GENRE CLASSIFICATION AND ANALYSIS OF MARATHI SONGS

Shreyas Nadkarni
Department of Electrical Engineering
Indian Institute of Technology Bombay
shreyasnadkarni@ee.iitb.ac.in

Preeti Rao
Department of Electrical Engineering
Indian Institute of Technology Bombay
prao@ee.iitb.ac.in

ABSTRACT

We investigate musical genre classification for three forms of Marathi vocal music each with a distinct socio-cultural context, namely devotional, poetic and folk dance. We present a dataset of songs covering the three genres, and discuss their musical and acoustic characteristics. We consider timbre and chroma based features for the 3-way classification. We specifically examine whether the source-separated vocal and instrumental accompaniment components can help improve genre recognition accuracy over that obtained with acoustic features extracted from the original mix audio track.

1. INTRODUCTION

Music has always been a distinctive form of entertainment and expression in Maharashtra. This work attempts to classify Marathi songs based on their genres and thereby analyze the variation seen in them despite being in the same language and from the exact geographical location. 3 genres from Marathi music: Bhaktigeet, Bhavgeet, and Lavani, have been chosen. The contributions of this work are as follows:

1. A new dataset of over 150 Marathi songs spanning 3 genres, with accompaniment, vocal and mix (original) tracks, spread over multiple artists.
2. A new approach using source separation into vocal and accompaniment components to study acoustic differences between genres
3. An analysis of audio features in these songs with machine learning to find which attributes are most prominent in determining the genre

The most widely used features in musical genre recognition are timbre features that capture instrumentation differences via spectral energy distributions. These include the MFCCs, spectral centroid and spectral roll-off. Kini et al. [1] to classify Bhajans and Qawwalis, two devotional forms of North Indian music using Support Vector Machines and Gaussian Mixture Models, making use of timbral features. Vidhwans et. al [2] classified Hindustani, Carnatic and Turkish Music using newly defined audio features based on energy and microtonality, such as the "Gamak Measure", and validated the dependence of features on melodic cues. H. Bahuleyan [3] presents a comparative study of different classification methods for identifying features which contribute most towards distinguishing genres in western music. The standard approach which has worked for classifying music is using convolutional neural networks on spectrogram images derived from the audio clips to build classifier models, as done by Pelchat et al. [4]. A detailed discussion of methods to extract meaningful features from audio for analysis can be found in the thesis by P. Grosche [5].

2. DATASET DESCRIPTION

A total of 154 songs were collected across the three genres, covering both male and female artists. Each of these songs is approximately 3-4 minutes long in duration. The number of songs and the total duration of audio across the three genres are given in Table 1. The audio format of these tracks was ’mono’ and the sampling rate was 44.1 kHz.

Genre	No. of Songs	Total Duration
Bhaktigeet	53	188 min 38 sec
Bhavgeet	53	185 min 11 sec
Lavani	48	173 min 33 sec

Table 1: Details of the dataset

The Bhaktigeet, a genre of devotional songs, has a relaxed melodic style with steady notes with vibrato, aiding in mindfulness for the listener. Another prominent feature in Bhaktigeet is the ‘meend’ which denotes a glide from one note to another. The Bhavgeet has a higher number of notes in a given time duration than a Bhaktigeet, and a fast tempo is usually maintained for a playful or a romantic mood, similar to an Indian filmy song. The local fluctuation in pitch is seen maximum in the Lavani, which is a folk dance genre, with a tempo which is usually fast and the focus being primarily on the rhythm rather than melody.

3. METHODS

3.1 Feature Extraction

Each song was split into 30s excerpts such that the starting time of two consecutive excerpts is separated by 10s.

Attribution: S. Nadkarni, and P. Rao, “Genre Classification and Analysis of Marathi Songs”, in *Extended Abstracts for the Late-Breaking Demo Session of the 23rd Int. Society for Music Information Retrieval Conf.,* Bengaluru, India, 2022.

From the 154 songs, we got a total of 2913 excerpts. These excerpts were used to extract several audio features such as chroma stft, root mean square value, spectral centroid, spectral bandwidth, spectral roll-off, zero-crossing rate, and Mel Frequency Cepstral Coefficients (MFCCs): 20 in number for each audio clip, resulting in a feature vector of length 26. The values taken were the mean values of the parameters over the entire 30s excerpt. For the short term analysis, a window length of 46.44 ms (2048 samples, which was also the DFT length) and a hop length of 512 samples were used. 'Chroma_stft' denotes the mean of the normalised energies for all chroma bins at all frames. Spectral centroid is a measure of the 'brightness' of the sound. The bandwidth measures the spread of frequencies present in the audio signal. The spectral roll-off is the frequency below which a specified percentage of the total spectral energy (here 85%) lies. The zero crossing rate is a measure of the signal's noisiness and a measure of frequency content. The MFCCs of a signal are a small set of features (here 20 in number) that concisely describe a spectral envelope's overall shape.

Further, using the Spleeter [6] construct, each of these excerpts was source separated into a vocal track and an accompaniment track. These gave two more parts in the dataset that were used similarly to extract the audio features. The dimension of the whole dataset hence was 2913 × 26 × 3. The features were extracted using Librosa [7].

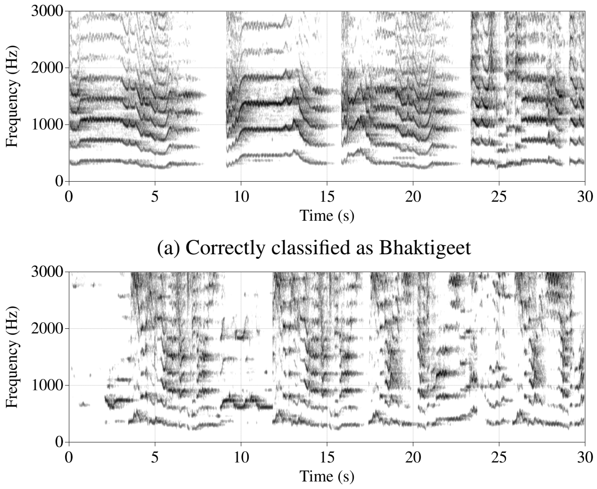
3.2 Classification

The model hyper-parameters for the support vector machine, trained using Scikit-Learn [8], were selected by cross-validation on the training data. Three different models were trained: one on the data generated from accompaniment track, the second from vocal track and the third from the original (mix) track of the excerpts. Each of the three parts in the data was split into training and testing parts in the ratio 4:1 using a group-shuffle split to ensure that there was no song with excerpts in both datasets, in order to evaluate the performance of the model(s) on unseen songs. The confusion matrices obtained for the three models are presented in Figure 1.

		Accompaniment			Vocals			Mixed		
		PREDICTED			PREDICTED			PREDICTED		
		Bk	Bv	L	Bk	Bv	L	Bk	Bv	L
TRUE	Bk	24	42	22	67	2	19	46	42	0
	Bv	35	75	31	26	74	41	12	83	46
	L	7	13	66	28	3	55	4	0	82

Figure 1: Confusion Matrices for the three models: Bk: Bhaktigeet, Bv: Bhavgeet, L: Lavani. The values shown are for excerpt-level classification.

The excerpt-level accuracy values obtained on the test data for these models were: 52% for Accompaniment, 62% for Vocals and 67% for Mix. For each song in the test data, the genre predicted for a majority of the excerpts of that song was considered as the predicted genre for the whole song. Using this approach, the song-level performance was found. Out of 16 songs in the test data, the accompaniment model misclassified 7 songs, the vocals model misclassified 5 songs, while the mix model misclassified 6 songs. The songs in the test data were analysed to find which excerpts in each were classified correctly.



(a) Correctly classified as Bhaktigeet

(b) Wrongly classified as Lavani

Figure 2: Spectrograms of two vocal excerpts from the same Bhaktigeet song classified differently. It is seen that the excerpt with steady notes (top) is classified as Bhaktigeet while the one with high fluctuations (bottom) in pitch is classified as Lavani.

4. RESULTS AND DISCUSSION

The spectrograms were plotted for some selected excerpts from the test dataset using Praat [9]. It is seen that the Bhaktigeet genre has more of steady notes, more than that in Bhavgeet, and it is the least in Lavani. The Lavani genre has more of fluctuating sound intensities and fast-changing pitch patterns. Development of a feature which captures the dynamics (constancy and steadiness in notes) will be instrumental in distinguishing the three genres from each other. Two sample waveforms obtained from Praat on two excerpts from the test data are presented in Figure 2.

The vocal audio excerpts give better results than their accompaniment counterparts which validates the fact that the major difference between these genres stems from the difference in the singing style (such as steadiness of notes). The relatively low performance of the models can be attributed to the dataset's small size, the model's simplicity, the overlap between the genres themselves across songs and across excerpts within the same song, and the imperfect splitting by Spleeter obtained on Marathi songs. Spleeter has been developed for western music and it does not handle Indian instruments such as the sitar well.

Future work can be based on experimenting deep learning methods for classification, development of new acoustic features for distinguishing these genres, and developing a new method for source separating Indian music. We hope that the constructed dataset and this analysis can help future research on the acoustics of Marathi music.

5. REFERENCES

[1] S. Kini, S. Gulati, P. Rao, "Automatic Genre Classification of North Indian Devotional Music", Proceedings of the National Conference on Communications (NCC), Jan 2011, Bangalore, India

[2] A. Vidwans, P. Verma, P. Rao, "Classifying Cultural Music using Melodic Features", Proceedings of SP-COM 2020, I.I.Sc Bangalore, July 2020

[3] H. Bahuleyan, "Music Genre Classification using Machine Learning Techniques", ResearchGate, 2018

[4] N. Pelchat, C. Gelowitz, "Neural Network Music Genre Classification", Canadian Journal of Electrical and Computer Engineering, 2020

[5] P. Grosche, "Signal Processing Methods for Beat Tracking, Music Segmentation, and Audio Retreival", Thesis, ResearchGate, January 2013

[6] Spleeter Source Separation, by Deezer: https://github.com/deezer/spleeter

[7] Librosa (Python package for Audio analysis): https://librosa.org/doc/latest/index.html

[8] Scikit Learn Documentation: https://scikit-learn.org/0.21/documentation.html

[9] Boersma, Paul & Weenink, David (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.08, retrieved 5 February 2022 from http://www.praat.org/