Genre Classification and Analysis of Marathi Songs
Summer Undergraduate Research Project

Shreyas Nadkarni
Department of Electrical Engineering
IIT Bombay
Mumbai, India
19D170029@iitb.ac.in
Under the guidance of Prof. Preeti Rao, Department of Electrical Engineering, IITB

***Abstract*—This project has been done under the Summer Undergraduate Research Project (SURP) Program, by UGAC IIT Bombay. It is an attempt to train a genre classifier model on Marathi songs from 3 genres, and check its performance on various songs, thereby analysing the acoustic differences between the genres. The SVM construct has been used with the help of the scikit-learn library of Python. A method of source separating the audio into vocals and accompaniment and training the model on them separately has also been experimented. This report presents an overall summary of the motivation behind the work, the experiments and results and the insights gained along with some related work that has been done and can be taken up in future. The data of features extracted and the code for analysis can be found in the following repository:** [1]

I. Introduction

Music has always been a remarkable form of entertainment and expression in the state of Maharashtra. Marathi music with its various genres has got recognition from almost all parts of India as well as many parts of the World. This project is an attempt to classify Marathi songs based on their genres and thereby do an analysis of the variation which is seen in them in spite of being in the same language and from the same geographical location. By doing so, it might be possible to explore the differences in the acoustic features of songs from these different genres and find which components of the audio signal work best in distinguishing them. For this study, 3 genres from Marathi music, namely Bhaktigeet, Bhavgeet and Lavani have been chosen. This is a new approach of analysis by using Machine Learning to classify the songs and then analysing the mis-classifications, trying to understand the reasons behind them. 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. These are genres on which no work has been done before and the dataset can be extended and used for further research on Marathi music.
2) Presenting information about the chosen genres about which very less documentation is available as of now.
3) An analysis of the audio features in these songs and comparing their values for the different genres with an attempt to use machine learning to gain insights about which components of a song are most promiment in determining the genre

II. Related Work

Music Genre Classification has been a very popular topic in the field of Music Information Retrieval. Classifying songs based on their genres is extremely useful for applications like music recommender systems. In [1], we see an approach by S. Kini et. al to classify Bhajans and Qawwalis, two devotional forms of North Indian music using Support Vector Machines and Gaussian Mixture Models, making use of timberal features. A. Vidhwans et. al in [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 in [3] presents a comparative study of different classification constructs including linear regression, random forests, support vector machines, gradient boosting and convolutional neural networks, in identifying features which contribute most towards distinguishing genres in Western music. N. Maddage et. al in [4] used a multi-layer SVM model for a genre classification task, which involved using two levels of SVMs in series. The standard approach which has worked for classifying music is the approach of using Convolutional Neural Networks on Spectrogram images derived from the audio clips to build classifier models, as can be seen in [5] along with comparison with other approaches. Music Information Retrieval has been a popular application of several signal processing techniques. A detailed discussion of methods to extract meaningful features from audio for analysis, can be found in the thesis by P. Grosche [6]. Several online blogs offer interesting insights into the machine learning techniques and related work, and can be found at [7].

[1] https://github.com/Shreyas789/Music-Genre-Recognition

III. About the Genres Chosen

- Bhaktigeet: It means 'devotional song'. The use of tabla and cymbals (there are various varieties like zanj, taal, etc) is very common. Melody instruments like harmonium, flute, sitar, etc are also used. The rhythm pattern ('taal') is usually 'Bhajni-theka' or 'Teen-taal'. It is very similar to 'Bhajans'. There are by both male and female artists but female singers are more common. Lord Vitthal is a highly idolized deity in some parts of Maharashtra (like Pandharpur, Alandi), and many verses, written by great saints of Maharashtra, have been composed as prayers to Him and in these the use of a traditional rhythm instrument called 'chipli' is also popular. The kind of music involved is mainly one that aids concentration, meditation or prayer (such as low-pitched beats and steady notes) and allows oneself to connect with Him. The tempo is varying across songs. The bhajni-theka is used very effectively in slow as well as fast tempos. Overall, this genre is closest to Hindustani classical music among the three genres.
- Bhavgeet: They are songs expressing emotion also called as "light music". They are songs which do not belong to any movie but are a part of albums, etc, composed on a certain theme such as love. Instruments such as guitar, violin, flute are used maximum in this genre (compared to the other two genres). Some songs (especially recent compositions) have a touch of Western music too. Rhythms belong to Indian Classical taals (major instrument: 'Tabla') or Western Rhythms and there is lot of variety. The tempo is usually slow, especially in songs having a high emotional value to be delivered through the words. There are both female as well as male artists, and they are almost equal in number.
- Lavani: They are songs almost always sung by female singers. They were very popular in traditional shows called 'tamashas' which were a major form of entertainment, before the advent of cinema and television. One of the most important and celebrated instruments used in this genre is the 'Dholak', 'Pakhawaj' and the 'Ghoongroo' (worn by dance performers near their ankles). The tempo is usually fast and the focus is on rhythmic beats rather than melody. A lot of compositions have 'komal swaras' (sombre notes) and ragas which are used to portray a romantic ('shrungara') theme. Several compositions start with a dholak artist playing for some time (in a loud volume usually to signal the start of a program/song) and then the main melody of the song. Melody is incorporated mainly through the harmonium.

Some Remarks

1) These are not exhaustive genres, nor are the boundaries between these genres strict. There are many other genres such as 'Powada', 'Bharud', 'Gondhal', 'Koligeet', etc, popular in specific parts of Maharashtra. Also, the instruments used are of a great variety and some instruments even have a lot of variants. The lack of research and documentation of instruments, their working and artists, especially from parts of rural Maharashtra makes it very tough to find and explore these. It is also difficult to identify such instruments clearlt upon listening a musical piece.
2) In recent times due to the advent of electronic instruments and the general technological development, many of these traditional instruments have become obsolete or are combined into a single electronic instrument like octopad or synthesizer/keyboard, which can be a reason of difference between old and recent songs (their audio features) since the sound from an electronic instrument is unavoidably different from that of an actual instrument.
3) Out of all the different genres we chose Bhaktigeet, Bhavgeet and Lavani for the genre classification task. These three genres are suitable because one can easily find songs of a similar duration (around 3-4 minutes). In other genres like Powada and Bharud , it is difficult to find such songs. Also, these contain some plain narration without melody which does not quite fit into an analysis of musical features. These three genres have a lot of songs and are probably more popular than the rest. Also, there are some overlapping features which can lead to some misclassifications, and on analysing these, one can get insights into the characteristics of the songs. Genres like Gondhal and Koligeet could have been selected however these are comparatively less popular and finding many songs of a similar duration is a difficult task. Due to these, we decided to work on Bhaktigeet, Bhavgeet and Lavani, and carry 3-genre classification.

IV. Implementation Details

A. The Dataset

A total of 154 songs were collected from Youtube videos. Each of these songs is approximately 3-4 minutes long in duration.

- Total length of Bhaktigeet audio (total 53 songs): 11318 sec = 188 min 38 sec. Average duration of song: 213.55 sec
- Total length of Bhavgeet audio (total 53 songs): 11111 sec = 185 min 11 sec. Average duration of song: 209.64 sec
- Total length of Lavani audio (total 48 songs): 10413 sec = 173 min 33 sec. Average duration of song: 216.94 sec

Audio format: Mono, Sampling rate: 44.1 kHz

Each song was split into 30s excerpts such that the starting time of two consecutive excerpts is separated by 10s (leading to an 'overlap' of 20s). The nomenclature followed for the

songs is "xxxx-yyy-zzz-30s.wav" where 'xxxx' is a four-digit ID for the excerpt (unique ID for each excerpt in the database) ,'yyy' encodes the genre and the song (with the first digit indicating the genre: 1 for Bhaktigeet, 2 for Bhavgeet and 3 for Lavani, and the other two indicating the serial number of that song in that genre) while zzz denotes the starting timestamp of the excerpt in the song in seconds. So an excerpt named "0023_102_050_30s.wav" contains a 30s excerpt which is the excerpt starting at timestamp 50s with ID 102 (song 02 from Bhaktigeet genre (1)) and this excerpt is numbered 0023 in the list of all excerpts (a total of 2913 excerpts, indexing starting from 0000). These excerpt audio files were loaded onto a Python notebook, and several audio features were extracted from them. The mean of the feature values was taken over the 30s excerpt resulting in one value per feature per excerpt. These features include 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. This provided a 'ready-to-use' dataset for implementing a genre classification task. Further, using the Spleeter construct developed by Deezer, each of these excerpts was source separated into a vocal track and an accompaniment track and stored as wav files separately. These gave two more parts in the dataset which were used in a similar fashion to extract the audio features. The whole data was split into training and testing datasets in the ratio 4:1 using a group-shuffle split in order to ensure that there was no song which had excerpts in both datasets (so all excerpts of any particular song were in either the training or the testing data). This was to check the performance of the model(s) on songs which it had not seen during training.

B. The Classification Task

The model hyper-parameters for the support vector machine were selected by cross-validation on the training data. Three different models were trained: one on the data generated from accompaniment tracks, the second one on that from vocal tracks and the third one from the original (mix) tracks of the excerpts. The excerpt level accuracies obtained on the test data for these models were: 52% for Accompaniment, 62% for Vocals and 67% for Mix. The confusion matrices obtained for the three models have been presented here:

Accompaniment			Vocals			Mix		
24	42	22	67	2	19	46	42	0
35	75	31	26	74	41	12	83	46
7	13	66	28	3	55	4	0	82

TABLE I: Confusion Matrices for the three models: Vertical Axis: True Genre, Horizontal Axis: Predicted Genre; 1: Bhaktigeet, 2: Bhavgeet, 3: Lavani

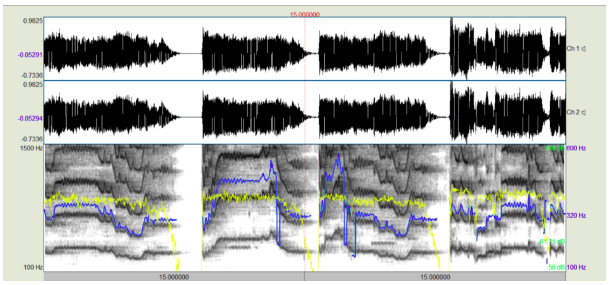
For each song in the test data, the majority of predictions towards a genre were considered, to predict a genre for the whole song. Using this approach, the song level performance was found. Out of 16 songs in the test data, the accompaniment model misclassified 7 songs, the vocals model misclassified 5 songs, while the mix model misclassified 6 songs. The variance in the results was quite high because our dataset is small and the accuracies can shoot or drop depending upon which songs fall into the testing data, if the experiment is repeated with a different random seed. However, the aim of the classification was to get a general idea of the differences between the genres and between the vocal and accompaniment components. The songs in the test data were analysed in detail to find how many excerpts in each were classified correctly.

Another task which was implemented was using the permutation importance method on the three models to find which features contributed the most in each model, in distinguishing the genres. The permutation importance score for a feature is defined to be the change in model score when that feature columns is randomly shuffled. In this implementation, for each of the three models, each feature was shuffled 10 times and the score was evaluated. The mean and standard deviation of the 10 scores was obtained. The mean is taken as an indicator of the importance of that feature in that model. Then an overall mean permutation score was calculated as the mean of the scores obtained in each of the three models, and this could be used as an estimator to get a rough idea of how each feature had performed in the classification task.

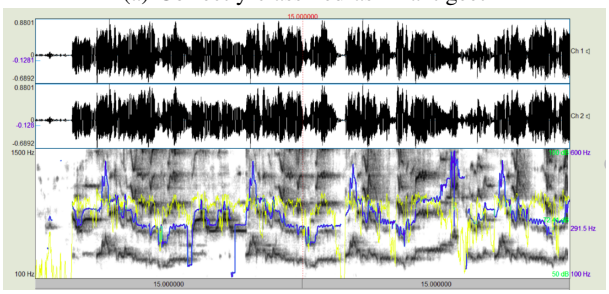
C. Observations

The pitch waveforms were plotted for some selected excerpts from the test dataset using Praat. It is seen that the Bhaktigeet genre has more of steady notes, which is more than that in Bhavgeet, and it is the least in Lavani. As listeners would agree, the Lavani genre has more of fluctuating sound intensities and fast-changing pitch patterns, which often makes it difficult artists to sing. The Bhaktigeet being a genre of devotional songs has more of a peaceful melodic style having steady notes with vibrato, aiding in mindfulness for the listener. The Bhavgeet might be considered to be midway as far as the melodic style and presence of steady notes is concerned, since the songs have more number of notes in a given time duration than a Bhaktigeet, and a fast speed is usually maintained for a playful or a romantic mood, similar to a 'filmy' Indian song. Another feature which is prominent in Bhaktigeet is the 'meend' which is a glide from one note to another, without breaking the voice (which might also be an attempt to aid in the continuity of the song for mindfulness). The local fluctuation in pitch is seen maximum in the Lavani, the tempo of which is also usually fast since they are essentially composed for a dance performance. The same general pattern can be observed for the sound intensity, with Bhaktigeet having a relatively steady intensity level of the voice across major parts of the song, and Lavani having the least because of various phonetic sounds which are expressive in nature, and also occasional breaks from the singing for chunks of plain speech with the accompaniment on in the background. If one is able to develop a feature which captures the steadiness in notes and in the intensity, it will be extremely useful in distinguishing the three genres from each other. Some

sample waveforms obtained from Praat on two excerpts from the testdata have been presented here:



(a) Correctly classified as Bhaktigeet



(b) Wrongly classified as Lavani

Fig. 1: Two vocal excerpts from the same Bhaktigeet song, but 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 was classified as Lavani

For the task of observing the performance of each feature using permutation scores, it was observed that the results varied across the three models. This can be attributed to the fact that each feature is responsible for capturing a particular musical attribute which is varying in accompaniment, vocals and the mixed songs. For example, the performance in the accompaniment model was completely dominated by the MFCCs (which seems probable since MFCCs are responsible for capturing the 'timbre' of the music, and the difference in instruments used in the three genres is the key distinguishing factor). The overall scores for the three models as well as the overall mean scores for the features across all three are shown in the following figure:

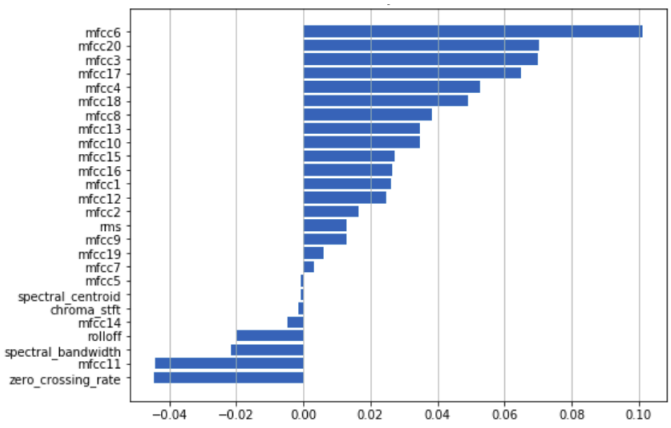


Fig. 2: Permutation scores for accompaniment model

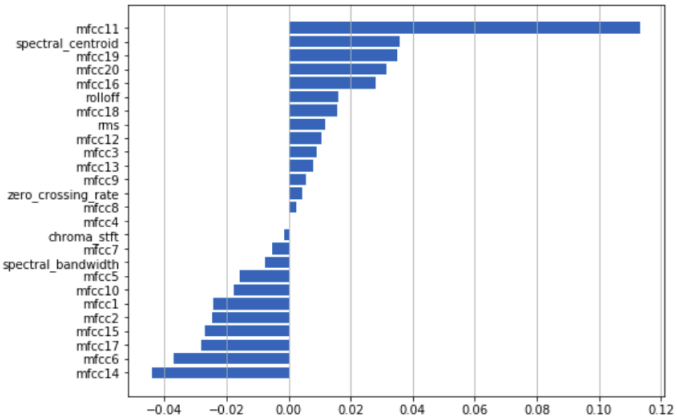


Fig. 3: Permutation scores for vocals model

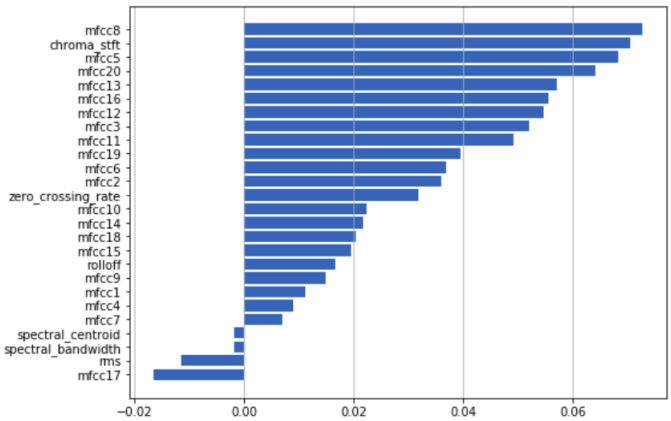


Fig. 4: Permutation scores for mix model

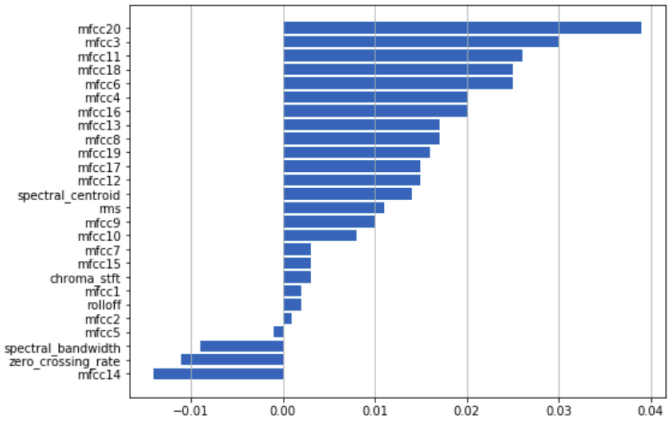


Fig. 5: Average permutation scores

V. Discussion and Conclusion

A genre classification approach on Marathi music has been presented in this paper. The analysis of the misclassifications provides interesting results as seen through the waveforms of the excerpts obtained through Praat. It is seen that 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 singing style (such as steadiness of notes) and melody. Another reason for this can be the fact that heavily melodic instruments such as the 'Sitar' have gone into the vocal audio rather than the accompaniment because Spleeter has been trained on Western music which does not have these melodic instruments, which are one of the most important feature distinguishing the Marathi genres. The relatively low performance of the machine learning models can be attributed to the small size of the dataset, simplicity of the model (using SVMs rather than GMMs and CNNs or other deep learning methods), the overlap between the genres themselves across songs as well as across excerpts within the same song, and possibly the imperfect splitting by Spleeter obtained on Marathi songs. The analysis of feature importance gives an overview of which features we can expect to work better in characterising a genre. For increasing the robustness of the models, there is a need of analysis over a much larger dataset, perhaps having well defined subcategories to lessen the variance in the music within each genre, so that a real time classifier can be constructed and deployed in applications such as recommender systems. Marathi (and other Indian) music being quite different from Western music and vibrant in itself can be a great research topic. We hope that the constructed dataset and the analysis presented in this paper can help to facilitate future research on acoustics of Marathi music.

Acknowledgements

I, Shreyas Nadkarni (main author), thank Prof. Preeti Rao, Department of Electrical Engineering, IITB, for her ideas and constant guidance through this project. This project would not have taken shape without her suggestions and her expertise in the field of Music Information Retrieval. I also thank Rohit M A and Nithya Shikharpur, M.Tech RAs, DAP-Lab, EE IITB for the their help in the implementation and programming caveats. We thank the research club of IIT Bombay: EnPoWER, Undergraduate Academic Council (UGAC), IIT Bombay for organising the Summer Undergraduate Research Project Program, 2021, to encourage UG students to pursue research projects. We also thank Arif Ashfaque Ahmad, Department Research Coordinator and UG Student, Department of Electrical Engineering, IIT Bombay for coordinating with us and UGAC through this project.

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 SPCOM 2020, I.I.Sc Bangalore, July 2020

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

[4] N. Maddage, C. Xu, X. Shao, F. Cao, Q. Tian "Musical genre classification using support vector machines", Conference Paper in Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on · May 2003

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

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

[7] Online Blogs:

1) P. Pandey, "Music Genre Classification with Python", https://towardsdatascience.com/music-genre-classification-with-python-c714d032f0d8
2) S. Afrulbasha, "Music Source Separtion with Spleeter on Google Colab", https://towardsdatascience.com/music-source-separation-with-spleeter-on-google-colab-9161dc72d1e8
3) C. Hermez, "Music Genre Classification with Tensorflow", https://towardsdatascience.com/music-genre-classification-with-tensorflow-3de38f0d4dbb
4) H. Marius, "Multiclass Classification with Support Vector Machines, Dual Problem and Kernel Functions", https://towardsdatascience.com/multiclass-classification-with-support-vector-machines-svm-kernel-trick-kernel-functions-f9d5377d6f02
5) P. Ippolito, "SVM: Feature Selection and Kernels", https://towardsdatascience.com/svm-feature-selection-and-kernels-840781cc1a6c

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

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

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