A KNOWLEDGE BASED SIGNAL PROCESSING APPROACH TO TONIC IDENTIFICATION IN INDIAN CLASSICAL MUSIC

Ashwin Bellur
Department of
Electrical Engineering,
IIT Madras, India
ashwinbellur@gmail.com

Vignesh Ishwar
Department of
Computer Science & Engineering,
IIT Madras, India
vigneshishwar@gmail.com

Xavier Serra
Music Technology Group
University of Pompeu Fabra,
Barcelona, Spain
xavier.serra@upf.edu

Hema A Murthy
Department of
Computer Science & Enginnering,
IIT Madras, India
hema@cse.iitm.ac.in

ABSTRACT

In this paper, we describe several techniques for detecting tonic pitch value in Indian classical music. In Indian music, the rāga is the basic melodic framework and it is built on the tonic. Tonic detection is therefore fundamental for any melodic analysis in Indian classical music. This work explores detection of tonic by processing the pitch histograms of Indian classic music. Processing of pitch histograms using group delay functions and its ability to amplify certain traits of Indian music in the pitch histogram, is discussed. Three different strategies to detect tonic, namely, the concert method, the template matching and segmented histogram method are proposed. The concert method exploits the fact that the tonic is constant over a piece/concert. template matching method and segmented histogram methods use the properties: (i) the tonic is always present in the background, (ii) some notes are less inflected and dominant, to detect the tonic of individual pieces. All the three methods yield good results for Carnatic music ($90-100\%$ accuracy), while for Hindustani music, the template method works best, provided the *vādi samvādi* notes for a given piece are known (85%).

1. INTRODUCTION

Melody is a fundamental element in most music traditions. Although melody is a common term that is used to categorize certain musical elements, each tradition has specific differences. Indian classical music is an example of a tradition with specific melodic traits, especially when compared to that of western classical music.

In western classical music, a melody is normally defined as a succession of discrete tones, tones that belong to a given scale and tonality context. Most melodic studies use the symbolic representation of the music and use concepts like notes, scales, octaves, tonality and key signatures. Also, given that western classical music uses equal temperament tuning, melodic analysis of a western piece of music is normally based on a quantized representation of pitches and durations within a well defined framework of possible relationships.

Melody in Indian classical music relates to the concept of *rāga*. This has little to do with the western concepts of tones and scales. A *rāga* also prescribes the way a set of notes are to be inflected and ordered. Most Indian instruments do not have a specific tuning, and if any, it is more related to just intonation than to equal temperament [1]. This music tradition has been preserved and has evolved as an oral tradition in which notation plays a very little role.

A fundamental concept in Indian Classical music is the one of tonic. The tonic is the base pitch chosen by a performer, used as a reference throughout a performance. Melodies are defined relative to the tonic. All the instruments accompanying a lead performer tune to that tonic and all the rāgas performed use that tonic as the base note of the rāga. The reference note is the note *Sa* (also called *ṣaḍja*) [1]. A simplified view of the difference with western classical music would be that Indian music uses a fixed tonic while western classical music uses a movable tonic, since for each key a different reference tonic is used. On the other hand, in western music, a fixed frequency is used as reference for tuning, normally A4 (440 Hz), while in Indian music there is no reference tuning frequency, the reference is the tonic of the lead performer. The frequency of the tonic of male vocalists is normally in the range of 100 to 180 Hz, while that female singers is in the range of 160 to 280 Hz. The tonic of lead instruments varies from 140 Hz to 200 Hz.

The tonic is ubiquitously present in Indian Classical music. The drone is played by either the *Tanpura*, an electronic *Śhruti* box, or by the sympathetic strings of an instrument like the *Sitār* or *Viṇā*. The sound of the drone consists of the tonic plus other related tones ($\frac{4}{3}\times tonic, 1.5\times$

[1] ṣaḍja and tonic are used interchangeably in this paper.

$tonic, \frac{15}{8} \times tonic, 2 \times tonic$) This drone is the reference sound that establishes the harmonic and melodic relationships during a given performance.

As the tonic is chosen by the performer, the pitch (or note) histogram of the same *rāga* can occupy different pitch ranges (Figure 1). Figure 2 shows the histogram on the cent scale, evaluated after normalizing the pitch extracted with respect to the tonic [2] .

$$c = 1200(log_2(\frac{f_2}{f_1})) \qquad (1)$$

where c is the cent value of frequency f_2 with respect the *ṣaḍja*/tonic f_1. Given that the pieces are of the same *rāga*, similarities across histograms are evident in Figure 2 when compared to Figure 1.

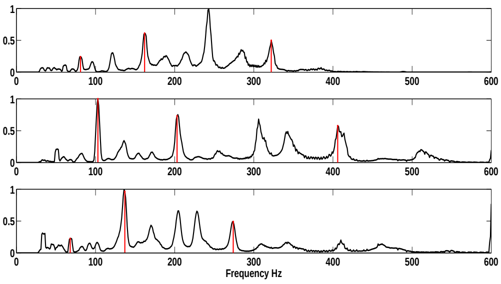

Figure 1. *Pitch Histogram of three performances of rāga kāmbōji by three different artists. The solid red line denotes the ṣaḍja*

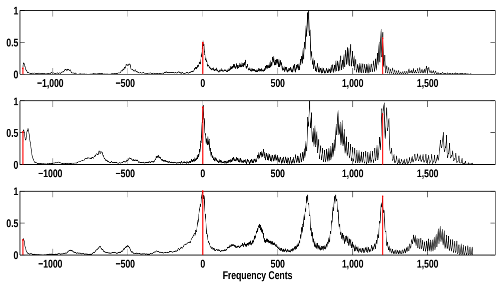

Figure 2. *Pitch Histogram of the three performances on the cent scale, after normalizing with respect to the tonic. ṣaḍja can be seen at -1200, 0 and 1200 cents.*

There have been various efforts to apply computational methods to analyze different aspects of Indian music using pitch [2] as the basic feature. In [3, 4], pitch class distribution and pitch dyads are employed for automatic *rāga* recognition. [5] employs a form of Hidden Markov Models with pitch as the basic feature to do the same. In [6] an attempt is made to study inflections/*gamakas* using pitch contours. [1] address the tuning issue in Indian music using pitch histograms.

Any melody based analysis of Indian music requires the identity of the tonic pitch value. In [3] and [4], tonic is manually identified. Serra et al. [1] use an interval histogram of the notes to remove the effect of tonic variation across pieces. Ranjani et al. [7] have attempted automatic tonic detection for Carnatic music by modeling pitch histograms using semi-continuous Gaussian Mixture Models (SC-GMM). Ranjani et al. exploit the presence of *gamakas* to detect tonic. *The term gamaka*, refers to meandering around a note rather than playing/singing the absolute note. The fact that note *ṣaḍja* and the note panchama at $1.5\times$ *ṣaḍja*, and their corresponding lower and higher octave equivalents, are less inflected compared to the other notes, is used to detect tonic. Assuming that any musician can utmost span three octaves, and that there are 12 semitones per octave, 36 mixture GMMs are used. A set of rules involving the variance and responsibility measures of each of the mixtures is attempted to detect tonic. The work [7] reports results *rāga Ālāpanas* alone on a small data set [3] .

In this paper, an attempt is made to perform tonic identification on an entire piece (including lead vocal, instrument and, accompanying instruments). Signal processing techniques are employed to determine the tonic. These techniques for detecting tonic are attempted on a large, varied dataset to test the robustness of the methods. In Section 2, we discuss the process of obtaining pitch histograms which form the basic representation of a music piece. The pitch histograms are further processed using group delay functions. The need for post processing of histograms and motivation to use group delay histograms is also explained in this Section. In Section 3, three different methods for automatic tonic identification from an entire piece are discussed, each exploiting some underlying characteristic that enables tonic detection in Indian music. A variation of the methods is proposed for detecting tonic in Hindustani music. Finally, the conclusions are presented in Section 4.

2. PITCH EXTRACTION AND GROUP DELAY HISTOGRAMS

In this work, Yin [8] has been used to extract pitch information. In [7], the authors have dealt with pitch extracted on the rāga ālāpana alone. Though it is indeed difficult to work with percussion due to discontinuities arising in the single pitch extracted using Yin, nevertheless, since percussion is also tuned to the same tonic, retaining segments with percussion does aid in tonic detection. Figure 3, shows a histogram evaluated on the pitch extracted from a piece with just *mṛdaṅgaṁ* [4] . Two clear peaks at the *ṣaḍja* of the middle and lower octave can be seen, indicating that percussion can aid in the detection of the tonic.

In this work. pitch histograms are processed using group delay functions to aid tonic detection. Group delay based features have been used extensively in the area of speech processing [9]. The group delay function is defined as:

$$\tau(\omega) = -\frac{d\phi(\omega)}{d\omega} \qquad (2)$$

[2] cent is an unit of measure used for musical intervals on the logarithmic scale.

[3] Ālāpana is a melodic improvisation without percussion.

[4] Percussion in a Carnatic music (Hindustani music) concert is provided by the instrument called *mṛdaṅgaṁ* (*tablā*).

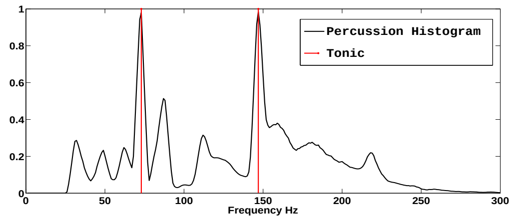

Figure 3. *Percussion Histogram*

where $\phi(\omega)$ is the phase of the Fourier Transform of a signal. The deviation of the group delay function from a constant corresponds to the nonlinearity of the phase as a function of frequency. When the group delay function is used as a model to represent the vocal tract, the peaks correspond to the poles of the transfer function, while the valleys correspond to the zeros of the transfer function. The peaks in the group delay function are inversely proportional to the bandwidth of the group delay function. Further, non-model based minimum phase group delay functions are very useful in resolving closely placed formants in speech, owing to their additive and high resolution properties. In this work, the histogram is treated as a power spectrum, with closely spaced peaks of histogram analogous to the closely spaced formants. Each peak in the histogram can be thought of as the impulse response of a pair of complex conjugate poles:

$$H(z) = \prod_{i=1}^{n} \frac{1}{(1 - z_i z^{-1})(1 - z_i^* z^{-1})} \tag{3}$$

where n corresponds to the number of peaks in the histogram. The group delay function of $H(z)$ is given by:

$$\tau_h(\omega) = \sum_{i=1}^{n} \tau_{z_i}(\omega) + \tau_{z_i}^*(\omega) \tag{4}$$

Modeling $H(z)$ using the allpole model as in Equation 3, requires that the order of the model be known.

Alternatively, for minimum phase signals, the group delay function can be computed as the Fourier transform of the weighted real cepstrum $c[n]$ [9]:

$$\tau_h(\omega) = \sum_{n=1}^{\infty} nc[n]cos\omega n \tag{5}$$

In Equation 5, the cepstrum can be obtained from the power spectrum or rather in the current context, the pitch histogram as:

$$c[n] = IDFT(\log P_H(\omega)) \tag{6}$$

where $P_H(\omega)$ corresponds to that of the pitch histogram (treated as a power spectrum) and IDFT corresponds to the *Inverse Discrete Fourier Transform*

We replace the $\log$ operation by $|.|^\gamma$ as in [9],i.e:

$$c_r[n] = IDFT(|P_H(\omega)|^\gamma) \tag{7}$$

The advantage of this form of the cepstrum (Equation 7) over that of Equation 6 in the context of pitch histograms

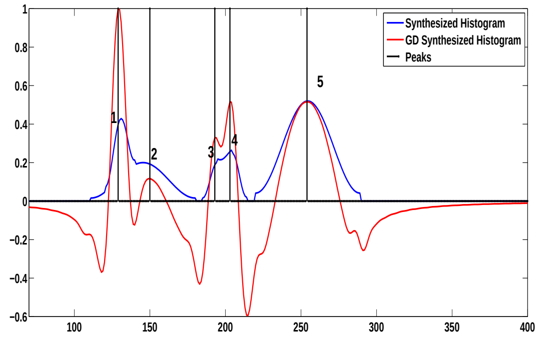

Figure 4. Illustration of the resolving power of group delay functions

is that, for values of the parameter $\gamma < 1$, even small peaks in the histogram can be resolved. The heights of the peaks are inversely proportional to the bandwidth, thus emphasizing the less inflected *ṣaḍja*. Figure 4 shows the effect of group delay processing on a synthetic histogram. Observe that the third and fourth peak are resolved very well in the group delay processed histogram. Also the first peak with a narrower bandwidth gets accentuated. We shall refer to the group delay processed histogram as Group Delay histogram (GD histogram). It is important to note that in computing the group delay histogram no effort is made to model the number of peaks in the pitch histogram.

3. TONIC IDENTIFICATION USING PROPERTIES OF THE ENTIRE WAVEFORM

In this section, three methods of tonic detection are explored. Features are extracted from the raw waveform to detect tonic. No effort is made to remove silences, applauses, noise, etc. In each method a specific property of Indian music is exploited to detect the tonic. The techniques are based on processing the pitch histograms using the relevant domain knowledge. In the following sections, we discuss different methods for identifying tonic for Carnatic music. These techniques are then applied to Hindustani music. Owing to the differences between Carnatic and Hindustani music, appropriate changes to the algorithms are suggested.

3.1 Method 1 - Concert based method

The database for Indian music, is generally in the form of audio CDs or recordings of concerts. A concert or an audio CD can be considered as a unit by itself. A Carnatic music concert or an audio CD, consists of a number of pieces in different *rāgas*. The *rāgas* are seldom repeated. Although the *rāgas* are different, the tonic in which they are rendered is kept constant. In addition to this, every *rāga*, contains the *ṣaḍja*, along with a subset of the 12 semitones that make up the *rāga*.

The basic idea of the approach for tonic identification proposed in this Section, is to identify the tonic for every concert. To detect the tonic of the concert the following algorithm is used:

1. Compute the GD histograms of all individual pieces, namely, $GDP_i, 1 \leq i \leq n$ in a concert, where n corresponds to the number of pieces in a concert.

2. Compute $\prod_{i=1}^{n} GDP_i$.

Since the *rāgas* of different pieces are different, with *ṣaḍja* being present across all pieces, the peak corresponding to that of the tonic must dominate in Step 2. Figure 5 shows the GD histograms for four pieces taken from the same concert. The fifth row in Figure 5 is the product of the four GDP_is evaluated on the four pieces performed in the concert. The dominant peak is the tonic used in the concert. It must be noted, that for a given individual song, the most dominant peak might not be the tonic, (first row in Figure 5) other notes may dominate the individual histograms. But with the *ṣaḍja* present in the percussion and drone, and with every *rāga* having the *ṣaḍja*, a prominent peak for the *ṣaḍja* in the histogram and GD histogram is guaranteed. Tonic identification thus reduces to determining the frequency of the peak that has the maximum value.

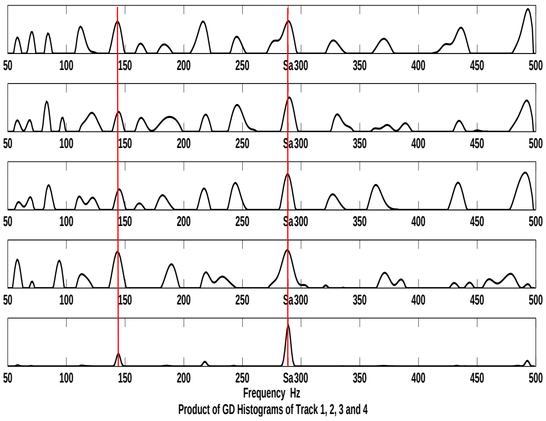

Figure 5. *Concert method (cent scale)*

3.2 Method 2 - Template matching

Although the previous method can be used for normalizing pitch values for large number of pieces in a concert, it will not work, when only individual pieces are available. The objective is to perform tonic identification when provided with individual pieces.

In this method, less inflected nature and the fixed ratio between *ṣaḍja* and *panchama* are exploited ($panchama = 1.5 \times ṣaḍja$). This method is comparable to that [7], where they attempt to exploit the same characteristics using SC-GMM. While in [7] five different rules are explored, in this work, *ṣaḍja-panchama* templates are used on the histogram and GD histograms. The procedure to detect tonic is as follows:

- Compute histograms and GD histograms.

- Let f_i, $i \in [1, N]$, correspond to the frequencies of the N peaks of the histogram.

- Let L be a vector such that: $L[k] = v_i$ for $k \in [f_i ... f_N]$, v_i is the height of the peak at f_i and $L[k] = 0, elsewhere$

- Each peak location is a candidate *ṣaḍja*. Now let f_j be the frequency of a candidate *ṣaḍja*, say S_j. $j \in [1, N]$.

- Given the frequency of S_j, the expected frequencies of *ṣaḍja* and panchama across the 3 octaves under consideration are

$$\begin{bmatrix} 0.5(f_j) & 0.75(f_j) & f_j & 1.5(f_j) & 2(f_j) & 3(f_j) \\ S_{j_{lower}} & P_{j_{lower}} & S_j & P_j & S_{j_{higher}} & P_{j_{higher}} \end{bmatrix}$$

$$E = [0.5\ 0.75\ 1\ 1.5\ 2\ 3] f_j$$

- Let T_j be the template vector for a test piece such that: $T_j[k + \delta : k - \delta] = 1$ for $k \in E$; δ allows for a leeway of δ around the expected peak, $T_j[k] = 0, elsewhere$

- $C_j = L^T T_j$

- $tonic = \underset{j}{argmax}\ C_j, j \in [1, N]$

This is a template matching procedure, where different templates are used for different candidate *ṣaḍjas*. GD histograms work well for template matching, when compared to histograms. It can be seen in Figure 6, even though GD histogram flattens the histogram, local peaks at the panchama get accentuated (property of the note and group delay functions), thus leading to trivial peak picking and template matching.

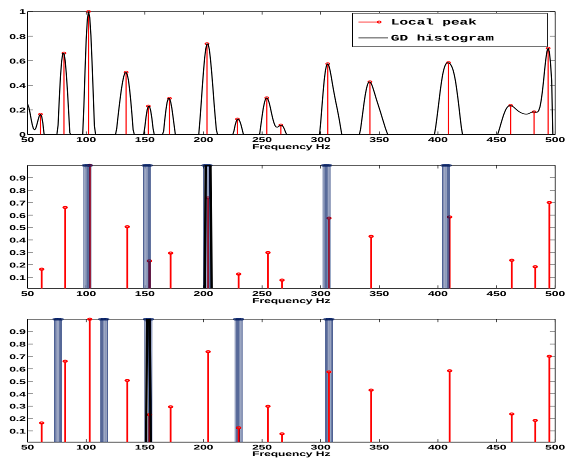

Figure 6. *Illustration of the template matching procedure. plot 1 shows the local peaks in Gd histograms. Plot 2 and 3 show the template matching procedure for two different cases. The black strip represents the local peak assumed as the tonic and the blue strip represents the corresponding template. Plot 2 being the correct estimate of the ṣaḍja, a better template match is obtained.*

3.3 Method 3 - Segmented histograms

In the method illustrated in the previous section, there are a few issues. The assumption is that the peak at which the template fits best is the *ṣaḍja*. There are a few drawbacks of this method. Since the template is basically using the fact that the *panchama* is $1.5\times$ *ṣaḍja* with respect to the *ṣaḍja*, there might be a perfect template fit for another set of notes with the same template. It is also evident that this method might fail for *rāgas* without the *panchama*, since the background drone (tuned to *ṣaḍja*, *panchama*) is seldom picked up by any single pitch extraction algorithm, when the lead musician dominates. To address these issues, another method for tonic identification was devised using piecewise histograms.

Figure 7, shows the histogram and GD histogram of a single four minute Carnatic piece. The note marked "*" on the X axis is most frequented, whereas "+" is the *ṣaḍja*. Global peak picking would have resulted in wrong estimation of *ṣaḍja*. As an attempt to detect *ṣaḍja* inspite of it not being the most dominant note even in the GD histogram, a given music piece is segmented into units of duration of $\approx$ one minute. The histograms and GD histograms are calculated on the pitch extracted from the segmented pieces. As mentioned before, the presence of the drone and the *mṛdaṅgaṁ* ensure that the segmented histograms will always show a local peak at the *ṣaḍja*. In Figure 8, plots 1-4 show the histograms and GD histograms computed on the segmented pieces. It can be seen that a local peak at the *ṣaḍja* is always present. Figure 8 also illustrates the ability of group delay function to accentuate peaks with narrow bandwidth. The ṣaḍja peak becomes prominent in each of the segment GD histograms. Similar to the method employed in the concert based method, the product of the segmented GD histograms is obtained. This is followed by picking the global peak to determine the tonic.

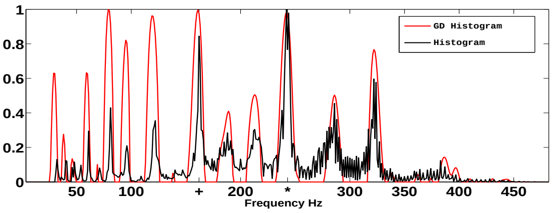

Figure 7. *Histogram and GD histogram of a single four minute Carnatic piece. The histogram bin marked with a "*" on the X axis is the most frequented note, whereas the bin marked "+" is the ṣaḍja*

3.4 Experiments and Results

The above mentioned methods were tested on a fairly heterogeneous data set. The dataset contains a mix of:

- Studio recordings released as audio compact discs.
- Professionally recorded concerts released as compact discs.
- A private collection of live concert recordings.
- Cassette recordings converted to digital audio.

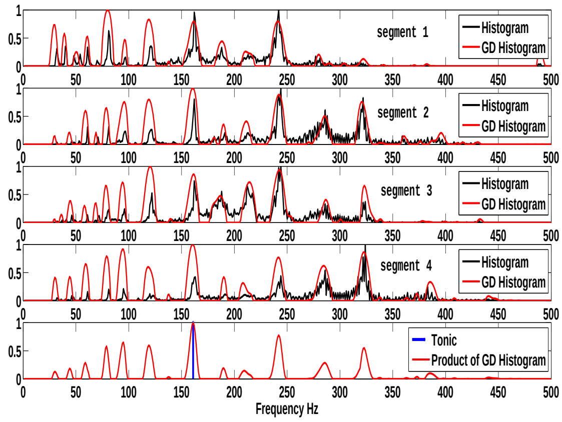

Figure 8. *Plots 1-4 show segment wise histogram and GD histogram. Plot 5 is the product of GD histograms in plot 1-4, with ṣaḍja as the global peak*

3.4.1 Carnatic Music

For Carnatic music, a set of 78 concerts (44 male and 13 female artists, 21 instrumental leads) were randomly chosen from a personal collection. The 78 concerts selected comprised of 722 pieces. Tonic was then estimated for the concerts and the individual pieces manually (by a professional musician) against which the performance of the three methods were tested. All the three methods described are used to detect the tonic for Carnatic music.

- Method 1 - In Concert/recording based tonic identification, considers a recording/concert as a unit. Tonic identification is performed using both histograms and GD histogram.
- Method 2: Tonic identification of individual pieces using templates. Two different templates were used (Template 2 is the same as Template 1, except that S_m is not used). [5] $: Template1[2] = [S_m]\ P_m\ S\ P\ S_t\ P_t$
- Method 3 – Tonic identification of individual pieces using segmented histograms.

Method 2 and 3, use the Male/Female/Instrumental information to restrict the range within which tonic is estimated. Table 1 summarizes the results for Carnatic music. The concert-based method was 100% successful, while for the piece based methods, the segmented GD histograms give the best performance.

3.4.2 Hindustani Music

The performance of Method 1 used for Carnatic music, when attempted for Hindustani music was rather poor. This is because in most Hindustani music concerts, the number of pieces performed is two or three. The melodies (or *rāgas*) chosen are based on the time of the day. It often

[5] m corresponds to that *mandara stāyi* (lower octave), t corresponds to that *tāra stāyi* (upper octave).

Method	GD Histogram	Histogram
Method 1	100%	100%
Method 2 (T_1)	95 %	92.17%
Method 2 (T_2)	92.37 %	91.66%
Method 3 (US)	90.69 %	87.67%
Method 3 (S)	95.28 %	87.67%

Table 1. *Accuracy of tonic recognition methods for Carnatic music. T_1 = Template 1, T_2 = Template 2, US = Unsegmented, S = Segmented*

happens that *rāgas* chosen may belong to the same *chalan*[6]. Any of the common notes across pieces may dominate on taking the product of the histograms instead of the *ṣaḍja*. The other factor is, in *Hindustani music*, the notes are less inflected compared to that of Carnatic music. Techniques discussed in Sections 3.2 and 3.3, which rely on the prominence of *ṣaḍja* and the ability of GD histogram to emphasize less inflected nature of *ṣaḍja* and panchama relative to the other notes, do not work for Hindustani music as can be seen in Table 2. On the other hand, it was observed that every *rāga* has *vādi* and *samvādi* notes. These are essentially notes that are most dominant in a given *rāga*. Therefore, in Hindustani music, in addition to the *ṣaḍja-panchama* template, the use of *vādi* and *samvādi* based template was explored. Templates used in Method 2 were modified to include the *vādi* and *samvādi* notes. A number of different templates were used based on the *rāga* of the piece.

$$TemplateVS(T_{VS}) = [S\ vadi\ samvadi\ S_t]$$

For example, for *rāga Darbari*, the template is:

$$[S\ R_2\ P\ S_t]$$

The note R_2 and P are the *vādi* - *samvādi* notes. The methods were tested on 126 pieces of Hindustani music. Table 2, summarizes the results of tonic identification for Hindustani music. The performance is reported for GD Histograms. A marked improvement in the accuracy of tonic detection can be seen in Table 2 on using the modified template. The drawback of this approach is that the knowledge of the *rāga* is required.

Method	Accuracy (GD histogram)
Method 2 (T_1)	66 %
Method 2 (T_{VS})	84.9 %
Method 3 (S)	62%

Table 2. *Accuracy of tonic recognition methods for Hindustani music. T_1 = Template 1, T_{VS} = Template VS, S = Segmented.*

4. CONCLUSION

A knowledge-based signal processing approach is proposed to perform tonic identification for Indian music, using pitch histograms as primary form of representation. Group delay processing of pitch histograms, necessitated by the presence of inflected notes are shown to improve the performance of tonic detection. The results estimated on a large varied dataset, indicate that the proposed methods are highly accurate for detecting tonic/*ṣaḍja* for Carnatic music. In Hindustani music,the notes being relatively less inflected compared to that of Carnatic music, the use of the dominant *vādi* and *samvādi* note information is shown to be vital in detecting the *ṣaḍja*.

[6] Notes of different melodies having a common subset or similar phraseology.

5. ACKNOWLEDGEMENTS

The authors would like to thank Prof. M V N Murthy for useful discussions on the concept of tonic and the concept of *vādi* and *samvādi* notes in Hindustani music. This research was partly funded by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

6. REFERENCES

[1] J. Serra, G. K. Koduri, M. Miron, and X. Serra, "Tuning of sung indian classical music," *In Proc. of ISMIR*, pp. 157–162, 2011.

[2] A. Krishnaswamy, "Application of pitch tracking to south indian classical music." *In Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 557–560, 2003.

[3] P. Chordia and A. Rae, "Raag recognition using pitch-class and pitch-class dyad distributions," *In Proc. of ISMIR*, pp. 431–436, 2007.

[4] P. Chordia, J. Jayaprakash, and A. Rae, "Automatic carnatic raag classification," *Journal of the Sangeet Research Academy (Ninaad)*, 2009.

[5] G. Pandey, C. Mishra, and P. Ipe, "Tansen: A system for automatic raga identification." *Indian International Conference on Artificial Intelligence*, pp. 1350–1363, 2003.

[6] A. Krishnaswamy, "Inflexions and microtonality in south indian classical music," *Frontiers of Research on Speech and Music*, 2004.

[7] H. G. Ranjani, S. Arthi, and T. V. Sreenivas, "Shadja, swara identification and raga verification in alapana using stochastic models," *In 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 29–32, 2011.

[8] A. D. Cheveigne and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *Journal of the Acoustical Society of America*, p. 111(4):1917–1930, 2002.

[9] H. A. Murthy and B. Yegnanarayana, "Group delay functions and its application to speech processing," *Sadhana*, vol. 36, no. 5, pp. 745–782, November 2011.