International Journal of Research in Engineering and Innovation (IJREI)
journal home page: http://www.ijrei.com
ISSN (Online): 2456-6934

An efficient method for tonic detection from south Indian classical music

Unnikrishnan G

School of Computer Sciences, Mahatma Gandhi University, Kottayam, Kerala, India

Abstract

This paper proposes a novel method to identify the tonic value of Carnatic Music recordings. For the recognition and classification of ragas (Scales), we need to first transcribe or extract the different notes constituting those ragas. Transcription of music is the process of analyzing an acoustic musical signal to obtain the musical parameters of the sounds that occur in it. Sa is the tonic or basic note, based on which all other notes are derived in Indian Classical Music. Hence in order to identify the raga of an Indian classical music performance, identification of Sa is necessary. The proposed method proved successful with monophonic and polyphonic recordings which is a major advancement from earlier methods

Keywords: Pitch Estimation, Tonic detection, Sruthi, Raga, Octaves, Relative Pitch Ratio

1. Introduction

Computational musicology is an interdisciplinary research area focusing on the investigation of musicological questions with computational methods. It takes contribution from both computer science and musicology. The main objective of Computational Musicology is to represent a musical problem in terms of algorithms and corresponding data structures. The focus of research in Computational Musicology is not to study music as such, but to design methods to retrieve musical information from the acoustical signals of music recordings. Tasks involved in Computational Musicology include genre classification, raga recognition, melody extraction, artist recognition, song recommendation etc.

Carnatic music is the classical music of the southern states of India. Carnatic music compositions (called *kritis*, *keerthanas* etc) are based on *ragas*. The rendering of a composition typically start with *alapana* (improvisation) of the raga in which the composition is made, followed by the *kriti*. There are thousands of Carnatic *ragas*.

A *raga* is a melodic concept. *Matanga Muni*, in his text *Brihaddeshi*, defines raga as 'that which colours the mind of good through a specific *swara* and *varna* (literally colour) or through a type of *dhwani* (sound)' [1]. A definition of raga from a computational perspective is given by Chordia and Rae [2]. They define raga as a melodic abstraction which can be defined as a collection of melodic phrases. These phrases are sequences of notes or *swaras* that are often inflected with various micro-pitch alterations and articulated with an expressive sense of timing. Longer phrases are built by joining these melodic atoms together. Features of a raga are a set of constituent notes (*swaras*), their progressions (ascent/descent or *arohana/avarohana*), the way they are intonated using various movements (*gamakas*), and their relative position, strength and duration.

The constituent notes of a raga relate themselves to the base note or tonic called *Adhara Shadjam* denoted by *Sa*. The tonic is the base frequency selected by an artist for comfortable rendering. It may vary from artist to artist and performance to performance. The sequence of the constituent notes of a *raga* starting with *Sa* in one octave and ending with *Sa* in the next higher octave is called the *Arohana* (Ascent). Similarly the sequence starting from the upper *Sa* to lower *Sa* is called the *Avarohana* (Descent). The ascent and descent patterns can also be *vakra* containing subsidiary ascents and descents. In addition to the *Arohana* and *Avarohana*, a raga normally has *sanchara prayogas* which are melodic phrases peculiar to that raga. The *sanchara prayogas* will adhere to the *Arohana-Avarohana* pattern of the raga. They are distinguished by *gamakas* (microtonal ornamentations).

Raga recognition and classification is a central topic in Indian music theory, inspiring rich debate on the essential

Corresponding author: Unnikrishnan G
Email Id: ukgkollam@gmail.com

characteristics of *ragas* and the features that make two *ragas* similar or dissimilar [3]. Automatic *raga* recognition is the process of identifying *ragas* using computational methods. The core problem is to correctly identify the *raga* of a recording by analyzing the recording using computational methods. Automatic *raga* recognition has tremendous potential use in various areas including Music Information Retrieval, Teaching and learning of music, Practicing music, Multimedia Databases, Interactive Composition, Accompaniment Systems etc.

2. Pitch Intervals and Musical Notes

Efforts to describe and measure the properties of music date back to antiquity. Ancient Vedic texts on music mention the notion of octave equivalence and divide an octave into *'swarasthanams'* (Pitch intervals). There are seven basic swaras, known as *Sapta Swaras*. They are *Shadjam* (Denoted by '*Sa*'), *Rishabham* (*Ri*), *Gandharam (Ga), Madhyamam (Ma), Panchamam (Pa), Dhaivatham (Da) and Nishadam (Ni)*. *Sa* is the tonic or *Adhara Shadjam*, based on which all other notes are derived. A series of swaras, beginning with *Sa* and ending with *Ni*, is called a *Sthayi* or Octave [Table 1].

The frequency of a note in an octave will be twice the frequency of the same note in the previous octave.

Out of the seven *swaras*, '*Sa'* and '*Pa*' are constant. They are called *Achala Swaras* (fixed notes). The remaining five *swaras* have varieties and they are called *Chala Swaras* (varying notes) [Table 2]. For extraction of notes, the relative pitch of each note with respect to '*Sa'* in an octave is considered [3]. A method for extraction of notes should satisfy the Relative Pitch Ratio (RPR) [Table 2].

Observe that in rows 3,4,10 & 11 of Table 2, two different names are used to denote the same note position. The notes at those positions have the same RPR. The pairs of notes having this property are *Ga1* & *Ri2, Ri3* & *Ga2, Ni1* & *Da2* and *Da3* & *Ni2*. Thus there are a total of 16 note names even though the note positions are 12 in number. This naming convention of using two different names to denote the same note is a unique feature of Carnatic music and it allows certain combinations of notes, which would have been impossible otherwise [4].

Table 1: Three Octaves in Indian Classical Music

Mandra Sthayi							Madhya Sthayi							Thara Sthayi						
Sa	*Ri*	*Ga*	*Ma*	*Pa*	*Dha*	*Ni*	Sa	Ri	Ga	Ma	Pa	Dha	Ni	Sa	Ri	Ga	Ma	Pa	Dha	Ni

For example, the combination *Ri1-Ga1* or *Sudha Rishabham* and *Sudha Gandharam* becomes allowed only under this convention. Otherwise, the combination would have been *Ri1-Ri2*, which is not allowable, because both are variations of *Ri (Rishabham)*.

Various combinations of the notes discussed above constitute different *ragas* in Carnatic Music. Each *raga* will have a unique sequence of notes with uniformly increasing frequency in the ascent (called *Arohanam*) and decreasing frequency in the descent (called *Avarohanam*) that determines the characteristic of the *raga*. In general, all music compositions and other forms of musical improvisations based on a *raga* must contain the notes constituting that *raga*. The ascent or the descent of a *raga* should generally contain at least 4 notes. The common forms of *raga* scales are pentatonic or '*Audava*' scales, that is, those containing five notes including '*Sa*', hexatonic or '*Shadava*' containing six notes and heptatonic or the complete scale called the '*Sampurna*' containing seven notes [5]. There are such 72 *Sampurna* ragas which constitute the *Melakartha Raga* ystem in Carnatic music [Table 3]. These ragas are also called *Janaka* (Parent) *Ragas* as all other *ragas* in Carnatic music are considered to be generated from these ragas by various rearrangements of notes. Such generated *ragas* are called *janya* (Child) ragas.

Table 2: Musical notes and RPR values in Carnatic System

No	Symbol		Relative Pitch Ratio (RPR)	Decimal Value of RPR
1	Sa		1	1
2	Ri1		16/15	1.07
3	Ri2	Ga1	9/8	1.13
4	Ga2	Ri3	6/5	1.2
5	Ga3		5/4	1.25
6	Ma1		4/3	1.33
7	Ma2		17/12	1.42
8	Pa		3/2	1.5
9	Dha1		8/5	1.6
10	Dha2	Ni1	5/3	1.67
11	Ni2	Dha3	9/5	1.8
12	Ni3		15/8	1.88
13	Sa'		2	2

Table 3: The Melakartha Raga System in Carnatic Music

RG Combination	DN Combination	M1		M2	
		No.	Raga	No.	Raga
R1G1	D1N1	1	Kanakangi	37	Salagam
	D1N2	2	Ratnangi	38	Jalarnavam
	D1N3	3	Ganamoorthi	39	Jhalavarali
	D2N2	4	Vanaspathi	40	Navaneetham
	D2N3	5	Manvathi	41	Pavani
	D3N3	6	Thanaroopi	42	Raghupriya
R1G2	D1N1	7	Senavathi	43	Gavambodhi
	D1N2	8	Hanumathodi	44	Bhavapriya
	D1N3	9	Dhenuka	45	Subhapantuvarali
	D2N2	10	Natakapriya	46	Shadvidhamargini
	D2N3	11	Kokilapriya	47	Suvarnangi
	D3N3	12	Roopavathi	48	Divyamani
R1G3	D1N1	13	Gayakapriya	49	Dhavalambari
	D1N2	14	Vakulabharanam	50	Namanarayani
	D1N3	15	Mayamalavagowla	51	Kamavardhani
	D2N2	16	Chakravakam	52	Ramapriya
	D2N3	17	Suryakantham	53	Gamanasrama
	D3N3	18	Hatakambari	54	Viswambhari
R2G2	D1N1	19	Jhankaradhwani	55	Syamalangi
	D1N2	20	Natabhairavi	56	Shanmukhapriya
	D1N3	21	Keeravani	57	Simhendramadhyamam
	D2N2	22	Kharaharapriya	58	Hemavathi
	D2N3	23	Gowri Manohari	59	Dharmavathi
	D3N3	24	Varunapriya	60	Neethimathi
R2G3	D1N1	25	Mararanjini	61	Kanthamani
	D1N2	26	Charukesi	62	Rishabhapriya
	D1N3	27	Sarasangi	63	Lathangi
	D2N2	28	Harikamboji	64	Vachaspathi
	D2N3	29	Sankarabharanam	65	Mechakalyani
	D3N3	30	Naganadini	66	Chithrambari
R3G3	D1N1	31	Yagapriya	67	Sucharithra
	D1N2	32	Ragavardhani	68	Jyothiswaroopini
	D1N3	33	Gangeyabhooshani	69	Dhathuvardhani
	D2N2	34	Vagadheeswari	70	Nasikabhooshani
	D2N3	35	Soolini	71	Kosalam
	D3N3	36	Chalanatta	72	Rasikapriya

3. Tonic Detection

For the recognition and classification of ragas, we need to first transcribe or extract the different notes constituting those ragas. Transcription of music is the process of analyzing an acoustic musical signal to obtain the musical parameters of the sounds that occur in it. It can be seen as transforming an acoustic signal into a symbolic representation. Sa is the tonic or Adhara Shadjam, based on which all other notes are derived. Hence in order to identify the raga of a carnatic music performance, first we have to find the frequency of Sa (or the base frequency called tonic in which the performance is rendered). This is the first phase of any raga recognition process.
The proposed method for tonic detection is based on the following peculiar properties of musical notes.

3.1 Critical Bands and Dissonance

When we try to study about the frequencies of a musical scale, there is an important concept called the critical bands. When sound enters the ear, it causes vibrations on the basilar membrane within the inner ear. Different frequencies of sound cause different regions of the basilar membrane and its fine hairs to vibrate. This is how the brain discriminates between various frequencies. However, if two frequencies are close together, there is an overlap of response on the basilar membrane – a large fraction of total hairs set into vibration are caused by both frequencies. When the frequencies are nearly the same, they can’t be distinguished as separate frequencies. Instead an average frequency is heard. If the two frequencies are 440 Hz and 450 Hz, for example, we will hear 445 Hz. If the lower frequency is kept at 440 Hz and the higher one is raised slowly, then there will come a point where the two frequencies are still indistinguishable and there is just a roughness to the total sound. This is called dissonance. It would continue until finally the higher frequency would become distinguishable from the lower. At this point, further raising the higher frequency would cause less and less dissonance. When two frequencies are close enough to cause the roughness or dissonance described above, they are said to be within a critical band on the basilar membrane. For much of the audible range, the critical band around some "central frequency" will be stimulated by frequencies within about 15% of that central frequency [David R. Lapp].

In the study of musical notes and musical scales, critical bands play an important role. Two frequencies that stimulate areas within the same critical band on the basilar membrane will produce dissonance which is undesirable in music.

3.2 Consonance

The opposite of dissonance is consonance – pleasant sounding combinations of frequencies. In the previous section, the simultaneous sounding of a 440 Hz with a 450 Hz was discussed. If the 450 Hz is replaced with an 880 Hz (2 x 440 Hz), you would hear excellent consonance. This especially pleasant sounding combination comes from the fact that every crest of the sound wave corresponding to 440 Hz would be in step with every other crest of the sound wave corresponding to 880 Hz. So doubling the frequency of one tone always produces a second tone that sounds good when played with the first. This interval between two frequencies is called a diapason. Diapason means literally "through all." 440 Hz and 880 Hz sound so good together, in fact, that they sound the same. As the frequency of the 880 Hz tone is increased to 1760 Hz (2x880Hz or 4x440Hz), it sounds the same as when the frequency of the 440 Hz tone is increased to 880 Hz. This feature has led widely different cultures to historically use an arbitrary frequency and another frequency, exactly one diapason higher, as the first and last notes in the musical scale. As mentioned above, frequencies separated by one diapason not only sound good together, but they sound like each other. So an adult and a child or a man and a woman can sing the same song together simply by singing in different diapasons. And they'll do this naturally, without even thinking about it. The same applies to a vocalist and his supporting instrumentalist [7].

The above mentioned feature has been used as the underlying principle in the proposed tonic detection method. This method calculates the tonic based on a sample taken from the recording under study. This sample may contain either the vocal part or a supporting instrument like violin or a combination of both. Normally the base frequency of vocal and a supporting instrument like violin will have a difference of one diapason. That is, the frequency of a note generated from the violin will be twice the frequency of the same note generated by the vocalist. However, as mentioned above, due to consonance, two notes separated by a diapason will sound alike. This is the reason a violinist and a vocalist are able to perform in unison. Based on this fact, it is hypothesized that the tonic identification can be independent of the medium of performance. That is, we can identify the tonic from the sound of violin or from the sound of vocalist or from a combination of these two. In all these cases, the detected tonic can be used to identify the *raga* from the violin portion as well as from the vocal portion or from a combination of these two. It is also hypothesized that, since the tonic is medium independent, it can also be used to identify *raga* from portions containing polyphonic music, for example, from portions where the sound of *mridangam* (a percussion instrument in Carnatic music) or some other accompanying instrument is also present. These hypotheses have been successfully proved through experiments. This is a major advancement from earlier works where the tonic was found either by tuning an oscillator and noting the value in Hz [2] or by categorizing instruments as either male or female and asking explicitly for the tonic of the performer [8].

3.3 The Proposed Method

First of all, the wave form of the recording was analysed using any wave editor such as wavepad. By observing the lower amplitude portions which indicates the ending portions of *raga visthara* (elaboaration of a *raga* accompanied by the *thanpura* and sometimes the violin) or any other finishing portions. From this portion, a small piece was chosen for tonic detection. This musical piece from which the tonic *Sa* was to be extracted was stored as a 'wav' file with a sampling frequency of 44.1 KHz. The musical signal contained in the 'wav' file was first decomposed with a frame size of 25 ms. Pitch estimation was performed for each frame and the corresponding frequencies were obtained. Autocorrelation method was used for pitch estimation.

The extracted frequencies included groups of nonzero frequency values separated by zeros. The musical piece may contain frequencies other than the tonic frequency indicating, probably, the presence of other notes or even noise. Hence, as a criteria for separating the tonic frequency, it was assumed that more than one zero value coming together indicated a note boundary. That is, when more than one zero occurred together, it indicated the gap between two notes. So the nonzero frequencies up to that point represented a note. In order to fix the correct frequency of the note, all the nonzero frequencies up to that point were grouped and analyzed. Most of these frequencies were having only slight differences in their values and hence an average of these frequencies seemed to be an immediate choice for the frequency of the actual note.

However, it was observed that there existed some very high and very low frequencies among these extracted frequencies. This could be due to the various noises that can occur during a real performance. Due to the presence of these highly variant frequencies, the average differed highly from most of the frequencies. Obviously, average was not a good choice. In order to obtain a frequency that represented most of the extracted and grouped frequencies and to filter out the highly variant abnormal frequencies, another statistical measure median was chosen. Median of the frequencies in the group was computed and it was fixed as the frequency of the candidate (probable) tonic of that frame. The same process was repeated to find the candidate tonics from the subsequent frames. The process terminated when all the extracted frequencies were examined. The result is an array containing the resultant candidate tonic values. Majority of these candidate tonics were almost the same with only negligible differences. Again, the median of these candidate tonics was taken as the tonic of the audio recording under analysis.

3.4 Algorithm

1. Get the input wav file.
2. Decompose the whole signal contained in the input file into frames.
3. Estimate the pitch of the signal contained in each frame and obtain the corresponding frequencies.
4. Look for more than one zeros coming together, to identify a note boundary.
5. Evaluate median of the nonzero frequencies up to that boundary and fix it as a candidate tonic value.
6. Repeat from step 4.
7. Stop when all the extracted frequencies have been examined.
8. Fix the median of the candidate tonic values as the resultant tonic.

4. Results & Conclusions

Studies were conducted on recordings of musical performances by 63 renowned Carnatic musicians in 91 *ragas*. In the case of *Melakartha ragas*, atleast two performances in each *raga*, by different musicians were included. Also, as part of cross-verification to test the effectiveness of tonic detection, performances in 70 (out of 72) *melakartha ragas* by a single musician were also included. In this case, the tonic value was obtained from only one performance among these 70 performances. For the rest, the same tonic was assumed since the performances were of similar nature (*melakartha raga* demonstrations in 70 ragas by Nookala Chinna Satyanarayana). This assumption proved to be totally correct as there was a success rate of around 92% with this assumed tonic. Also, the tonic values were found to be independent of the medium of rendering. Hence tonic extracted from violin also suited for the vocalist and vice versa.

The method successfully detected the *ragas* from polyphonic recordings which is a major advancement from earlier methods. Out of the seventy recordings in *melakartha ragas*, 44 were having strong presence of *mridangam* and violin along with vocal. Also, the recordings used in this study were of varying qualities. Still, the *ragas* were detected correctly.

4.1 Results Summary

Table 4: Various *ragas*-Various performers

Average Sample Duration (in seconds)	35	100	180
Number of *Ragas* Tested	70	70	70
No. of Correctly Identified *Ragas*	60	66	67
Success Rate	85.7	94.3	95.7

Table 5: Various ragas-Same performer

Average Sample Duration (in seconds)	35	100	180
Number of *Ragas* Tested	70	70	70
No. of Correctly Identified *Ragas*	58	63	64
Success Rate	82.9	90	91.4

4.2 Waveforms

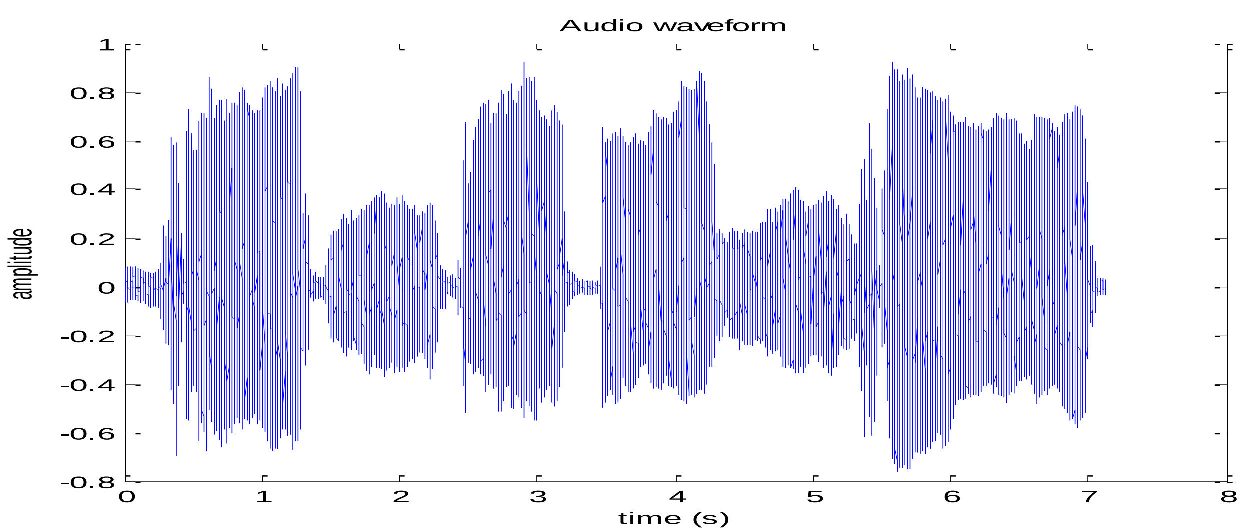

Figure 1: Raga: Hamsadhwani (Arohana: SaRi2Ga3PaNi3Sa’, Avarohana: Sa’Ni3PaGa3Ri2Sa).

'*Hamsadhwani* ' is an *'audava'* or pentatonic *raga* having five notes *Sa, Ri2, Ga3, Pa and Ni3.*

Table 6: Note frequencies extracted from the wave form in Figure 1

Note	Extracted Frequency (Hz)	Computed RPR	Ideal RPR
Sa	209	1	1
Ri2	235	1.12	1.125
Ga3	260	1.24	1.25
Pa	313	1.49	1.5
Ni3	390	1.86	1.875

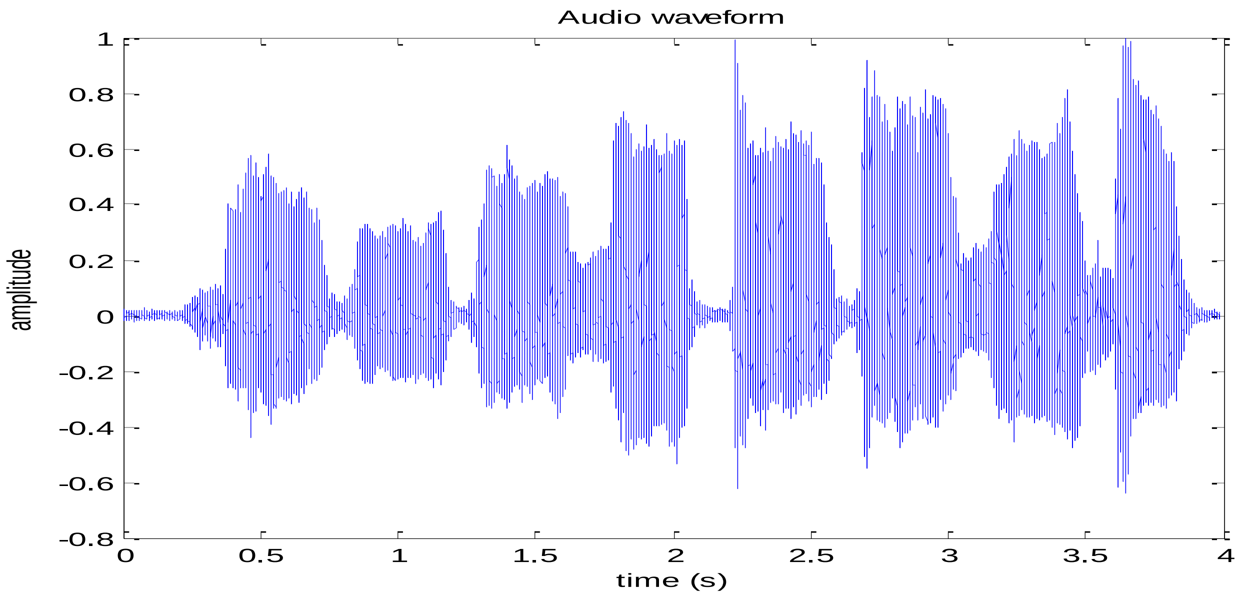

Figure 2: Raga: Mayamalavagowla (Arohanam: SaRi1Ga3Ma1PaDa1Ni3Sa', Avarohanam: Sa'Ni3Da1PaMa1Ga3Ri1Sa).

'*Mayamalavagowla*' is a *Sampoorna* (Heptatonic) *raga* having seven notes *Sa, Ri1, Ga3, Ma1, Pa, Da1, and Ni3.*

Table 7: Note frequencies extracted from the wave form in figure 2

Note	Extracted Frequency (Hz)	Computed RPR	Ideal RPR
Sa	110	1	1
Ri1	119	1.08	1.07
Ga3	136	1.23	1.25
Ma1	146	1.32	1.33
Pa	168	1.52	1.5
Da1	173	1.57	1.6
Ni3	210	1.91	1.875

Figure 1 shows the wave form of the *arohana* of *Hamsadhwani.* Six notes *Sa, Ri2, Ga3, Pa, Ni3* and the upper (*Thara sthayi*) *Sa* can be observed in the figure. Similarly, Figure 2 shows the wave form of the *arohana* of *Mayamalavagowla.* Eight notes Sa, Ri1, Ga3, Ma1, Pa, Da1, Ni3 and the upper (*Thara sthayi*) *Sa* can be observed in the figure. Table 6 and 7 show the extracted frequency values from the waveforms shown in Figure 1 and Figure 2 respectively, using the proposed method. They also show the computed RPR values using the extracted frequencies and the ideal or theoretical RPR values. It can be observed that the computed RPR values are very close to the ideal values which shows the accuracy of the proposed method.

References

[1] 1. Brihaddesi of Sri Matanga Muni Edited and Translated by Prem Lata Sharma.

[2] Chordia, P., Rae, A., Raag Recognition using Pitch-Class and Pitch-Class Dyad Distributions, Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR), Vienna, Austria, 2007

[3] Arvindh Krishnaswamy, "On the Twelve Basic Intervals in South Indian Classical Music", Audio Engineering Society Convention Paper presented at the 115th Convention, 2003 October 10-13, New York.

[4] K. Balasubramanian, "Combinatorial Enumeration of Ragas (Scales of Integer Sequences) of Indian Music", Journal of Integer Sequences, Vol. 5 (2002).

[5] Arvindh Krishnaswamy, "Application of Pitch Tracking to South Indian Classical Music", In Proc. IEEE ICASSP 2003, Apr 6-10, Hong Kong

[6] Anssi P. Klapuri, "Automatic Music Transcription as We Know it Today", Journal of New Music Research 2004, Vol. 33, No. 3

[7] David R. Lapp, "The Physics of Music and Musical Instruments", Wright Center for Innovative Science Education, Tufts University Medford, Massachusetts

[8] James K N, "Realtime Raga Detection and Analysis using Computer", Ph.D Thesis, CUSAT, Kochi.