

Article

Music Melodic Pattern Detection with Pitch Estimation Algorithms

Makarand Velankar^{1,*}, Amod Deshpande² and Dr. Parag Kulkarni³

¹ Faculty Cummins College of Engineering and Research Scholar PICT, Pune, Maharashtra, India

² Founder Consonance Acoustics, Aurangabad, Maharashtra, India

³ CEO, Iknowlation Research labs Pvt. Ltd. Pune, Maharashtra, India

* Correspondence: makarand.velankar@cumminscollege.in

Abstract: Music acoustics is an interdisciplinary field and mathematics is the basis in the music art form. Music and mathematics correlation exist since the inception of music. Various philosophers, scientists, mathematicians and musicians have expressed their views about this relationship. This paper attempts to explore this association with focus on melodic pattern identification. Mathematics in Indian Classical music with raga as the basis and just intonation tuning system is discussed. Indian vocal music clips are used for different pitch estimation algorithms in the experimentation. Harmonic product spectrum and autocorrelation algorithms are tested for accurate pitch estimation. Enhanced autocorrelation function using audio segmentation is compared with other approaches for effective pitch extraction. Results indicate pitch extraction with enhanced autocorrelation function provides accurate results as compared with other approaches tested.

Keywords: music acoustics, melody discovery, pitch tracking, Indian classical music

1. Introduction

Mathematics is traditionally being referred as mother of all engineering and science courses. Music however is considered as one of the art form. Generally art forms such as music, drawing, dancing etc. are not considered to be correlated with mathematics. It is being said that mathematics is present everywhere [1]. "Mathematics is not about numbers, equations, computations or algorithms; it is about understanding" as referred by Willam Paul Thurstan. Mathematics is not just number manipulation but extracting relationships and patterns among different phenomenon around us. "Mathematics is ultimately a study of patterns" as quoted by Hannah Fry. Patterns are everywhere and our life is full of patterns such as our schedules, predictions, planning etc. One can observe and identify patterns in different art forms such as paintings, poems and music.

Russel compared mathematician with musician as "The pure mathematician, like the musician, is a free creator of his world of ordered beauty". Musicians compose the music with specific objectives using some specific melodic and rhythmic patterns. The correlation of music and mathematics is explained by Pythagoras as "There is geometry in the humming of strings, there is music in the spacing of spheres". Considering the quotes by many philosophers, mathematicians, scientists and great music composers it can be revealed that there is a strong relationship of music and mathematics. Mathematics is being used for reasoning various phenomenon to build the theoretical mathematical model. Mathematical models are developed with specific objectives either to understand something or to solve specific problems in real world. Mathematical model usually describes the system in the form of variables and equations. It uses scientific approach to understand dynamics and predict behavior of the system.

Mathematical model for parallel computation was proposed by Karp et.al. [2] with the study of different programming languages. Nelson [3] discussed about the mathematical modeling for computer performance measure using probability, stochastic processes and queuing theory. Similarly mathematical models are being developed for various systems as a strong foundation.

Mathematical model using K Nearest neighborhood and Gaussian mixture was proposed by Li et. al. [4] intended for music genre classification. Bamberger [5] explored use of music in learning mathematical concepts such as ratio, proportion, fractions, and common multiples. Patterns in the compositions developed by famous musician Bach were examined for notes, scales, chord progression by Siddharthan [6]. Study of few compositions by Bach and others provided strong possible support for numerical analysis of musical composition using fractal geometry of music as observed by Hsu et.al. [7]. Study of intervals, pitch relationships, scales and tuning system for different musical cultures was carried out by Burns [8]. Sturm [9] explored a metaphor created that links together sound configuration and classical mechanics, with the quantum mechanical notion of particles acting as waves which blurs the distinction between science and art. Benson [10] covered use of mathematics in different aspects of music in detail such as sound production and Fourier theory, orchestration and use of wave equations for sound generation in different instrument along with acoustics, association between consonance and dissonance and simple integer proportions of frequencies, different scales and harmony, digital storage and compression techniques, synthesis and modulations, symmetries etc.

Song An et.al. [11] examined the effect of classroom activities with integration of music and mathematics and observed improved performance in learning mathematical concepts. Study of Indian songs and harmonic structure for about 10 years was carried out by Fillmore [12] and it was observed that Indian songs use mostly same western intervals. Instrumental and vocal music has subtle difference with former type i.e. vocal music precedes all instrumental music by an immeasurable interval as per Fillmore. Farrell [13] explored different elements of Indian music such as sounds or ideas of patterns as a reflection in jazz and pop music. In an article on science and music, Trainor [14] focused on our basic encoding system to listen and understand the music. Our interpretation of the musical rhythm leads to tap or dance. Pitch perception is another critical factor and listeners can perceive the pitch information differently. Indian classical music is based on raga patterns and the seasoned listeners can identify the raga by listening to the melodic patterns quickly.

This paper explores the correlation of music and mathematics related to melodic patterns. The paper is organized in following manner. Section 2 provides overview of Indian classical music raga form and section 3 covers tuning system used in music to fix notes and frequencies using just intonation scale. Pitch extraction using different approaches is covered in section 4. It covers mathematical foundation used in melodic pattern recognition for music from the computational perspective. Different pitch extraction algorithms are experimented for accurate pitch estimation of vocal rendition in Indian Classical Music. Results of the algorithms experimented are compared in section 5 along with the conclusions.

2. Overview of Indian Classical Music

The fundamental theory of Indian Classical music (both Carnatic and Hindustani) revolves around the concept of Raga. A Raga is a melodic composition of specific notes allowed in the composition of particular raga. Tonic or fundamental note which is referred as Sa or shadaj is always present in any raga. It also includes two other notes called Vadi Swar and Samvadi Swar. The performer stresses them the most during a presentation of the Raga. Another aspect is the ascending scale and the descending scale of the Raga called Aarohan and Avarohan respectively. It is important to note that the ascending scale and the descending scale of a Raga can have different musical notes. For example, Raga Des ascending and descending scale is as shown in figure 1. The notes in ascending and descending scale are different as noticed here. Thus raga is composition of allowed notes in ascending and descending melodic patterns. Motifs or catch phrases are another important aspect in the identity of a Raga also termed as 'Pakad' of raga. It is sequence of notes played in specific manner. The trained musicians or seasoned listeners can identify the Raga by listening to these catch phrases. Raga improvisation by the performer is another unique aspect in Indian Classical music performance.

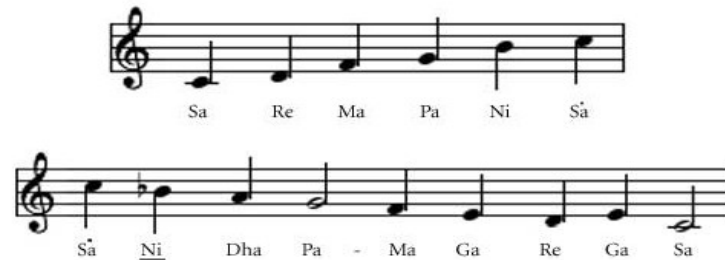


Figure 1. Raga Des Ascending and Descending note pattern.

Indian Classical has been a presentation art. Specific prominence is given to live performance and creativeness. The performer is expected to be confined by the rules of the Raga and extemporize within the limitations. Every performance has a 'drone' constantly accompanying the performance. The notes of the drone are defined for every Raga. A 'Bandish' is the next part of the performance wherein the rhythmic instrument (the Tabla being the most popularly used) gives structure to the performance after the mood is set by the Alap. Drone is constant in both the Alap and the Bandish. A peculiar thing to note about the Bandish is the changing tempo. The Bandish usually starts with a slow tempo ('vilambit lay') after the Alap. It then continues to medium tempo ('madhyam lay') and then fast tempo ('drut lay') towards the end. A Bandish can be compared to the flow in a canal where it is controlled but the twists and turns are the will of the performer.

Various approaches are proposed and attempted for automatic identification of raga. The first order Markov assumption states that the probability of a state depends on the probability of its immediate predecessor as experimented by Pandeet. al. [15]. This is the basis of the Hidden Markov Models (HMM) algorithm. Thus the probabilities of states are obtained and an identity is assigned to it. HMMs are relevant in case of identifying Raga as Raga can be assumed as sequential data and thus be modeled as Markov chains and can be modeled as patterns recognition problem as per Bishop [16]. Raga identification using repetitive notes pattern for Carnatic classical music was experimented by Sreenivas [17] using perspective notations. Phrase-based raga recognition using vector space modeling was another approach used by Gulati [18]. Swara histogram based pattern analysis used by Pranay Dighe et.al. [19] for identification of raga. Music melodies based on the notation system developed by various musical traditions across the world. Each tradition has own notation representation with different tuning systems. Tuning system represents the notes and their positions or frequencies in the octave.

3. Tuning system for Music

Since ancient period, music learning had been accomplished with learn by ear approach with oral presentation. Later on the need of music documentation was felt by musicians for better learning, recall, performing etc. It guided the development of different music data representation systems suitable to their own traditional and folk music forms. Western music composers developed sheet music with staff notations. Sheet music is music data representation using musical symbols to represent notes, scales, rhythms, chords, lyrics etc. Sheet music guides performers to read musical notations while presenting the songs, however understanding sheet music requires music notation literacy. One can notice music sheets in stands near the performers during orchestral performances such as sonatas, choir etc. Western music is based on equal tempered scale and the subsequent notes are calculated using same. Frequencies used in western notations are fixed for each note.



Figure 2. Sheet music data representation

Sheet music representation for single line of song 'Jingle bell' is shown in the Figure 2. The melody is represented using staff notations with symbolic representations for scale and notes. Vertical lines or bars shown indicate time intervals. The notation sequence as 'E E E, E E E, E G C D' represents notes played in specific time intervals.

Many non-western musical traditions have developed their own musical notation systems such as swarlipi for Indian classical raga music, shakuhachi for Japanese music, Chinese musical notations etc. Different traditions use several musical scales. Scale is a combination of music notes derived from the fundamental frequency with different ratios or distances. Equal tempered scale in western music, just intonation scale in Indian classical music or Arabic scale with maqam is a representation of different tuning systems and scales used in music. Detail discussion on various notation systems and scales for music is beyond scope here. Discussion about Indian Classical music notation system is presented here with the use of mathematics in the tuning systems.

Indian Classical music has been mainly vocal centered; however different artists have made a significant contribution with the instrumental music as well. The Indian Classical music tuning system is closer to the just intonation tuning system as per study from Schmidt-Jones [20]. It is a system to calculate the other notes from the tonic using ratios. The ratios used in just intonation system are as shown below to calculate the frequencies of the subsequent notes in the octave.

Ratios in Just Intonation scale

[16/15, 9/8, 6/5, 5/4, 4/3, 45/32, 3/2, 8/5, 5/3, 16/9, 15/8, 2]

Thus octave is divided into 12 notes with the frequencies are adjusted using the just intonation scale. Another peculiar concept in Indian Classical Music is the concept of 'Shrutis'. The octave is divided between 22 such Shrutis which are intervals smaller than that of the semitones [20]. The details about shrutis and their ratios are available on Wikipedia [22] for interested readers. All the notes are respective to the tonic in Indian classical music. Tonic is not fixed and depending on singer or instrument tonic can be different and further notes are fixed according to the tonic. Identifying the note patterns is done using pitch extraction. Arvinth [23] explained applications of pitch tracking for South Indian Classical Music. Pitch is perceived frequency in Hertz. Accuracy of pitch extraction is extremely important for the proper identification of notes and melodic patterns in the music played.

4. Materials and Methods

Pitch is an auditory sensation represented by two characteristics namely the tone height and the Chroma. The tone height showcases the rise in pitch as the frequency rises and the Chroma showcases the perceptual similarity between two notes separated by an octave as observed by Gelfand [24]. While extracting pitch from an audio sample, both the dimensions should be considered. It should be kept in consideration that both the dimensions at a certain time can be obtained by extracting the fundamental frequency at that time. Due to this, the algorithms which primarily detect the fundamental frequency are often called pitch extraction algorithms.

Pitch extraction algorithms are often used in speech prediction and music information retrieval. In the context of this paper, the audio samples in consideration were all monophonic vocal audio samples. While extracting pitch from audio, care needs to be taken with regards to voiced and unvoiced sounds. The spectral analysis of vocal audio shows formants influencing the audio sample in that even though the fundamental frequency of two vowels is same, the harmonics might have different energies. This, along with several other reasons creates difficulties in the extraction of pitch. There exist some robust algorithms which are discussed in this section and their performances are also compared in different cases.

According to the context established earlier, the methods chosen for detecting the fundamental frequency of a segment of the audio did not need to be extremely complicated as the audio samples

are monophonic with good signal-to-noise ratio. Therefore, two methods have been discussed and implemented here, Harmonic Product Spectrum (HPS) and Autocorrelation.

4.1. Harmonic Product Spectrum

A pitch detection algorithm proposed by A. Micheal Noll [25] is designed for human speech. The principle behind this algorithm is that the product of the harmonics of a fundamental frequency of a frame would be maximum at the peaks in the spectrum. In other words, it is a measure of the coincidence of the harmonics [26].

$$Y(\omega) = \prod_{n=1}^K |X(\omega n)|, \quad [1]$$

$$\hat{Y} = \max_{\omega_i} \{Y(\omega_i)\}. \quad [2]$$

In Equation [1], ω is the frequency and K is the number of harmonics to be considered. This equation measures the coincidence of the harmonics of each frequency. The maximum coincidence is found out by using Equation [2]. Thus, segmenting the signal and applying Equation [2], it is possible to estimate the fundamental frequency of the segment. Figure. 3 explains the process intuitively. It can be seen that the product of the amplitudes of 183 Hz and its harmonics will be maximum.

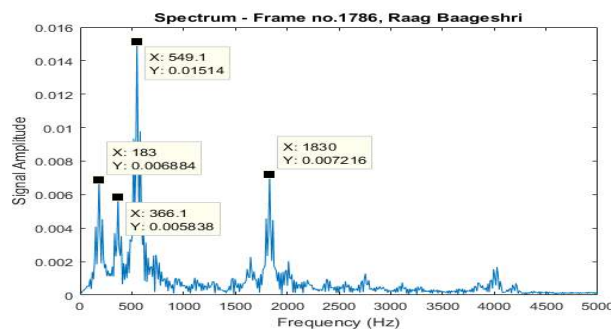


Figure 3. Spectrum of a frame which depicts the amplitudes of the fundamental frequency and its harmonics

The implementation consisted of firstly dividing the audio signal into 10 millisecond segments without overlap in the first case and 40 milliseconds with 75% overlap in the second case. The segments were windowed using Gaussian window function. These segments or frames further underwent Fast Fourier Transformation with the number of points of the FFT were set to 4096 to ensure high resolution. It was observed in this and other pitch extraction algorithms that the fundamental frequency was usually contained within 100 Hz to 500 Hz. Thus, all other frequencies were ignored in the further analysis. Subsequently, Equations [1] and [2] were implemented on all the remaining frequencies. The number of harmonics to be considered was taken to be 5. A comparison between the overlapped and non-overlapped frames is shown in Figure 4 as an example. In context of the HPS algorithm, it can be said that an overlap reduces bias towards a frequency. Pitch was being determined for every 10 seconds.

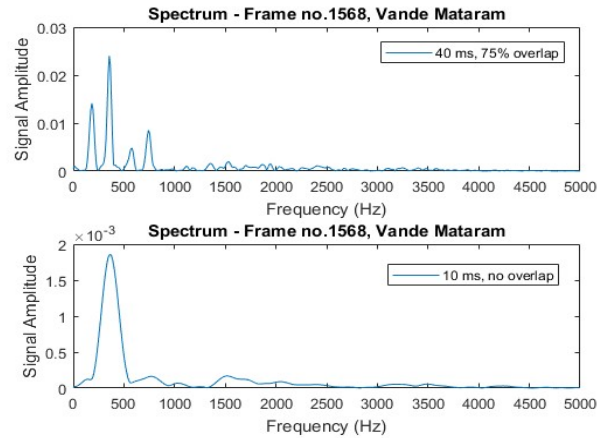


Figure 4. Comparison between overlapped and non-overlapped segments for pitch extraction

4.2. Autocorrelation

An autocorrelation function (ACF) is used to display structure in a waveform. Therefore, if there is periodicity in the autocorrelation function, it indicates periodicity in the signal [27]. An autocorrelation function of a signal $x(n)$ is given by Rabiner [27] as shown in equation 3.

$$\phi_x(m) = \lim_{K \rightarrow \infty} \frac{1}{2K+1} \sum_{n=-K}^K x(n)x(n+m), \quad [3]$$

K is the size of the frame being analyzed in equation 3. Segments of 40 milliseconds with 75% overlap were used for the analysis. Thus in a segment, the peaks in the autocorrelation function will indicate periodicity. The first peak will therefore indicate the fundamental frequency of the segment. Figure 5 showcases the ACF of an example frame in Raga Lalit. The first peak denotes that the signal correlates with itself after some lag. Therefore when the sampling frequency is divided by the lag, the fundamental frequency is obtained as the output.

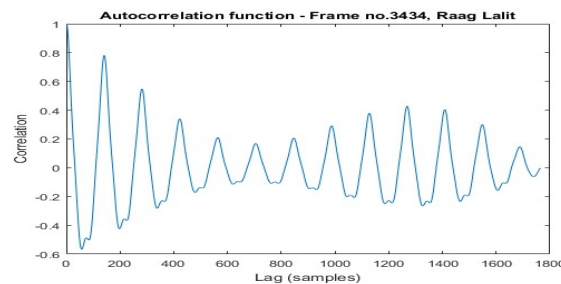


Figure 5. Autocorrelation function of a frame in Raga Lalit

The implementation of the ACF for detecting pitch consisted of firstly filtering the signal using a low pass filter with the pass band frequency at 900 Hz and stop band frequency at 1 kHz with 65 dB as the stop band attenuation level. This was done “to partially eliminate the effects of higher formant structure on the autocorrelation function” as per Rabiner [27]. Subsequently, the signal was segmented in 40 millisecond frames with 75% overlap and the segments were windowed using a Hamming window. The ACF of each frame was then determined and the first peak located to determine the fundamental frequency of that frame.

4.3. Autocorrelation used in tool 'Praat'

Paul Boersma [28] proposed that accuracy in pitch detection can be increased by having a small modification in the traditional pitch detection by autocorrelation. The traditional method involving the short term autocorrelation function (ACF) of signal would not account for the autocorrelation function of the window used while segmentation. Thus, accuracy was increased by dividing the ACF of the windowed signal by the ACF of the window [28],

$$\phi_x = \frac{\phi_{ws}}{\phi_w}, \quad [4]$$

Where ϕ_{ws} is the ACF of the windowed signal, ϕ_w is the ACF of the window and ϕ_x is the ACF of the original signal.

This algorithm is effectively implemented in the software tool called 'Praat'. Praat is an open source tool developed mainly for speech processing. Praat was used for extracting pitch of voiced melody in the experimentation done here. Figure 6 depicts the pitch extraction implemented in Praat. Praat was successfully used by Ramesh for effective music data analysis [29].

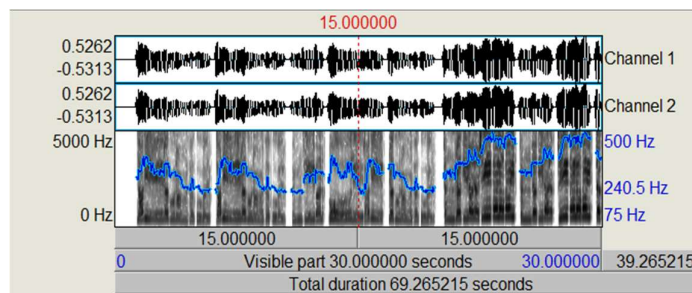


Figure 6. A 30 second excerpt of Raga Lalit in Praat

5. Results and Conclusion

The methods used for autocorrelation are compared to find the best method for the raga audio samples used in the experimentation. All the methods were experienced to be almost equivalent when compared to the computational time required for each. The noise robustness was not tested for the implementations of the Harmonic Product Spectrum and Autocorrelation as the audio samples for analysis were recorded in a noise-free environment. A big concern was the accuracy of the algorithms for pitch extraction of voiced samples.

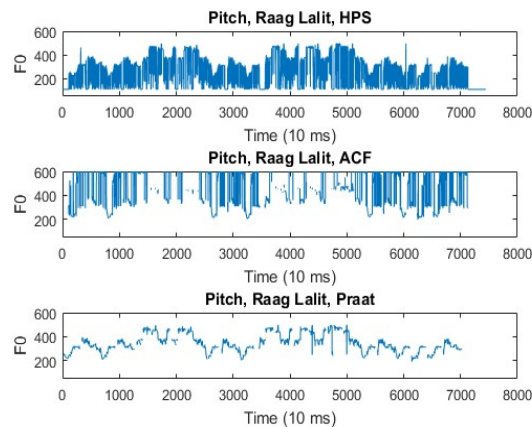


Figure 7. Comparison of algorithms for pitch extraction

It can be observed from Figure 7, both the Harmonic Product Spectrum (HPS) and the Autocorrelation (ACF) algorithm exhibit inaccurate responses. Although same trend of the fundamental frequency is followed by all three algorithms, the pitch extracted with the help of Praat is the most accurate extraction. Thus for voiced audio samples, praat was found out to be the most suitable tool for pitch extraction. Pitch information extracted can be further used for various applications. Melodic pattern identification using different algorithms such as pitch class distribution (PCD), Hidden Markov Model (HMM) etc as explained by Makarand et.al. [30] during the application of query by humming.

Music melodic patterns are identified from the pitch information generated from pitch extraction algorithm used. Real challenge to achieve computational intelligence in musicology is music mathematical modeling. We humans ourselves have not able to judge the entire musical aspects in true sense to model it for further processing. Modeling human perception and fast feature learning scalable parallel algorithms will lead further progress in the domain. Bridging semantic gap between user preference and music audio signal will be crucial for the successful systems. Advancements in the musicology domain and machine learning with artificial intelligence agents will lead to better understanding of music and associative musical pattern matching methodologies. Interesting end to end applications for various tasks using more advanced music recognition systems are likely to dominate in coming years.

Funding: This research received no external funding

Acknowledgments: This research was made possible by the efforts of vocalists Payal Sarkate, Rohit Dasharathi, Asawari and Atul Khandekar. They helped us record the Raag samples and songs required for analysis. We extend our acknowledgements to Prof. Hari Sahasrabuddhe, his knowledge and support was extremely helpful while tackling a critical issue in the research.

Conflicts of Interest: The authors declare no conflict of interest

References

1. <https://www.youtube.com/watch?v=V9O4d0e7qdo> accessed 25 Oct 2018.
2. Karp, R. M., & Miller, R. E. (1967, October). Parallel program schemata: A mathematical model for parallel computation. In *Switching and Automata Theory, IEEE Conference Record of the Eighth Annual Symposium on* (pp. 55-61).
3. Nelson, R. (2013). *Probability, stochastic processes, and queuing theory: the mathematics of computer performance modeling*. Springer Science & Business Media.
4. Li, T., & Ogihara, M. (2004, May). Content-based music similarity search and emotion detection. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. V-705).
5. Bamberger, J. (2003). Music as embodied mathematics: A study of a mutually informing affinity, *International Journal of Computers for Mathematical Learning*, 8(2), 123-160.
6. Siddharthan, R. (1999). Music, mathematics and Bach. *Resonance*, 4(5), 61-70.
7. Hsü, K. J., & Hsü, A. J. (1990). Fractal geometry of music. *Proceedings of the National Academy of Sciences*, 87(3), 938-941.
8. Burns, E. M. (1999). Intervals, scales, and tuning. In *The Psychology of Music (Second Edition)*, pp. 215-264.
9. Sturm, B. L. (2001). Composing for an ensemble of atoms: the metamorphosis of scientific experiment into music. *Organized Sound*, 6(2), 131-145.
10. Benson, D. J. (2008). Music: a mathematical offering. *The Mathematical Intelligencer*, 30(1), 76-77.
11. An, S. A., & Tillman, D. A. (2015). Music Activities as a Meaningful Context for Teaching Elementary Students Mathematics: A Quasi-Experiment Time Series Design with Random Assigned Control Group. *European Journal of Science and Mathematics Education*, 3(1), 45-60.
12. Fillmore, J. C. (1899). The harmonic structure of Indian music. *American Anthropologist*, 1(2), 297-318.
13. Farrell, G. (1988). Reflecting surfaces: The use of elements from Indian music in popular music and jazz. *Popular Music*, 7(2), 189-205.
14. Trainor, L. (2008). Science & music: the neural roots of music. *Nature*, 453(7195), 598.

15. Pandey, G., Mishra, C., & Ipe, P. (2003). TANSEN: A System for Automatic Raga Identification. In IICAI, pp. 1350-1363.
16. C. M. Bishop. (2006) Pattern Recognition and Machine Learning, Springer Science Business Media LLC.
17. Sreenivas, T. V. (2017). Raga Identification using Repetitive Note Patterns from prescriptive notations of Carnatic Music. arXiv preprint arXiv:1711.11357.
18. Gulati, S., Serra, J., Ishwar, V., Sentürk, S., & Serra, X. (2016, March). Phrase-based rāga recognition using vector space modeling. In Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on (pp. 66-70).
19. Pranay Dighe, Harish Karnick, Bhiksha Raj (2013) Swara histogram based structural analysis and identification of Indian Classical ragas, International Society for Music Information Retrieval Conference
20. Schmidt-Jones, C. (2011). Indian classical music: Tuning and ragas. Connexions.
21. Serra, J., Koduri, G. K., Miron, M., & Serra, X. (2011, October). Assessing the Tuning of Sung Indian Classical Music. In ISMIR (pp. 157-162).
22. [https://en.wikipedia.org/wiki/Shruti_\(music\)](https://en.wikipedia.org/wiki/Shruti_(music)) accessed 15 June 2018.
23. Arvinth Krishnaswamy (2003) Application of Pitch tracking to south Indian Classical Music, ICASSP.
24. S. A. Gelfand (2010) Hearing: An Introduction to Psychological and Physiological Acoustics, Informa Healthcare.
25. A. M. Noll (1969) Pitch Determination of Human Speech by the Harmonic Product Spectrum, the Harmonic Sum Spectrum, and a Maximum Likelihood Estimate, in Symposium on Computer Processing in Communications Polytechnic Institute of Brooklyn.
26. P. d. l. Cuadra, A. Master and C. Sapp, Efficient Pitch Detection Techniques for Interactive Music, Centre for Computer Research in Music and Acoustics, Stanford University.
27. L. R. Rabiner (1977) On the Use of Autocorrelation Analysis for Pitch Detection, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vols. ASSP-25, pp. 24-33.
28. P. Boersma (1993) Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound, in IFA Proceedings 17.
29. Ramesh, V. M. (2008, July). Exploring Data Analysis in Music using tool praat. In Emerging Trends in Engineering and Technology, ICETET'08. IEEE International Conference, pp. 508-509.
30. Makarand, V., & Parag, K. (2018). Unified Algorithm for Melodic Music Similarity and Retrieval in Query by Humming. In Intelligent Computing and Information and Communication Springer, Singapore, pp. 373-381.