Recognition and Classification of Accompanying Audios of Kathak Dance

Anirban Saha and Indranil Sarkar

August 1, 2021

Recognition and Classification of Accompanying Audios of Kathak Dance

Anirban Saha*, Indranil Sarkar†
*Department of Computer Science and Engineering,
Indian Institute of Technology Kharagpur, India.
†Department of Information Technology,
National Institute of Technology Raipur, India.
Emails: anirbansaha@iitkgp.ac.in*, isarkar.phd2018.it@nitrr.ac.in†

***Abstract*—Here a novel method is been implemented to recognize and classify the accompanying audios used in Kathak Dance. To achieve this Ladi dataset have been used. This audio dataset has been analysed and feature extraction is done. Thus audio and speech processing plays an important part here. Using this dataset beat detection is carried out to detect onset of the beats and then beat tracking is done. Then recognition and classification of Ladi audios is done based on rhythms and mfccs into 7 classes namely L2, L3, L4, L5, L6, L7 and L8 using machine learning based approach.**

***Index Terms*—Kathak dance, Machine Learning, Audio Analysis, Feature Extraction, Audio Recognition, Audio Classification, Audio and Speech Processing, K-Nearest Neighbors.**

I. Introduction

Here an initiative have been taken to develop a method to automatically recognize and classify the accompanying audios used in Kathak Dance [1]. To achieve this Ladi dataset has been used. This Ladi dataset contains accompanying audio files of pre-recorded kathak dance in mono format. The length of each audio file is minimum of 1 minute. Then a preliminary pre processing is performed to remove the unwanted noise from each of the audio files. Then using this processed dataset beat detection is performed first [2]. This beat detection is based on an efficient technique of detecting onset. This onset refers to the very instant of time that marks the starting of the transient sound part, or the moment that is the earliest at which a brief sound can be reliably detected [16]. Then beat tracking is performed [3]. After that recognition and classification of Ladi audios based on rhythms and mfccs into 7 classes namely L2, L3, L4, L5, L6, L7 and L8 is carried out using K Nearest Neighbors. approach [4] [17].

II. Literature Survey

A. Similar dance form - Bharatanatyam

It is an Indian classical dance form, which is a symbol of the rich and vibrant cultural heritage of India. Recognition and analysis of such vibrant dance forms are very important for the preservation of such cultural heritage [13]. Like in many traditional dance forms, a bharatanatyam dancer performs dance in synchronization with rhythmic structured music, which is called sollukattu, that is comprised of instrumental beats with vocalized utterances which is called bols to create a rhythmic music structure [14]. Thus the analysis of bharatanatyam, requires an indepth structural analysis of these sollukattus which are very difficult to identify. Speech processing techniques has been used for recognizing bols [5]. This is done by exploiting the already defined structures of the sollukattus and the bols that has been detected and recognized. Tempo period has been measured by using two methods [6]. Lastly, full annotation of the audio signal is been generated through beat marking. Information of the beats that has been detected from onset envelope of sollukattu signal has been used [7]. For training and testing, Sollukattus dataset has been created and has been annotated. 85% of accuracy has been detected in recognition of bols, 95% accuracy in recognition of sollukattus, 96% accuracy in estimation of tempo period, and more than 90% in case of beat marking has been achieved [19]. This is one of the first attempt to perform full and structural analysis of music of a classical Indian dance form by using audio speech and signal processing techniques for marking beats.

B. Estimation of time signature, micro time and tempo from percussive music

Time signature, micro time and tempo are some of the basic rhythmic and important features for a whole wide range of musical analysis [15]. So automatic extraction of these features from the music are required for a large number of applications like in entertainment applications, automatic transcription in music information retrieval systems, musicological analysis, authoring tools or in educational purposes [8]. The presented analysis procedure estimates time signature, micro time and tempo from percussive music [9]. From a total of 117 excerpts each of eight seconds in length, 73.5% of the time signatures, 84.6% of the tempo values and 83.8% of the micro time were been estimated accurately [18]. A large amount of research work is concerned with the extraction of meta data automatically from the musical audio signals.

C. Automatic Detection of Hindustani Talas

A novel approach has be implemented to develop an automatic recognition system of hindusthani talas that can be easily trained by construction of labeled corpus consists of hindusthani songs accompaniment by tabla [10]. Detection of talas can be achieved from the bol patterns . Here transcription of bols in the context of polyphonic can be an important area of research [11]. The weka platform was very helpful and useful for the part dealing with the machine learning although the python library used for mixture models, pymix are a great source of alternative when gaussian mixture models been used [12].

III. Methods used in Audio Signal Processing

A. Feature Extraction

Some of the common audio feature extraction techniques have been applied like spectral centroid, spectral roundoff, spectral bandwidth, zero crossing rate, Mel-Frequency Cepstral Coefficients (MFCCs), chroma feature etc for visualization purpose. Some of these features will be used for the classification and recognization of acompanying audios for kathak dance.

B. Beat Detection

Beat detection refers to the process of using computer hardware or computer software to detect and spot the beat of a musical composition. There are a lot of methods available for beat detection and it is always a compromise between speed and accuracy. In media player plugins beat detectors are very commonly used for visualization purpose. It uses many statistical model algorithms which are based on sound energy or it may involve comb filters or other means. They can be fast enough so that the can run in real time or be slow and are only be able to analyze small section or length of songs.

C. Novelty Function

To perform the detection of note onsets, sudden changes in the audio signal must be located which marks the starting of transient regions. Most of the times an onset candidate is detected by an increase in signal's amplitude envelope. But sometimes that is not the scenario because note onset might change from one particular pitch to another but without affecting or changing the amplitude, for example a guitar being played in slurred notes. Novelty functions indicate local changes in properties of signal like spectral content or energy. Here spectral based novelty function is used.

D. Spectral Novelty Function

We computed a spectral based novelty function by -
1. Compute and calculate the log amplitude spectrogram.
2. Within each and every frequency bin denoted by 'k', compute and calculate the energy based novelty function by evaluating
   (i) 1st-order difference,
   (ii) Half-wave rectification.
3. Perform sum across all the frequency bins which is denoted by 'k'.

We used Onset strength function to compute and calculate the novelty based function using spectral flux.

E. Peak Picking

As the name suggests it is the process of identifying peaks in an audio signal. Lets for example say we may want to detect and identify peaks in a novelty function when we are performing onset detection. These detected peaks would co-relate to the onsets of the music.

F. Onset Detection

Automatic recognition and identification of musical events in an audio signal is one of the primary and important tasks in musical information retrieval system. Here, it has been shown how to recognise/detect/identify an onset in accompanying audios of kathak dance, from the very instant that marks the starting of the transient part of a audio signal, or the earliest moment at which a transient part can be genuinely detected.

G. Onset Detection Algorithm

The steps are -
1. Compute spectral based novelty function.
2. Find peaks that exists in the spectral based novelty function.
3. Perform backtracking from each and every peak to a preceding local optimum which is the local minimum here.

Backtracking are very helpful for finding the segmentation points when the onset just occurs shortly after the staring of the segment.

H. Onset Based Segmentation

We used onset based segmentation with backtracking. We used the onset detection algorithm previously discussed to find peaks in a spectral based novelty function. But sometimes, these peaks may not really coincide with the preliminary rise in energy. So to perceive the starting of a musical note backtracking has been used from every peak to a preceding local optimum (minimum). Backtracking are often very useful for finding out the segmentation points so that the onset occurs just shortly after the starting of the segment. By this way glitches have been avoided by backtracking from the detected onsets.

I. Beat Tracking

It is the process of deriving a sequence of beat instants from a musical audio signal that might correspond to when a human listener would tap his foot for example.

J. Tempo Analysis

The rate of musical beat is called tempo. It indicates the speed of musical piece. It is given by the reciprocal of the beat period. It's unit is BPM i.e beats per minute.

K. Tempogram

There may be a great variation of tempo locally within a piece. Thus tempogram is being used as a feature matrix to show the prevalence of some specific tempi at each sequence in time. Fourier tempograph is basically the magnitude spectrogram of the novelty function. Auto-correlation tempogram is used wherever the auto-correlation is high and is a good candidate of the beat period. The auto-correlation is useful for finding repeated patterns in a signal. For example, at short lags, the auto-correlation can tell us something about the signal's fundamental frequency. For longer lags, the auto-correlation may tell us something about the tempo of a musical signal. This principle of auto-correlation to estimate the tempo at every segment in the novelty function has been used i.e a short-time auto-correlation of the (spectral) novelty function is used.

L. Rhythm Detection

Rhythm is any regular recurring motion/sound or symmetry marked by controlled succession of powerful and light elements of different or opposite conditions. This general meaning of regular pattern or recurrence in time can be applied to a wide range of cyclical inherent phenomena having a frequency or periodicity of anything between microseconds in lower side to that of several seconds in higher side.

IV. Proposed Method

The proposed method is based on beat detection and beat tracking. We computed the spectral based novelty function for each audio data files. Then we performed peak picking and count of peaks from spectral based novelty function. Sometimes peaks might not always coincide with the initial energy rise. Thus, to perceive the initial beginning of the musical note, backtracking has been used from each and every peak to the preceding local optimum which is the local minimum in this case. Subsequently, we perform onset detection followed by tempo estimation. For visualization purpose these are shown in from figure 1 to figure 5. In order to classify and recognise the accompanying audios in kathak dance, we finally select the peak count and MFCCs as features for our K-Nearest Neighbors machine learning algorithm. We computed first 20 MFCCs from each sampled audio data. From these 20 MFCCs we find the MFCCs mean and the co-variance. Then we constructed the feature vector by taking the MFCCs mean and co-variance matrix together. The heart of KNN is it's distance function which calculates the distance between feature vectors. We computed the distance of each feature vector. Its nearest neighbors are identified using this distance metric. First we performed a training and testing split. Then we performed a K fold cross validation on training set where k is taken as 10.

V. Results And Discussions

In this paper, Ladi dataset has been used. It consists of 182 audio files. These 182 audio files can be classified into 7 classes which we call Ladis. These Ladis has been categorized as L2 to L8. Categorization and classification has been done using a supervised machine learning algorithm called K-Nearest Neighbors (KNN). We have generated synthetic dataset from the original ones to increase the number of audio files in the dataset from 182 to 1400. After-that we have performed 3 types of observation based on training and testing split ratio. These splits are 90:10, 80:20 and 70:30 where the first part of the ratio is for training set and latter is for testing set. In first case we got 95.2% and 93.4% training and testing accuracy. Similarly for second case we got 90.9% and 88.4% and third case we got 85.5% and 82.7% training and testing accuracy. In case of 80:20 or 70:30 split some of the L7 variations were classified as L8 class for example L7V1D1R1.wav which is a L7 class audio file in Ladi dataset is falsely classified as L8 class, though the L2 to L6 predictions were accurate. As the number of beats increased the rate of misclassification also got increased. We use different combination of features in order to classify the audio data into respective 7 different classes. These features include peak count, zero crossing rate, chroma feature, spectral centroid, spectral roll-off, spectral contrast, spectral bandwidth, root mean square energy (RMSE), Mel-Frequency Cepstral Coefficients (MFCCs) to name a few. But we got the best result when we use MFCC from 1 to 20 i.e first 20 MFCC coefficients.

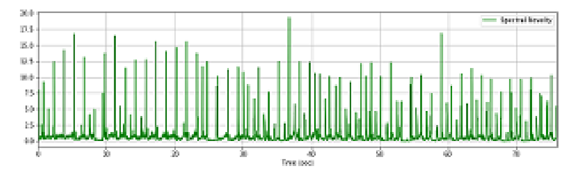

Fig. 1. Spectral novelty function plot against the time domain of a sample audio file named L2V1D1R1.wav from the Ladi dataset.

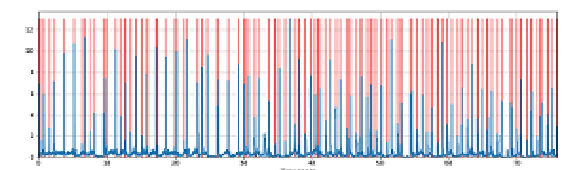

Fig. 2. Onset envelope plot with peak of a sample audio file named L2V1D1R1.wav from the Ladi dataset.

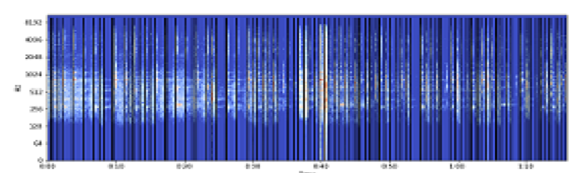

Fig. 3. Spectrogram plot during onset based segmentation of a sample audio file named L2V1D1R1.wav from the Ladi dataset.

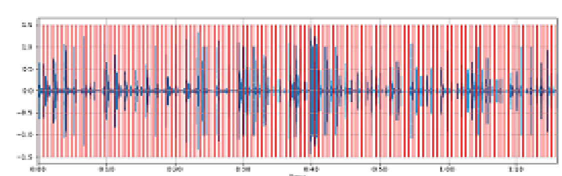

Fig. 4. Tempo estimation plot of a sample audio file named L2V1D1R1.wav from the Ladi dataset.

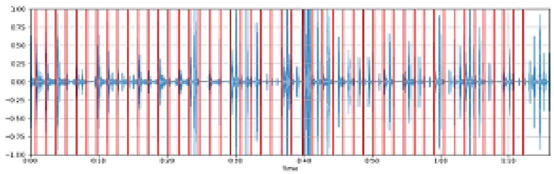

Fig. 5. Beat location shown over wavelength plot of a sample audio file named L2V1D1R1.wav from the Ladi dataset.

Spectral novelty function plot, onset envelope plot with peak, spectrogram plot during onset based segmentation, tempo estimation plot and beat location over wavelength plot has been shown in fig. 1 to fig. 5, respectively, for visualization purpose. As depicted in fig. 6, which is the best result we got in a 90:10 ratio training and testing split. We get a training accuracy of 95.2%. Validation accuracy is near about 93.5%. Testing accuracy is 93.4% respectively as previously mentioned using the above mentioned machine learning approach. Since here we are classifying the data into 7 classes so the value of K has been chosen as 7.

Here python 3.7 has been used to generate all the plots as shown from Fig 1. to Fig 6. respectively.

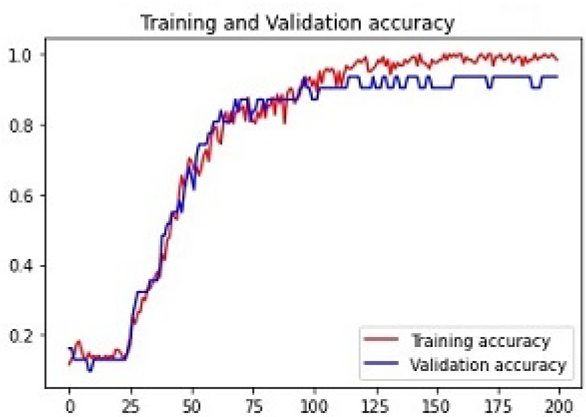

Fig. 6. Training vs validation accuracy in case of 90:10 training and testing split of the dataset.

VI. Conclusion

Here we have come up with an efficient approach in automatic classification and recognition of accompanying audios of Kathak dance with very high accuracy which is generally quiet difficult to achieve. This is due to its complicated nature of different taals and bols which needs to be classified accurately into 7 classes namely L2, L3, L4, L5, L6, L7 and L8. We achieve this using a supervised learning technique called K-Nearest Neighbors on Ladi dataset. There is also scope for future improvement on this work where each classes can again be classified into sub classes based on number of bols and taals in the accompanying audios of kathak dance. Using Convolution neural network (CNN), more robust result can be achieved which is in experimental phase as of now.

Acknowledgment

We are very thankful to Dr. Partha Pratim Das, Professor, Department of Computer Science and Engineering, Joint Principal Investigator, National Digital Library of India Project and Editor-in-Chief of Journal of The Institution of Engineers (India): Series B, Indian Institute of Technology Kharagpur and to Mrs. Saptami Ghosh, Phd scholar of Indian Institute of Technology Kharagpur for guiding through this research work.

References

[1] Chakravorty, J., B. Mukherjee, and Ashok Kumar Datta. "Some Studies in Machine Recognition of Ragas in Indian Classical Music." Journal of the Acoust. Soc. India 17.3 and 4 (1989).

[2] Rao, Vishweshwara, Pradeep Gaddipati, and Preeti Rao. "Signal-driven window-length adaptation for sinusoid detection in polyphonic music." IEEE transactions on audio, speech, and language processing 20.1 (2011): 342-348.

[3] Schedl, Markus, Emilia Gómez Gutiérrez, and Julián Urbano. "Music information retrieval: Recent developments and applications." Foundations and Trends in Information Retrieval. 2014 Sept 12; 8 (2-3): 127-261. (2014).

[4] Kumar, Vijay, Harit Pandya, and C. V. Jawahar. "Identifying ragas in indian music." 2014 22nd International Conference on Pattern Recognition. IEEE, 2014.

[5] Hu, Ming-Kuei. "Visual pattern recognition by moment invariants." IRE transactions on information theory 8.2 (1962): 179-187.

[6] BARTSCH, MA. "Using Chroma-based Representations for Audio Thumb-nailing." Proceeding of WASPAA'01 (2001).

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] Dighe, Pranay, Harish Karnick, and Bhiksha Raj. "Swara Histogram Based Structural Analysis And Identification Of Indian Classical Ragas." ISMIR. 2013.

[9] Felix Weninger, Noam Amir, Ofer Amir, Irit Ronen, Florian Eyben, Bjorn Schuller, Robust feature extraction for automatic recognition of vibrato singing in recorded polyphonic music, Proceed. 37th International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, pp. 85–88, (2012).

[10] Miron, Marius. "Automatic detection of hindustani talas." Master's thesis, Universitat Pompeu Fabra, Barcelona, Spain (2011).

[11] Srinivasamurthy, Ajay, and Xavier Serra. "A supervised approach to hierarchical metrical cycle tracking from audio music recordings." 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014.

[12] Tzanetakis, George, and Perry Cook. "Musical genre classification of audio signals." IEEE Transactions on speech and audio processing 10.5 (2002): 293-302.

[13] Weiss, Christof, and Meinard Müller. "Tonal complexity features for style classification of classical music." 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015.

[14] Mitri, George, Alexandra L. Uitdenbogerd, and Vic Ciesielski. "Automatic music classification problems." Proceedings of the 27th Australasian conference on Computer science-Volume 26. 2004.

[15] Chew, Elaine, Anja Volk, and Chia-Ying Lee. "Dance music classification using inner metric analysis." The next wave in computing, optimization, and decision technologies. Springer, Boston, MA, 2005. 355-370.

[16] Bhatt, Malay, and Tejas Patalia. "Neural network based Indian folk dance song classification using MFCC and LPC." Int. J. Intell. Eng. Syst 10.3 (2017): 173-183.

[17] Vidwans, Amruta, and Preeti Rao. "Identifying Indian classical music styles using melodic contours." Proc. of Frontiers of Research on Speech and Music, Gurgaon, India (2012).

[18] Uhle, Christian, and Juergen Herre. "Estimation of tempo, micro time and time signature from percussive music." Proc. Int. Conference on Digital Audio Effects (DAFx). 2003.

[19] Mallick, Tanwi, Partha Pratim Das, and Arun Kumar Majumdar. "Beat Detection and Automatic Annotation of the Music of Bharatanatyam Dance using Speech Recognition Techniques." arXiv preprint arXiv:2004.08269 (2020).