Automatic Classification of Musical Pieces Into Global Cultural Areas

Anna Kruspe[1], Hanna Lukashevich[1], Jakob Abeßer[1], Holger Großmann[1], and Christian Dittmar[1]

[1]*Fraunhofer IDMT, Ilmenau, 98693, Germany*

Correspondence should be addressed to Anna Kruspe (krusae@idmt.fraunhofer.de)

ABSTRACT

Music Information Retrieval (MIR) has a large variety of applications. One aspect that has not gathered a lot of attention yet is the application to non-western music ("world music"). In a task comparable to genre classification, this work's goal is the classification of musical pieces into their corresponding cultural regions of origin. As a basis for such a classification, a three-tier taxonomy based on musical and geographic properties is created. A database consisting of approximately 4400 musical pieces representing the taxonomical classes is assembled and annotated. Based on rhythmic, tonal, and timbre-related audio features, different classification experiments are performed. We achieved an accuracy of approx. 70% for the classification of musical pieces into nine large world regions. Twelve recently introduced features that are especially suited for non-western music were implemented as well. They improve the classification result slightly. For the purpose of comparison, we carried out a listening test with musical laypeople with an average accuracy of 52%.

1. INTRODUCTION

1.1. Motivation

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 [6]. 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 [27]. 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 [19]. Discovering distinct properties of each genre is another unsolved problem.

Non-western music (or "world music") is a field of application for MIR that has only started to receive attention in the past few years. Related to genre recognition, 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 MIR techniques. Tzanetakis et al. have coined the term *Computational Ethnomusicology* [28]. To our knowledge, almost no improvements specific to non-western music were made to the feature extraction algorithms yet. This is one of the goals of this paper. Additionally, the classification process is applied to a larger collection of music, with the aim of using music from all cultural areas of the world. Finding these areas in the first place is not trivial, as there are no clear definitions of the term and no clear limits between musical cultures, mirroring the problems of genre recognition.

Despite its difficulties, such an approach could prove useful for a range of applications. As an example, existing music recommendation and search engines could be improved in order to be able to use music from all over the world. In addition, completely new possibilities in the field of musicology are opened up. Automatic classification of non-western music may, for example, facilitate the work of ethnomusicologists, making large amounts of musical pieces of one style quick to find and easy to compare. This also offers new options for music analysis by highlighting relations between different recordings or complete datasets, or even by recognizing influences of other geographic regions.

All music classification tasks follow roughly the same procedure. First of all, a music database is compiled. It is then annotated with arbitrary labels. 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 pieces can be classified using this model. This allows for an evaluation of the model. We used this sequence of steps and will describe their details in the following.

The remainder of this paper is organized as follows. In Sec. 1.2, we give a brief overview over related work. The features used for classification are described in Sec. 2, the actual classification process is explained in Sec. 3. We present the concrete conditions for the various experiments in Sec. 4 before showing and discussing our results in Sec. 5. Finally, Sec. 6 contains a conclusion of our work and ideas for future developments.

1.2. State of the Art

As the application of MIR techniques to non-western music is still a relatively new idea, only a few approaches can be found in literature. Genre classification in general is a quite popular task, though. Standardized data sets, such as the ones for the *MIREX*[1] contest exist, and various researchers have obtained results between 61% and 83.5% (accuracy) using these data sets [21].

Experiments using non-western music were made, for example, in [19], [13], [12], [18], [17], and [25]. All of them have some properties in common: They all make use of low- and mid-level features from the three musical domains rhythm, timbre, and tonality, and they all utilize Support Vector Machine (SVM) classifiers. However, their results cannot be directly compared as they all use different data sets. These data sets range between a variety of folk music genres to only a handful of specific styles from various regions. None of these papers use music from all over the world. Consequently, no standardized data set (such as the one for *MIREX*) exist.

Additionally, hardly any of these approaches implement features specific to world music. Instead, only previously existing, non-specific features are used.

In this paper, an attempt is made to collect a wide variety of non-western and western genres to guarantee a better representation of the musical cultures of the world. We assume that the implementation of specialized features will improve classification results for the given data set. The results are compared against the ones achieved using conventional audio features.

To obtain a general understanding of how well world music can be classified by human subjects, a small listening test is conducted.

2. FEATURE EXTRACTION

A large set of previously existing low- and mid-level features is extracted from the musical pieces. These are mostly "commonplace" features such as the ones described in [22] and in the MPEG-7 standard.

As mentioned above, a set of new features was implemented additionally. These features are intended to be especially suited for the classification of non-western music. Most of them can be characterized as mid- to high-level features since they take musical background knowledge into account. Before the actual feature extraction, an algorithm for the frame-wise recognition of fundamental frequencies (f_0) is applied to the musical pieces (based on MR-FFT [7]). Some of the implemented features make use of these f_0 trajectories, while the resulting pitch class histograms are used for others. In the following, the previously existing features as well as the new features based on the f_0 frequencies, the ones based on the pitch class histograms, and the ones based on neither of them are described.

2.1. Previously existing features

We utilize a broad palette of low-level acoustic features and several mid-level representations [3] [22]. 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 dimensional-

[1] `http://www.music-ir.org/mirex/wiki/MIREX_HOME`, Last visited 03/10/11

ity and decorrelate them as described in [5].

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 [29] and the Envelope Cross-Correlation (ECC). Further mid-level features [5] 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 [11].

Tonal features Tonality descriptors are computed from a Chromagram based on Enhanced Pitch Class Profiles (EPCP) [16]. 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 [8] are derived from the Chromagram as mid-level features. Their usefulness for audio description has been shown in [10].

2.2. New features based on fundamental frequencies

Vibrato and Tremolo In this context, vibrato describes the (slight) frequency fluctuation of one note, while tremolo describes its amplitude fluctuation. Using the extracted fundamental frequencies, a vibrato value is calculated according to Regnier and Peeters [26]. The same algorithm is applied to extract a tremolo value. In order to do this, the original wave data is included as well to obtain the varying amplitudes of the fundamental frequency sinusoidals. Regnier and Peeters used these values to differentiate between vocal and purely instrumental pieces. This distinction is a useful concept for different world music styles as well. Additionally, vibrato and tremolo are used as stylistic elements in certain non-western music genres, e.g. in some Eastern European or Middle Eastern styles.

Roughness The empirically motivated formula from [31] is used to obtain information about the roughness of the musical piece. Roughness describes a musical property that can be characterized as “harsh, raspy, hoarse” [14] and is caused by the interference of two sinusoidal components with a slight difference in their frequencies. It occurs at amplitude fluctuation rates between 20 Hz and 75-150 Hz, while the sensation of beating occurs below 20 Hz. Some non-western styles require their instruments to be detuned against each other on purpose in order to achieve this sensation. The most prominent example are probably Indonesian Gamelan orchestras.

Finally, the proportion of non-voiced and voiced parts of the musical piece is calculated as well by comparing the number of frames where f_0 frequencies were detected and the ones where this is not the case.

2.3. New features based on the pitch class histogram

Pitch class histograms are calculated by generating a histogram over the fundamental frequencies, collapsing it into one octave and scaling its values to lie between 0 and 1. The frequency resolution of the histograms is 100 steps per semitone (or 1 step per 1 Cent).

Histogram Flatness and Crest Flatness and crest values similar to the commonly known spectral flatness (SFM) and crest (SCM) measures [22] are calculated for the pitch class histograms. These measures characterize how “discrete” the scale of the musical piece is, i.e. how clearly the different scale steps are separated. A piece played on a piano, for example, would generate low values due to the distince note frequencies, while a piece with lots of slides produces high values since this technique results in a continuous progression of note frequency over time.

Equal-Tempered Deviation The pitch class histogram can be used to extract further information about the musical scales of a piece, which is one of the most characteristic differences between different styles of non-western (and western) music. The *Equal-Tempered Deviation (ETD)* feature presented in [13] signifies the deviation between the peaks of the histogram and the peaks expected in western music. As shown in Fig. 1, western music such as Austrian folk music usually has pitches with multiples of semitones (=100 Cent) between them, while non-western music such as Japanese classical music does not necessarily adhere to this convention.

Scale Correlations Another interesting information is the scale itself. There is a vast array of different scales from all over the world. We collected 137 of them, mostly from the appendix of [14]. The algorithm from [9] was then expanded to be able to recognize these scales. It performs a template matching between pitch class histograms. Afterwards, the highest correlation of the piece with all scales of a geographic region is selected for each region. This vector represents the likelihood of the piece's scale coming from each of the regions.

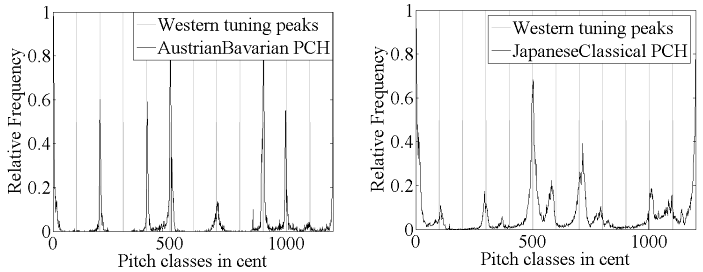

Fig. 1: Pitch class histograms (PCH) for an Austrian piece (left) and a Japanese Classical piece (right) with the expected peaks for equal-tempered tuning overlaid (vertical lines)

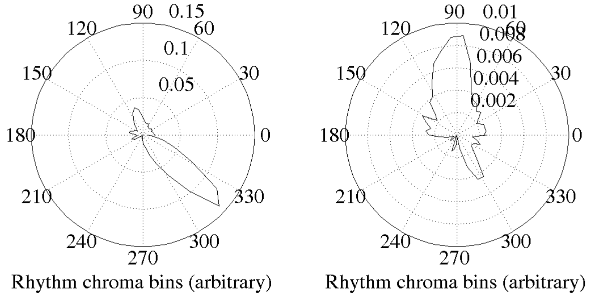

Fig. 2: Rhythm chromagram for a techno piece (left) and for a classical piece (right)

2.4. Other new features

Harmonic-to-Attack Ratio Some non-western styles make extensive use of percussive elements (e.g., most Latin American styles), while others hardly use any percussive instruments at all (e.g., Japanese classical music or Sephardic music). In order to quantify these differences, the *H2A (Harmonic-to-Attack Ratio)* feature is calculated according to [24]. The algorithm includes a filtering of the spectrogram using vertical and horizontal edge detectors to find percussive and harmonic components and then setting the energy of both components in relation to each other.

Rhythm Chromagram To analyze the rhythmic structure of a piece more in-depth, a rhythm chromagram was implemented following [15]. This chromagram presents information about the prevalence of certain tempi in one piece and their relation to each other, all collapsed into one "tempo octave". Additionally, some statistical measures such as mean, sum and crest are calculated over the rhythm chromagram. Examples for the chromagram are shown in Fig. 2. This figure demonstrates that the rhythm chromagram already allows for a distinction between different western styles, as techno music produces one sharp main peak, signifying a single strong underlying beat with a constant tempo. Classical music, on the other hand, does often not contain such a single dominant rhythm.

3. CLASSIFICATION

3.1. Feature Selection

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 is closely related to Fisher's LDA. This principle was first introduced in [23]. 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.

3.2. Classification Algorithm

As stated above, the SVM 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 annotations. 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 feature space and finds separating hyper-planes with maximum margins 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 [30].

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.

4. EXPERIMENTAL SETUP

4.1. Music Collection

In order to guarantee a wide range of musical styles from all over the world, a large taxonomy was created. Three ethnomusicologists were consulted about their opinions and extensive literature research was carried out in order to define the taxonomy. During this research, no existing taxonomy for non-western music was encountered. Therefore, the index of the "Garland Encyclopedia of World Music" [20] was used as a starting point. This list comprises nine very large regions of the world, such as Africa or the Middle East, plus one category for western music. From a very rough point of view, the musical styles inside each one of these regions share similar properties. A somewhat finer representation was achieved by splitting up these large regions into smaller sub-regions. Finally, a number of representative musical genres for each sub-region were compiled. The complete taxonomy is available under `http://www.idmt.fraunhofer.de/eng/kruspe/regions_final.pdf`. Of course, such a classification is always imperfect. Music as a cultural concept cannot simply be defined by such strict borders and genres. Musical genres are also inherently ill-defined [27] and in a constant state of change. Additionally, these genres (and even the broader musical regions) influence each other all the time. The biggest factor here is the immense influence of western music on a large proportion of non-western music. "Fusion" styles, however, were not taken into account in this publication. All in all, 81 non-western genres were selected for 9 large world regions, with an additional 5 western genres for comparison. For clarification, 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.

Built upon this taxonomy, approximately 4400 musical pieces were collected. They are divided into at least 50 pieces per genre. No special attention was paid to the duration of the pieces, as this may also be a differentiating factor between genres. However, no piece is shorter than 20s or longer than 20min. The songs are all in stereo and in MP3 format with a bitrate of 128kbps and a sampling frequency of 44.1kHz. An overview over all regions, their number of genres, and their number of musical pieces is shown in Table 1.

Region	No. of genres	No. of pieces
Africa	12	600
South America	13	656
North America	9	457
Southeast Asia	3	150
South Asia	5	256
Middle East	10	511
East Asia	4	215
Europe	20	1071
Australia & Oceania	4	203
Western Music	5	279

Table 1: The ten large regions of the music collection

4.2. Listening Test

A small listening test was conducted to approximate the accuracy with which laypeople are able to classify non-western music. Snippets with a duration of 30s were prepared from a selection of songs from all non-western genres to serve as testing data. 20 test subjects were then asked to listen to all of these pieces and to state where they believed each of the songs to come from. All participants came from a European background and had no special education concerning world music. No training was performed previously, but the participants received an overview over the contained world regions and the typical genres chosen for them. They were allowed to listen to the snippets in whichever order they preferred and as often as they liked.

4.3. Automatic Training and Classification

Prior to the training, the low- and mid-level features described above were extracted from the training songs. The low-level features were extracted with a resolution of 10ms, while the mid-level features have a resolution of approximately 5s with a hopsize of 2.5s. Furthermore, an automatic segmentation algorithm based on Mel-Frequency Cepstral Coefficients (MFCCs) and the Bayesian Information Criterion (BIC) was used to extract a segmentation of each musical piece. This segmentations represents the musical structure of the piece (e.g. choruses and verses). The new mid- and high-level features were then extracted segment-wise. Both sets of features were combined for the training and evaluation processes. All features are averaged over 5s frames. Therefore, a matrix with approximately 37,000 rows (time frames) and 1,303 columns (feature dimensions) is generated.

As described above, IRMFSP and SVM implementations are used for feature selection and classification. The SVM implementation is based upon the *LIBSVM* library [4] and utilizes the RBF kernel. The error margin parameter C and the kernel parameter γ are optimized automatically during each test run via a grid search. The maximum number of features selected by the IRMFSP algorithm is set to 256.

To guarantee meaningful results, a five-fold cross-validation is performed in every experiment. The five data sets are arranged in such a way that they do not share artists or albums, if at all possible. This reduces the influence of the album effect [32]. The classification is done segment-wise. Its results are weighted with the duration of each segment to obtain the most likely genre. Since the musical pieces are annotated on the genre level and the results for all 86 genres are rather hard to interpret, a simplification is carried out afterwards: The results for all genres of one region are summed up. Thereby, a confusion matrix for all combinations of the regions is generated.

5. RESULTS AND DISCUSSION

5.1. Listening Test

A confusion matrix for the results of the listening test is shown in Table 2 (rows: originally annotated labels, column: labels found by the participants). The resulting mean accuracy is 52.4%. This yields several interesting insights. As expected, musical pieces that the participants felt unsure about were often classified as “European” - a region the participants supposedly were familiar with. On the other hand, hardly any Southeast Asian pieces were classified correctly. These were often grouped into the East Asian region. The Southeast Asian set contains a number of musical styles western listeners may not have heard of, such as Gamelan music or Tuvan Overtone music. We assume that the participants therefore did not have a good conception of Southeast Asian music, but felt more sure about East Asian music (which contains Chinese and Japanese music).

Another interesting phenomenon is the occurence of some of the same misclassifications as in the automatic system. As an example, the human subjects often confused African and South American music, just like the automatic classifier. These musical regions share lots of similar characteristics (e.g. similar rhythms or a percussion-dominated instrumentation).

Many participants stated that they focused a lot on the vocals' language if there was singing in a piece. The automatic classifier does not have any means to do the same. This is an interesting approach and might form a motivation to integrate language recognition into music classification in the future.

It must be stated that the results of the listening test are not directly comparable to the ones from the automatic classification experiments. The participants were not given a previous training and could only listen to segments of the musical pieces. The experiment does, however, show that world music classification is not an easy task for humans either. Without any previous training, the participants seem to have a hard time to find differentiating features between the various regions. Of course, this also serves as another motivation for the improvement of automatic classification.

	01Africa	02LatinAmerica	03NorthAmerica	04SoutheastAsia	05SouthAsia	06MiddleEast	07EastAsia	08Europe	09AustraliaOceania
01Africa	**.44**	.35	.	.	.02	.02	.01	.14	.02
02LatinAmerica	.03	**.63**	.	.	.	.02	.	.30	.01
03NorthAmerica	.07	.03	**.52**	.	.01	.02	.	.32	.03
04SoutheastAsia	.08	.	.02	**.15**	.05	.03	.47	.03	.17
05SouthAsia	.12	.05	.03	.11	**.32**	.24	.08	.03	.02
06MiddleEast	.03	.03	.02	.04	.12	**.55**	.05	.18	.01
07EastAsia	.08	.04	.08	.08	.05	.16	**.43**	.06	.04
08Europe	.03	.05	.11	.02	.03	.07	.02	**.65**	.02
09AustraliaOceania	.23	.19	.01	.	.01	.03	.	.14	**.40**

Table 2: The confusion matrix for the listening test, results grouped into regions

5.2. Automatic Classification using low- and mid-level features only

The automatic classification using only the previously existing low- and mid-level features produced the confusion matrix shown in Table 3 (rows: annotated labels, columns: labels found by the classifier)[2]. The average accuracy is 68.9%. This is already a fairly good result and comparable to the percentages achieved in the *MIREX* contest.

Southeast Asian music achieved the lowest accuracy. This can be explained by the fact that this data set was the smallest one and the taxonomy contained the fewest genres of all regions. Therefore, misclassified pieces were less likely to be accidentally classified as one of the other genres of the correct region, which would mark them as "correct" when summming up the results for all regions. The Southeast Asian genres do not have a lot of musical properties in common either. Additionally, Gamelan music is among these genres. Gamelan instruments often possess overtones that are not integer multiples of the fundamental frequency. This may cause the f_0 extraction algorithm used for the feature calculation to fail.

Unsurprisingly, western music produced the highest accuracy. The features used in this experiment were mostly developed with western music in mind. We assume that they were optimized accordingly.

	01Africa	02LatinAmerica	03NorthAmerica	04SoutheastAsia	05SouthAsia	06MiddleEast	07EastAsia	08Europe	09AustraliaOceania	10Western
01Africa	**.70**	.11	.02	.	.01	.02	.01	.06	.03	.03
02LatinAmerica	.09	**.69**	.02	.	.03	.03	.	.08	.02	.03
03NorthAmerica	.06	.06	**.56**	.01	.04	.04	.01	.16	.03	.03
04SoutheastAsia	.03	.04	.05	**.48**	.10	.05	.09	.10	.03	.03
05SouthAsia	.03	.03	.06	.04	**.57**	.10	.02	.11	.02	.02
06MiddleEast	.04	.05	.03	.01	.03	**.71**	.02	.07	.02	.02
07EastAsia	.01	.02	.03	.11	.03	.04	**.64**	.11	.01	.
08Europe	.03	.04	.06	.01	.01	.03	.01	**.78**	.01	.02
09AustraliaOceania	.07	.04	.05	.02	.05	.03	.02	.05	**.63**	.04
10Western	.04	.02	.05	.	.01	.02	.	.03	.01	**.82**

Table 3: Confusion matrix for a cross-validation using only the pre-existing features

5.3. Automatic Classification using all features

Table 4 shows the results that are achieved when the new features are added[3]. The average accuracy is now at 70.2%. The improvement of 1.3% is statistically significant with $p < 0.0001$. (However, the improvement is not significant for the genre-wise classification with $p > 0.05$, therefore this result must be considered cautiously). The results in the confusion matrix show no peculiarities - the result for each region is improved slightly (except for South Asia). East Asia is still the most problematic region for the reasons described above. Specific tests were run on smaller subsets of the music collection to find out which features were selected by the IRMFSP algorithm. Using a limit of 256 dimensions, only three novel features were selected: Two statistical

[2] This table shows the aggregated results for each region; the complete genre-wise results are available under `http://www.idmt.fraunhofer.de/eng/kruspe/result_lowAndMidLevelFeatures.pdf`

[3] This table shows the aggregated results for each region; the complete genre-wise results are available under `http://www.idmt.fraunhofer.de/eng/kruspe/result_allFeatures.pdf`

measures of the Rhythm Chromagram and the roughness. Using 512 dimensions, the pitch class histogram crest and flatness measures and the vibrato and tremolo measures were added.

	01Africa	02LatinAmerica	03NorthAmerica	04SoutheastAsia	05SouthAsia	06MiddleEast	07EastAsia	08Europe	09AustraliaOceania	10Western
01Africa	**.71**	.11	.02	.	.01	.02	.01	.05	.03	.03
02LatinAmerica	.10	**.71**	.03	.	.02	.03	.01	.07	.01	.03
03NorthAmerica	.05	.06	**.58**	.01	.04	.04	.01	.15	.02	.04
04SoutheastAsia	.03	.06	.05	**.49**	.10	.07	.09	.09	.	.03
05SouthAsia	.02	.04	.05	.04	**.57**	.10	.01	.12	.02	.03
06MiddleEast	.03	.06	.03	.01	.03	**.73**	.01	.06	.02	.02
07EastAsia	.01	.01	.03	.10	.03	.05	**.65**	.10	.02	.
08Europe	.02	.03	.06	.01	.01	.03	.01	**.80**	.01	.02
09AustraliaOceania	.07	.03	.06	.02	.05	.01	.02	.04	**.65**	.04
10Western	.03	.02	.05	.	.01	.01	.	.03	.01	**.83**

Table 4: Confusion matrix for a cross-validation using all features

5.4. Discussion

Considering that the new features apparently produced satisfying results when tested separately, it is surprising that they were discarded by the feature selection algorithm. It is suspected that this is caused by the discrepancy between the IRMFSP and the SVM algorithms. As explained above, the feature selection handles all classes at once all of the time, while the SVM training algorithm works on a one-vs-one basis. This means that the IRMFSP algorithm generally prefers features that show differences over many classes, even if they are small, to those that discriminate sharply between only a few classes. Most of the new features offer such a strong binary or ternary distinction rather than a weak multi-class differentiation (e.g. ETD, H2A, Vibrato, and Tremolo). The features for which this is not necessarily true (Rhythm Chromagram, Roughness, and Histogram Crest and Flatness) were among those to be most likely selected.

We assume that this feature selection effect might generally tend to exclude high-level features. These features are based on human perception and musical cognition, which make attempts to group different musical styles into categories (e.g. rhythmical or scale-based ones). IRMFSP stands in strong contrast to this, as it bases its decisions on small differences between classes rather than on strong ones between groups of classes. This criterion is more likely to be achieved by low- and mid-level features that have a stronger relation to the acoustic signal.

Additionally, timbre is often the sole distinguishing criterion between two musical styles (e.g., when different typical instruments are used). Most timbral properties are already covered very well by low- and mid-level features (e.g. MFCCs). This supposed connection between feature selection and high-level features could be a new topic of research. The new features could also be used for other purposes, such as similarity searches or a tree-like classification structure.

Finally, it is suspected that such a classification task has an inherent upper limit (possibly related to the "glass ceiling" mentioned in [2]). As explained above, it may be impossible to draw hard borders between different world music genres or even between musical regions. This is reflected by the results of the listening test, too.

6. CONCLUSION AND OUTLOOK

In this paper, non-western musical pieces were classified into their regions of origin. In order to do this, a taxonomy with three layers was compiled. According to this taxonomy, 4400 musical pieces were collected and annotated. There is a number of problems with such an approach, mainly because the regions and genres are ill-defined and not easy to separate.

Using a large set of common low- and mid-level audio features, an accuracy of 68.9% was achieved over the nine defined large musical regions and a western music category. Twelve world music specific features were implemented. They show good results on their own, but improve the over-all classification result by only 1.3%. The feature selection algorithm might be the reason why the improvement is not greater than this.

Additionally, a listening test was performed with 20 laypeople. The test produced an average accuracy of 52% over nine large regions, demonstrating the necessity for an automatic classification system.

Considering the problems of defining a strict taxonomy, an approach using a similarity search or clustering methods looks promising. Furthermore, the systematic errors caused by the interference of different genres are confronted in [19] by introducing a new technique called "Multi-Domain Labeling". Using this technique, the musical piece is split up "horizontally" into time frames and "vertically" into its musical domains timbre, rhythm, and tonality. A single genre can then be assigned to each field in this two-dimensional grid. This breaks the problem of multi-class labeling down into smaller classification

problems with single classes and could also be a new approach to handle the classification of the data set.
Results using another feature selection algorithm or other classifiers would be of interest. A thorough examination of one-vs.-one selection and training in contrast to their multiclass counterparts could help future experiments.
Many of the participants of the listening test paid special attention to the language of music that contained singing. It is therefore believed that the addition of language recognition algorithms could improve the automatic classification results. Finally, further listening tests could be conducted, e.g., with experts in ethnomusicology or with a training done in advance.

7. ACKNOWLEDGEMENTS

This work has been partly supported by the German research project *GlobalMusic2One*[4] funded by the Federal Ministry of Education and Research (BMBF-FKZ: 01/S08039B). We would also like to thank the ethnomusicologists from HfM Weimar and from the "Afro-Karibisches Musikarchiv Köln" for their help in creating the taxonomy.

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.-J. Aucouturier and F. Pachet. Improving timbre similarity: How high's the sky? *Journal of Negative Results in Speech and Audio Sciences*, 1(1), 2004.

[3] 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.

[4] C.-C. Chang and C.-J. Lin. LIBSVM - a library for support vector machines, 2001. Last visited: 03/11/11.

[5] 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.

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

[7] K. Dressler. Sinusoidal extraction using an efficient implementation of a multi-resolution FFT. In *Proc. of the 9th Intl. Conference on Digital Audio Effects (DAFx-06)*, 2006.

[8] 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.

[9] A. C. Gedik and B. Bozkurt. Automatic classification of Turkish traditional art music recordings by Arel theory. In *Proc. of the Conf. on Interdisciplinary Musicology*, 2008.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] H. Helmholtz. *On the Sensations of Tone*. Dover Publications, 1954.

[15] E. Humphrey. Automatic characterization of digital music for rhythmic auditory stimulation. In *Proc. of the Intl. Conf. on Music Information Retrieval (ISMIR)*, 2010.

[4] `http://www.globalmusic2one.net`, Last visited 03/13/11

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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: 03/10/11.

[21] I. Panagakis, E. Benetos, and C. Kotropoulos. Music genre classification: A multilinear approach. In *Proc. of the Intl. Conf. on Music Information Retrieval (ISMIR)*, 2008.

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

[23] 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.

[24] T. Pohle, P. Knees, K. Seyerlehner, and G. Widmer. A high-level audio feature for music retrieval and sorting. In *Proc. of the 13th Intl. Conference on Digital Audio Effects (DAFx-10)*, 2010.

[25] 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.

[26] L. Regnier and G. Peeters. Singing voice detection in music tracks using direct voice vibrato detection. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2009.

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

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

[29] 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.

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

[31] P. N. Vassilakis. *Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance*. PhD thesis, University of California, Los Angeles, CA, USA, 2001.

[32] 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.