Tonic Note Extraction In Indian Music using HPS and Pole Focussing Technique

Conference Paper · July 2019

DOI: 10.1109/ICCCNT45670.2019.8944486

3 authors, including:

Rahul Kodag
University of Mumbai
2 PUBLICATIONS **4** CITATIONS
SEE PROFILE

Mukesh Patil
Ramrao Adik Institute of Technology
16 PUBLICATIONS **229** CITATIONS
SEE PROFILE

Tonic Note Extraction In Indian Music using HPS and Pole Focussing Technique

Chandrakant J. Gaikwad
Electronics & Telecommunication Department,
Ramrao Adik Institute of Technology, Navi Mumbai
Email:cjgaikwad@gmail.com

Rahul B. Kodag
Electronics & Telecommunication Department,
Ramrao Adik Institute of Technology, Navi Mumbai
Email: kodagrahul@gmail.com

Mukesh D. Patil
Electronics & Telecommunication Department,
Ramrao Adik Institute of Technology, Navi Mumbai
Email: mukesh.rait@gmail.com

***Abstract*—In Indian music, tonic note extraction is important for computational analysis and related problems. It is a reference note to which other notes are related by the scale used by the singer. In this work, pole focussing technique is proposed for extraction of tonic note and it is shown that the accuracy of methods available in literature can be further increased by using the proposed technique. The methods available in literature for tonic extraction are compared and through simulations, it is shown that the proposed method performs better than the methods available in literature.**

***Index Terms*—Tonic Note, Pole focussing, Indian music, Histogram, Highest peak**

I. Introduction

Carnatic and Hindustani music are the two important constituents of Indian music. In these two traditions, raga forms the fundamental melodic framework [1]. In literature, we find many works on fundamentals of Indian music [2], [3]. The essential basics are discussed below.

The fundamental swars (Notes) in Indian music system are Shadja, Rishabha, Gandhara, Madhyama, Panchama, Dhaivata and Nishadha. Often, short notations are used for these notes. These are Sa, Re, Ga, Ma, Pa, Dha and Ni respectively. Svar Sa is also called as a tonic note [1], [4], [5]. Further, these notes have either sharp or flat variations, they are called as teevra and komal versions. The notes Sa and Pa do not have teevra or komal version. Thus, for practical analysis of music, in this work, we consider 12 notes in an octave. In Indian music, the pitches of the notes are not fixed. Once tonic note (Sa) is fixed, the pitches of other notes gets fixed according to the scale used by the singer. There are various scales used by the singers [4]. Thus, the tonic note refers to the base pitch chosen by the singer on which raga or all melodies will be based [1], [2], [3], [4].

The tonic note extraction is an important problem in Indian music. It is important in many applications like Raga identification, Music Information Retrieval problems and more [6], [7], [8], [9], [10].

In this work, we consider tonic extraction using monopitch algorithms. There are many works based on monopitch algorithms. These algorithms typically, use autocorrelation function [1], Average Magnitude Difference Function (AMDF)[11], etc. The AMDF based pitch detection algorithm, also called as YIN algorithm, is discussed in [1], [11], [12]. Praat is also used in some of the works, the method is described in [1], [13].

The accuracy of the tonic extraction algorithms discussed above, is not sufficiently high [11]. In this work, Harmonic Product Spectrum (HPS) and pole focussing based techniques are combined to develop tonic extraction algorithm. The proposed algorithm gives improved results as compared to other algorithms based on monopitch analysis.

The paper is organised as follows. The proposed algorithm is discussed in detail in Section 2. In Section 3, results of various algorithms are presented and discussed and Section 4 concludes the paper.

II. Proposed Method based on HPS and Pole focussing Technique

In this section, pole focussing technique and HPS method of pitch extraction are explained and in subsequent subsection, proposed method is developed that is based on these two techniques.

A. Pole Focussing Technique

In human voice production, glottis opens and closes in a periodic fashion, indicating a periodic variations. The pitch frequency can be estimated as a reciprocal of measured time interval between sequential openings of glottis. In music signal, we observe fundamental frequency along with series of harmonic frequencies. The Z transform of signal $x[n]$ is defined as

$$X(z) = \sum_{n=-\infty}^{+\infty} x[n]z^{-n} \tag{1}$$

If we evaluate Z transform on unit circle, we get Discrete-Time Fourier Transform (DTFT), that can be written as

$$X(\omega) = \sum_{n=-\infty}^{+\infty} x[n]re^{-j(\omega n)} \tag{2}$$

where $r = 1$.

Since the vocal tract can be considered as a stable filter system, in which the poles of the vocal tract system must all lie

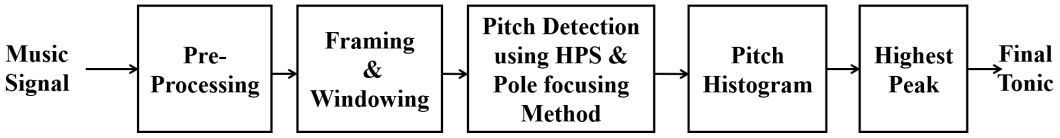

Fig. 1. Proposed method for tonic extraction

within the unit circle. Therefore, in pole focussing technique, z-plane search path is chosen such that $|z| < 1$, i.e. $z = re^{j\omega}$, $0 < r < 1$ [14]. This technique has been used in linear prediction (LP) based spectrogram and LP based vocoder [15] [16]. This enhances the effect of vocal tract poles on the vocal tract frequency response. The off-axis spectral peaks helps in capturing vocal tract pole locations.

B. Harmonic Product Spectrum (HPS)

This method of pitch detection is based on product of original spectrum of signal and series of down sampled versions of the spectrum [17], [18]. This algorithm makes use of various harmonics present in the signal. This estimates the highest match for harmonics using

$$N(\omega) = \prod_{l=1}^{L} |M(\omega l)|\,, \tag{3}$$

where L is the number of harmonics to be considered [12], [17]. The estimated pitch is given by

$$\widehat{N} = \max_{\omega_k} \{N(\omega_k)\} \tag{4}$$

where frequency ω_k is range of fundamental frequencies considered in the computation [17], [18].

This method is fast and works very well in wide applications but octave errors are one of the issue in this method [17]. In this work, we are addressing some of the issues in HPS method by utilizing pole focussing technique.

C. Proposed Method

The block diagram of the proposed tonic extraction method is shown in Figure 1. The main sections include preprocessing, framing and windowing, pitch detection and pitch histogram. The final tonic selection is done via highest peak method. The main feature used in tonic note extraction is pitch frequency. The first three blocks serves the purpose of estimating pitch frequency of music excerpt.

The estimated pitch frequencies are used to plot the pitch histogram to study pitch distribution of notes in the music excerpt. The pitch frequencies in the range of 30 -800 Hz are considered in plotting histograms. The frequency resolution used is 1 Hz.

In this work, the range of tonic note frequencies considered is 100-250 Hz so histograms in this range are plotted and studied. In music literature, it is well known that the tonic note occurs the most number of times in the histogram. This has been used in many works [1].

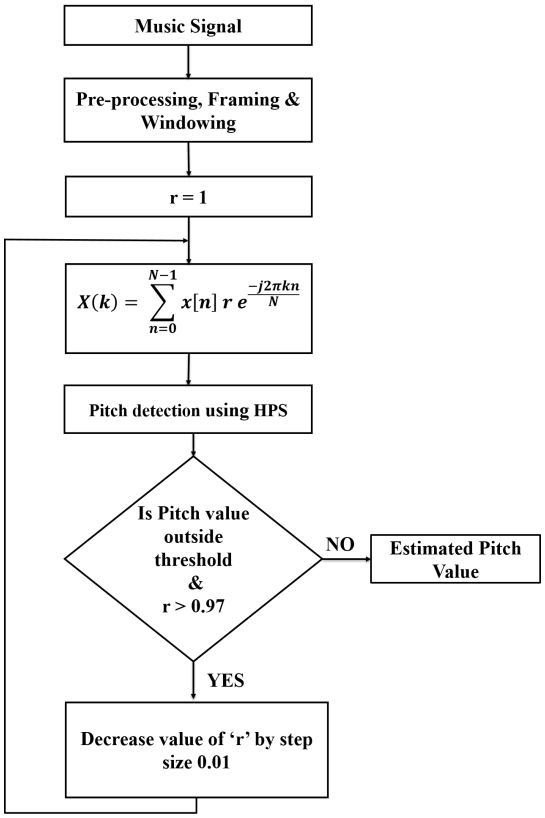

Fig. 2. Flow diagram of proposed method

The proposed method makes use of pole focussing technique and HPS in pitch estimation of music signal. Thus, the tonal features are obtained using combination of pole focussing and HPS. These pitch values are used to plot the pitch histogram. The peak in pitch histogram gives the pitch value of the frame. As shown in Figure 2, if estimated tonic pitch value is outside the threshold, value of r is decreased by 0.01, till the tonic note is in specified range. Usually, if estimated pitch value is outside the threshold for $r = 1$, it is found to be within threshold for $r = 0 \cdot 99$ or $r = 0 \cdot 98$. If it is not within the threshold for these values, the tonic detected for $r = 1$ is retained.

For switching to different value of r, the range is selected empirically. For male singers, switching to different r is done if tonic note is found to be outside the range of 100 Hz to 160 Hz and for female singers, this switching is done if tonic note is found outside the range of 160 Hz to 250 Hz.

The CompMusic music collection, a part of the CompMusic project dataset, CM2 is used [1]. The CM2 dataset contains

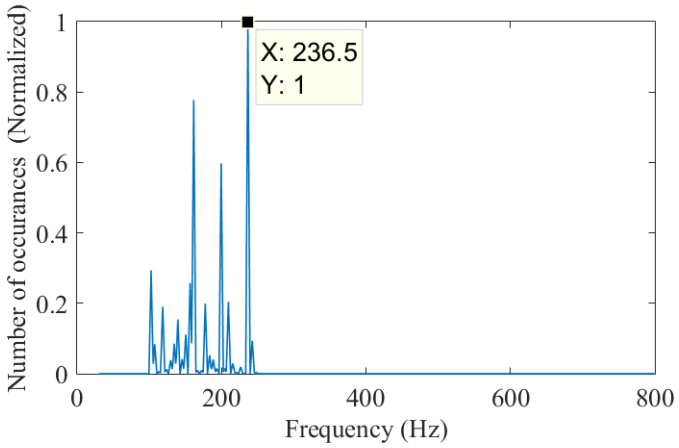

Fig. 3. Pitch Histogram of Male Singer 1 with r=1

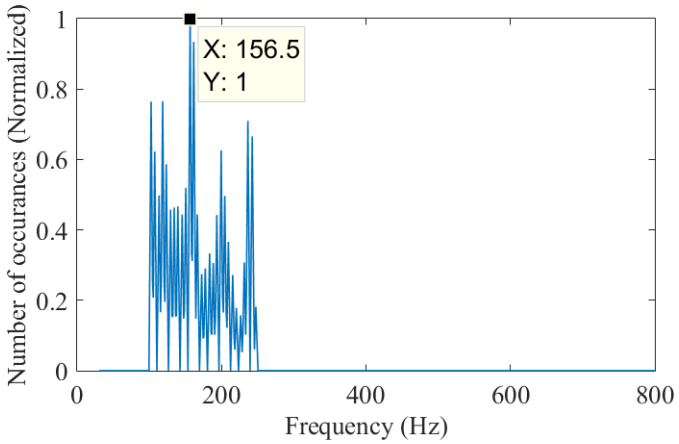

Fig. 4. Pitch Histogram of Male Singer 1 with r=0.99

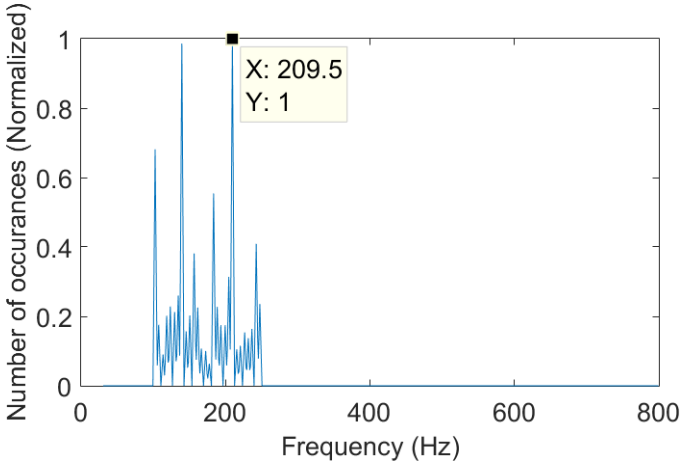

Fig. 5. Pitch Histogram of Male Singer 2 with r=1

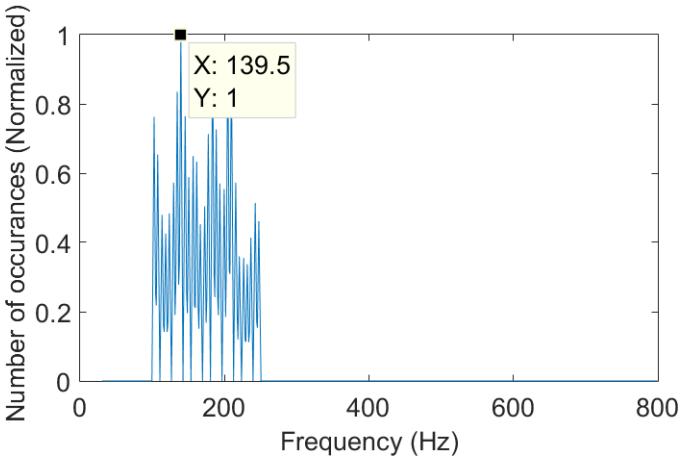

Fig. 6. Pitch Histogram of Male Singer 2 with r=0.99

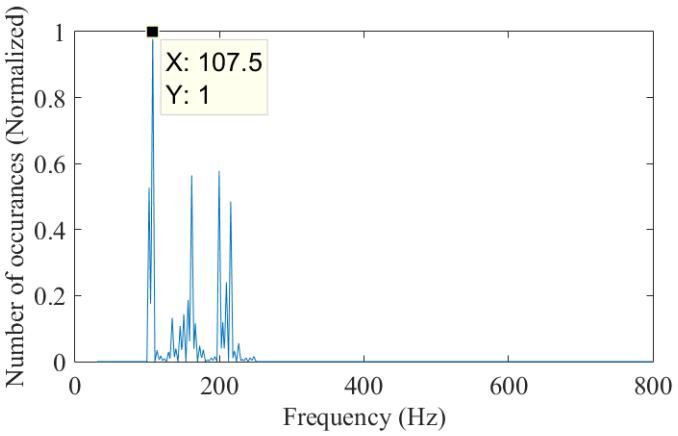

Fig. 7. Pitch Histogram of Female Singer 1 with r=1

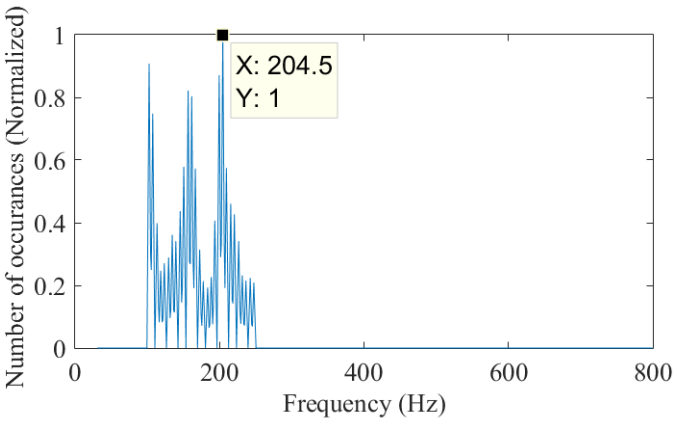

Fig. 8. Pitch Histogram of Female Singer 1 with r=0.99

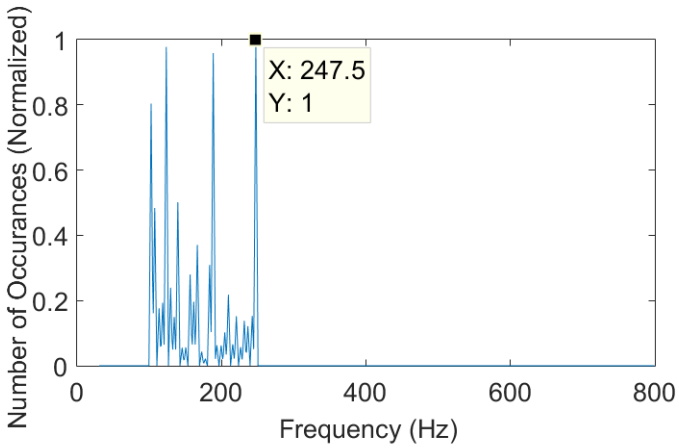

Fig. 9. Pitch Histogram of Female Singer 2 with r=1

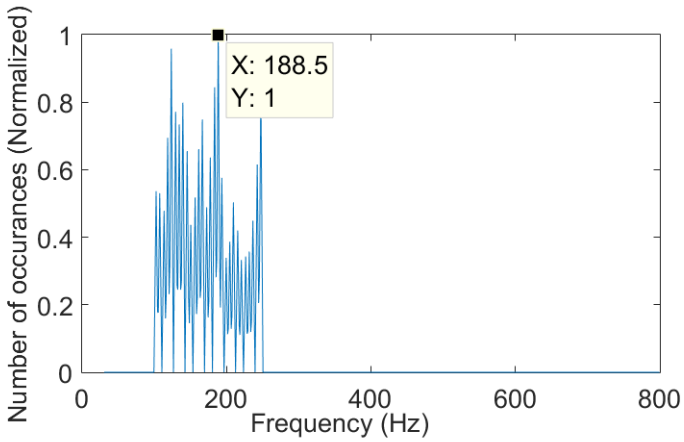

Fig. 10. Pitch Histogram of Female Singer 2 with r=0.99

935 excerpts of Indian music. This dataset comprises of both Carnatic music excerpts and Hindustani music excerpts. In this, nearly 55% samples are Carnatic and remaining samples are Hindustani. The duration of each sample is 3 minute long and is extracted from full performance. Overall 81 artists were involved in the experiment. Among this, male singers were 68% and remaining were female singers.

Histograms of male and female singers with various values of r are analyzed below. In Figure 3, male singer 1 histogram is shown with value of r to be 1. In this histogram, the tonic is incorrectly detected as 236.5 Hz. Actual tonic is 156 Hz. In Figure 4, with r value to be 0.99, it is correctly detected as 156.50 Hz. Similarly, Figure 5 and 6 shows correct and incorrect tonic note detected with r to be 1 and 0.99 respectively for male singer 2. Histograms of female singers are shown in Figures 7–10. Figures 7 and 9 shows incorrect detection with r=1 and corresponding correct detections are shown in Figures 8 and 10.

Table I compares performance of HPS with and without pole focussing technique and Table II compares performance of HPS and pole focussing technique with the various methods

TABLE I
RESULTS: PERFORMANCE OF HPS WITH AND WITHOUT POLE FOCUSSING TECHNIQUE

Method	Accuracy (male)	Accuracy (female)
HPS without Pole Focussing Technique	71.69%	88.62 %
HPS with Pole Focussing Technique	83.80%	90.96%
Overall Improvement	12.11%	2.34%

TABLE II
RESULTS: COMPARATIVE ACCURACY OF VARIOUS METHODS

Method	Overall Accuracy
HPS with Template Matching	77.11 %
Method based on YIN algorithm [1]	74.50 %
Method based on PRAAT [1]	69.60%
ACF with Template Matching	65.02%
ACF with Highest Peak	64.38 %
HPS with Pole Focussing Technique	87.38%

available in the literature. Tonic note extraction performance of autocorrelation based approach is around 65%. The Praat and YIN based approach improves these results to 69.60% and 74.50% respectively. Proposed HPS and pole focussing based approach further improves these results by removing the errors that occur in HPS with $r = 1$.

As shown in Table II, the accuracy of the proposed method is 87.38% for full dataset. Here, we assume that the tonic extraction is correct if error is less than 2 Hz. It has been observed that the accuracy of tonic extraction depends on frame size. According to our previous studies, optimal frame size is 80 ms and this value has been used in this work.

III. CONCLUSION

The tonic note extraction is important in many applications. In this work, a method based on HPS and pole focussing technique is proposed. The proposed method helps in improving tonic extraction using pole focussing technique. Proposed technique solves some of the problems in HPS. The results of proposed method are compared with methods available in literature. The comparison indicates the proposed method performs better than the methods available in literature.

ACKNOWLEDGMENT

The authors would like to thank CompMusic project, particularly Sankalp Gulati for sharing dataset with us.

REFERENCES

[1] Sankalp Gulati, et al. “Automatic tonic identification in Indian art music: approaches and evaluation.” Journal of New Music Research 43.1 (2014): 53-71.
[2] B.C. Deva, *Indian Music*, Indian council for cultural Realations, New Delhi, 1974.
[3] R.R. Menon, *The Penguin Dictionary of Indian Classical Music*, Penguin Books, India, 2008.
[4] C.J.Gaikwad and P. Sircar. On Determination of Tonic Note for Singers of Indian Music, IETE Technical Review, 2018, DOI: 10.1080/02564602.2018.1528897
[5] M. Ramesh. A gentle introduction to Carnatic music, Kindle edition, 2018.

[6] R. B. Kodag, M. D. Patil, and C. J. Gaikwad, "Harmonic product spectrum based approach for tonic extraction in Indian music," in Proceedings of International Conference on Computing, Analytics and Security Trends (CAST), Pune, 2016, pp. 622–626.
[7] Ranjani, H. G., S. Arthi, and T. V. Sreenivas. "Carnatic music analysis: Shadja, swara identification and rAga verification in AlApana using stochastic models." Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on. IEEE, 2011.
[8] Pradeep, R., et al. "Raga identification based on Normalized Note Histogram features." Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on. IEEE, 2015.
[9] Salamon, Justin, Sankalp Gulati, and Xavier Serra. "A multipitch approach to tonic identification in indian classical music." Gouyon F, Herrera P, Martins LG, Mller M. ISMIR 2012: Proceedings of the 13th International Society for Music Information Retrieval Conference; 2012 Oct 8-12; Porto, Portugal. Porto: FEUP Edioes; 2012.. International Society for Music Information Retrieval (ISMIR), 2012.
[10] Krishnaswamy, Arvindh. "Application of pitch tracking to south indian classical music." Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on. Vol. 5. IEEE, 2003.
[11] De Cheveign, Alain, and Hideki Kawahara. "YIN, a fundamental frequency estimator for speech and music." The Journal of the Acoustical Society of America 111.4 (2002): 1917-1930.
[12] A. Bellur and Hema A. Murthy, "A novel application of group delay function for identifying tonic in Carnatic music." Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European. IEEE, 2013.
[13] Boersma, Paul, and Vincent van Heuven. "Speak and unSpeak with PRAAT." Glot International 5.9-10 (2001): 341-347.
[14] G. Duncan and MA Jack,"Formant estimation algorithm based on pole focusing offering improved noise tolerance and feature resolution." IEE Proceedings, Vol. 135, Pt. F, No. I, 1988.
[15] Abraham, Anu, P. Vijayalakshmi, and T. Nagarajan. "Pole-focused linear prediction-based spectrogram for coarticulation analysis." 2010 IEEE Students Technology Symposium (TechSym). IEEE, 2010.
[16] Saranya A., Sripriya N. (2011) LPC VOCODER Using Instants of Significant Excitation and Pole Focusing. In: Nagamalai D., Renault E., Dhanuskodi M. (eds) Advances in Parallel Distributed Computing. PDCTA 2011. Communications in Computer and Information Science, vol 203. Springer, Berlin, Heidelberg
[17] De La Cuadra, Patricio, Aaron Master, and Craig Sapp. "Efficient pitch detection techniques for interactive music." Proceedings of the 2001 international computer music conference. (2001): 403-406.
[18] Noll, A. Michael. "Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate." Proceedings of the symposium on computer processing communications. Vol. 779. 1969.
[19] Klapuri, Anssi. "Multiple Fundamental Frequency Estimation by Summing Harmonic Amplitudes." ISMIR. (2006): 216-221.