Classification of Myanmar Ethnic Traditional Style Based Music Signals

May Thu Myint
University Of Computer Studies, Mandalay
mthumyint@gmail.com

Abstract

The large quantities of information in both audio and video is a great need for efficient ways of searching and retrieving relevant information. The goal of an audio indexing is to provide the capability of searching and browsing through audio content. This work presents an effort to discriminate Myanmar ethnic traditional style music based on the large number of audio features(timbre, rhythm, pitch) are evaluated for their suitability in such a classification task, including well-known physical and perceptual features, audio descriptors. For this purpose, we use the Multi Band Autocorrelation Peak (MBAP) features, extracted in multiple bands providing complementary information and give a better description of the audio signal and then use the classifier based on Training and Testing audio data to get class labels .The system have selected more popular cultural music styles of Myanmar ethnic music namely: Kachin, Kayin, Shan and Yakhine ethnic group songs for experiments.

1. Introduction

The systematic study of ethnic music has started in the late 19th century. From then on, field work has been conducted all over the world, and numerous sound recordings have been made and documented. Worldwide, thousands of cultures have developed their own musical culture, with specific qualities and specific purposes. This makes the field of ethnic music very broad, with a whole range of different timbres, moods, styles, instruments and musical characteristics. Most MIR-applications need to be redesigned in order to give a relevant contribution to the analysis and classification of ethnic music. Culture has a significant influence on music in term of creation, performance and interpretation. People from a certain cultural background often prefer music with a particular cultural style. Cultural style information is useful for music browsing, retrieval and recommendation.

Between the advent of audio recordings in 1877 and the emergence of Napster in 1999, most music collections were based on discrete storage units, whether wax cylinders, vinyl records, cassettes, or compact discs. In the 21st century, the confluence of high-quality audio compression, the growing storage capacity of hard drives, and the ubiquity of high-speed Internet have made it possible to transform large audio collections into digital archives on a home computer. The result is a need for tools to explore and organize these collections. In content-based audio analysis, a music piece is described by a set of features that are directly computed from its content, i.e., the audio signal is parameterized into suitable feature vectors, which should retain salient information while discarding unnecessary details. Therefore, the initial step in content-based audio analysis is to represent the audio signal in a low-dimension form which can be used to manipulate more meaningful information. In this paper, Myanmar is made up of several dozen ethnic groups, speaking their own languages and dialects, having distinct cultural traditions as well as music genres. Based on this observation, we hypothesize that music audio signals can be classified according to their cultural styles using simple searching algorithm. The system have selected more popular cultural music styles of Myanmar ethnic music namely: Kachin, Kayin, Shan and Yakhine ethnic group songs for experiments.

2. Related Work

Audio Content Analysis, also called Computer Audition or Machine Listening, deals with the extraction of information from sounds. It should be noted that, in order for an audio system to be regarded as performing Content Analysis, it should be able to provide “high-level” descriptions, such as spoken content for voice or melody and tempo for music, or even more abstract information like musical structure, genre or mood. Simple features like loudness or pitch provide” low-level” information and by themselves might not suffice for speaking of a true description of content.

Liu et al presented cultural style based music classification using 6 different cultures Western, Chinese, Japanese, Indian, Arabic and African. Timbral feature, wavelet feature and musicology-based features sets are extracted. Decision tree, OAO and OAA SVM and k-NN classifiers are applied for classification purpose. The highest accuracy of 86.50% is achieved by SVM-OVA using all features. Eerola, T., etal. (2001) presented a method for comparative analysis of folk music based on musical feature extraction and neural networks using a simple data-mining tool for databases that use a symbolic representation of melodic information and extracting the distributions of pitches, intervals and durations Self-organizing neural network (SOM) was used to visualize the feature vectors. Lydia MutiaraDewi (2012) referred to the dataset of Asian and European folk songs. Timbl will be used for classifying the musical pieces in accordance with East Asian and European classes and applies the combination of N-Gram, tf*idf weighting, and machine learning The means of accuracy for this combination of global features are 84.22% for both halves of the musical pieces and 83.92% for the whole musical piece. Chai, W., & Vercoe, B. (2001) describes the work on the music corpora of Irish, German and Austrian folk music in various symbolic formats were used as the HMM structures achieved 75%, 77% and 66% for 2-way classifications and 63% for 3-way classification using 6-state left-right HMM with the interval representation in the experiment.

3. Background

The fundamental music classification tasks are Musical data collection, feature extraction and machine learning. Firstly, the music data collection provides the instances (basic entities), Audio recordings, scores, cultural data is used to classify. In feature extraction steps, various features represent characteristic information about instances and must provide sufficient information to segment instances among classes (categories). Finally, the machine learning algorithm (“classifiers” or “learners”) is used to associate feature patterns of instances with their classes. The types of musical data is taken from Sampled sound, Wave, MP3, AAC for Audio recordings, musical instructions , music scores, MIDI, Humdrum are abstracted for symbolic recordings . For cultural information, cultural information are retrieved external to musical content itself (e.g. playlists, album reviews, Billboard stats, etc.) based on web searches, surveys, expert opinion, etc. Myanmar is an ethnically diverse nation with 135 distinct ethnic groups officially recognized by the Burmese government. These are grouped into eight "major national ethnic races" and it is Kachin, Kayah, Kayin, Chin, Mon, Bamar, Rakhine and Shan. These ethnic group's instruments reported as following:

- Instruments beaten with a padded hammer
- Instruments knocked with a hard or semi-hard hammer
- Hand-beaten instruments
- Plucked instruments
- Pulled instruments
- Bowed instruments
- Blown instruments
- Shaken instruments

4. Proposed System

In content-based audio analysis, a music piece is described by a set of features that are directly computed from its content, i.e., the audio signal is parameterized into suitable feature vectors, which should retain salient information while discarding unnecessary details. Therefore, the initial step in content-based audio analysis is to represent the audio signal in a low-dimension form which can be used to manipulate more meaningful information. The basic process for representing an audio signal can be defined by four steps: Data preprocessing, Segmentation, Feature extraction and Classification.

4.1. Pre-processing for Music Signal

In order to extract features from an audio signal, samples are often pre-processed using standard digital signal processing techniques in figure (1). The purpose of this is to help improve performance when features are processed by classification algorithms. One common technique is to normalize the amplitudes in an audio signal such that they are uniformly scaled for further manipulation. Normalization is helpful when applying classification algorithms as it controls variability in recording levels.

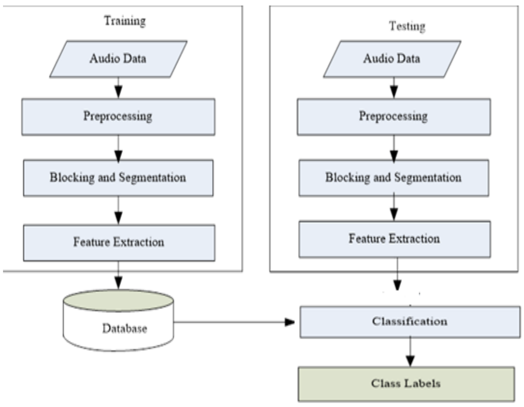


Figure 1. Architecture of the proposed system

Another common technique is down sampling, which reduces the sampling rate of an audio signal. It is useful for adjusting different audio sources to the same bandwidth and to reduce data size, which helps speed up the feature extraction process and the related classification algorithms. Breaking audio samples into shorter windows is a common technique for obtaining a degree of time localization when performing Fourier analyses. Two of the most commonly used windowing functions are Hamming windows and Hann windows, which are both moderate with respect to dynamic range. Their popularity is due to their applicability to a wide variety of wave types.

4.2. Blocking and Segmentation

The features extracted from each frame of an audio signal represent a significant reduction in terms of data size and dimensionality. To further reduce an audio signal to a lower dimensional form, instead of extracting features on a frame-by-frame basis, features can be calculated on a group of consecutive frames (an alysis windows). This process, known as segmentation, is accomplished by applying acoustic analysis to the frames audio signal and looking for significant transitions. In segmentation, the audio data is divided into variable-length units called segments and the data within a segment does not vary much. The positions in audio data where very sharp changes occur define the segment boundaries. The audio data and the query are divided into fixed-size blocks.

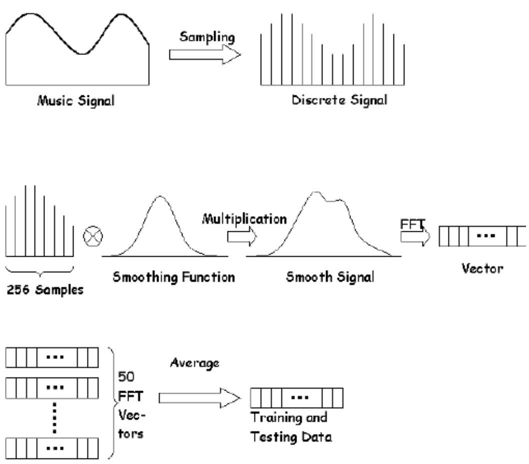


Figure 2. Preprocessing of Audio Clip to Get the Input Data

4.3. Feature Extraction

The feature extraction step transforms the input audio signal into a low-dimensional representation which contains the information necessary for classification or content analysis. The audio file is broken into small segments in time and for each of these segments a feature vector is calculated. They use information such as spectral or statistical variations in order to determine rhythm, pitch, tempo, melody, and timbre. The audio features are described as followings: Pitch, Timbre Texture Features and Rhythm Features. Pitch is an important perceptual feature and an attribute of every musical tone and the fundamental or first harmonic of any tone. For example pitch helps the human ear to distinguish between string instruments, wind instruments and percussion instruments such as the drums, table etc. These could be categorized into Time-domain and Frequency-domain analysis.

In a song, we hypothesize that there exist multiple such pitch trajectories of interest in different bands, that might provide discriminative information. An attempt to extract a single pitch from such a signal would lead to a noisy pitch estimate, possibly jumping between these multiple pitch trajectories. Under the assumption that these trajectories do not overlap, we try to measure this information by computing the fundamental frequency in different bands by modifying the ranges fmin and fmax above. Specifically we use the following bands (in Hz).

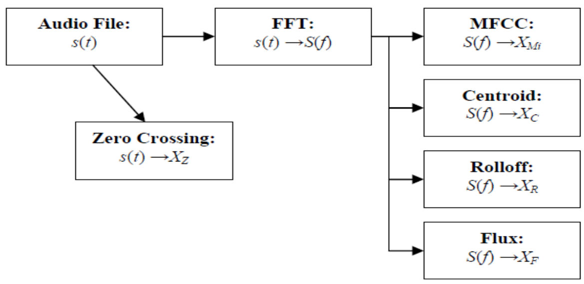


Figure 3. Dataflow Diagram for Timbre Features

The MBAP features fi are thus computed for the 8 bands in Table I. We do not consider any bands beyond these because any higher fundamental frequencies were found to be rare for the songs in our database.The sampling of MBAP features for the first 5 bands extracted for an orig song and its corresponding alt sample.The minimum value in each band is f_{min}. Also note the sparsity of MBAP features in higher bands because of absenceof higher fundamental frequencies.

Table 1. Bands in Which MBAP Features Are Extracted

i	f_{min}	f_{max}
1	5	50
2	50	100
3	100	150
4	150	200
5	200	250
6	250	350
7	350	450
8	450	550

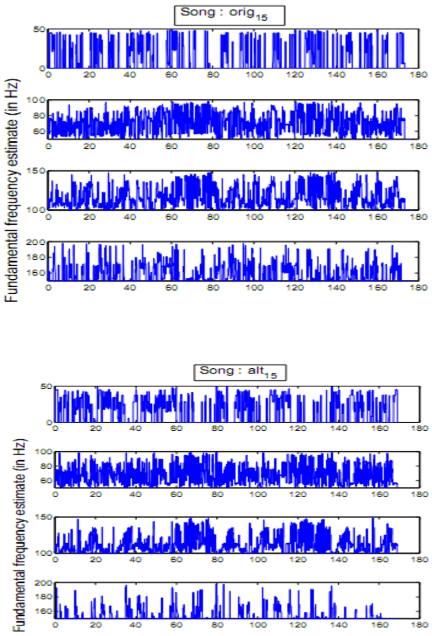


Figure 4. MBAP Features Computed in the First Five Bands for the Songs orig15 (above) and alt 15 (below)

5. Classification Scheme

Before diving into the details of classification, it is necessary to realize that the parallel clips in our dataset might not exactly be time aligned. This might be the case when the songs do not have the same tempo. This can cause variations between the two versions, necessitating the use of temporal modelling techniques to find an optimal alignment along time. This problem being similar to edit distance approach between two time series, our natural choice was to use Dynamic Time Warping (DTW). Thus, we use DTW to match two songs, using the alignment cost as a classification metric for tune similarity. The rationale is that dissimilar songs will require heavy warping to align, thereby incurring a large alignment cost.

After all the DTW alignments have been computed (Figure 7) classification comprises simply comparing the song alignmentcosts for all the song options. Let C_{ij} be the cost of aligning origin with altj ; i, j = 1 · · · 24. We classify origin as being similar to alt_k . Given two time series origi[m] and altj [n], DTW re- turns a sequence of index tuples : {(m1; n1); (m2; n2); : : :} corresponding to the two series defining the best alignment. Then, the best alignment cost C can be computed for the time-aligned signals, using a predefined distance function. By substituting an appropriate distance function this algorithm can be extended to multidimensional signals. For our work, we use Mahalanobis distance [19] which normalizes for unequal variances along different feature dimensions. This is useful when combining features with different ranges or units. For matching two sequences of length M and N each, we pre- compute this distance matrix of size M × N.As mentioned, classification is one of the possible applications of Audio Content Analysis. In a classification system, the input signal is analyzed and a label describing that signal is delivered at the output. There are many possible criteria upon which the signals can be separated, such as timbre content, pitch or musical tempo. But in the context of multimedia data managing and browsing, the most useful classification is the one that assigns labels describing the type or category of the signal, that is, if the analyzed sound is speech, noise, if it belongs to a certain music genre, etc.

$$k = \arg\min_{j} C_{ij}$$

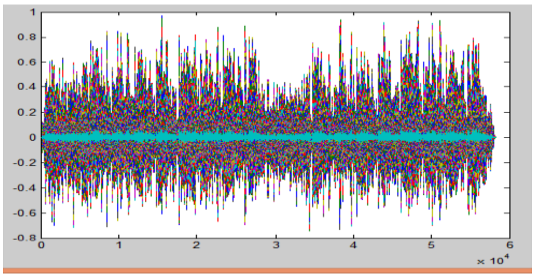


Figure 5. Testing for Hamming Window Function with One Song

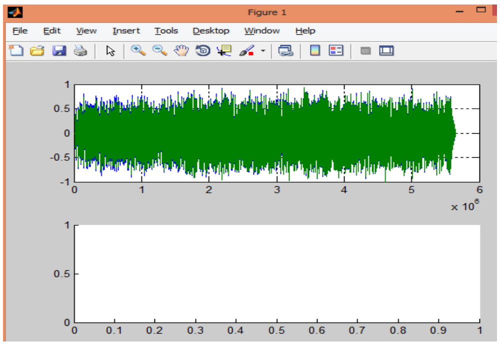
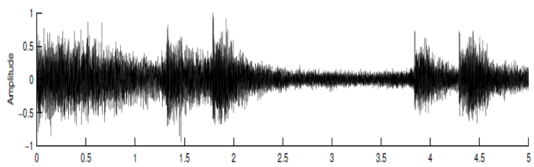
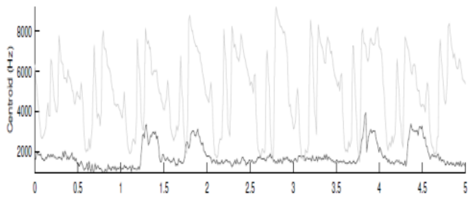


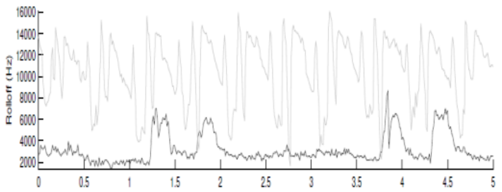
Figure 6. Testing for Preprocessing stage with One Song



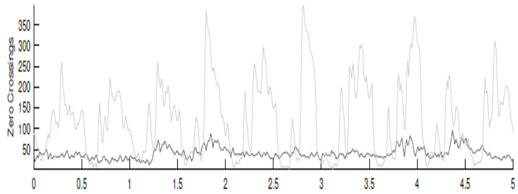
(a) Kachin song example



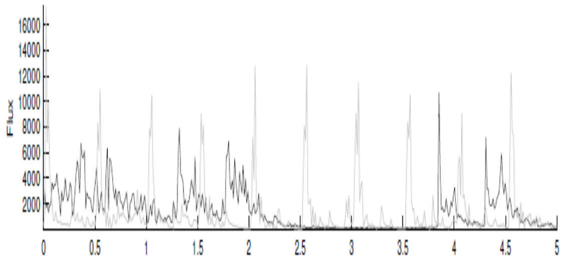
(b) Centroid



(c) Rolloff



(d) Zero crossings



(e) Spectral Flux from the Examples Plotted in Figure (a) and (b)

Figure 7. Song Examples and Timbral Features

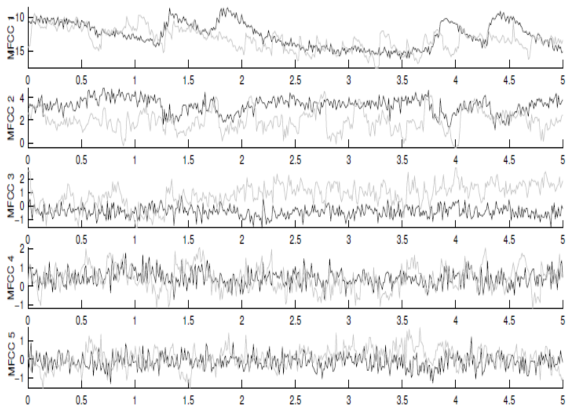


Figure 8. First Five MFCCs from the Kachin Song Examples Plotted in Figs.

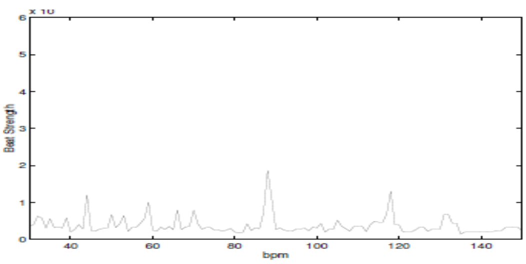
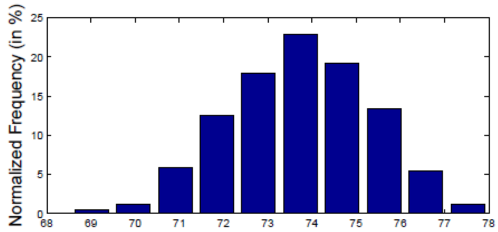


Figure 9. Beat Histogram of Kachin Song examples. High Peaks Correspond to High Beat Strength.



Percentage of energy retained in the power spectrum

Figure 10.Histogram Showing Effects of Feature Sampling on the Frequency Content of the Energy.

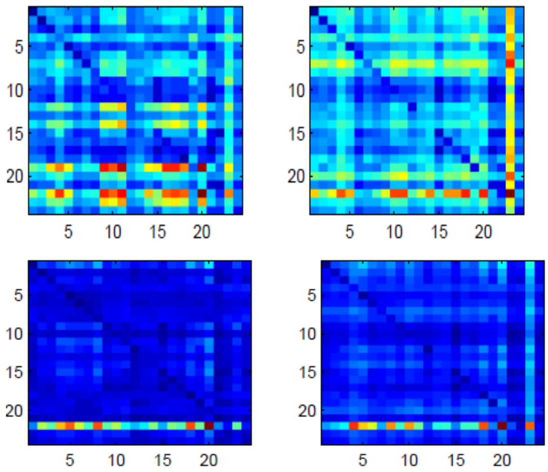


Figure 11. 24*24 DTW Alignment Cost Metrics for the Features.

6. Conclusion

In this paper, music classification system for Myanmar ethnic traditional songs is considered. The system is to describe Kachin, Kayin, Shan and Yakhine ethnic group songs are used for experiments. The categorization of features can be done on the basis of pitch, timbre, MBAP and rhythm and give a better description of the audio signal. From a signal processing point of view, the system use the Multi Band Autocorrelation Peak (MBAP) features, extracted in multiple bands providing complementary information which helps to improve the accuracy. Results obtained on a classification task suggest that these features can outperform traditional features which capture information from the entire spectrum. Alignment cost using the dynamic time warping algorithm was used a classification metric on a dataset of songs in this paper.

References

[1] Vidwans and P.Rao , “Identifying Indian Classical Music Styles using Melodic Contours” Department of Electrical Engineering Indian Institute of Technology Bombay, Mumbai 400076, India, Jan. 2012.

[2] C. Zhen and J. Xu."Multi-modal music genre classification approach" , In Intl. *Conf. on Computer Science and Information Technology*, vol. 8, (2010), pp. 398–402.

[3] C. Zhen and J. Xu."Solely tag-based music genre classification", In *Intl. Conf. on Web Information Systems and Mining*, (2011), pp. 20–24.

[4] J. Serra, E. G´omez, and P. Herrera, "Audio cover song identification and similarity: background, approaches, evaluation, and beyond," *Advances in Music Information Retrieval*, pp. 307–332, 2010.

[5] L. Mutiara Dewi , "Incorporating Global Information In a Folk Song Classification System "HAIT Master Thesis series nr. 11-005 (2012).

[6] R. Typke, F. Wiering, and R. Veltkamp, "A survey of music information retrieval systems", 2005.

[7] S. H. Mao and E. E. P. Myint, "Performance Comparison of Multi-Class SVM Classification for Music Cultural Style Tagging", *International Journal of Computer Theory and Engineering*, Vol. 5, No. 2, April 2013.

[8] T.Eeoral., Järvinen, T., Louhivuori, J. & Toiviainen, P. " A method for comparative analysis of folk music based on musical feature extraction and neural networks " *7th International Symposium on Systematic and Comparative Musicology 3rd International Conference* on Cognitive Musicology 2001 Jyväskylä, Finland .

[9] W.,Chai,&Vercoe, B., "Folk Music Classification Using Hidden Markov Model", Proceeding of *International Conference on Artificial Intelligence*, 2001.

[10] Y. Liu *et al., "*Cultural Style Based Music Classification of Audio Signals," in Proc. *IEEE international conference* on Acoustic Speed and Signal Processing, pp. 57-60, 2009.

[11] Z. Juhász, "Classification and Comparison of Different Folk Music Traditions Using Self Learning Algorithms" Research Institute for Technical Physics and Materials Science, Hungary. 2010.

[12] Z. W.Bond "Unsupervised Classification of Music Signals: Strategies Using Timbre and Rhythm" Virginia Polytechnic Institute and State University, Virginia, Nov. 2006.