Pitch Estimation of Notes in Indian Classical Music

Conference Paper · January 2010

DOI: 10.1109/INDCON.2009.5409364 · Source: IEEE Xplore

3 authors:

Siddarth Rai Mahendra
Indian Institute of Technology Bombay

Hemant Patil
Dhirubhai Ambani Institute of Information and Communication Technology

Narendra Kumar Shukla
Shiv Nadar University

Pitch Estimation of Notes in Indian Classical Music

Siddarth Rai Mahendra
Graduate Institute of Electrical Engineering, NTPU
Taipei-23741, Taiwan
s79682701@webmail.ntpu.edu.tw

Hemant A. Patil
DAIICT Gandhinagar
Gujarat-382007, India
hemant_patil@daiict.ac.in

Narendra Kumar Shukla
Aricent Technologies Pvt. Ltd.
http://www.aricent.com
narendra.shukla@ieee.org

***Abstract*—Music consists of various tones having a definite pitch. Thus the study of music from a signal processing perspective involves extraction and processing of the *pitch* information of these tones. In this paper, we have introduced a new concept of fundamental score and its significance in a musical octave. We have applied *autocorrelation* method for the pitch extraction of musical notes in Indian Classical Music (ICM). We have tested this method on the data collected from male speech (sung by one of the authors) and a musical instrument, *viz.*, Reed-Organ. The results were found to be quite *accurate* when compared with the standard pitch frequencies computed using the equal temperament musical tuning method. The advantage of the proposed scheme is that it does not require knowledge of the source (i.e., musical instrument) to estimate the musical notes.**

Keywords-Musical notes; fundamental score; autocorrelation method; Indian classical music (ICM)

I. INTRODUCTION

Musical sounds are voiced sounds and hence have a *definite pitch.* Representation of music involves identifying and recognizing the specific pitch frequencies and representing them with their corresponding symbols in the Indian classical music (ICM). *Musical transcription* is a way of representing music using notations that can be transcribed on paper. The notation consists of a series of musical scores. Musical scores are defined as the symbols given to sounds (voiced sounds having definite pitch) that are distinguishable and comparable to each other. This kind of representation helps in reproducing back the same music. Connoisseurs of music are known for their ability to recognize the specific scores just by hearing the music, and the students of music aim to do this. Automatic transcription of music involves extracting voiced segments from music signals, calculating their quasi-periodicity (i.e., pitch) and then giving a musical score corresponding to these pitch values. The ultimate motivation for this work is to be able to do the automatic transcription of any musical signal or song. Hence, the first step in achieving this is the automatic detection of pitch, which is quite difficult a task in its own right. There have been some efforts for solving this problem in the ICM literature. For example, Jairazhbhoy [1] and Levy [2] reported work on using special equipments for estimating the pitch. However, the equipment used by these two authors is now antiquated and hence not able to produce any detailed, high-resolution plots. Their focus was also restricted to North Indian music. While some of their inferences and observations were convincing, their work seems to have had little effect on the overall mindset of most Indian musicians and musicologists [3]. Recently, Krishnaswamy reported use of three pitch tracking algorithms for the analysis of South ICM [3], [4]. However, this work does not compare the results obtained with the standard pitch estimation methods like the equal temperament musical tuning method [5]-[7]. In this paper, the autocorrelation method for pitch estimation has been implemented and the pitch frequencies obtained for the various musical notes in ICM are approximately the same as compared to those of the standard pitch values computed using the equal temperament musical tuning method. The novelty of the proposed approach is that it does not require knowledge of the source (i.e., musical instrument) to estimate the musical notes.

II. THEORY OF MUSIC

For producing sound, three things are required: a *vibrating object* producing sound, a *medium* through which sound can propagate (e.g., air) and a *receiver* (transducer, ear or sensor) to perceive sound. The frequency of vibrations produced by the vibrating object defines the nature and characteristic of the audio signal. The frequency of vibration of acoustic signals are divided into four major types, *viz.*, constant frequency vibrations, variable frequency vibrations, steady or continuous vibrations and unsteady or discontinuous vibrations.

Music has been defined as the art of combining sounds which appeals to the human ear. Music is concerned with only two types of vibrations – constant frequency vibrations and steady or continuous vibrations. Hence, the sounds made of these vibrations can in turn be referred to as *Musical Sounds*. Therefore, in general sounds can also be classified into *Musical Sounds* and *Unmusical Sounds*. Musical sounds are referred to as *'Naada'* in the ICM.

Of the many frequencies of musical sounds (Naada), those frequencies which are distinguishable and comparable to the other frequencies are known as *Shruti(s)* in the ICM. In ICM, the musical sounds have been divided into twenty-two small divisions in an octave of frequencies, each division being referred to as a *Shruti*. Shrutis are the microtonal intervals of sound, which are very small units and are twenty-two in number [8]. It would be fulfilling to mention in the passing that it is a close to impossible task even for an experienced musicologist to identify these twenty-two Shrutis, let alone singing them. Hence from these twenty-two micro-tonal intervals, selected twelve are referred to as *Musical Notes*, which are clearly audible and distinguishable to a normal listener. Hence, Musical Notes are a *subset* of these Shrutis. The placement of these twelve musical notes on specific

978-1-4244-4859-3/09/$25.00 ©2009

Shrutis is shown in Table-I [8]. Different placement orders, giving the position of notes in the Shrutis, have been given by different musicologists at different times in the history of ICM. However, Table-I gives the most recent placement of the notes as agreed by the modern musicologists. The musical scale used here is in accordance with the *western chromatic scale notation,* i.e., *G#, A, A#, B, C, C#, D, D#, E, F, F#, G*. The respective transliterated notes in the Indian notation are: *Sa, Re-, Re, Ga-, Ga, Ma, Ma-, Pa, Dha-, Dha, Ni-, Ni* (The note for 'Sa' one octave higher is chosen as 'Saa').

TABLE I. MODERN PLACEMENT OF MUSICAL NOTES ON SPECIFIC SHRUTIS [8]

Shruti Number	Shruti Name (Sanskrit)	Shruti Symbols – with Western Names	Chromatic Musical Scale
1	Tivra	N.S.S.	-
2	Kumudati	N.S.S.	-
3	Mandra	N.S.S.	-
4	Chandovati	Shadaj (Sa) – Do	G#
5	Dayavati	N.S.S.	-
6	Ranjani	Rishabh (Re-)	A
7	Raktika	Rishabh (Re) – Re	A#
8	Raudri	Gandhar (Ga-)	B
9	Krodhi	Gandhar (Ga) – Mi	C
10	Vajrika	N.S.S.	-
11	Prasarini	N.S.S.	-
12	Priti	N.S.S.	-
13	Marjani	Madhyam (Ma) – Fa	C#
14	Ksiti	N.S.S.	-
15	Rakta	Madhyam (Ma-)	D
16	Sandipani	N.S.S.	-
17	Alapini	Pancham (Pa) – So	D#
18	Madanti	N.S.S.	-
19	Rohini	Dhaivat (Dha-)	E
20	Ramya	Dhaivat (Dha) – La	F
21	Ugra	Nishadh (Ni-)	F#
22	Ksobhini	Nishadh (Ni) – Ti	G

a. N.S.S. = Non-Specified Shruti

III. PITCH - IMPORTANCE AND ESTIMATION METHODS

A. *What is Pitch?*

Speech sounds in general can be broadly classified into two categories: *voiced sounds* and *unvoiced sounds*. Voiced sound is the quasi-periodic part of the speech and its spectrum is composed of a fundamental frequency along with the natural harmonics. In this paper, we are only interested in the voiced component of speech. The fundamental frequency of the voiced component of speech is known as the *pitch frequency*. Voiced sound is produced by the vibration of the vocal folds, hence characteristics of the vocal folds are reflected in the pitch information and the change in tension of the vocal folds with time decides the variation of the pitch [9], [10]. In fact, pitch is unique for a speaker and it depends upon the differences in the shape and size of the vocal folds, and also on the peculiarity of the accent of the speaker.

B. *Pitch Estimation*

Pitch estimation is a major and very important problem in speech research (e.g., speech analysis, speaker recognition, etc.) and especially in the studies of ICM [3], [4], [11]. The significance of this problem can be well understood by knowing the difficulties associated with its estimation.

Firstly, voiced sound is quasi-periodic in nature, i.e., though the period is clearly visible to the human eye, automatic detection through a computer is tough, as no part of the signal in any period is exactly the same. Secondly, in time-domain estimation it is difficult to detect the beginning and end of pitch period in a voiced segment. The other problem is the interaction of the voiced segment (generated by the vocal folds) with the formant frequencies (resonant frequencies) of the vocal tract. Finally, distinguishing between the voiced and unvoiced segment of the speech is in itself hard to detect [12]. Several pitch estimation methods have been proposed in the speech processing literature [12]-[14].

The first musical score *'Sa'* (in ICM) or *'Do'* (in Western classical music) helps in determining the pitch range of a song. Hence, this first musical score of any octave has been defined in this work as the *fundamental score*. Determining the fundamental score of any song gives us an idea of the range of frequencies of voiced sounds in a song. This is because the octave of frequencies can be defined to be the range: from the fundamental frequency to the double of the fundamental frequency. This fundamental score is also referred to as '*Shadaj*' or '*OM*' or '*Nada Brahma*' in the Indian music [8]. Also, every piece of music or song has a unique fundamental score associated with it. This is also the case with humans in general, that is, different people have their own unique fundamental scores that they find comfortable to sing in.

C. *A General Pitch Detector*

The basic task of any pitch detector algorithm or functional block is firstly to distinguish between the voiced and unvoiced components of speech and then if the speech is voiced, to estimate the pitch period of the voiced component, else ignore that particular speech frame. A general pitch detector block diagram is as shown in Fig. 1:

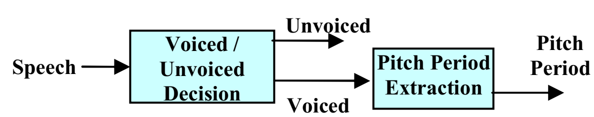

Figure 1. A general block diagram for pitch detection

IV. DATA COLLECTION AND CORPUS DESIGN

The database for this study was prepared from two sources, *viz.*, a male speaker and Reed-Organ (Instrument). The data was recorded using a close-talking microphone (i-ball i-342MV electret condenser microphone) in a closed small room. The whole data was collected within a period of 2 days for about 6~7 hrs a day with intervals of 1~2 hrs. Data for a particular pattern was collected continuously without any intervals. The data from the Reed-Organ was collected by placing the microphone near the Reed-Organ as shown in Fig.

2 and then periodically and slowly moving its *air-slab* to generate sound. The data from the male voice of the author was collected with the same close-talking microphone. The software used for audio recording is Sony Sound Forge. The data has been sampled at 44.1-kHz and stored as **.wav* files.

Figure 2. Experimental Setup for data collection from Reed-Organ

V. AUTOCORRELATION METHOD APPLIED TO PITCH EXTRACTION OF MUSICAL NOTES

The autocorrelation of any function *x(n)* is:

$$r_{xx}(l) = \sum_{n=-\infty}^{\infty} x(n+l)\cdot x(n)\,,\ l = 0, \pm 1, \pm 2, ... \qquad (1)$$

Here, l is called the *lag value*. Autocorrelation method can be used for the determination of the fundamental period of a periodic or quasi-periodic signal. This method has been used in the paper for the calculation of frequencies of musical notes. To calculate the pitch frequency, we divide the sampling frequency by the lag value between two consecutive maxima peaks (a standard technique from the sampling theory), i.e.

$$\text{Pitch Frequency} = \frac{\text{Sampling Frequency}}{\text{Lag value between two consecutive peaks}}$$

The analysis of this method under additive noise conditions is presented in Appendix I. The results obtained have been compared with the standard pitch values computed using the well known method of equal temperament musical tuning used to generate discrete frequencies (musical notes) [5]-[7]. These discrete frequencies are used to construct musical instruments that are tuned to these specific frequencies. In this method, a constant factor between each musical note is used. Starting from any note the frequency to other note may be calculated from its frequency by [6]:

$$F = (note) \times 2^{N/12}$$

Here, N=0 for the first frequency note and we can start tuning from any frequency. The pitch values of the musical notes obtained for the instrument Reed-Organ and from the male voice are shown in Tables II and III respectively. Some of the observations from the results are as follows:

- It is clearly seen from Tables II and III that the pitch frequencies of the male voice (Table III) are approximately one octave lower (one-half the frequency) than the pitch frequencies of the Instrument Reed-Organ (Table II).
- The frequency value of ‘Saa’ is approximately double of the frequency value of ‘Sa’, which satisfies the criterion, that the frequency values in a higher octave are double of that in the lower octave.
- The pitch frequency values obtained, match very closely to those of the pitch frequency values given in [5]-[7].

TABLE II. PITCH VALUES OF MUSICAL NOTES FROM REED-ORGAN

Musical Note	Lag-value(l)	Freq=F_s/l (in Hz)	[5] (in Hz)	[6] (in Hz)	[7] (in Hz)
Sa (G#)	191	230.89	233.08	233.08	233.082
Re- (A)	181	243.64	246.94	246.96	246.942
Re (A#)	170	259.41	261.63	261.64	261.626
Ga- (B)	161	273.91	277.18	277.20	277.183
Ga (C)	153	288.23	293.66	293.68	293.665
Ma (C#)	144	306.25	311.13	311.12	311.127
Ma- (D)	136	324.26	329.63	329.64	329.628
Pa (D#)	128	344.53	349.23	349.24	349.228
Dha- (E)	120	367.5	369.99	370.00	369.994
Dha (F)	114	386.84	392.00	392.00	391.995
Ni- (F#)	107	412.15	415.30	415.32	415.305
Ni (G)	102	432.35	440.00	440.00	440.000
Saa (G#)	95	464.21	466.16	466.16	466.164

TABLE III. PITCH VALUES OF MUSICAL NOTES FROM MALE VOICE

Musical Note	Lag-value (l)	Freq=F_s/l (in Hz)	[5] (in Hz)	[6] (in Hz)	[7] (in Hz)
Sa (G#)	393	112.213	116.54	116.54	116.541
Re- (A)	367	120.16	123.47	123.48	123.471
Re (A#)	341	129.32	130.81	130.82	130.813
Ga- (B)	320	137.8125	138.59	138.60	138.591
Ga (C)	307	143.64	146.83	146.84	146.832
Ma (C#)	288	153.125	155.56	155.56	155.563
Ma- (D)	275	160.36	164.81	164.82	164.814
Pa (D#)	255	172.94	174.61	174.62	174.614
Dha- (E)	240	183.75	185.00	185.00	184.997
Dha (F)	230	191.74	196.00	196.00	195.998
Ni- (F#)	215	205.11	207.65	207.66	207.652
Ni (G)	206	214.077	220.00	220.00	220.000
Saa (G#)	196	225.00	233.08	233.08	233.082

VI. SUMMARY AND CONCLUSIONS

In this paper, the method of autocorrelation is applied and analyzed for the extraction of pitch frequency of musical signals. The results obtained are also compared with the standard pitch frequencies reported in the literature. The major contributions of the paper are as under:

- Applying autocorrelation method for analysis of standard musical notes in ICM and analysis of this method for pitch extraction problem under additive noise.
- The new concept of *fundamental score* has been defined. Knowing the fundamental score, we can accurately know the positions of the other musical notes also and we can also judge or predict beforehand the range of frequencies that any musical signal or song may contain within it.

Some of the suggested future work includes:

- Specific determination of all the 22 Shrutis in ICM using signal processing techniques for pitch extraction. This helps in understanding the relationship between the 12 basic musical notes and the other 10 Shrutis.

- The *fundamental-score* plays a very crucial role in singing in ICM and is determined by various factors: the natural singing frequency range of the artist and the inherent frequency range of a song. Matching these two ranges is hence necessarily important. This specifically implies nothing but matching the fundamental-scores of the two ranges. Hence, our future work looks upon the topic of: *'Automatic Fundamental-Score switching of any Song'*.

- It is interesting to note that the proposed method of autocorrelation and the well-known method of equal temperament musical tuning are two distinct and opposite approaches in analyzing ICM. For example, the latter method helps in generating discrete musical notes whereas the former method extracts the musical notes from any given segment of music (even with the source instrument unknown). It is also a worthwhile problem to identify the source of music from the output of the pitch detector.

APPENDIX I

A. *Analysis of autocorrelation method under additive noise*

From (1), for l=0, we get the well-known property of auto-correlation that helps in finding the periodicity of signals [15]:

$$r_{xx}(l) \le r_{xx}(0) = E_x$$

Here, E_x is the energy of the signal *x(n)*. This means that, the autocorrelation sequence of a signal attains its maximum value at zero lag. This result is consistent with the notion of correlation, that a signal perfectly matches itself at zero shifts. Hence, the autocorrelation of a periodic signal attains its maximum value at shifts equal to its period. From this, the period of the signal can be calculated by measuring the difference of the lags between two consecutive peak values of autocorrelation and multiplying this lag value with the sampling time interval. This idea can also be extended to find the frequency values corresponding to the musical notes and the fundamental score. The reason why autocorrelation helps in finding the period of quasi-periodic signals is demonstrated below [15]:

Assume, that a signal *x(n)* is a perfectly periodic signal, but has become quasi-periodic by the interference of noise and other signals. Let us consider such a signal as:

$$y(n) = ax(n-d) + w(n)$$

Here, *'a'* is some multiplicative constant, *d* is the delay and *w(n)* is the additive noise. Taking the autocorrelation of *y(n)* yields the following result:

$$r_{yy}(l) = \sum_{n=-\infty}^{\infty} a^2 x(n-d)x(n-d-l) + \sum_{n=-\infty}^{\infty} w(n)\cdot w(n-l)$$
$$+a\cdot \sum_{n=-\infty}^{\infty} x(n-d)w(n-l) + a\cdot \sum_{n=-\infty}^{\infty} w(n)x(n-d-l)$$

Here, substituting *z(n)* for $x(n-d)$ gives:

$$r_{yy}(l) = a^2 \cdot r_{xx}(l) + r_{ww}(l) + a\cdot r_{zw}(l) + a\cdot r_{wz}(l)$$

From the R.H.S of the above equation, it is clear that:

- The first term is the autocorrelation of the signal *x(n)* which is same as the autocorrelation of *x(n-d)*. As a^2 is always positive, it only scales the term. Since *x(n)* is periodic, its autocorrelation sequence also has the same period and hence contains large peaks at positive integral multiples of the lag values corresponding to the period. The second term is the autocorrelation of the additive noise *w(n)*. As *w(n)* is a random sequence, its autocorrelation will only contain a peak at zero lag and with increasing lags is expected to move rapidly towards zero. The last two terms are the cross-correlation of *w(n)* and *x(n)* and are expected to be relatively small due to the expectation that *x(n)* and *w(n)* will be totally uncorrelated.

- Hence, only the first term which is the autocorrelation of *x(n)* will have large peaks at the lag values corresponding to the period. Calculating the consecutive lag values corresponding to the peaks and multiplying this with the sampling interval gives the period of the periodic signal *x(n)* buried in the interference of *w(n)*.

ACKNOWLEDGMENT

The authors would like to thank the authorities of DAIICT, Gujarat, India for their support in carrying out this research.

REFERENCES

[1] Jairazbhoy and Stone, "Intonation in present-day north Indian classical music," Bulletin of the School of Oriental and African Studies, vol. 26, pp. 119-132, 1963.

[2] Levy, Intonation in North Indian Music: A Select Comparison of Theories with Contemporary Practice, Biblia Impex Private Limited, New Delhi, India, 1982.

[3] Arvindh Krishnaswamy, "Application of Pitch Tracking to South Indian Classical Music," Proc. IEEE ICASSP, pp.V-557-V-560, 2003.

[4] Arvindh Krishnaswamy, "Pitch Measurements versus Perception of South Indian Classical Music," Proc. Stockholm Music Acoustics Conference (SMAC) 2003.

[5] http://www.phy.mtu.edu/~suits/notefreqs.html

[6] http://www.techlib.com/reference/musical_note_frequencies.htm

[7] http://www.sengpielaudio.com/calculator-notenames.htm

[8] Swatantra Sharma, "Fundamentals of Indian Music," Pratibha Prakashan Publications.

[9] B.S. Atal, "Automatic speaker recognition based on pitch contours," J. Acoust. Society of America, Vol. 52, pp. 1687-1697, July 1972.

[10] L.R. Rabiner, R.W. Schafer, "Digital Processing of Speech Signals," Prentice-Hall Signal Processing Series.

[11] M. S. Sinith, K. Rajeev, "Pattern Recognition in South Indian Classical Music using a hybrid of HMM and DTW," Int. Conf. on Computational Intelligence and Multimedia Applications, 2007.

[12] L.R. Rabiner, M.J. Cheng, A.E. Rosenberg, C.A. McGonegal, "A comparative performance study of several pitch detection algorithms," IEEE Trans. Acoust., Speech ,Signal Processing, Vol. ASSP-24, pp. 399-418, Oct. 1976.

[13] John D. Markel, "The SIFT algorithm for fundamental frequency estimation," IEEE Trans. Audio and Electroacoustics, Vol. AU-20, pp. 367-377, Dec. 1972.

[14] A.M. Noll, "Cepstrum pitch determination," J. Acoust. Soc. Am., Vol. 41, pp. 293-309, Feb. 1967.

[15] J.G. Proakis, D.G. Manolakis, "Digital Signal Processing – Principles, Algorithms and Applications," 3rd Edition, Prentice Hall India, 2005.