Automatic Music Transcription of Indian Classical Music into MIDI Data

Kalyani Akant[1], Rajesh Pande[2], and Shyamkant Limaye[3]

[1] Manoharbhai Patel Institute of Engineering and Technology, Gondia
[2] ShriRamdeobaba College of Engineering and Management, Nagpur
[3] Jhulelal Institute of Technology, Nagpur
kalyaniakant@gmail.com, panderaj@yahoo.com,
shyam_limaye@hotmail.com

Abstract. Musical transcription of sampled audio data is the process of taking a sequence of digital data corresponding to the sound waveform and extracting from it the symbolic information i.e. MIDI scores. In this work, two different approaches for transcription of Indian Classical Music are discussed, one with prior knowledge of Raga and another without it. For this purpose, pitch detection is done using Fourier of Fourier Transform with Parabolic Interpolation to spectral peaks. This method for pitch detection is very accurate and suitable for shrutis in Indian Classical Music. For polyphonic music, we performed source separation by classifying frames as monophonic/polyphonic one. These separated monophonic frames are handled by a monophonic transcription system. Finally results are shown with conversion of Raga audio sample into MIDI format.

Keywords: Musical Transcription, MIDI, Fourier of Fourier Transform, Parabolic Interpolation, Shruti, monophonic/polyphonic, Raga.

1 Introduction

The huge growth of digital music in recent years has led to a large number of musical recordings becoming available in digital form as sampled audio. Additionally, progress in electronic music production has resulted in a lot of symbolic music data being created. Sampled audio cannot be manipulated as easily as symbolic music formats, and symbolic formats lack the authenticity of real recordings. A key step towards combining the benefits of these two realms is the ability to automatically produce a symbolic representation of a sampled music recording. This process is referred to as musical audio transcription [1]. There exist various techniques which can accurately transcribe monophonic recordings (eg.YIN [2], TWM [3], and the correlogram [4]) [1]. Raga is the most important concept in Indian music, making accurate recognition a prerequisite to almost all musical analysis [5]. There has been insufficient scientific research work to analyze the recordings of maestros who sing Indian classical music (ICM). This is the main aim of this research work. Proposed method is accurate and computationally efficient. Estimation of pitch is done by

V.V. Das and J. Stephen (Eds.): CNC 2012, LNICST 108, pp. 347–350, 2012.

Fourier of Fourier Transform method (FFT2) [7]. The location of spectral peaks is further refined with parabolic interpolation [6]. The resulting accuracy is suitable for detecting microtones [7].

2 Algorithm for Harmonic/Inharmonic Classification

For polyphonic recordings, we performed source separation by classifying frames as monophonic/polyphonic one [8]. These separated monophonic frames are further classified as harmonic/inharmonic using following algorithm.

Step 1: FFT2 of frame of size N is computed as discussed in [7].
Step 2: Bin numbers of all peaks in FFT2 spectrum from 0 to N are stored in a vector V. V={V1,V2……..,Vn}.
Step 3: Bin number of the peak having maximum amplitude (except bin0) in the FFT2 spectrum is detected. Let's denote this by K. (If bin0 is at 1, as in case of Matlab, K should be considered K-1). This is first harmonic.
Step 4: Second harmonic will be in the vicinity of ((K-1)*2)-5 to ((K-1)*2) + 5. Hence nth harmonic will be in the vicinity of ((K-1)*n)-5 to ((K-1)*n) + 5. If these harmonic peaks appear in vector V, and their amplitudes are such that, amplitude of 1st harmonic > amplitude of 2nd harmonic > amplitude of 3rd harmonic……till significant harmonic, then signal for selected frame is harmonic.

3 Estimation of Pitch

To determine pitch at a certain time t, temporal frame centered at t is considered. Then it is checked if the frame is monophonic and harmonic. If yes, then pitch for that frame is estimated as discussed in [7]. Frame overlapping is done by selecting hop size such that there is 75% overlap of the frames. This procedure is repeated for each frame. Finally we get "time vs. pitch" graph.

4 Result

4.1 Automatic Music Transcription: Approach 1

Here music transcription is done with prior knowledge of Raga. The notes in Raga are known; hence the note sequence along with duration of each note can be obtained as intermediate step. Finally, this note sequence is converted into MIDI note sequence.

4.2 Automatic Music Transcription: Approach 2

Here music transcription is done without prior knowledge of Raga. The Raga of audio data is unknown; pitch value of audio is quantized to nearest note level. This quantized note sequence is finally converted into MIDI note sequence.

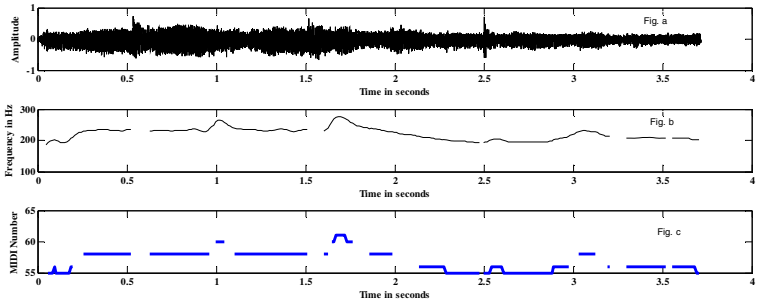

Fig. 1. Stages in music transcription of polyphonic ICM into MIDI data through approach 1

In Fig. 1, on the onset of tabla strokes near 0.5, 1.5 and 2.5 sec. the polyphonic and inharmonic frames are rejected to get smooth pitch graph. Fig. 2a shows audio waveform. Fig. 2b shows pitch graph. Fig. 2c shows pitch graph in MIDI numbers.

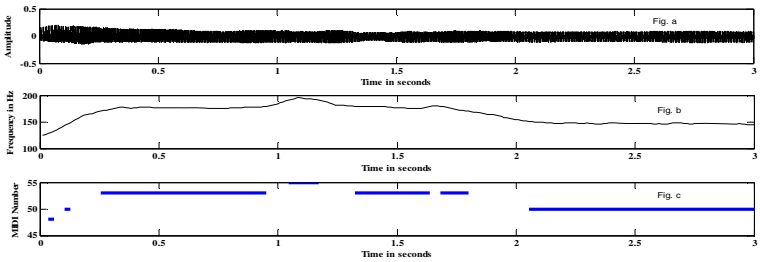

Fig. 2. Stages in music transcription of ICM into MIDI data through approach 1

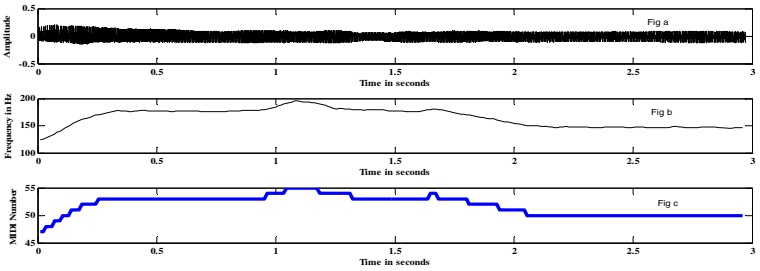

Fig. 3. Stages in music transcription of ICM into MIDI data thorugh approach 2

Fig.1, 2 and 3 show the progress of music transcription of ICM into MIDI data. Fig. 2a shows audio waveform. Fig. 2b shows pitch graph. Fig. 2c shows pitch graph in terms of MIDI numbers. MIDI note sequence for audio file in Fig.2 through approach 1 is 3C(34.8) 3D(23.2) 3F(696.6) 3G(127.7) 3F(441.2) 3D(1091.3). Here, 3C(50) means octave-3; MIDI note-C; note duration: 50 ms. MIDI note sequence for

audio file in Fig.3 through approach 2 is 3C(11.6) $3C^{\#}$(23.2) 3D(23.2) $3E^{b}$(34.8) 3E(58) 3F(696.6) $3F^{\#}$(69.7) 3G(127.7) $3F^{\#}$(127.7) 3F(313.5) 3E(23.2) 3F(116.1) 3E(116.1) $3E^{b}$(104.5) 3D(534.1). The audio file details are given in table 1.

Table 1. Details of audio files

Figure	Raga	Tonic Frequency	Audio file name	Male singer name	link
1	Alhaiya bilawal	155 Hz	03-KhayalinRaagAlhaiyaBilawal.mp3	RaghunandanPanshikar	www.cooltoad.com
2 and 3	Malkauns	146.8 Hz	Amit_Malkauns.mp3	Amit Mukherjee	http://paragchordia.com/data/GTraagDB/mp3/

5 Conclusion

Pitch estimation is carried out using Fourier of Fourier Transform with Parabolic Interpolation to spectral peaks. This reduces computational complexity to great extent. Automatic music transcription using approach 1 shows MIDI note sequence along with note duration for the notes which are there in the raga of audio sample. Whereas, in approach 2, MIDI note sequence is comprised of any note in the octave since here prior knowledge of Raga is not given. This note sequence can be further used for musical pattern recognition i.e. Raga identification. Raga identification can be treated as basis for music information retrieval of ICM and film songs based on ICM.

References

1. Sutton, C.: Transcription of vocal melodies in popular music. Report for the degree of MSc in Digital Music Processing at the University of London (2006)
2. de Cheveigné, A., Kawahara, H.: YIN, a fundamental frequency estimator for speech and music. J. Acoust. Soc. Am. 111(4), 1917–1930 (2002)
3. Maher, R.C., Beuchamp, J.: Fundamental frequency estimation of musical signals using a two-way mismatch procedure. J. Acoust. Soc. Am. 95(4), 2254–2263 (1994)
4. Slaney, M., Lyon, R.F.: A perceptual pitch detector. In: Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, pp. 357–360 (1990)
5. Chordia, P., Rae, A.: Raag recognition using pitch-class and pitch-class dyad distributions. In: Proceedings of the International Conference on Music Information Retrieval (2007)
6. Smith, J.O., Serra, X.: PARSHL: An Analysis/Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation. In: Proceedings of the 1987 International Computer Music Conference, pp. 290–297. International Computer Music Association, San Francisco (1987)
7. Akant, K.A., Pande, R., Limaye, S.S.: Accurate Monophonic Pitch Tracking Algorithm for QBH and Microtone Research. Pacific Journal of Science and Technology 11(2), 342–352 (2010)
8. Akant, K.A., Pande, R., Limaye, S.S.: Monophony/Polyphony Classification system Using Fourier of Fourier Transform. International Journal of Electronics Engineering 2(2), 299–303 (2010)