RAGA RECOGNITION USING MACHINE LEARNING

M.SANTHOSH KUMAR, M.SATHYA DEVI
CVR College of Engineering, Hyderabad, Telangana, India
Department of Computer Science and Engineering
kumarsanthosh213@gmail.com, satyamaraganti@gmail.com

ABSTRACT: - ***Classical Indian music is divided into two main forms: Hindustani and Carnatic, performed and played in northern and southern India. In Carnatic music research, one of the most daunting milestones is the identification of raga. Ragas is known as the foundation of classical Indian music. For raga recognition, a machine-learning method has been proposed to operate on partially noisy raga audio recordings. There are many obstacles in accurate raga detection, the most important barriers are music pitch and mood, extra tones, data translation, and the raga pace. Various classifiers can be used, such as the Naive bayes, K- nearest neighbour, where the performance is analyzed and features of raga recognition can be implemented using Python.***

Keywords: ***Raga, Carnatic, Naïve Bayes, k-nearest neighbor classifier.***

I. INTRODUCTION

The heritage of classical Indian Carnatic music is focused on 1500 BC. Classical Indian music's backbone is Ragas. Music is the heart of India and the other countries in ancient times. The source is very rich in Indian classical music.

Therefore many scientists are involved in research into musical information retrieval. Music is divided into various thaats on which the ragas are derived.

Music is an important part of our life and is continuously rising in development and consumption. In the last two decades, the consumption of music and its practise have changed dramatically. With digital music, a vast amount of music is available on demand, requiring new ways to organise the collections automatically. In addition to listening to an enriching and engaging experience of the quality of music, the engagement is increasing. This scenario presents an excellent opportunity to develop our experience of engaging with music.

Before the technical age, musicians used their hearing in a composition to distinguish notes and ragas. Musicians could then understand even less than 0.2Hz of the tonal variations. It consists of seven separate keys called Swaras viz in Carnatic music.

- Sa (Shadja)
- Ri (Rishabha)
- Ga (Gandhara)
- Ma (Madhyama)
- Pa (Panchama)
- Dha (Dhaivata)
- Ni (Nishada)

II. RELATED WORK

The following lacunae have been studied in previous techniques after reviewing many research works.

1) It is important to define key phrases[1], as they can extract maximum cases.

2) Compared with social actions even the attributes[1] should be considered.

3) The training subset may be used for pitch and mood recognition.

4) Identical crafted compositions and different patterns[2] should be separately defined.

5) Signal segmentation [3] at the same frequency must be observed.

There have been many attempts to describe a music note's raga. The transcription of raga directly in swaras at each interval of time and the classification by means of classifiers like K-NN or SVM[4] is one method for raga classification. Instead of using the absolute frequency [4] relative frequencies are used, as notes have a defined frequency ratio.

2.1 Note Transcription

The art of listening to a music piece and writing the sound notation of the sounds that constitute the piece is described in note transcription[6]. This means that an auditory signal is to be transformed into a symbolic portrayal, including musical events and their parameters. There are several attempts to transcribe the Notes. The method for automated transcription of western music has been proposed by Kalpuri[5]. In this case, signal processing techniques are implemented which solve various facets of the global challenge. A. Krishnaswamy[6] has identified a method on how to use Pitch Tracking for South Indian Classical Music Note transcription. The findings are seen with pitch trackers for samples of classical (Carnatic) South Indian music. He studied various musical notes used, their intonation and experimented with different pitch tracks and watched their success in the study of Carnatic music. The Swara Identification Framework for Classical South Indian Music was put forward by Rajeswari Sridhar and T.V.Geetha.[8] In a given Carnatic song it deals with the identification of the swaras. The segment frequencies are established and the exact tagging of the swara is carried out to determine the 7-swara combinations in the music signal given.

The relationship between the Swaras and their frequency is not set in Indian classical music. This is based on the basic frequency of the sounds known as 'Sa' or 'Si' of the album. All swaras are connected to Shruti with their basic frequencies, with a given ratio as shown below.

Ma2-- 729/612 Sa——-1
Re1 ——256/243 pa.
Re2--- 9/8 dha1--128/81 Re2-- 9/8 dha
Ga1-- 32/7 Dha2/16 Galaxy Galaxy Galeo.
Ga2-- 81/65 neither1---16/9
Ma1---------- 4/3 S2-----243/128

In Indian Classical Music (it can be different for different singers), Shruti is highly variable. Therefore, one important task in this process of defining the raga is to decide which fundamental frequency is the Swara 'Sa' frequency. The song scale is set to 240Hz in our system. Once the scale is identified it is mapped to 36 swaras of Mandra, Madhya and Taarasaptaka, according to the specified relationship.

2.2 Raga Identification

Pandey[8] has built a method to classify ragas Yaman and Bhupali automatically with a Markov model. On a two-target test of 31 samples, a success rate of 77 percent was recorded, although the methodology was not well recorded. A further phase in pursuit of unique pitch sequences increased efficiency to 87%. In an exploratory [3], Chordia rated 100 30 parts, each sixty seconds, from 13 ragas. Each segment was fitted with the Harmonic Pitch Class (HPCP) profile.
A K-NN classifier was used to achieve perfect results, which was used as featuring more advanced learning algorithms for pitch class distributions, PCDs, and pitch class dyad distributions (PCDDs). The classification accuracy of 94 percent with 10-fold cross-validation was achieved in a 17-fold aim test with 142 segments. The relevance of the findings was, however, limited by the database size of both cases. A system is developed to identify ragas based on PCDs and Pitch-Class Dyads measured directly from audio signal . A system is developed to identify ragas based on the PCDs. Here the training and testing of the system was carried out by 19 various performers in a huge, assorted database consisting of 20 hours of recorded performances at 31 different ragas. This was categorised with the Maximum A Posteriori (MAP) law, and the Random Forests . A maximum classification accuracy of 99.0% was achieved during a cross-validation experiment when classification was performed on 60 segments. In a tougher unseen experiment for generalisation, accuracy was 75%.

III. PROPOSED WORK

Time delayed melody surfaces with a k-nearest neighbour a variety of well-known distance measurements. The results recorded can be obtained from two versatile, diverse and representative collections of Carnatic and Hindustani music, one of which was originally introduced and published in this review. As far as we know, these are the largest data sets for recognising raga recordings, the number of ragas and the total audio length which are available to the public.

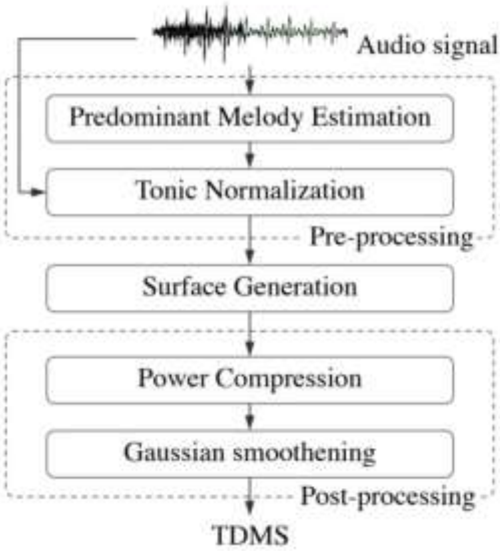

Figure1: Block diagram for the computation of TDMSs

- Critical assessment of current raga-recovery approaches and identification of some of their principal limitations.
- To conduct a comparative assessment under the same experimental conditions using the most effective state-of-the-art methods.
- A scalable raga-recognition Carnatic Music dataset which contains relevant metadata, annotations or calculated functions should be released publicly.

3.1 Time-delayed melody surface

The measurement of TDMS has three phases (Figure 1): pre-treatment, surface generation and after-treatment. A melody of an audio recording, normalised with the tonic or basic frequency of the piece, is obtained in pre-processing. We measure a two-dimensional surface based on the delay co-ordinate principle in surface generation.

3.2 Predominant melody estimation

The pitch of the prevailing melodic source reflects the melody of a sound note. The approach suggested by Salamon and Gomez[1] is used for the prevailing pitch calculation. this was used in a number of genres of music, including IAM (Indian Art Music), and was used in many other studies for a similar task [7,12,13] in MIREX 2011 (international MIR evaluations campaign). It is available to implement this algorithm, Essentia is an open source C++ library for both audio analysis and MIR content.

3.3 Tonic normalization

For an IAM melody the base frequency of the lead artist[10] is chosen as the tonic pitch to which all of the accompanying instruments are tuned. Thus, we normalise the prevalent melody of a recording for a musically significant role for raga recognition by considering its tonic pitch a reference frequency during the Hertz to cent conversion.

$$c_i = 1200 log_2\left(\frac{f_i}{w}\right)$$

The standardised ith sample of the dominant pitch(in cents) is 0 < i < N, where N is the total pitch samples, and the i sample of this predominant one (in Hz). For each recording, the tonic pitch is defined by the Gulati etal multi-pitch[10] approach. This technique has been documented to achieve cutting-edge results and has been used elsewhere[8,11] successfully. An additional majority voted on Pa (fifth) style error [10] in tonical values for various recordings of an artist.

3.4 Surface generation

The following step is to construct a two-dimensional surface on the basis of delay co-ordinating (also known as phase space integration)[15, 28]. In a double-dimensional map of Poincare[10], such a two-dimensional surface is visible as a discretized histogram of the components.

$$\bar{s}_{ij} = \sum_{t=r}^{N=1} I(B(c_t), i) I(B(c_{t-r}), j)$$

B is an octave wrapper integral binning operator where I are an indicator function such that I(x, y) = 1 iff x = y, I(x, y) = 0, elsewhere.

$$B(x) = \left\lfloor \left(\frac{\eta x}{1200}\right) mod \eta \right\rfloor$$

And µ is an index of time delay (in frames) left as parameter. Notice that the frames in which a prevailing pitch can not be obtained are excluded from any estimate, as stated above. We use ± = 120 for the scale of ̄s. This value is 10 cents per bin, stated with optimum pitch resolution.

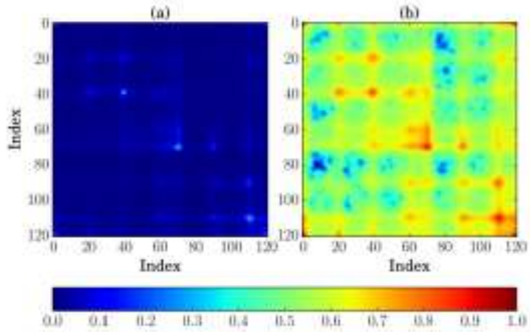

Figure 2.Music surface before (a) and after (b) post-processing (prescription to S and Å S respectively). Figure 2. Both matrices are scaled between 0 and 1. to make visualisation easier.

The prominent peaks in the surface correspond to the raga, they are steep and the surface has a high dynamic range. The essence of the melodies in those musical traditions can be attributed to that particularly in Carnatic music, which frequently includes long host swaras. The dynamic range is also high as the pitches are within the range of the swara in the stable swara regions, compared with the pitches in the transitional melodic regions. This means that in the stable regions the frequency values are mapped in a smaller set of containers, making the prominent peaks steeper.

3.5 Post-processing

Element-wise power compression is used to accentuate the values of the melody 's transitional regions and to reduce the surface dynamic range.

$$\bar{S} = \bar{S}^{\propto}$$

Where α is a parameter exponent left. When a more compact surface is obtained (in the dynamic range),

Gaussian smoothing is applied circularly convolving $\bar{S}$ with a two-dimensional Gaussian kernel. The cyclic (or octave-folded) nature of the TDMS (Eqn) that imits the cyclical nature of pitch groups is chosen to cause the circular convolution. The standard deviation of this kernel is σ bins (samples).

The kernel length is reduced to 8,01% bins, after which value (below 0,01% of the full amplitude of the kernel) is negligible proceedings must be standardised on the calculated surface ÁS. Divide S according to the regular L1 matrix:

$$S = \bar{S} / \left\| \hat{S} \right\|_1$$

This also results in S values, the final TDMS, which can be represented in discrete probabilities.

Figure 2(b) shows the result after the surface is after post-processing with power compression and gaussian smoothing. Accentuated values for the non-diagonal components. Figure 2(b) offers a visual inspection of some musical perspectives into the melodic aspects of the recording. The diagonal rates (60.60) and the diagonally (70.60). The highest salience in the index (110,110) is that of the Ni Swara, which is the Vadiswara, i.e.the raga, which is musically the most prominent swara. The diagonal asymmetry in the matrix refers to the asymmetric aspect of the up and down swara pattern of the raga (compare, among other things, the salience at indexes (70,90) to indices (90,70), though the former is more prominent than the latter). The tetrachord structure of the raga is characterised by the similarity between indices (20,20) and (70,70) and matrix between indices (70,70) and (120,120). Finally , an important aspect of TDMSs is that the mean total across its columns and rows provides a PCD representation.

3.6 Classification and distance measurement

The task of classifying audio recordings according to their raga mark is to demonstrate the capacity of the TDMS to capture raga characteristics. K-nearest neighbour (kNN) classifier is chosen to perform classification. There are many reasons why we choose. First of all, the kNN classification with its well-studied performance and architecture relationships with other classificators is well understood. Secondly, it is quick to accelerate testing or retrieval, with virtually no training and with known technologies. Thirdly, it only has one parameter, k blindly set to a relatively small value or can be optimised easily during the training period. Finally, it is an easy-to -implement classifier whose effects can be interpreted and quickly repeated.

A kNN classification 's efficiency depends heavily on the distance calculation used to collect k-neighbors. With TDMS features S(n) and S(m), three separate measurements are considered for the calculation of distance between two records n and m. The difference between S(n) and S(m) is considered in the Frobenius Standard,

$$D_F^{(n,m)} = \|S_n - S_m\|_2$$

symmetricKullback-Leibler divergence

$$D_{KL}^{(n,m)} = D_{KL}\left(S^{(n)}, S^{(m)}\right) +$$

$$D_{kl}(S^{(m)}, S^{(n)}) D_{kl}(X,Y) = \sum X log\left(\frac{X}{Y}\right)$$

where element-oriented operations and sum are carried out on the resulting matrix over all elements. Finally, which states that the PCD-based role for the same task does not perform other distance measurements,

$$D_B^{(n,m)} = -log\left(\sum \sqrt{S^{(n)}.S^{(m)}}\right)$$

Element-wise operations and sum over all the elements are performed on the resultant matrix.

IV. CONCLUSION

Raga is a very complex framework in Indian music. The notes used to play the songs are raga-based. The series of Raga identification notes is examined here. The Raga detection technique is precise with a range of obstacles. The key problems include the dynamic parameters of music pitch and mood, which miss additional tones. A TDMS captures a melody's tonal features as well as the short duration of time. They come from the tonically normalised pitch of the main melodic audio source. We identify audio recordings according to their raga labels to demonstrate the capabilities of TDMSs in the capture of raga features. For this reason, we used large Carnatic music collections of more than 120 hours. In addition, we analysed the effect of different parameters on the accuracy obtained by TDMSs and concludes that it is essentially invariant with different parameter values. The featured modern systems in raga recognition are outperformed with the closest Neighbor Classifier. A review of the classification errors revealed that there are confusions between musically related ragas, which share a common collection of swaras with related melodic phrases. Future work involves proposing a well-defined and robust raga recognition machine learning technique, giving maximum accuracy compared to other classifiers and identifying individuals, following the development of the data set to many more ragas.

REFERENCES

1) *J. Salamon and E. G´omez. Melody extraction from polyphonic music signals using pitch contour characteristics. IEEE Transactions on Audio, Speech, and Language Processing, 20(6):1759–1770, 2012.*
2) *K. K. Ganguli, A. Rastogi, V. Pandit, P. Kantan, and P. Rao. Efficient melodic query-based audio search for Hindustani vocal compositions. In Proc. of Int. Soc. for Music Information Research Conf. (ISMIR), pages 591–597, 2015.*
3) *Parag Chordia, "Automatic raaga classification of the pitch- tracked performances using pitch-class and pitch-class dyad distributions", In Proceedings of International Computer Music Conference, 2003.*
4) *Preeti Rao and Anand Raju, Building a melody retrieval system, National Conference on Communications, 2002.*
5) *AnssiKlapuri, "Signal Processing Methods for the Automatic Transcription of Music", Thesis for the degree of Doctor of Technology, USA, 2004.*
6) *Krishnaswamy, A, "Application of pitch tracking to South Indian classical music", Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on 19-22 Oct. 2003*
7) *Rajeswari Sridhar, and T.V. Geetha, "Swara Identification for South Indian Classical Music", 9th International Conference on Information Technology (ICIT'06), (2006), pp 143-144.*
8) *Gaurav Pandey, Chaitanya Mishra, and Paul Ipe, "Tansen: A system for automatic raga Identification", In*

Proceedings of the 1st Indian International Conference on Artificial Intelligence, pp 1350–1363, 2003.

9) *Parag Chordia, "Automatic raaga classification of the pitch- tracked performances using pitch-class and pitch-class dyad distributions", In Proceedings of International Computer Music Conference, 2006*

10) *S. Dutta, S. PV Krishnaraj, and H. A. Murthy. Raga verification in Carnatic music using the longest common segment set. In Int. Soc. for Music Information Retrieval Conf. (ISMIR), pages 605–611, 2015.*

11) *S. Gulati, A. Bellur, J. Salamon, H. G. Ranjani, V. Ishwar,H.A.Murthy,andX.Serra.Automatictonic identification in Indian art music: approaches and evaluation. Journal of New Music Research, 43(1):55–73, 2014.*

12) *X. Serra. A multicultural approach to music information research. In Proc. of Int. Conf. on Music Information Retrieval (ISMIR), pages 151–156, 2011.*