Accurate Monophonic Pitch Tracking Algorithm for QBH and Microtone Research

K. A. Akant, M.Tech.[1]; Rajesh Pande, Ph.D. [2]; and S. S. Limaye, Ph.D.[3]

[1]Manoharbhai Patel Institute of Engineering & Technology, Gondia, India.
[2]Shri Ramdeobaba Kamla Nehru Engineering College, Nagpur, India.
[3]Jhulelal Institute of Technology, Nagpur, India.

E-mail: kalyaniakant@gmail.com
panderaj@yahoo.com
shyam_limaye@hotmail.com

ABSTRACT

An algorithm is proposed to detect the pitch of a singing voice accurately for Query By Humming (QBH) and microtone research applications. Computational efficiency and accuracy are the important features of this algorithm, which are principle requirements for these applications. The traditional method for finding the fundamental frequency is to plot the Fourier Transform of the sound segment and pick up the highest peak. The difficulty with this method is that the first peak (i.e., the peak corresponding to fundamental frequency) is often suppressed and it may not be dominant. So the detected frequency is often multiple of the actual.

Another popular technique for pitch determination is autocorrelation and it suffers from the same drawback. In our proposed method, we detect the periodicity of the Fourier transform by again taking its Fourier transform to obtain the “Fourier of Fourier transform”. Then the resolution of frequency measurement is improved much beyond the FFT bin size by employing Parabolic Interpolation.

To track the pitch over time, the music sample is divided into overlapping windows of 46.4 ms duration. This algorithm is efficient for harmonic sounds and was tested on natural as well as synthetic signals, including male vocal, female vocal samples and mathematically generated harmonic signals. Our method outperforms the existing methods significantly by eliminating possibility of octave error and improving performance by accurate estimation of pitch. Accuracy in frequency estimation was compared with other methods; FT[1], Autocorrelation, and Fourier of Fourier.

(Keywords: QBH-Query by humming, Microtone Research, Fourier of Fourier Transform, Parabolic Interpolation, Pitch, FT[1], Autocorrelation)

INTRODUCTION

Pitch estimation is the problem of determining the fundamental frequency present in the signal. It is inherently related to the detection and estimation of sinusoids. The estimation and tracking of single or multiple sinusoids is a fundamental problem in many branches of applied sciences, so it is less surprising that the topic has also been deeply investigated in statistics. However, ideas from statistics seem to be not widely applied in the context of musical sound analysis, with only a few exceptions [20], [21] who present techniques for very detailed analysis of musical sounds with particular focus on decomposition of periodic and transient components [26], [2].

This paper presents a computationally efficient algorithm for monophonic pitch tracking to be applied for query by humming and microtone research. Varieties of methods are available for tracking the fundamental frequency of harmonic sound in the literature. Some primarily use time domain analysis (e.g., Time-Autocorrelation), some primarily use frequency-domain analysis (e.g., power spectral density), and others use a combination of both [29]. We have used Fourier of Fourier transform [24] for the estimation of pitch.

With the rapid development of multimedia applications, we are enjoying more entertainment activities in our life especially the music songs. In order to meet the needs, more and more music information retrieval (MIR) systems have been developed. Some of them were based on text retrieval models by entering music names, genre, artists etc., while the others were based on content-based retrieval models by humming the melody or singing the lyrics [13]. Recently, some studies have placed emphasis on statistical analysis in Music Information Retrieval (MIR).

People usually recall a song by its melody rather than its title or composer. Naturally, retrieval based on humming a piece of the song is much more convenient for music retrieval [4], [7]. Systems able to find a song based on a sung, hummed, or whistled melody are called query by humming, or QBH, even though humming is not always the input [27]. Monophonic pitch tracking is required for QBH applications. Several query-by-humming techniques have been proposed for content-based music retrieval. Reliable pitch extraction from humming is critical for such music retrieval systems to work well [28]. Important acoustic features of the query are those related to the tune or melodic pitch contour. The signal processing front-end of the QBH system transcribes the vocal query into a sequence of note pitches and durations (inter-onset intervals) to be matched against previously stored transcriptions of database melodies [19].

Another application of pitch tracking is automatic music transcription. Once pitch contour is obtained, note segmentation and note labeling is done. Note labeling deals with music content of audio signal. For each note event, an estimation of pitch measured in Hertz must be converted into music note labels; this process should be carried out with respect to a musical scale. This process is critical because intonation is never absolute and singers always introduce errors during a real performance [7].

Most research into music information retrieval thus far has only examined music from Western tradition using equi-tempered scale. In this scale, an octave is divided into 12 notes whose frequencies are in geometrical progression. In other words, these notes are logarithmically equidistant in frequency. This division has been accepted as a compromise for ease of manufacturing musical instruments. However, purists argue that for ensuring perfect harmony, the note frequencies should be related by simple ratios rather than be in a geometric progression. This gives rise to the "Just tuned" scale. The notes in this scale slightly deviate from the equi-tempered scale notes. These minute variations are called "microtones".

Very little work has taken place in the area of applying techniques from computational musicology and artificial intelligence to the realm of Indian Classical Music (ICM) [8]. Digital sound processing tools offer new possibilities to the analysis of musical structures, the modeling of the acoustic characteristics and the musical pattern comparison and recognition.

Raag is a melodic abstraction around which almost all Indian classical music (ICM) is organized. A raag is most easily explained as a collection of melodic gestures and a technique for developing them. The gestures are sequences of notes 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 [18], [5]. In other words, Raag is a typical musical pattern observed in ICM. Raag identification from the audio sample is nothing but musical pattern recognition, which would be another application of our method. [14] and [15] propose a scheme for the recognition of such predefined musical patterns in a monophonic environment in the context of South Indian Classical Music. Raag identification can also be used as a good basis for music information retrieval [22].

Music of other origins often conforms to different tuning systems. Therefore there are problems both in representing this music as well as finding matches to queries from these diverse tuning systems [10]. In the context to Indian Classical Music microtones are termed as *Shruti.* There are 22 *Shruti*s in an octave [9]. There are pitches inflexions used sometimes that are not mere ornamentations they are essential to the correct rendition of certain notes. Though these inflexions can be viewed as different versions of a particular note, they are certainly not equivalent to constant-pitch intervals like Just Intonation intervals, semitones or quartertones [3].

In order to find the positions of these microtones, we need to analyze the sound samples from the well known maestros. For a QBH application we should be able to estimate the pitch to the nearest semitone. The semitones are separated by a factor $2^{(1/12)}$ (i.e., roughly 6%). For microtone research, the expected accuracy is 1%. Performance of proposed method in terms of accuracy is very much beyond expectation. QBH is also real time and there is need for computationally efficient algorithm. In view of these applications accuracy and computational efficiency are the mainly required features, which are fulfilled by the proposed method.

FOURIER OF FOURIER TRANSFORM

In our analysis we have used two Fourier transforms in sequence referred as Fourier of Fourier Transform. This method works very well in the case harmonic sounds (i.e., sounds rich in harmonics). It is not suited for pure sinusoids. Fourier transform (first Fourier transform of the signal) of a typical musical sound has a series of peaks in its magnitude spectrum corresponding to the harmonics of the sound, at frequencies close to multiples of the fundamental frequency *F*. The peak showing fundamental frequency may not always be dominant. Hence single Fourier transform is inefficient to identify correct peak.

Fourier of Fourier Transform is of great interest in locating this peak, which helps to overcome the possibility of octave error. To find out Fourier of Fourier Transform, we compute magnitude spectrum of the Fourier transform of singing voice. Magnitude spectrum of the Fourier transform of the above magnitude spectrum is then computed. Figure 1 shows the first Fourier transform of male vocal singing C# of upper octave. This first Fourier transform has a series of uniformly spaced peaks as shown in Figure 1, corresponding to the harmonics of fundamental frequency. Fundamental frequency of the signal under consideration was also evaluated by autocorrelation method, Praat [17] as 271.3 Hz.

We can clearly see that, peak corresponding to fundamental frequency is not dominant. If fundamental frequency is F, the distance between two consecutive peaks corresponds to a period of Δ_1 bins where:

$$\Delta_1 = N_1 \frac{F}{Fs} \quad (1)$$

N1 : size of the first Fourier transform
Fs : sampling frequency

The first peak is at bin 0 and it corresponds to the DC level. The difference between second peak and the first peak is Δ_1 bins.

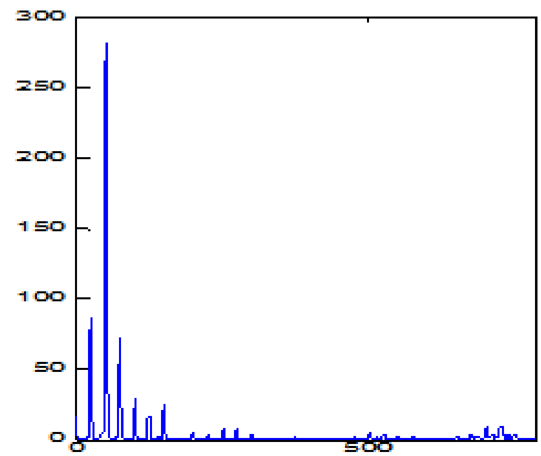

Figure 1: Fourier Transform of Male Vocal Singing C# of Upper Octave.

Figure 2 shows the spectrum of Fourier of Fourier Transform of male vocal singing C# of upper octave.

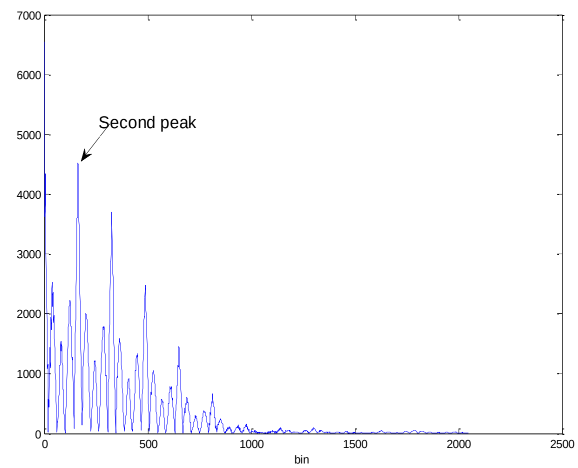

Figure 2: Fourier of Fourier Transform of Male Vocal Singing C# of Upper Octave.

In this spectrum of Fourier of Fourier Transform, there are series of peaks. Here also, the first peak is at bin 0 and it corresponds to the DC level. The distance between two consecutive peaks corresponds to a period of Δ_2 bins where:

$$\Delta_2 = \frac{N_2}{\Delta_1} \quad (2)$$

N_2: size of the second Fourier transform

From Equation (1) and Equation (2), we get:

$$\Delta_2 = \frac{N_2}{\left(\frac{(N_1)F}{Fs}\right)} \quad \textbf{(3)}$$

If size of first and second Fourier transform is same ($N_2 = N_1$), Fundamental frequency F is given by,

$$F = \frac{Fs}{\Delta_2} \quad \textbf{(4)}$$

PITCH TRACKING ALGORITHM

Using Fourier of Fourier Transform

To determine pitch at a certain time t, temporal frame centered at t is considered. The first step is to multiply temporal frame with Hann analysis window. For good harmonic resolution, at least four periods of the harmonic signal under the window are required. Magnitude spectrum of the Fourier transform of this temporal frame is evaluated (First Fourier transform). The Fourier transform size N_1 is normally chosen to be the first power of two that is at least twice the window Length M, with the difference N_1–M filled with zeros (zero padding).

The reason for increasing the Fourier transform size and filling in with zeros is that zero-padding in the time domain corresponds to interpolation in the frequency domain, and interpolating the spectrum is useful in various ways. First, the problem of finding spectral peaks which are not exact bin frequencies is made easier when the spectrum is more densely sampled. Second, plots of the magnitude of the more smoothly sampled spectrum are less likely to confuse the untrained eye [11].

This spectrum shows series of peaks at harmonic frequencies. After that, magnitude spectrum of Fourier transform of the magnitude spectrum of First Fourier transform is evaluated. The size of second Fourier transform N_2 was chosen same as N_1. This spectrum also shows series of peaks, where, first peak is at bin 0. In the spectrum of this second Fourier transform, difference between second peak and first peak in terms of bins (Δ_2) is found out.

With the knowledge of Δ_2 and N_2 we get Δ_1 from Equation (2). Δ_1 is the difference between the peaks in First Fourier transform. Here first peak is at bin 0, hence location of the second peak can be calculated. There may be error of few bins in this calculation. Hence the exact peak position is searched in the vicinity of evaluated peak location in the spectrum of first Fourier transform.

This procedure eliminates the possibility of selection of wrong peak and thus renders the results octave error free.

Exact Peak Detection

The maximum frequency resolution is limited to bin size. In our study, the sampling frequency is 44100 Hz, window size is 2048 and size of first as well as second Fourier transform is 4096. This yields a bin size of Fs/N, which is equal to 10.76 Hz. This resolution is inadequate for micro pitch estimation, which requires accuracy of at the most half a semitone or preferably less than that.

For QBH applications, accuracy of at least a semitone is required; hence resolution of 10.76 Hz. is not sufficient for QBH applications as well. For an octave starting at lower C (120 Hz), semitone corresponds to 5 Hz. Microtone research will require resolution of at the most 2.5 Hz. This shows that accurate estimation of pitch is principle requirement of QBH and microtone research applications.

Frequency resolution can be improved by increasing window size but we will lose time resolution and computational effort will also increase. In our method frequency resolution is reduced to 1 Hz without changing bin size, using parabolic interpolation strategy [11]. This ensures accurate estimation of pitch without increasing computational complexity.

In this technique, the Fourier spectrum lobe assumed to be a parabola as shown in Figure 3. Coordinate system for the parabolic interpolation is shown in Figure 4.

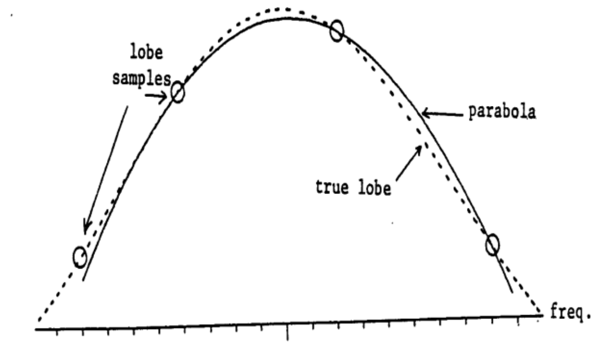

Figure 3: Parabolic Interpolation of the Highest Three Samples of a Peak.

β is the highest spectrum value located at bin k. If the signal frequency happens to be exactly located at this location, then the spectrum amplitude on either side of the bin k will be equal. However if it lies between the frequency samples, then the amplitudes will be unequal. Let α be the spectrum value at (k-1)th bin and γ be the spectrum value at (k+1)th bin.

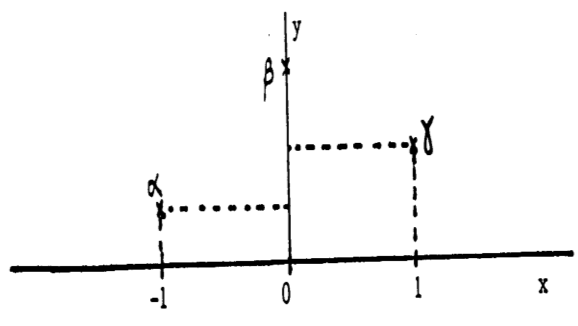

Figure 4: Coordinate System for the Parabolic Interpolation.

Solving for the parabola peak location *p*, we get,

$$p = \frac{1}{2}\frac{\alpha - \gamma}{\alpha - 2\beta + \gamma} \qquad \textbf{(5)}$$

Here, p will be a fraction ranging from -1 to +1. Estimate of the true peak location in bins will be,

$$k^* = k + p \qquad \textbf{(6)}$$

Estimation of Fundamental Frequency using Proposed Method

Once we find out the correct location of second peak in first Fourier transform, the difference Δ_1 between this second peak and first peak location (bin 0), in terms of bins is calculated. According to Equation (1), fundamental frequency F of the harmonic sound at time t will be,

$$F = \frac{\Delta_1 \mathrm{Fs}}{N_1} \qquad \textbf{(7)}$$

The temporal variations of spectral components are investigated by advancing a frame by one hop. Hop size is selected such that there is 75% overlap of the frames. Then the whole procedure is repeated to find out fundamental frequency for new frame. This is continued till end sample of the signal to get time vs. pitch graph. In case of fast pitch fluctuations, hop size should be reduced without changing frame size.

Estimation of Frequency using Fourier of Fourier Transform

Temporal frame selection, windowing, finding Fourier of Fourier transform is done as stated above. In the spectrum of Fourier of Fourier transform, first peak is at bin 0. Location of second peak in terms bins is found out. Here also parabolic interpolation strategy is applied to find out exact location of this second peak. Evaluating Δ_2, (difference between exact location of second peak and first peak) the fundamental frequency F can be calculated using Equation (4) as, $F=Fs/\Delta_2$

Estimation of Frequency using order 1 Fourier Transform (FT^1) [16], [25]

Fourier transform of the derivative of the signal is equal to angular frequency times Fourier Transform of the signal. Hence fundamental frequency F will be equal to (1/2π) times Fourier transform of the derivative of the signal divided by Fourier transform of the signal.

Following steps are followed to find out fundamental frequency.

1. Correct location of the second peak in the magnitude spectrum of Fourier

transform of the signal is found out as explained below.

2. Signal is differentiated.

3. Correct location of the second peak in the magnitude spectrum of Fourier transform of the derivative of the signal is identified as explained below.

4. Spectral values at the evaluated exact peak locations in the magnitude spectrum of Fourier transform of the signal as well as magnitude spectrum of Fourier transform of the differentiated signal are found out. Using parabolic interpolation, these spectral values are estimated as,

$$y(p) = \beta - \frac{1}{4}(\alpha - \gamma)p \qquad \textbf{(8)}$$

5. Where, p is calculated by Equation (5). Spectral value $y(p)$, p, α, β, γ are separately calculated for Fourier transform of the derivative of the signal and Fourier transform of the signal.

6. (1/2π) times $y(p)$ of the differentiated signal divided by $y(p)$ of the signal is fundamental frequency at time t.

7. Now, frame is advanced by one hop and whole procedure is repeated to find out fundamental frequency for new frame.

8. This is continued till end sample of the signal to get time vs. pitch graph.

RESULTS

Accuracy

The accuracy of the algorithm was tested by finding out % Error for mathematically generated harmonic signals ranging from 86-1600 Hz. Harmonic signal was considered as fundamental (with amplitude 0.75) and its first harmonic (with amplitude 0.25) with sampling frequency 44100 Hz. Mathematical signal is generated by the Equation (9) given below.

$$y = 0.75 \sin(2 \text{ pi } F_{ref} \, t)+0.25 \sin(2 \text{ pi } (2F_{ref}) \, t) \qquad \textbf{(9)}$$

Size of first and second Fourier transforms (N_1 and N_2) was 4096. Number of samples per analysis frame (window size) was 2048.
% Error is calculated as follows.

$$\%Error = \frac{|Fcal - Fref|}{Fref} \times 100 \qquad \textbf{(10)}$$

Where, *F*ref be the exact fundamental frequency and *Fcal* its measured value.

Figure 5 shows % Error vs. Frequency of the harmonic signal using proposed method. Calculation of fundamental frequency is done for harmonic signal generated using mathematical formula given in Equation (9).

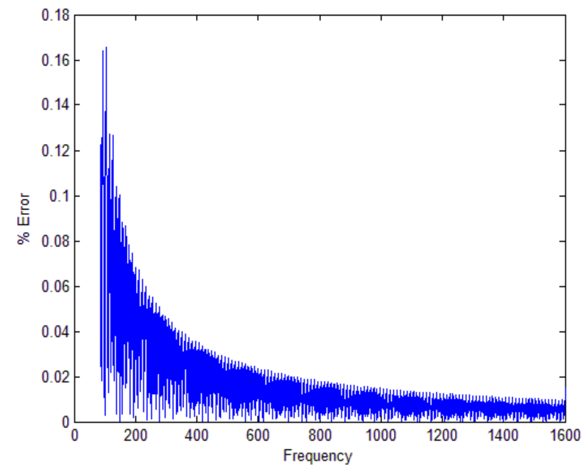

Figure 5a: % Error vs. Fundamental frequency for the Proposed method.

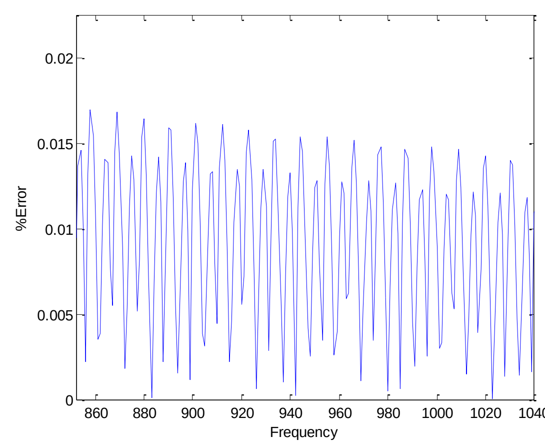

Figure 5b: Magnified Plot of Graph in Figure 5a.

Lowest value of frequency considered is 86 Hz. Estimation of fundamental frequency is done as explained below. Estimated frequency and actual frequency of the signal may not be same. The error in estimation of fundamental frequency is calculated by Equation (10).

Now, frequency of the signal in mathematical equation was advanced by 1 Hz and % Error is calculated. This way, calculation of % Error is done for all the frequencies till 1600 Hz. Variation of % Error vs. Frequency of the signal is plotted. Calculation of fundamental frequency using Fourier of Fourier transform and FT^1 are discussed below. % Error for both these methods are calculated for the same harmonic signal in Equation (9). Plot of % Error vs. Frequency for Fourier of Fourier transform and FT^1 methods are shown in Figure 6 and Figure 7, respectively.

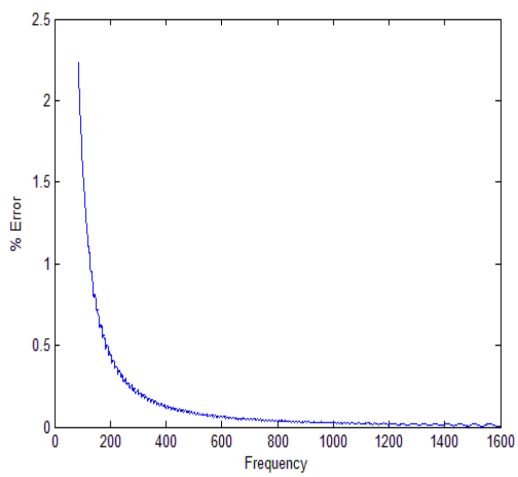

Figure 6: % Error vs. Fundamental Frequency for the Fourier of Fourier Transform Method.

Comparative analysis of Fig.5, Fig.6 and Fig.7 gives us following conclusions:

% Error is more for lower frequencies for Fourier of Fourier transform method; nearly 13 times larger for lowest frequency 86 Hz than in proposed method. % Error is lowest in FT^1 method for lower frequencies, but it increases for higher frequencies. The above analysis shows that proposed method is best suitable for higher as well as lower pitches. Maximum % Error is 0.17% for lowest frequency 86 Hz and it goes on reducing to 0.01% at highest frequency 1600 Hz. Method proposed by Marchand [24] shows 1% to 6% Error for 440 to 1660 Hz, where as our method shows 0.03% to 0.01% Error for the same range, which is very much appreciable in view of microtone research and QBH applications.

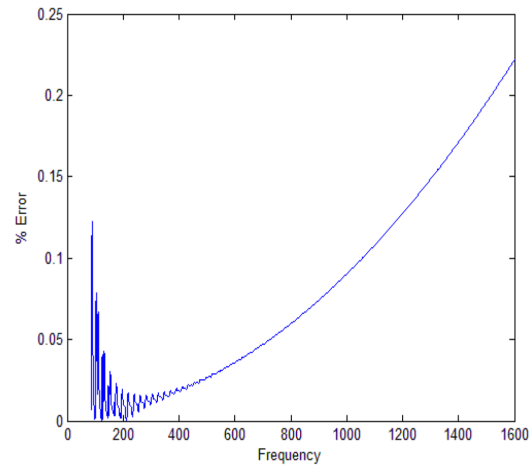

Figure 7: % Error vs. Fundamental Frequency for FT^1 Method.

Performance

This method requires computation of two Fourier transforms. It is much faster than well known autocorrelation method. Also it is similar in principle to Cepstrum [1] but computationally more efficient as logarithm is not required.

Robustness

Frequency resolution in Autocorrelation method is very good but it shows octave errors. Results were compared with autocorrelation methods Praat [17] and Reduced Autocorrelation method [23] for various vocal samples.

Figures 8 and 9 show the pitch graphs for the same music sample (male vocal singing C). We can clearly see the octave error in Figure 8, which is not seen in Figure 9.

Many audio samples were analysed with methods in [17] and [23]. One of them is song Dnyaniyancha Raja Guru Maharav for 10 sec. to 13 sec. available at www.dhingana.com. It was analysed to get pitch graph by Praat method. It gives octave error as shown in Figure 10. For the same audio sample, pitch graph was obtained using our method with no octave error (Figure 11). Hence it is clear from Figures 8, 9, 10, and

11 that octave error is one of the most important drawbacks of autocorrelation method. Proposed algorithm obviates the octave error as the crux of this algorithm is the ability of detecting correct peak corresponding to the fundamental frequency.

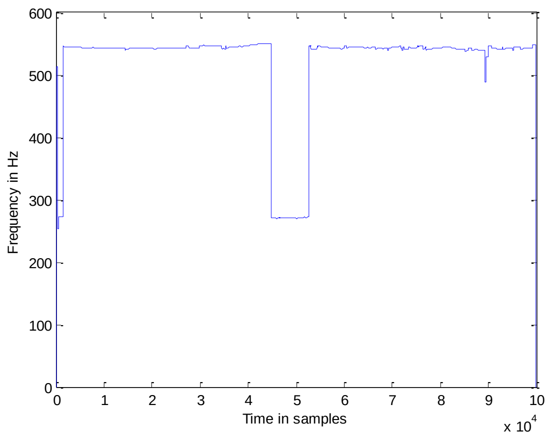

Figure 8: Pitch Graph for Male Vocal Singing C#, using Reduced Autocorrelation Method Showing Octave Error.

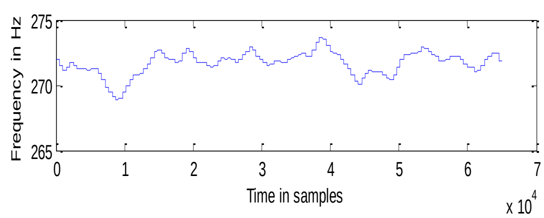

Figure 9: Pitch Graph for Male Vocal Singing C#, using Proposed Method Showing no Octave Error.

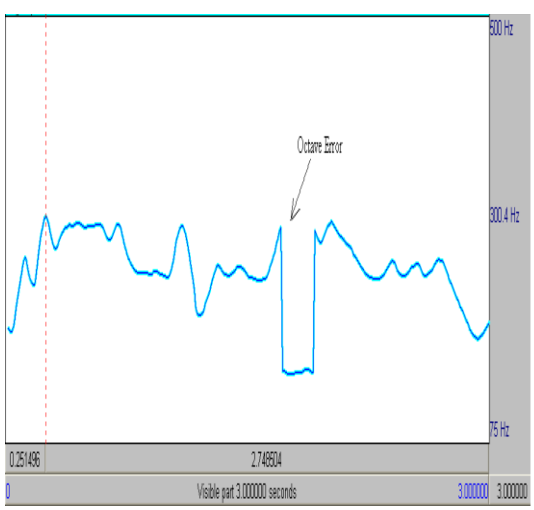

Figure10: Pitch Graph for Song *Dnyaniyancha Raja Guru Maharav* for 10 - 13 sec. available at www.dhingana.com by Praat [22] Method Showing Octave Error.

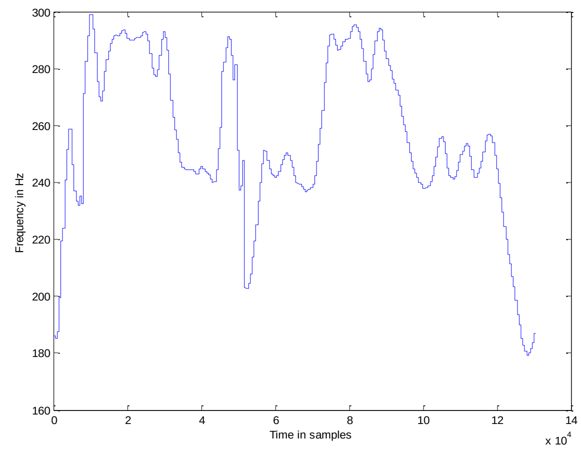

Figure 11: Pitch Graph for the Same Sample in Figure 10 above by Proposed Method showing no Octave Error.

Pitch graph for the song *Happy Birthday to You* for all four lines is shown in Figures 12, 13, 14, and 15. This song was sung by first author (female vocal) in monophonic environment.

This algorithm is tested for monophonic male vocal, monophonic female vocal as well as monophonic instrumental audio samples and proved to be accurate and octave error free.

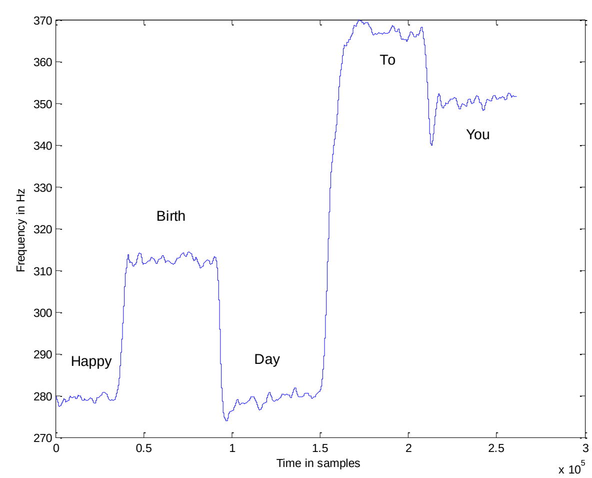

Figure12: Pitch Graph for first line of *Happy Birthday to You*.

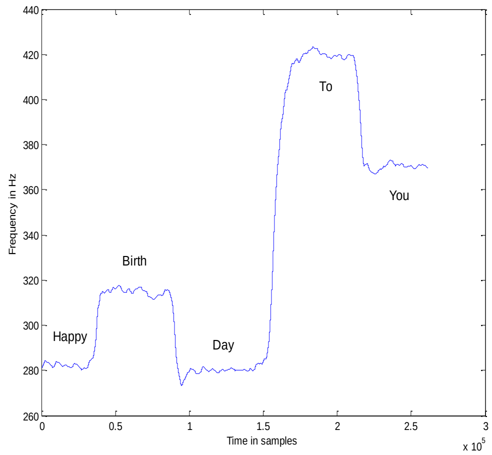

Figure 13: Pitch Graph for Second line of *Happy Birthday to You.*

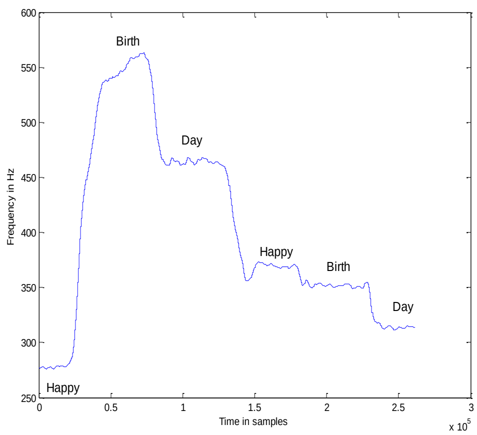

Figure 14: Pitch Graph for Third line of *Happy Birthday to You.*

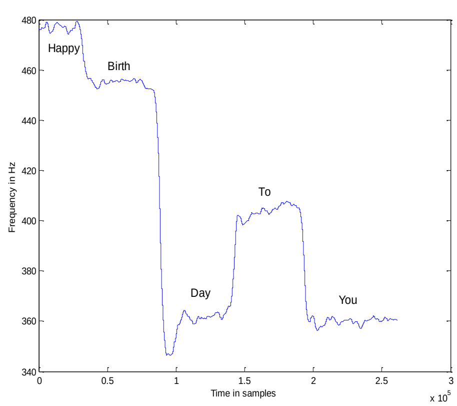

Figure 15: Pitch Graph for Forth line of *Happy Birthday to You.*

CONCLUSION

There are two major difficulties, namely, octave errors and pitch estimation accuracy, which most pitch detection algorithms have to deal with [6]. In this article we have proposed a method of pitch tracking which gives accurate estimation of pitch and results with no more octave errors.

The computational complexity in the algorithm is not very large as it is in the other pitch detection methods. For applications such as QBH, microtone research, musical pattern recognition and musical genre classification, the first step is to identify musical note through pitch tracking. High accuracy in pitch estimation is the main requirement for such cases. This method is suitable in the realm of Indian Classical Music, where, there is concept of Shruti (microtones).

Once pitch contour is obtained, next steps would be pitch quantization and note segmentation. Finally pitch vector for the audio sample under consideration would be obtained which can be used for various applications of pitch tracking.

Our algorithm is limited to frame level analysis. Its input is a frame and outputs are frequency and average amplitude. In future, the amplitude value of the signal will be used by the higher layer to detect silence zones, note onsets and note duration. The higher layer will convert the frequency to a MIDI note number and quantized note duration. Still higher layer will operate on the note sequence and do higher level tasks like HMM [12], genre identification, database search, etc.

Tracking the fundamental frequency of monophonic music in the presence of interfering sounds is one of the needs of QBH applications. Further research would include pitch tracking in polyphonic environment which would give multiple uniformly spaced series of peaks corresponding to different pitches in the sound.

REFERENCES

1. Noll, A.M. 1967. “Cepstrum Pitch Determination”. *Journal of Acoustical Society of America*. 41(2):293-309.

2. Cemgil, A.T., H.J. Kappen, and D. Barber. 2003. “Generative Based Polyphonic Music Transcription”. *Proc. of IEEE WASPAA*. New Paltz, NY. IEEE Workshop on Applications of

Signal Processing to Audio and Acoustics. October 2003.

3. Krishnaswamy, A. 2004. "Application of Pitch Tracking to South Indian Classical Music". *Proceedings of IEEE, ICASSP*. Hong Kong, China, April 2003. 49-52.

4. Liu, B., Y. Wu, and Y. Li. 2003. "Linear Hidden Markov Model for Music Information Retrieval Based on Humming". *Proceedings of ICASSP'03*. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. 5:533-536.

5. Chordia, P. and A. Rae. 2007. "Raag Recognition using Pitch-Class and Pitch-Class Dyad Distributions. *Proceedings of International Conference on Music Information Retrieval*.

6. Dziubinski, M. and Kostek, B. 2004. "High Accuracy and Octave Error Immune Pitch Detection Algorithms". Archives of Acoustics. 9:3-24.

7. Pollastri, E. 2002. "A Pitch Tracking System Dedicated to Process Singing Voice for Music Retrieval". *IEEE International Conference on Multimedia and Expo, 2002. ICME '02.* 1:341-344.

8. Pandey, G.T. 2003. "A System for Automatic Raga Identification". *Indian International Conference on Artificial Intelligence 2003*. Hyderabad, India.

9. Generative Model Based Polyphonic Music. http://www.22shrutiharmonium.com

10. Iman, S., H. Suyoto, and A.L. Uitdenbogerd. 2004. "Exploring Microtonal Matching". *ISMIR 2004, 5th International Conference on Music Information Retrieval.* Barcelona, Spain. October 10-14, 2004.

11. Smith, J.O. and X. Serra. 1987. "PARSHL: An Analysis/Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation". *Proceedings of the 1987 International Computer Music Conference, International Computer Music Association*, San Francisco, CA. 290 - 297.

12. Rabiner, L.R. 1989. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition". *Proc. IEEE*. 77(2):257-286.

13. Fu, L. and X. Xue. 2005. "A New Spectral Based Approach to Query by Humming for MP3 Songs Database". *World Academy of Science, Engineering and Technology.* 4:117-121.

14. Sinith, S.M. and K. Rajeev. 2007. "Pattern Recognition in South Indian Classical Music Using a Hybrid of HMM and DTW". *IEEE Conference on Computational Intelligence and Multimedia Applications*. Dec. 2007. 2:339-343.

15. Sinith, M.S. and K. Rajeev. 2006. "Hidden Markov Model based Recognition of Musical Pattern in South Indian Classical Music". *IEEE International Conference on Signal and Image Processing*. Hubli, India.

16. Desainte-Catherine, M. and S. Marchand. 2000. "High Precision Fourier Analysis of Sounds Using Signal Derivatives". *Journal of the Audio Engineering Society*. 48(7/8):654-667.

17. Boersma, P. and D. Weenink. "Praat: Doing Phonetics by Computer". Institute of Phonetic Sciences, University of Amsterdam. http://www.praat.org

18. Chordia, P. and A. Rae. 2007. "Understanding Emotion In Raag : An Empirical Study of Listener Responses". *Proceedings of the 2007 International Computer Music Conference*. ICMC 2007.

19. Kumar, P., M. Hariharan, S. Dutta-Roy, and P. Rao. 2007. "Sung Note Segmentation for a Query-by-Humming System". *Proc. of Music-AI (International Workshop on Artificial Intelligence and Music)*. IJCAI, 2007.

20. Irizarry, R.A. 2001. "Local Harmonic Estimation in Musical Sound Signals". *Journal of the American Statistical Association*. 96:357-367.

21. Irizarry, R.A. 2002. "Weighted Estimation of Harmonic Components in a Musical Sound Signal". *Journal of Time Series Analysis*. 23:29-48.

22. Sridhar, R. and T.V. Geetha. 2009. "Raga Identification of Carnatic Music for Music Information Retrieval". *International Journal of Recent Trends in Engineering*. 1.

23. Uppgård, S. 2001. Master of Science thesis project in signal processing at KTH, Stockholm, Sweden.

24. Marchand, S. 2001. "An Efficient Pitch-Tracking Algorithm using a Combination of Fourier Transforms". *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01)*, Limerick, Ireland, December 6-8, 2001.

25. Marchand, S. 1998. "Improving Spectral Analysis Precision with an Enhanced Phase Vocoder Using Signal Derivatives". *Proceedings of the Digital Audio Effects (DAFx) Workshop*. Barcelona, Spain, November 1998. 114–118.

26. Cemgil, T., H.J. Kappen, and D. Barber. 2006. "A Generative Model for Music Transcription". *IEEE Transactions On Audio, Speech, And Language Processing*. 14(2):679- 694.

27. Birmingham, W., R. Dannenberg, and B. Pardo. 2006. "Query by Humming with the Vocal Search System". *Communications of the ACM*. 49:49-52.

28. Zhu, Y., and M.S. Kankanhalli. 2003. "Robust and Efficient Pitch Tracking for Query-by-Humming". *Fourth Pacific Rim Conference on Multimedia and Proceedings of the 2003 Joint Conference of the Fourth International Conference on Information, Communication and Signal Processing*. 3:1586-1590.

29. Zhao, Z. and L.J. Brown. 2003. "Musical Pitch Tracking using Internal Model Control Based Frequency Cancellation". *42nd IEEE Conference on Decision and Control*. 5:5544- 5548.

ABOUT THE AUTHORS

K.A. Akant, M.Tech. is a researcher at Manoharbhai Patel Institute of Engineering and Technology in Gondia, India.

Rajesh Pande, Ph.D. is a faculty member at the Shri Ramdeobaba Kamla Nehru Engineering College in Nagpur, India.

S.S. Limaye, Ph.D. is a faculty member at the Jhulelal Institute of Technology in Nagpur, India. The authors have research interests in practical applications of data storage, retrieval, and interpolation.

SUGGESTED CITATION

Akant, K.A., R. Pande, and S.S. Limaye. 2010. "Accurate Monophonic Pitch Tracking Algorithm for QBH And Microtone Research". *Pacific Journal of Science and Technology*. 11(2):342-352.

Pacific Journal of Science and Technology