and tonality. The approach is evaluated for non-western music. The idea of domain-based classification is taken up in this paper as well.

2.2 Artist filtering

The album effect is a known phenomenon in classification tasks. It was first described in [27] and further explored in [3] and [18]. The task performed in all three papers is artist classification. The authors notice strong differences in their classification results depending on the ways their data sets were assembled. If songs from the same album appeared in both the training and the evaluation data sets, the classifier produced good results. These results took a sharp decline when this was not possible. This indicates that the features extracted from the audio data and used by the classifier do not necessarily represent the auditory properties of the artist only. Instead, the distinct properties of each album seem to have a large influence as well. Such properties may be caused by the production quality, preferences of the producer, effects that were in fashion at the time of production, and many more [12].
Therefore, the album effect may cause results that would possibly not hold up in real world applications. Obviously, this is a problem in genre classification as well. Additionally, a similar effect comes into existence with regards to the artists in the data set. Analogous to the album effect, the genre classification result improves when the same artist appears in both the training and the evaluation data set - a sort of "artist effect". In this case, it is possible that the features represent properties of both the artist and the album rather than the genre. Pampalk, Flexer, and Widmer observed this effect in [20] and proposed the use of a so-called "artist filter" to prevent it. This technique is further described in [6].
Depending on the data set, it may be hard to even differentiate between the two effects. We paid no special attention to this fact. We however suspect that the influence of the "artist effect" in the conducted experiments is larger than the album effect, mostly because our training database usually contains more than one album by each artist. In the following, the composition of the data sets is determined only with regards to the artist. Since we used no sampler albums, this composition inherently depends on the album as well, but the experiments do not allow any further analysis of the album effect.

[1] `http://www.music-ir.org/mirex/wiki/MIREX_HOME`, Last visited 05/11/11

3. EXPERIMENTAL SETUP

3.1 Audio material

In [13], we present a large collection of music from all over the world. It was created using a taxonomy based on the "Garland Encyclopedia of World Music" [19] and is comprised of ten large geographic regions which are further subdivided into 86 genres. The music collection itself contains 4398 musical pieces in MP3 format. Their durations range between 30s and 20min. Each genre is represented by at least 50 pieces by at least 5 artists. This criterion enables us to subdivide the collection into five subsets in a way that the artists do not appear in two different sets. This is necessary for the cross-validation used in the presented experiments. An overview of the regions, their number of genres, and the ratio of artists to genres is shown in table 1. In this taxonomy, the "Western" music category contains genres such as Rock, Pop, and Classical music, while the North American and European categories contain genres that have stronger ties to these geographic regions and originated there. Examples are Native American music, Inuit music, and Blues for Northern America and Celtic music, Musette, or Flamenco for European music.

Region	No. of genres	Average no. of artists per genre
Africa	12	6.5
South America	13	6.0
North America	9	5.2
Southeast Asia	3	6.7
South Asia	5	7.8
Middle East	10	5.8
East Asia	4	7.5
Europe	20	6.4
Australia & Oceania	4	5.0
Western Music	5	5.0

Table 1: The ten large regions of the music collection

For our experiments, we used this music database as a whole to quantify the influence of artists on genre classification. We also ran separate experiments on the four largest regional data sets. These are:

- Africa (with the subregions Western Africa, Central Africa, Eastern Africa and Southern Africa)
- South America (with the subregions Latin America and the Caribbean)
- Europe (with the subregions Celtic cultures, Scandinavia, Central Europe, Eastern Europe and Mediterranean Europe) and
- Western music (genres include Classical music, Pop, Rock, Electronic, and Urban music)

3.2 Features

We utilize a broad palette of low-level acoustic features and several mid-level representations [2]. To facilitate an overview, the audio features are subdivided into three categories covering the timbral, rhythmic and tonal aspects of sound.

Timbral features Although the concept of timbre is still not clearly defined with respect to music signals, it has proved to be very useful for automatic music signal classification. To capture timbral information, we use Mel-Frequency Cepstral Coefficients, the Audio Spectrum Centroid, the Spectral Flatness Measure, the Spectral Crest Factor, and the Zero-Crossing Rate. In addition, modulation spectral features [1] are extracted from the aforementioned features to capture their short term dynamics. We applied a cepstral low-pass filtering to the modulation coefficients to reduce their dimensionality and decorrelate them as described in [4].

Rhythmic features All rhythmic features used in the current setup are derived from the energy slope in excerpts of the different frequency-bands of the Audio Spectrum Envelope feature. These comprise the Percussiveness [25] and the Envelope Cross-Correlation (ECC). Further mid-level features [4] are derived from the Auto-Correlation Function (ACF). In the ACF, rhythmic periodicities are emphasized and phase differences are annulled. Thus, we also compute the ACF Cross-Correlation (ACFCC). The difference to ECC again captures useful information about the phase differences between the different rhythmic pulses. In addition, the log-lag ACF and its descriptive statistics are extracted according to [9].

Tonal features Tonality descriptors are computed from a Chromagram based on Enhanced Pitch Class Profiles (EPCP) [14].The EPCP undergoes a statistical tuning estimation and correction to account for tunings deviating from the equal-tempered scale. Pitch-space representations as described in [7] are derived from the Chromagram as mid-level features. Their usefulness for audio description has been shown in [8].

3.3 Feature selection and classifier

Feature selection is often performed before applying training algorithms to the data set. It reduces the amount of data, removes redundant and unnecessary information, and puts more weight on the features that produce better discrimination between the classes.

From a variety of feature selection algorithms, the IRMFSP algorithm was chosen. IRMFSP stands for "Inertia Ratio Maximization Using Feature Space Projection" and this principle was first introduced in [21]. The algorithm consists of two steps that are iterated until a set threshold is reached. These two steps are:

1. The ratios r_i between the between-class inertia for all features and the total-class inertia are calculated. Since high values of r_i indicate a good separation between classes for a feature f_i, the feature with the highest r_i is selected. This guarantees the selection of relevant features.
2. The feature space is projected onto the selected feature. The new feature space is then used for the calculation of the ratio in the next iteration. This step serves to avoid the selection of redundant features.

A gain value calculated for each selected feature after each iteration may serve as a stopping criterion. That is to say, if the feature selected last does not contribute a lot to class separation, the algorithm is stopped. It is assumed that all subsequent features would contribute even less.

For training, the SVM (Support Vector Machine) algorithm has become state-of-the-art in MIR tasks. Therefore, this algorithm was used for training the models and for classifying new songs.

In all classification algorithms, models are trained using a set of training feature vectors and their associated genre labels. The algorithm then tries to find discriminating properties between the sets of feature vectors corresponding to each distinct annotation label, making up the classes. The SVM algorithm virtually transforms the feature vectors into a higher-dimensional space and finds separating hyper-planes between the classes. These hyper-planes are characterized by their support vectors (hence the name) and their offset. New, unknown feature vectors can then be classified using the found configuration [26].

Additionally, the input data is often transformed using a kernel function. We used the RBF (Radial Basis Function) kernel. Both the kernel parameter γ and the SVM's error parameter C are optimized using a grid search.

The basic SVM algorithm can only be used for two-class problems. For multi-class problems such as the classification of non-western music, several one-vs-one trainings are performed and the resulting set of classifiers is used for the classification of new data. This causes a bit of a discrepancy between the feature selection and the classification, since the IRMFSP algorithm handles all classes at once, rather than on a one-vs-one (or one-vs-all) basis.

3.4 Experiments

To quantify the influence of (erroneous) artist classification on genre classification, we ran each of our experiments twice. First, we split up the data set randomly. Therefore, the same artists could appear in the training and the evaluation data sets. For the second part of the experiment, we made sure that this was not possible anymore. In [20], this is called "artist filtering". To guarantee a better generalization, a five-fold cross validation was performed in each experiment. We then conducted three different sets of experiments:

Experiment A In this experiment, we first performed trainings and evaluations on the whole data set using all features.

Experiment B We then repeated this process three times. Each time, we only used features belonging to one of the three domains presented above: Timbre, rhythm, and tonality.

Experiment C All features were used again in this experiment. However, we only used the four subsets of the music database presented above. Our question was whether some genres of music would be more susceptible to the artist effect than others.

4. RESULTS

4.1 Results using the whole music collection and all features (Experiment A)

Our very first experiment was the comparison of the results for the data sets with and without artist filtering using the complete music collection and all features. The genre classification accuracies for each region were summed up to make them easier to interpret. In doing this, a classification result was counted as "correct" whenever a genre from the correct region was chosen. The results are shown in tables 2 (without artist filtering) and 3 (with artist filtering).

	01Africa	02LatinAmerica	03NorthAmerica	04SoutheastAsia	05SouthAsia	06MiddleEast	07EastAsia	08Europe	09AustraliaOceania	10Western
01Africa	**.77**	.08	.02	-	.01	.02	-	.02	.02	.04
02LatinAmerica	.08	**.80**	.02	-	.01	.01	-	.03	.01	.05
03NorthAmerica	.05	.04	**.71**	.01	.02	.01	.01	.10	.02	.04
04SoutheastAsia	.04	.03	.02	**.59**	.05	.07	.07	.05	.02	.06
05SouthAsia	.03	.02	.03	.02	**.65**	.06	.03	.06	.03	.06
06MiddleEast	.02	.05	.02	.01	.03	**.77**	.01	.03	.01	.05
07EastAsia	.01	.01	.03	.04	.02	.01	**.79**	.06	.01	.02
08Europe	.02	.03	.04	.01	.02	.03	.01	**.79**	.02	.03
09AustraliaOceania	.07	.05	.05	.01	.01	.02	-	.03	**.68**	.07
10Western	.04	.04	-	-	-	-	-	-	-	**.92**

Table 2: Confusion matrix for the complete collection using all features, without artist filtering; Accuracy=0.75

	01Africa	02LatinAmerica	03NorthAmerica	04SoutheastAsia	05SouthAsia	06MiddleEast	07EastAsia	08Europe	09AustraliaOceania	10Western
01Africa	**.67**	.15	.03	.01	.01	.02	.01	.03	.04	.05
02LatinAmerica	.10	**.69**	.03	-	.01	.02	-	.07	.02	.06
03NorthAmerica	.08	.07	**.53**	.01	.03	.05	.01	.14	.04	.04
04SoutheastAsia	.03	.04	.05	**.41**	.05	.09	.10	.09	.02	.12
05SouthAsia	.02	.04	.06	.05	**.51**	.10	.03	.10	.03	.06
06MiddleEast	.06	.07	.04	.01	.05	**.66**	.02	.06	.02	.04
07EastAsia	.01	.02	.05	.10	.03	.04	**.59**	.11	.02	.03
08Europe	.03	.05	.05	-	.02	.03	.01	**.74**	.02	.04
09AustraliaOceania	.10	.05	.06	.01	.04	-	.01	.06	**.58**	.08
10Western	.07	-	.03	-	-	-	-	.04	-	**.86**

Table 3: Confusion matrix for the complete collection using all features, with artist filtering; Accuracy=0.62

The difference between the two results is 13%, confirming Flexer's observations about the large influence of the "artist effect" [6]. Notably, all ten regions suffer from the effect. Most of the results experience a decrease of around 10 to 15%. Lower exceptions are Western music with only 6% and European (traditional) music with only 5%. On the other side of the spectrum, the results for Southeast Asia mark a decline of 18%, those for North America decrease by 18%, and the results for East Asia (China and Japan) are even 20% worse.

4.2 Results of the domain-wise comparison (Experiment B)

We then ran the same experiment using the same data sets, but only sub-groups of the features. The results are shown in table 4.

Domain	No AF	With AF	Diff.	Diff. %
All	.75	.62	.13	.21
Timbre	.78	.62	.16	.25
Rhythm	.73	.60	.13	.21
Tonality	.52	.46	.06	.14

Table 4: Classification accuracies for all domains (AF = Artist Filtering, Diff. = Difference). The last column shows the difference in % of the accuracy for the result with artist filtering.

The results show that the effect appears in all three domains. The strongest influence seems to be caused by timbral features, while tonal features contribute the least.

4.3 Results of the region-wise comparison (Experiment C)

In this experiment, the results were not summed up over each region. Instead, the accuracy of the genre classification was evaluated directly for each regional data set. In contrast to Experiment A, the results do not represent the classification of all regions against each other, but rather the classification of specific genres of one region against each other. The results are shown in table 5. The result for the complete data set (Experiment A) is shown again in the first row for comparison.

Data set	No AF	With AF	Diff.	Diff. %
All	.75	.62	.13	.21
Africa	.69	.57	.12	.21
South America	.81	.66	.15	.23
Europe	.79	.68	.11	.16
Western	.87	.81	.06	.07

Table 5: Classification accuracies for all data sets (AF = Artist Filtering, Diff. = Difference). The last column shows the difference in % of the accuracy for the result with artist filtering.

As in Experiment A, western music stands out with very good results compared to the geographic regions. Africa and South America produce results which are slightly higher than or equal to the ones from Experiment A, while the European result is slightly lower.

5. DISCUSSION

5.1 General influence of the two effects (Experiment A)

The results from Experiment A confirm Flexer's observations and show an influence of the artist effect on non-western music as well. Looking at the different regions, the accuracy for western music is already quite good with artist filtering and does not change a lot when the artist filter is obmitted. We suppose that this is because the "common" features were developed with western music in mind and optimized accordingly. Therefore, they represent its characteristics quite well. European music has some similar qualities, so its result does not change a lot either. Additionally, the European category constitutes the largest one of the music collection and covers lots of genres. It is therefore somewhat more likely that a given musical piece will be classified as "European", which is also demonstrated by the high number of incorrect classifications into this category.

On the other side of the spectrum, the resulting differences for North American, Southeast Asian, and East Asian music are much higher than the average. In the case of North American music, we suspect that this is caused by the strong musical relation between North American and European or Western music. The chosen features do not seem to be able to represent the small musical differences between them. Thus, when trying to train models to distinguish between them, the features representing the properties of the artists are more likely to be picked up by the feature selection algorithm. The European and Western categories are not affected because of the reasons given above.

In the case of Southeast Asian music, the result is already fairly bad without artist filtering. We assume that this is because of the specific musical properties of this region. The category contains, for example, Gamelan music. Gamelan instruments do not have integral overtones and may therefore cause some feature extraction algorithms based on a preliminary f0-detection to fail, since they depend on several assumptions derived from the harmonic structure of most Western instruments. In consequence, the features do not consistently represent the musical characteristics of the region anymore. Similar to the North American example, features representing the artist and album properties are preferred instead. The same principle may be at work for East Asian music, which also has some properties that may not be well-represented by the features (e.g., gongs in Chinese Opera). Both East Asian and Southeast Asian music possess rhythmic and tonal traits that are quite dissimilar to western music. These may be the reasons why they are often confused with each other and why the artist and album effects are so prevalent.

5.2 Influence in each musical domain (Experiment B)

Looking at the results for Experiment B, it does not come as a surprise that the features based on timbre contribute most to the album and artist effects. As described in [12], the album effect is often caused by specific mixing and equalization preferences over the whole album, recording conditions, or effects used on that album. Similarly, the artist effect is likely caused by the spectral properties of the artists' voices or instruments. All of these conditions are reflected in the timbral domain.

What does however surprise us, is the non-neglectable influence of the rhythmic features. The possible causes listed above should not be picked up in the rhythmic domain. There are two possible reasons behind this: On the one hand, both effects might not only be caused by the specific spectral properties of one artist or album. Conceivably, each artist could have their distinct rhythmic style. On the other hand, the spectral differences could influence the extraction of the rhythmic features as well. Since some algorithms are based on the signal spectrum (e.g., rhythm event detection), this is quite possible. The contributions of the tonal features can be explained in a similar fashion. Further research is required to confirm or deny the two possible reasons.

Interestingly, all three feature subgroups produce acceptable results on their own, especially the timbre and rhythm groups. The differences between the music of the ten regions seem to be present in all three domains.

5.3 Influence on music from different regions (Experiment C)

Finally, the results of Experiment C mirror the good classification results for Western music from Experiment A. Again, this could be caused by the fact that the features were developed with Western music in mind. The next-best region is Europe, although the artist effect has a stronger influence here. The differences between the various African and South American styles are possibly not represented as well by the features as they are in Western music.

6. CONCLUSION

6.1 Summary

We performed various experiments to find out more about the album effect and the artist effect. One experiment showed a strong influence on the classification of musical pieces into their regions of origin. The influence of these effects depends on the music in each category and its relation to the music of other categories. If two regions share similar music, the results are more prone to be influenced by the effects (e.g., North American and European music). Western music produces good results and does not seem to be as susceptible to the artist effect as many non-western genres. We suspect that this is because many features were developed and tested using western music. They are therefore well-suited to represent the properties of western genres.

When training models with subsets of the features, the results using only timbral features showed the highest sensitivity to the album and artist effects. We expected this result since the specific properties of artists and albums are often reflected in the spectrum. However, rhythmic and tonal features also contributed to the effects.

6.2 Future work

As mentioned at the beginning, we do not differentiate between the album effect and the artist effect. Our data sets did not allow for such a detailed analysis. In the future, it would be interesting to find out how large each of these effects' influence on the classification results is.

We only experimented with two influences on the artist effect. Since both this effect and the album effect are such prominent phenomena, it might be interesting to explore other parameters that could influence them, e.g. the used feature selection and classification algorithms. Furthermore, a more in-depth analysis of the reasons behind these effects would be welcome. This could help to avoid the effect in future research.

7. ACKNOWLEDGMENTS

This work has been partly supported by the German research project *GlobalMusic2One*[2] funded by the Federal Ministry of Education and Research (BMBF-FKZ: 01/S08039B).

8. REFERENCES

[1] L. Atlas and S. S. Shamma. Joint acoustic and modulation frequency. *EURASIP Journal on Applied Signal Processing*, pages 668–675, 2003.

[2] J. P. Bello and J. Pickens. A robust mid-level representation for harmonic content in music signals. In *Proc. of the Intl. Conf. on Music Information Retrieval (ISMIR)*, London, UK, 2005.

[3] A. Berenzweig, D. P. W. Ellis, and S. Lawrence. Using voice segments to improve artist classification of music. In *Proc. of the 22nd AES Conference on Virtual, Synthetic and Entertainment Audio*, 2002.

[4] C. Dittmar, C. Bastuck, and M. Gruhne. Novel mid-level audio features for music similarity. In *Proc. of the Intl. Conf. on Music Communication Science (ICOMCS)*, Sydney, Australia, 2007.

[5] J. S. Downie. Music information retrieval. In B. Cronin, editor, *Annual Review of Information Science and Technology 37*, chapter 7, pages 295–340. 2003.

[6] A. Flexer. A closer look on artist filters for musical genre classification. In *Proc. of the Intl. Conf. on Music Information Retrieval (ISMIR)*, 2007.

[7] G. Gatzsche, M. Mehnert, D. Gatzsche, and K. Brandenburg. A symmetry based approach for musical tonality analysis. In *Proc. of the Intl. Conf. on Music Information Retrieval (ISMIR)*, Vienna, Austria, 2007.

[8] M. Gruhne and C. Dittmar. Comparison of harmonic mid-level representations for genre recognition. In *Proc. of the 3rd Workshop on Learning the Semantics of Audio Signals (LSAS)*, Graz, Austria, 2009.

[9] M. Gruhne, C. Dittmar, and D. Gärtner. Improving rhythmic similarity computation by beat histogram transformations. In *Proc. of the Intl. Conf. on Music Information Retrieval (ISMIR)*, Kobe, Japan, 2009.

[10] E. Gómez, M. Haro, and P. Herrera. Music and geography: Content description of musical audio from different parts of the world. In *Proc. of the Intl. Conf. on Music Information Retrieval (ISMIR)*, 2009.

[11] E. Gómez and P. Herrera. Comparative analysis of music recordings from western and non-western traditions by automatic tonal feature extraction. *Empirical Musicology Review*, 3(3):140–156, 2008.

[12] Y. E. Kim, D. S. Williamson, and S. Pilli. Towards quantifying the "album effect" in artist identification. In *Proc. of the Intl. Conf. on Music Information Retrieval (ISMIR)*, 2006.

[13] A. Kruspe, H. Lukashevich, J. Abeßer, H. Großmann, and C. Dittmar. Automatic classification of musical pieces into global cultural areas. In *Proc. of the 42nd AES Conference on Semantic Audio*, 2011.

[14] K. Lee. Automatic chord recognition from audio using enhanced pitch class profile. In *Proc. of the Intl. Computer Music Conf. (ICMC)*, New Orleans, USA, 2006.

[15] T. Lidy, C. N. Silla Jr., O. Cornelis, F. Gouyon, A. Rauber, C. A. A. Kaestner, and A. L. Koerich. On the suitability of state-of-the-art music information retrieval methods for analyzing, categorizing and accessing non-western and ethnic music collections. *Signal Processing*, (90):1032–1048, 2010.

[16] Y. Liu, Q. Xiang, Y. Wang, and L. Cai. Cultural style based music classification of audio signals. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2009.

[17] H. Lukashevich, J. Abeßer, C. Dittmar, and H. Großmann. From multi-labeling to multi-domain-labeling: A novel two-dimensional approach to music genre classification. In *Proc. of the Intl. Conf. on Music Information Retrieval (ISMIR)*, 2009.

[18] M. I. Mandel and D. P. W. Ellis. Song-level features and support vector machines for music classification. In *Proc. of the Intl. Conf. on Music Information Retrieval (ISMIR)*, 2005.

[19] B. Nettl, R. M. Stone, J. Porter, and T. Rice, editors. *The Garland Encyclopedia of World Music*. Garland, 1998. `http://glnd.alexanderstreet.com/`, Last visited: 05/11/11.

[20] E. Pampalk, A. Flexer, and G. Widmer. Improvements of audio-based music similarity and genre classification. In *Proc. of the Intl. Conf. on Music Information Retrieval (ISMIR)*, 2005.

[21] G. Peeters and X. Rodet. Hierarchical gaussian tree with inertia ratio maximization for the classification of large musical instrument databases. In *Proc. of the 6th Intl. Conference on Digital Audio Effects (DAFx-03)*, 2003.

[22] P. Proutskova and M. Casey. You call *that* singing? Ensemble classification of musical audio from different parts of the world. In *Proc. of the Intl. Conf. on Music Information Retrieval (ISMIR)*, 2009.

[23] N. Scaringella, G. Zoia, and D. Mlynek. Automatic genre classification of music content: A survey. *IEEE Signal Processing Magazine*, 23(2):133–141, 2006.

[24] G. Tzanetakis, A. Kapur, W. A. Schloss, and M. Wright. Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, 1(2):1–24, 2007.

[25] C. Uhle, C. Dittmar, and T. Sporer. Extraction of drum tracks from polyphonic music using independent subspace analysis. In *Proc. of the 4th Intl. Symposium on Independent Component Analysis (ICA)*, Nara, Japan, 2003.

[26] V. N. Vapnik. *Statistical learning theory*. Wiley, 1998.

[27] B. Whitman, G. Flake, and S. Lawrence. Artist detection in music with Minnowmatch. In *Proc. IEEE Workshop on Neural Networks for Signal Processing*, pages 559–568, 2001.

[2] `http://www.globalmusic2one.net`, Last visited 05/11/11