Automatic Carnatic Ragam Identification

A Laboratory Oriented Project under the guidance of
Dr. K. K. Gupta

Sreekar Raghotham 2012A8TS346P
Vinod Subramanian 2012A8TS297P

Birla Institute of Technology and Sciences, Pilani

November 30, 2015

Contents

Abstract

Two algorithms for Carnatic *ragam* identification have been developed. One algorithm identifies the ragam based on the scale input and the Harmonic Pitch Class Profiles of the song input. The second algorithm constructs a Hidden Markov Model based on Gaussian Mixture Models to generate a sequence of *swaras* which aid ragam identification.

1 Introduction

Though Indian classical music is a major form of music, known for its exacting standards of technical expertise, the field of music information retrieval in the context of Indian music is still in its nascent stage. Compared to Western Music, Indian classical music's technical finesse has not been utilised to its fullest.

Essential components of Indian Classical Music, both Carnatic and Hindustani, are the *ragam* and the *talam*. Identifying the ragam of the song is essential for MIR. A ragam is a essentially a string of notes, called *swaras*. In this project we attempt to develop a scale dependant ragam identification system and model a Hidden Markov Model with Gaussian Mixture Models providing the emission probabilities.

2 Carnatic Music Theory

Indian Classical Music and the vibrant music scene that it finds itself a part of, can be classified into two schools - Hindustani and Carnatic. Hindustani music is prevalent in north India while Carnatic music is popular in south India. Raga(m) and Tala(m) are the two important components of any Indian classical music piece on the macro level. While the ragam is a note sequence, the talam is a time sequence. The ascending order of swaras in a raga is called *arohanam* and the descending order is called *avarohanam*

2.1 Rāgam

Ragam, the word, is derived from the Sanskrit for 'colour or hue', indicating that a ragam has the potential to colour one's mood. A ragam is usually made of five to nine notes or swaras over which a melody is woven. This, however does not restrict an artist's performance as (s)he is allowed to add a personal touch to each performance of a ragam based on comfort levels.

Swaras are placeholder names for a set of frequencies and unlike Western music, the frequency a swara indicates is not absolute but is relative. There are seven basic swaras: Sa, Ri, Ga, Ma, Pa, Da, Ni. These are related to each other by the fixed ratio of their frequencies. Usually, a 12 or 22 note system is used to indicate minor frequency variations in the swaras. In the 12 note system, Ga2, Ma2, Da2 and Ni2 are added. They represent half-step variations from 'Sa', the swara used to calibrate a frequency scale.

Depending on the number of swaras and their sequence, ragams are classified into various sub-divisions. If a ragam has all seven major swaras in it and the swaras are in ascending order of half-step variations from Sa in the arohanam and descending in the avarohanam, with the last swara being the Sa of the next octave, it is a *Melakarta* ragam or *Janaka* ragam. There are seventy-two such ragas. The term Janaka ragam indicates that these are parent ragas and that all other ragas can be derived from these.

3 Scale Dependant *Ragam* Identification using Harmonic Pitch Class Profiles (HPCP)

In this method *ragam* is identified by using Harmonic Pitch Class Profiles(HPCP). HPCPs are also known as chroma and are an important tool in audio signal processing to split an audio signal into a set number of bins. Each bin corresponds to a particular frequency along with its harmonics. For example, if the frequency of the bin is set as 16 Hz, all the signals at 32, 48 Hz etc. are added to the bin. The value of each bin is determined by summing up the magnitude of the various signals found near the frequency of the bin along with the harmonics.
In this project 12 bins are used. Each bin corresponds to each of the 12 notes found in music, C, C♯, D, D♯, E, F, G, G♯, A, A♯ and B. However, while mapping the frequencies found in the FFT of the audio file it is easily spotted that the frequencies do not correspond exactly to the frequencies of the bins or their harmonics. For this, the various frequencies in the FFT are mapped to the bins using the equation:

$$x_1 = round(12xlog_2(\frac{f}{fref}))|12$$

fref: Frequency of any bin or their harmonics. In this case the frequency taken is 261.63 Hz.
f: All frequencies found in the FFT.
round: Function which approximates the value of the expression to the nearest integer.
$|$: Gives the remainder when the function is divided by 12
x_1: Array with the same number of elements as found in the original audio file.

Once the various frequencies are assigned a bin each part of the signal that corresponds to a particular frequency is taken and the magnitudes are added. When the process is complete for all bins a 12 X 1 matrix is left. Each bin contains a magnitude, which relative to each other gives an indication of how often the note appeared in the audio file.

The audio files used in the project are all *Melakarta ragams* which mean that they contain 7 of the 12 notes. Therefore to identify the raagam of the audio clip if the 5 bins with the least magnitude can be compared to the 5 missing notes in the ragam the answer is found.

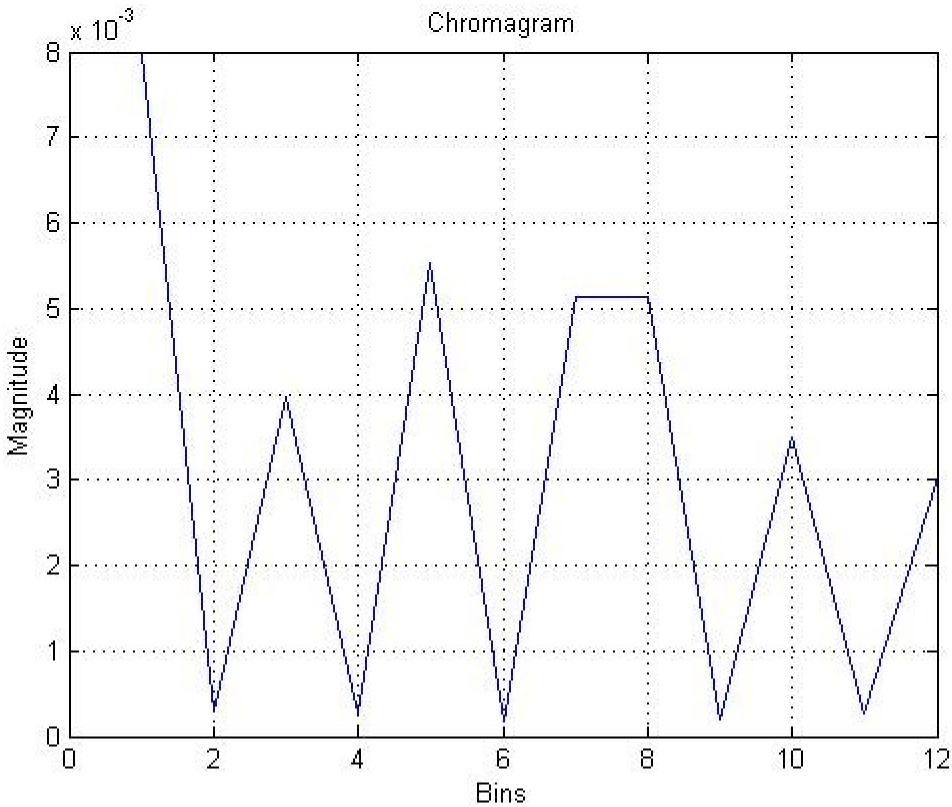

Fig 1. Chromagram for a single-frame audio input

4 *Ragam* Identification using GMM based Hidden Markov Models

In this part of the project, we try to model a Hidden Markov Model along the lines of [1]. The MIRToolbox[2] is used to extract required features in the MATLAB environment.

Initially, 12 chromas or Harmonic Pitch Class Profiles are extracted, which correspond to the twelve semitones in western classical music. A chromagram is a visual representation of the energies of each semitone and its harmonics. Since the energies of the harmonics are added to the fundamental semitones, computing the chromas results in a 12-dimensional vector for each frame of music. The music input to the function first undergoes a framing process, where the signal is divided into smaller frames to ease computation. In this project, each frame is 0.2 seconds long with a frame to frame overlap of 0.025 seconds.

Figure 1 represents a chromagram of a 80 second snippet of a song sung in *Kalyani* raga. Darker colours represent higher energies and the darkest section in a time frame corresponds to the swara in that time frame.

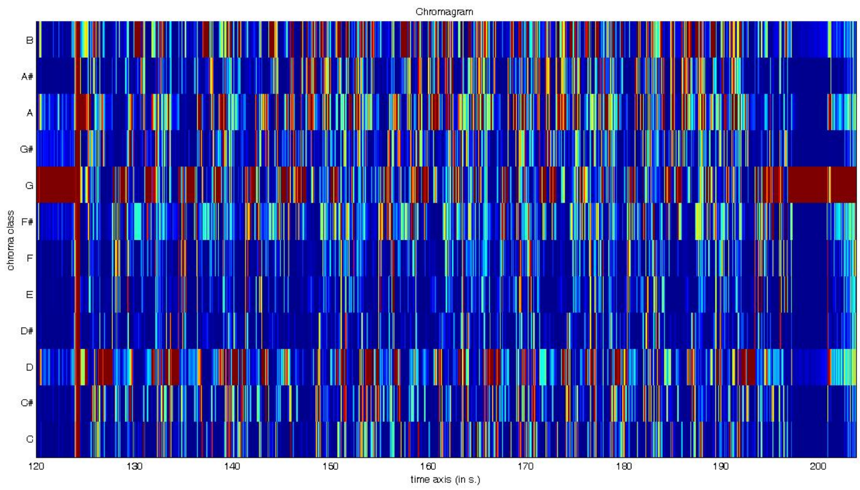

Fig 2. Chromagram for a piece in Kalyani Raga

4.1 Gaussian Mixture Model

A Gaussian Mixture Model is a parametric probability density function represented as a weighted sum of Gaussian Densities. The probabilty function is given by the equation

$$p(x|\lambda) = \Sigma\omega_i g(x|\mu_i, \Sigma_i); i = 1, ..., M$$

where x is a D-dimensional continuous valued data vector , ω_i are the weights and $g(x|\mu_i, \Sigma_i)$ are the Gaussian densities.

For this project, the 12 dimensional chroma data is divided into another 12 datasets per dimension, i.e. to obtain 12 Gaussian densities. The mean of these datasets is computed to obtain a 3-Dimensional array. Similarly, a 3-Dimensional covariance array is computed. Using these, a Gaussian Mixture distribution is obtained and a Gaussian Mixture Model is obtained by creating a multivariate random number matrix from the mean and the covariance arrays.

4.2 Hidden Markov Model

A Hidden Markov Model is a finite-state system whose output depends only on the current state where the state sequence is not directly observable. To build a HMM, we need a set of states, in this case, the 12 chroma features, the transitions probabilities i.e. the probability of moving from one state to another, and the emission probabilities, in this case, given by the Gaussian Matrix Model.

Since, in this project, the states are in the know, and the emission probabilities obtained, the transition probabilities need to be obtained. Initially assuming that the transitions are equiprobable, we generated a sequence of states of length l. Using this sequence, the transition and emission probabilities were improved upon and finally calculated the most-likely path through the Hidden Markov model. This path gives us the sequence of states or chromas which can be used to identify the ragam.

5 Shortcomings of the Project

In the first method used to identify the ragam, a major drawback of this method is that the chromagram needs to be calibrated to identify which of the bins acts as 'Sa'.

In the second method, only a model has been developed and has not been tested for lack of clean vocal samples of ragas, which brings us to another shortcoming of this project - it can only be used for monophonic sound signals and not polyphonic inputs.

6 References

1 Dighe, P., Agrawal, P., Karnick, H., Thota, S., Raj, B. *Scale Independent Raga Identification using Chromagram Features and Swara Based Features.* IEEE International Conference on Multimedia and Expo Workshops (ICMEW), 2013

2 Lartillot, O., Toiviainen, P.,*A Matlab Tool for Musical Features Extraction From Audio*, International Conference on Digital Audio Effects, Bordeaux, 2007

3 Reynolds D., *Gaussian Mixture Models*, MIT Lincoln Laboratory.

4 Rabiner L.R. *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*, Proceedings of the IEEE 77(2), 1989