Identification of Carnatic Music Raagas using Fisher-Jenks Classification

1st Aditya D.S.
Department of Electronics and Communication Engineering
IIITDM Kancheepuram
Chennai, India
esd18i001@iiitdm.ac.in

2nd Ganesh T S
Department of Electronics and Communication Engineering
IIITDM Kancheepuram
Chennai, India
esd18i006@iiitdm.ac.in

***Abstract*—This work aims to classify audio input according to the *raaga* being played. The fixed ratio of all the *swaras* in a *raaga* relative to the *shadja* is used. The Discrete Fourier Transform(DFT) and some filtering techniques are used to remove the noise in the audio. The Fisher-Jenks Natural Breaks Optimization Algorithm is used for clustering the frequencies, after which the highest frequency in a cluster is considered as dominant frequency, and all dominant frequencies are then normalised with lowest dominant frequency. The Euclidean Distance between thus obtained ratios and database is compared, and the lowest distance value entry is identified as *raaga*.**

***Index Terms*—Carnatic Music, *raaga*, Discrete Fourier Transform, Machine Learning, Fisher-Jenks Natural Breaks Algorithm, Filters, Threshold Noise Filter, Bandpass Filter, Euclidean Distance**

I. Introduction

The analysis of music using scientific techniques has been prevalent in the West. This was done initially to make music learning easy, and has since then evolved into the large field of musicology. Music signal processing is a field which uses signal processing to study the complex music signals.Indian Classical Music in general and Carnatic Music in particular have always been highly exclusive in nature. Indian Classical Music, was never meant to be learned easily, and the techniques were fiercely guarded by the various schools of music, which were patronised by the rulers. This caused the absence of standards, and it is for standardisation that analysis is done.

Over time, certain standards on the lines of Western music were made and learning Indian Classical Music became comparatively easier, and so did its analysis. Prominent physicist Sir CV Raman is one of the trailblazers in the field, with work on the acoustics of various instruments including the *tabla*,*ektara* among many others. These can be accessed through [7]. Music signal processing on Indian Classical Music is also an active field now.

The number of listeners of music in general and Carnatic Music in specific is increasing day by day in India and abroad due to the prevalence of streaming services. Most of them may not have formal training in the discipline, making the identification and classification of *raagas* and other compositions difficult. Initially, before they are able to perceive the *raagas* on their own, an accurate aid is required.

In olden days, this aid was in form of a person who was well versed in Carnatic music. This causes the task of finding a well versed person and that person agreeing to help you very difficult. If a wide spectrum of people are able to listen to this genre, it is due to the reduced costs of recording and broadcasting, the latter initially through radio/TV and nowadays over the internet. With progresses made in technology, this aid can now be a computer. This will increase the accuracy of prediction as well, as there will be no human error. Apart from the layman listener, this can also be used by the professional musicians to perfect their skills and analyse their performance.

II. Carnatic Music Theory

The classical music of India can be generally divided into broadly two categories, Carnatic and Hindustani. Both these branches have roots in the recitation of Vedas, but the differences are prominent in the way of presentaion.

Swara is one of the most fundamental concepts of Indian Classical Music. This is analogous to what is in western music called the note. There are seven fundamental *swaras*, collectively known as the *Saptaswaras*, namely, *Shadja*, *Rishabha*, *Gandhara*, *Madhyama*, *Panchama*, *Dhaivatha* and *Nishadha*. With the exception of *Shadja* and *Panchama*, all other *swaras* can take up multiple positions called *swarasthanas*. The distinct names associated with those *swarasthanas* and the ratios relative to *shadja* are given in table I. Unlike western music where the frequency of notes is fixed, in the *swaras* only the frequency of *Shadja* is fixed, and all other *swaras* have frequencies based on that particular frequency.

Shruti is the Carnatic analog for pitch, and is in layman terms an indicator for a *swara* to be "high" or "low". The span of *saptaswaras* is called *sthayi*. There are in total 5 *sthayis*, with the reference *sthayi* called *madhya*. The next higher spans are called *taara* and *ati taara*. The lower spans are called *mandra* and *anumandra*. This can be considered to be an analogue for octave.

A *Raaga* is defined as *Swara varna visheshena dhwani bhedaya va punah, Ranjyate yena yan kashichit sa ragah samsthatham*, which means that a *Raaga* is a kind of sound composition which is adorned with *swaras* in some special way, which is capable of generating feelings in anyone. This definition is supposed to have been given by the sage *Matanga*.

In Carnatic music, the major division of *raagas* is as *janaka* or *melakarta* and *janya*. *Janaka* in Sanskrit means father/originator, and the *melakarta raagas* are like the parents of the *janya raagas* which are derived by either adding or deleting *swaras*. The key features of a *raaga* are the set of *swaras*, their ascending progression and descending progression which is called *aarohana* and *avarohana* respectively and beautification elements called *gamaka*. A *melakarta raaga* must satisfy the conditions that it should have *saptaswaras*, the same set of *swaras* must be there in *aarohana* and *avarohana* and the final rule is that all *swaras* must be increasing in frequency *aarohana* and decreasing in frequency *avarohana*. There are 72 *melakarta raagas* which are grouped into 12 *chakras* named *Indu*, *Netra*, *Agni*, *Veda*, *Bana*, *Rithu*, *Rishi*, *Vasu*, *Brahma*, *Disi*, *Rudra* and *Aditya*.

TABLE I
FREQUENCY RATIOS OF SWARAS

Swara	***Ratio***
Shadja(S)	1.000
Shuddha Rishabha(R1)	1.067
Chatushruthi Rishaba(R2)	1.125
Shatshruthi Rishaba(R3)	1.200
Shuddha Gandhara(G1)	1.125
Saadharana Gandhara(G2)	1.200
Anthara Gandhara(G3)	1.250
Shuddha Madhyama(M1)	1.333
Prati Madhyama(M2)	1.416
Panchama(P)	1.500
Shuddha Dhaivatha(D1)	1.600
Chatushruthi Dhaivatha(D2)	1.667
Shatshruthi Daivatha(D3)	1.800
Shuddha Nishadha(N1)	1.667
Kaisiki Nishadha(N2)	1.800
Kakali Nishadha(N3)	1.875

III. PREVIOUS WORK

A lot of work has been done in the general field of music signal processing. In this particular field of research related to Indian Classical Music, most of the work has come out in the last decade. With ML(machine Learning)/DL(Deep Learning) becoming a buzzword, lot of publications(mostly conference papers) can be found. These three specific publications which were found to have been referred by a lot of researchers were studied and analysed as follows.

In [2], *swara* intonation information is used to identify Hindustani *raagas*. Vocal music samples for the various *raagas* were taken. At regular intervals, the vocal pitch was extracted and written to a pitch contour file. The probability of a pitch value occurring in the duration of a segment is given by pitch distribution. Folding of this distribution was done into one octave, and a folded pitch distribution(FPD) was computed. After this, Pitch Class distributions, which represent the probability of occurrence of the *swaras*, were constructed from the FPD's. *Swara* features like the mean, mode, standard deviation and probability of occurrence were then found from FPD. A combination of Euclidean distance and Kullback Leibler distance was used on the *Swara* features to classify *raagas*.

Auto-correlation analysis is performed to extract pitch contour from input audio signal in [4]. The probability density function of pitch values is then obtained from the pitch contour by kernel density estimation method. Parameters for comparison are extracted from the PDF curve, the parameters being frequencies of peaks, their mean and variance. This was followed by normalisation and then given as input to an ANN.

[4] takes into account the fact that the *swara* frequencies are all dependent on the base *shadja* frequency. The analysis is restricted to *melakarta raagas* . Frame wise are pitch frequencies are obtained from Praat software. Parzen window method is then used to find a density estimate, after which the *shadja* is identified. The variance and weights of each *swara* are then found using a semi continuous Gaussian Mixture Model. These weights are then used to find which variant of the *rishabha*, *gandhara*, *madhyama* and *dhaivata* is there in the input audio. This is then compared with existing corresponding *swaras* and if it is found to match it is said to be verified.

In this work also, the fact that the *swaras* in a *raaga* are in a fixed ratio relative the *shadja* is used, but instead of taking only particular *swaras*, the whole set of *swaras* in a *raaga* are used. The Discrete Fourier Transform(DFT) and some filtering techniques are used to remove the noise in the audio. ML algorithm is used to cluster similar frequencies and identify the dominant frequency in the cluster. The Fisher-Jenks Natural Breaks Optimization Algorithm is used for clustering the frequencies and as per the literature survey, there seems to be no precedent for usage of this method for classifying the frequencies.

IV. DATASET DETAILS

The dataset of 10 audio files was obtained by asking a friend, Mr Harish Murali to play the aarohana and avarohana of 10 Carnatic raagas on the violin. The music was recorded using a mobile microphone and was stored in the .aac format at a sampling rate of 48kHz. The 10 ragas and their *aarohana-avarohana* structure is mentioned in table II. These *raagas* are a mix of both *melakarta* as well as *janya raagas*.

V. METHODOLOGY

The block diagram of the proposed method is shown in 1. The method relies on the fact that every *swara* irrespective of the *shruti* is always at a fixed position relative to *madhyama sthayi Sa*, i.e, irrespective of the *shruti*, the *swaras* in a *raaga* will be in a fixed ratio and this ratio is unique to each *raaga*. Thus, if one is capable of successfully retrieving the dominant frequencies in a recording, these frequencies can be

TABLE II
AAROHANA AND AVAROHANA OF RAAGAS IN DATASET

Raaga	***Aarohana***	***Avarohana***
Aboghi	$S\ R_2\ G_2\ M_1\ D_2\ \dot{S}$	$\dot{S}\ D_2\ M_1\ G_2\ R_2\ S$
Hindola	$S\ G_2\ M_1\ D_1\ N_2\ \dot{S}$	$\dot{S}\ N_2\ D_1\ M_1\ G_2\ S$
Kalyani	$S\ R_2\ G_3\ M_2\ P\ D_2\ N_3\ \dot{S}$	$\dot{S}\ N_3\ D_2\ P\ M_2\ G_3\ R_2\ S$
Madhyamav-athi	$S\ R_2\ M_1\ P\ N_2\ \dot{S}$	$\dot{S}\ N_2\ P\ M_1\ R_2\ S$
Mayamalava-gowla	$S\ R_1\ G_3\ M_1\ P\ D_1\ N_3\ \dot{S}$	$\dot{S}\ N_3\ D_1\ P\ M_1\ G_3\ R_1\ S$
Mohana	$S\ R_2\ G_3\ P\ D_2\ \dot{S}$	$\dot{S}\ D_2\ P\ G_3\ R_2\ S$
Panthuvarali	$S\ R_1\ G_3\ M_2\ P\ D_1\ N_3\ \dot{S}$	$\dot{S}\ N_3\ D_1\ P\ M_2\ G_3\ R_1\ S$
Shankarabha-rana	$S\ R_2\ G_3\ M_1\ P\ D_2\ N_3\ \dot{S}$	$\dot{S}\ N_3\ D_2\ P\ M_1\ G_3\ R_2\ S$
Shanmukha-priya	$S\ R_2\ G_2\ M_2\ P\ D_1\ N_2\ \dot{S}$	$\dot{S}\ N_2\ D_1\ P\ M_2\ G_2\ R_2\ S$
Vasantha	$S\ G_3\ M_1\ D_2\ N_3\ \dot{S}$	$\dot{S}\ N_3\ D_2\ M_1\ G_3\ R_1\ S$

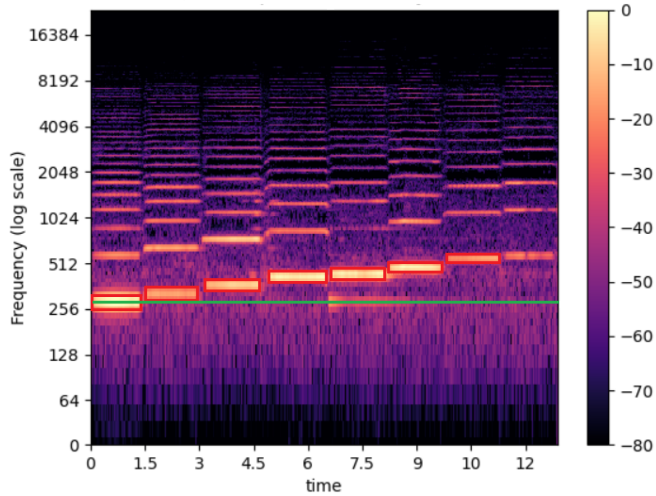


Fig. 2. Power Spectrogram of an audio file. The red boxes highlight each *swara* in the *raaga* and the number of red boxes is $N_s = 7$. The green line is the approximate frequency $f_0 \approx 290Hz$ of *madhyama sthayi Sa*

normalized and matched against a database of these ratios to identify the *raaga* in the recording. More details regarding the database is given later. The different stages in the block diagram are explained below.

A. Pre-processing

To identify a *raaga*, all the *swaras* in the *raaga* must be played at least once. This means that either the *aarohana* or the *avarohana* alone must be retained. Hence, the audio files were trimmed and only the relevant portion of the audio was retained. The trimmed audio files were exported to .wav format. The audio trimming operations were performed using *Audacity*, an open source software. The sampling rate was left unaltered.

Next, time-frequency analysis is performed on the audio file and a *Power Spectrogram* of the audio is plotted (figure 2). From this plot, the number of unique *swaras* N_s in the *raaga* and the approximate frequency f_0 of *madhyama sthayi Sa* can be determined. This information will be useful later on in identifying the *raaga*.

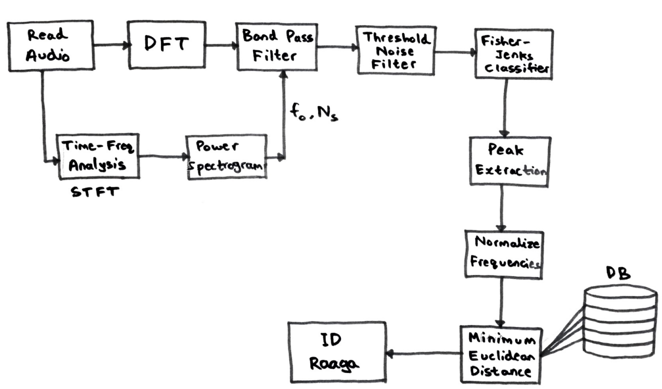


Fig. 1. Block diagram of proposed method

B. Reading the Audio File

Audio file is read at the native sampling rate of the audio file (48kHz). This is done to ensure that no information is lost. The audio samples are returned as a 1-dimensional numpy array.

C. Discrete Fourier Transform

The audio time series data is transformed to the frequency domain using the Discrete Fourier Transform (DFT) with the number of points being equal to the length of the audio data. If the audio data is a power of 2, then the *Radix-2 Fast Fourier Transform(FFT) algorithm* is used, else *Bluestein's algorithm* is used to compute the DFT. The efficiency of the FFT algorithm is $O(n \log(n))$ whereas the efficiency of Bluestein's algorithm is never worse than $O(n \log(n))$.

The magnitude spectrum is plotted (figure 3) which consists of several peaks which indicate the different *swaras* present in the *raaga*. The spectrum also consists of noise frequencies and overtones which have to be suppressed. This is done in the next step.

D. Filtering

Two filters are applied. The first filter is a Bandpass filter whose corner frequencies are chosen to be $f_0 - \delta$ and $2(f_0 - \delta)$ where δ is some *guard frequency*. This is done to prevent f_0 from getting filtered. Typically, it is around 5-10 Hz. The output of this bandpass filter is the FFT only in the region of interest.

The second filter is a *Threshold Noise Filter* which is used to suppress the noise. A threshold value for the FFT magnitude is chosen as a fraction of the peak magnitude. Frequencies corresponding to FFT magnitudes less than the threshold are considered to be noise frequencies and are discarded. Further, the amplitudes less than the threshold are zeroed out and discarded as well. What is now left is a FFT only in the region of interest with most of the noise frequencies removed (figure 4).

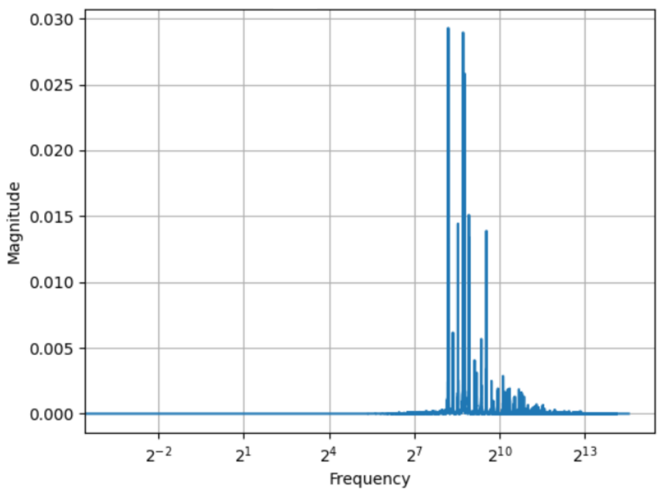


Fig. 3. Magnitude spectrum of the audio file. Logarithmic scale has been used for the x-axis.

E. Clustering Frequencies and Peak Extraction

Although most of the noise is removed, the peaks in the magnitude spectrum are not at a single frequency. Rather, a very close collection of peaks around a dominant frequency can be observed. Dominant frequencies from the spectrum can be extracted by having variable width windows and looking at the peak only inside that window (figure 5). This is essentially a *clustering problem* where the task consists of dividing the frequencies into different groups/clusters such that frequencies in the same cluster are similar to one another.

To get around this problem, a 1-dimensional clustering algorithm called *Fisher-Jenks Natural Breaks Algorithm* is used. This algorithm accepts an array of scalars and computes the bounds for each cluster. The number of clusters to be formed must also be specified and is equal to N_s.

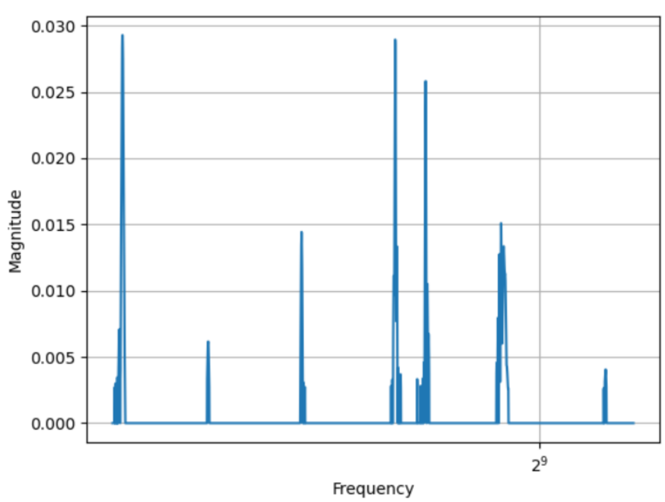


Fig. 4. Magnitude spectrum after using a bandpass filter and threshold noise filter. The frequencies of the *swaras* are clearly visible now.

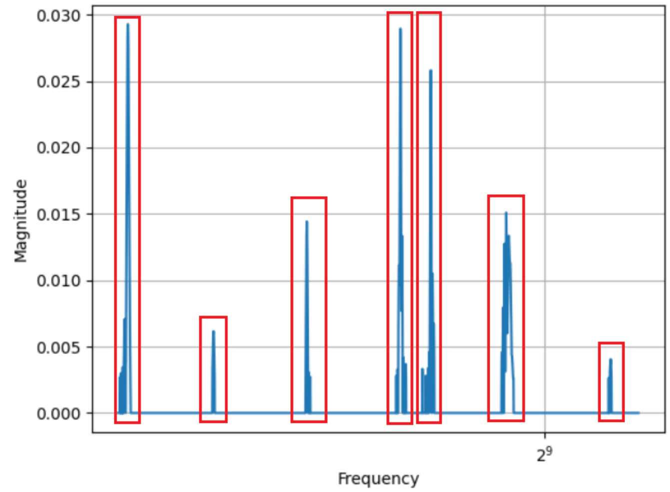


Fig. 5. Different clusters of frequencies in the magnitude spectrum are highlighted in red.The frequencies corresponding to the peaks in each cluster is what is of interest in this approach.

Once the bounds for each cluster are obtained, the frequency of *swaras* can be extracted by computing the frequency with the highest amplitude (dominant frequency) within the cluster. After extracting the dominant frequency from each cluster, these frequencies are normalized to the smallest dominant frequency. This list of normalized frequencies is enough to identify a *raaga*.

F. Database

The database is a simple *JavaScript Object Notation* (JSON) file with the following structure:

```
[
    {
        "raaga": "mayamalavagowla",
        "ratios": [1.0, 1.067, 1.250, 1.3
            33, 1.500, 1.600, 1.875]
    },
    ...
]
```

JSON is an open standard file format, and data interchange format, that uses human-readable text to store data objects consisting of key–value pairs. Each entry in the JSON file consists of the name of the *raaga* and the ratios. The ratios array contains the theoretical ratios of the *raaga* which can be computed from table II.

G. Raaga Identification

For each entry in the database, the similarity between the ratios array and the array of normalized frequencies is observed. The similarity is quantified by computing the *Euclidean distance* between the two arrays. The Euclidean distance between two 1-dimensional arrays u, v of length m is computed as follows:

$$dist(u, v) = \sqrt{\sum_{i=0}^{m-1} (u(i) - v(i))^2} \tag{1}$$

Entry in the database corresponding to the least distance is identified as the *raaga*. To calculate the confidence with which the *raaga* was predicted, the following technique was used:

$$dist(u, v) \leq \sqrt{m\big((\mu_u - \mu_v)^2 + (\sigma_u + \sigma_v)^2\big)} \tag{2}$$

where, μ_u is the mean of u, μ_v is the mean of v, σ_u is the variance of u, σ_v is the variance of v

Thus, the confidence percentage can be computed as:

$$100\left(1 - \frac{\sqrt{\sum_{i=0}^{m-1} (u(i) - v(i))^2}}{\sqrt{m\big((\mu_u - \mu_v)^2 + (\sigma_u + \sigma_v)^2\big)}}\right) \tag{3}$$

VI. Classification Results

TABLE III
SUMMARY OF THE CONFIDENCE PERCENTAGES FOR ALL AUDIO FILES

Raaga	**Identified?**	**Distance**	**Confidence%**
Aboghi	Yes	0.028	88.32%
Hindola	Yes	0.053	85.11%
Kalyani	Yes	0.041	90.52%
Madhyamavathi	Yes	0.012	96.43%
Mayamalavagowla	Yes	0.078	82.29%
Mohana	Yes	0.021	92.03%
Panthuvarali	Yes	0.063	85.33%
Shankarabharana	Yes	0.047	89.30%
Shanmukhapriya	Yes	0.070	80.58%
Vasantha	Yes	0.066	87.04%

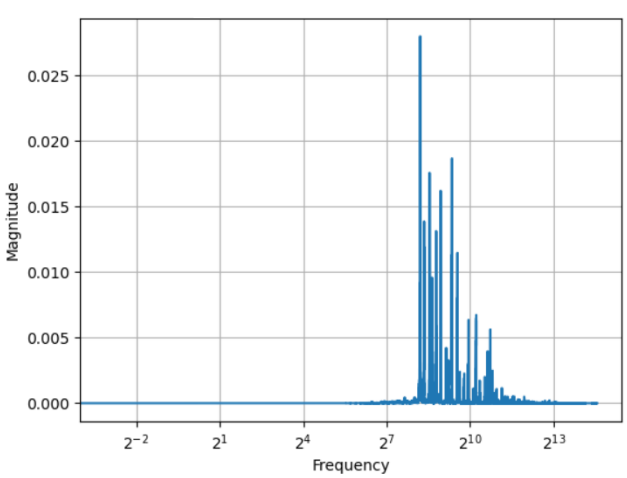


Fig. 6. Magnitude spectrum of sankara.wav

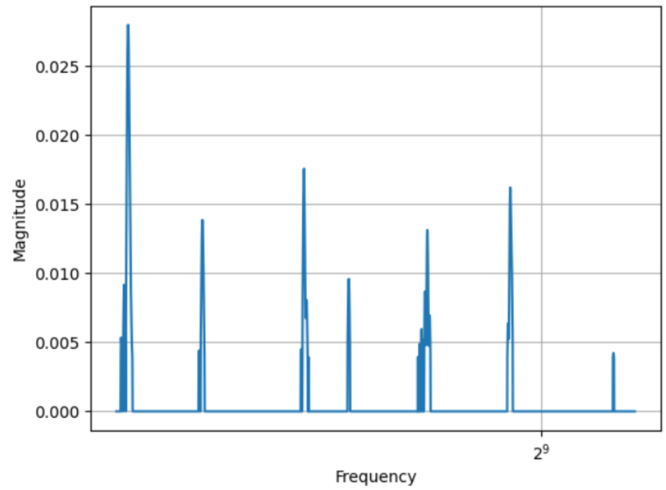
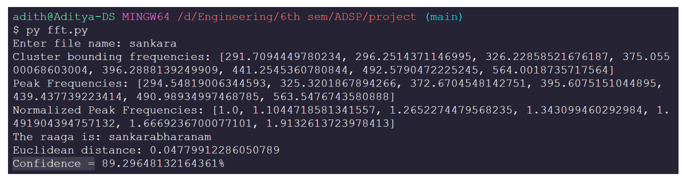


Fig. 7. Magnitude spectrum of sankara.wav after filtering



```
adith@Aditya-DS MINGW64 /d/Engineering/6th sem/ADSP/project (main)
$ py fft.py
Enter file name: sankara
Cluster bounding frequencies: [291.7094449780234, 296.2514371146995, 326.22858521676187, 375.055
00068603004, 396.2888139249909, 441.2545360780844, 492.5790472225245, 564.0018735717564]
Peak Frequencies: [294.54819006344593, 325.3201867894266, 372.6704548142751, 395.6075151044895,
439.437739223414, 490.98934997468785, 563.5476743580888]
Normalized Peak Frequencies: [1.0, 1.1044718581341557, 1.2652274479568235, 1.343099460292984, 1.
491904394757132, 1.6669236700077101, 1.9132613723978413]
The raaga is: sankarabharanam
Euclidean distance: 0.04779912286050789
Confidence = 89.29648132164361%
```

Fig. 8. Terminal output. Percentage confidence = 89%.

Figures 6, 7 and 8 show the output of our program when the sankara.wav audio file is provided as input. The *raaga* is correctly identified as *Shankarabharana* with a very high confidence of 89%. Figures 6 and 7 show the Magnitude spectrum before and after filtering. Drastic reduction in noise frequencies can be observed.

The results for all audio files have been summarised in table III.

As can be seen, all ragas have been identified successfully with high confidence. It can be seen that the confidence percentage varies between 80 to 96%. There are several factors which contribute to this large variation which include factors like:

1) Poor recording equipment.
2) Slight inaccuracies in playing a *swara*.
3) Existence of too much noise leading to formation of incorrect clusters.
4) Selection of incorrect threshold for Threshold Noise Filter.

VII. Acknowledgements

We would like to heartily thank our friend Harish Murali from the Department of Computer Engineering, IIITDM Kancheepuram, for recording a dataset of *raagas* on the violin and also for providing technical assistance regarding Carnatic Music. We also like to thank Dr. Asutosh Kar, our instructor for the course Advanced Digital Signal Processing, for allowing us to work on this topic.

VIII. Conclusion and Future Scope

In this work, an attempt was made to classify *raagas* using the ratios of the *swaras* with respect to *shadja* and comparing them with existing data of *raagas*.A dataset of *aarohana* and *avarohana* of a mix of *janya raagas* and *janaka raagas* played on the violin was used. The frequencies were obtained from the audio signal and the Fisher-Jenks Natural Breaks Optimization Algorithm was used for clustering the frequencies and identifying the dominant frequency in a cluster. It can be seen that all the *raagas* were predicted with a confidence percentage greater than 80%.

As the whole set of *swaras* in a *raaga* are used, predicting certain *janya raagas* having the same set of *swaras* may not be possible using this technique. All the *janya raagas* and *janaka raagas* used here had a unique *swara* set, thus meaning that an approach must be found for those *janya raagas* which have non unique *swara* sets. The dataset here had only the violin, for which satisfactory results were seen. It must now be replicated with various other instruments like flute, *veena* etc., first individually and subsequently maybe in combination. This can gradually then be extended to use on existing performance recordings, songs etc. Further extension would mean a real time implementation of this technique in live *kacheris*, which would help the layman listener to get to know what *raaga* is being played.

Currently, the user is required to identify the frequency of the *shadja* and also input the number of unique *swaras* from the Power spectrogram. This can be avoided by making use of some pitch detection algorithms.

References

[1] P. Upadhyaya, Suma S. M. and S. Koolagudi, "Identification of allied raagas in Carnatic music," in 2015 Eighth International Conference on Contemporary Computing (IC3), Noida, India, 2015 pp. 127-131. doi: 10.1109/IC3.2015.7346666

[2] Belle, S., R. Joshi and P. Rao, 2009. Raga identification by using swara intonation. J. ITC Sangeet Res. Acad., 23: 1-7.

[3] H. G. Ranjani, S. Arthi and T. V. Sreenivas, "Carnatic music analysis: Shadja, swara identification and rAga verification in AlApana using stochastic models," 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 2011, pp. 29-32, doi: 10.1109/ASPAA.2011.6082295..

[4] Suma S.M., Koolagudi S.G. (2015) Raga Classification for Carnatic Music. In: Mandal J., Satapathy S., Kumar Sanyal M., Sarkar P., Mukhopadhyay A. (eds) Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing, vol 339. Springer, New Delhi.

[5] N.S Ramachandran, The Ragas of Karnatic Music, University of Madras, 1938

[6] National Institute of Open Schooling, Textbook for Carnatic Music Secondary Level Course Theory

[7] http://dspace.rri.res.in/handle/2289/1473