CMMR/FRSM 2011, Springer LNCS 7172, pp. 34-43, 2012.

METER DETECTION FROM AUDIO FOR INDIAN MUSIC

Sankalp Gulati, Vishweshwara Rao and Preeti Rao

Department of Electrical Engineering
Indian Institute of Technology Bombay, Mumbai 400076, India
sankalp.gulati@gmail.com, {vishu, prao}@ee.iitb.ac.in

Abstract. The meter of a musical excerpt provides high-level rhythmic information and is valuable in many music information retrieval tasks. We investigate the use of a computationally efficient approach to metrical analysis based on psycho-acoustically motivated decomposition of the audio signal. A two-stage comb filter-based approach, originally proposed for double/ triple meter estimation, is extended to a septuple meter (such as 7/8 time-signature) and its performance evaluated on a sizable Indian music database. We find that this system works well for Indian music and the distribution of musical stress/accents across a temporal grid can be utilized to obtain the metrical structure of audio automatically.

Keywords: Meter detection, Indian music, complex meter, comb filtering.

1 Introduction.

All music, across geographies and cultures, comprises of events occurring at regular time intervals. Meter is a hierarchical temporal framework consisting of pulses at different levels (time-scales), where pulses represent regularly occurring musical events [1]. Perception of meter is an innate cognitive ability in humans. Meter provides useful rhythmic information essential in understanding musical structure and is useful in various music retrieval tasks like similarity based music classification [2], beat tracking and tempo estimation of music [3]. In this study we investigate automatic meter detection for Indian music.

1.1 Previous work on meter detection

Considerable research has been directed towards extraction of low-level rhythmic information like onset detection and beat tracking [4]. However, less attention has been paid to higher-level metrical analysis. Most of the earlier work on meter analysis concentrated on symbolic data (MIDI). The system proposed by Goto and Muraoka[8] is considered as being the first to achieve a reasonable accuracy for the meter analysis task on audio signal. Their system was based on agent based architecture, tracking competing meter hypotheses and operated in real time. Meter detection requires tempo independent information about the rhythmic structure. And hence tempo normalization becomes a crucial stage in the meter detection system. In the approach proposed by Gouyon and Herrera [9] the beat indices are manually extracted and then an autocorrelation function, computed on chosen low level features (energy flux, spectral flatness, energy in upper half of the first bark band) is used to detect meter type. Also, in this approach the meter detection problem was simplified by restricting the result to double (2/4, 4/4) and triple (3/4, 6/8) meter. Metrical analysis of non-Western music using the scale transform for the tempo normalization is proposed by Holzapfel and Stylianou [2]. A more detailed description of previous work on meter analysis from audio can be found in [1].

1.2 Meter in Indian music

Meter, from a perspective of Indian music, is discussed in depth by Clayton[10]. Rhythmic organization in Indian Classical Music is described by the Tāl system [10]. Tal can be viewed as a hierarchical structure organized on three temporal levels, the smallest time unit 'matra', the section 'vibhag' and the complete rhythmic cycle 'avart'. Matra may be interpreted as the beat in most cases. Automatic metrical analysis from audio of Indian music is a relatively unexplored area despite the well established Tal framework of rhythmic organization. There are multiple Tals containing a given number of beats in a rhythmic cycle but which differ from each other in terms of sectional divisions and distribution of stressed/unstressed beats. In the current work we do not discriminate between the different possible sectional structures within a cycle but restrict ourselves to obtaining higher metrical level information by mapping the number of beats in a cycle to a meter type. This is similar to considering 3/4 and 6/8 metrical structure to both belong to triple meter [11].

In the current work, we implement the meter detection system proposed by Schuller, Eyben, and Rigoll[11] in which the tatum duration is extracted to establish the temporal grid on which metrical analysis is then implemented. Tatum can be defined as that regular time division which coincides most highly with all notes onsets [12]. This approach does not explicitly use any knowledge about the note onsets, beat positions or downbeat locations. We evaluate the above system on a previously used database of ballroom dance music and also a new database of Indian music. The latter, in addition to songs having double or triple meter, also includes songs in a complex meter, in this case septuple meter (7 beats in a cycle).

2 System implementation

The meter detection system is described in Figure 1. The method relies on finding the tatum duration and how well the integer multiple of this duration resonates with the sizable segment of the song. We follow the implementation procedure described in [11]. As can be seen in Figure 1 whole system can be divided into three stages. The implementation of each of these stages is described next.

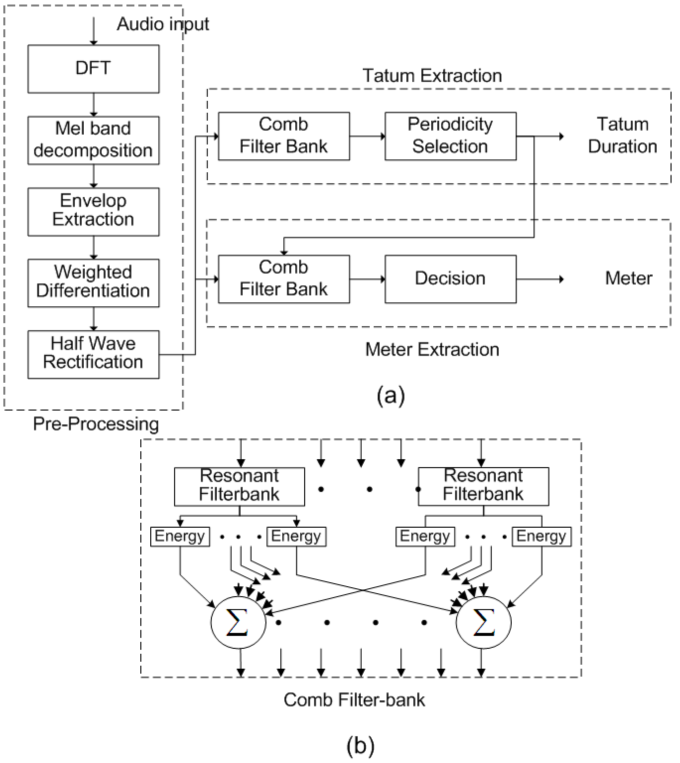

Fig. 1.(a) Block diagram of meter detection system, (b) Comb filter bank

2.1 Pre-processing

The input audio signal is down sampled to 16 kHz and converted to mono channel. The data is split into 32 ms frames with a hop size of 5 ms and corresponding frame rate of 200 Hz. A Hamming window is applied to each frame and a 512-point FFT is computed. By using 12 overlapping triangular filters, equidistant on the Mel-Frequency scale, these DFT frequency bins are reduced to 12 non-linear frequency bands. The band envelope is then converted to log scale (dB) and low pass filtered by convolving with a half-wave raised cosine filter of length 15 frames (75 ms). This filters out the noise and high frequencies in the envelope signal without diminishing the fast transient attacks. From this envelope a weighted differential dwtd is computed according to

$$d_{wtd}(i) = (o_i - \bar{o}_{i,l}) \cdot \bar{o}_{i,r} \quad (1) \qquad (1)$$

where, o_i is the sample at position (frame) i, $\bar{o}_{i,l}$ is the moving average over one window of 10 samples to the left of the sample and $\bar{o}_{i,r}$ of the window of 20 samples to the right of the sample o_i.

2.2 Tatum extraction

Tatum can be defined as that regular time division which coincides most highly with all notes onsets [12]. It is the lowest metrical level of the song. The tatum extraction method used by [11] uses a comb filter bank based approach, originally proposed by Scheirer[13]. The comb filter bank is implemented with delay varying from 0.1-0.6 sec consisting of 100 filters. The filter bank processes the extracted differential signal for each Mel band and the total energy over all bands of the output of each filter is computed. These values for each comb filter forms the tatum vector. The location of the maximum peak of this function is the delay corresponding to the tatum duration.

2.3 Meter extraction

The meter vector $\vec{m}$ is computed from the extracted differential signal by setting up narrow comb filter banks around integer multiples of tatum duration. The number of comb filters per filter bank is equal to twice the integer multiple of the tatum duration plus one to compensate for the round off factor of the tatum duration. For each filter bank that filter with the highest output energy is selected and the total energy of this filter over all Mel bands is taken as the salience value in the meter vector at the position of the integer multiple. In the current implementation multiples from 1-19 are considered. An example of meter vectors for different meters is shown in Figure 2.

$$S_2 = [\vec{m}(4) + \vec{m}(8) + \vec{m}(16)] \cdot \tfrac{1}{3} \tag{2}$$

$$S_3 = [\vec{m}(3) + \vec{m}(6) + \vec{m}(9) + \vec{m}(18)] \cdot \tfrac{1}{4} \tag{3}$$

$$S_7 = [\vec{m}(7) + \vec{m}(14)] \cdot \tfrac{1}{2} \tag{4}$$

The final meter value is determined from $\vec{m}$ using a simple rule based approach. For each possible meter i.e. double, triple and septuple, we calculate a salience value as in Eq. 2, 3, 4 respectively. The maximum of S2, S3, S7 determines the final meter of the song.

3 Experimental evaluation

3.1 Database

We have used two databases in the evaluation of the above system. The first is the well-known ballroom dance database containing 698 30-sec duration audio clips [14]. The audio is categorized by 8 different ballroom dance styles (Jive, Quickstep, Tango, Waltz, Viennese Waltz, Samba, Cha chacha and Rumba). Each of these styles belongs to either double or triple meter category. We have annotated them as such. The total duration of this database is 5 hrs 49 min.

The second database includes 620 30-sec duration audio clips from Indian film songs. Most of the songs from old Indian films tend to rigidly follow the tal framework and use mostly acoustic instruments whereas the songs from new movies also contain drum loops and electronic instrumentation. In this database we have included an equal number of popular songs from both old as well as new films. These audio clips belong to three different metrical structures most commonly found in Indian film music. 470 clips belong to double meter (4/4, 2/4 time signature), 109 triple meter (3/4, 6/8) and 41 follow septuple meter (7/8 time signature). The total duration of the database is 5 hrs 10 min. The ground truth meter values for the database have been annotated by the authors.

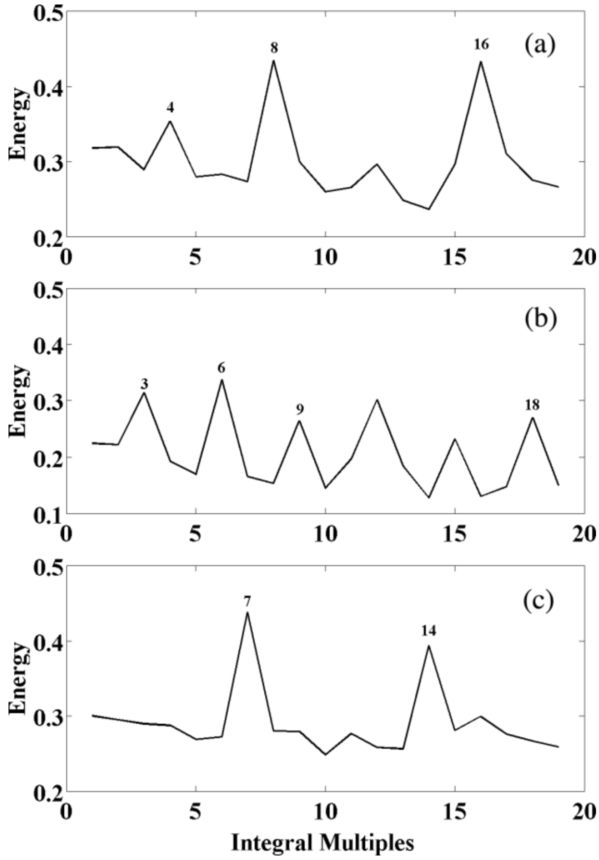

Fig. 2. Meter vector for (a) double meter, (b) triple meter and (c) septuple meter

3.2 Evaluation and results

The performance accuracy of the meter detection system for both databases Indian music database (IMDB) and ballroom dance database (BDDB) is summarized in Table 1 in the form of a confusion matrix. It is to be noted that although database 1 did not have any audio clips in the septuple meter category, this category was still included as a possible output of the meter detection system. Removing this category from the system naturally increases system accuracy for this dataset.

We note that although the overall accuracies for both datasets are quite high, the performance of the system for the triple meter for both datasets is quite low. The performance for the double meter, for both databases, and the complex meter (septuple), for database 2, are equally high. The overall accuracy for the meter extraction over both databases is 87.1%.

Table 1Confusionmatrixformeterdetection and performanceaccuraciesforboththedatabases BDDB and IMDB

Database	**Annotated Meter**	**Detected**			**Overall Accuracy (%)**
		Double	**Triple**	**Septuple**	
BDDB	Double	482	24	17	92.16
	Triple	33	116	26	66.29
	Septuple	-	-	-	-
	Total				85.67
IMDB	Double	443	25	2	94.26
	Triple	39	69	1	63.3
	Septuple	3	0	38	92.68
	Total				88.7

3.3 Discussion

As seen from Table 1 the maximum number of errors is encountered in the detection of triple meter with large confusion between triple and double meters in both datasets. An analysis of the misclassified cases revealed that for many songs the error was due to incorrect estimation of tatum duration. Such errors in tatum estimation for triple meter songs are more often found to occur in songs with fast tempo. Here periodicity at metrical levels higher than the tatum, such as half-rhythm cycle, fall within the search range of tatum delays (0.1-0.6 sec). Peaks in the tatum vector at such locations have saliences comparable to that at the true tatum duration leading to incorrect tatum detection. Thus it is mainly octave errors in tatum estimation that lead to incorrect meter detection.

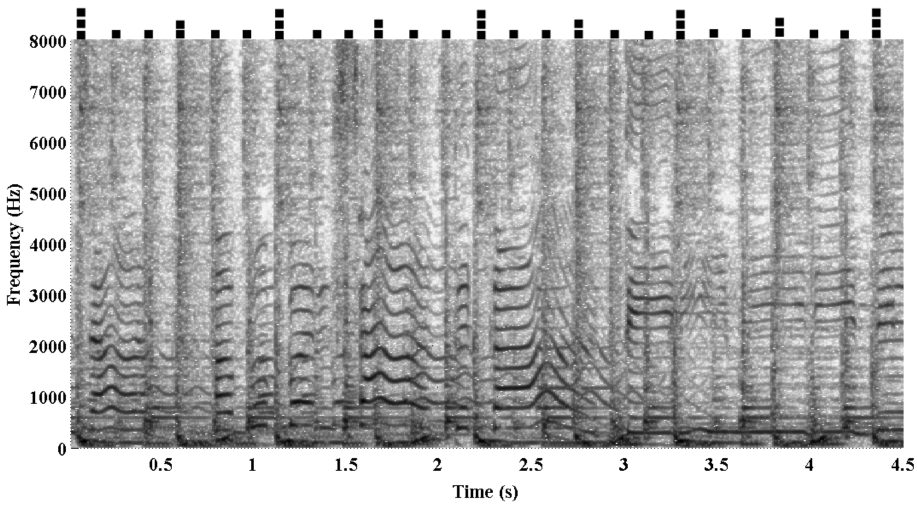

Fig. 3. Spectrogram of a song in triple meter. Rhythm cycle boundaries are indicated on top boundary by three boxes, half cycle by two and tatum boundaries by single box.

This phenomenon can be observed in Figure 3, which shows a spectrogram of a song in triple meter. In this figure we notice that the percussion strokes at periodicity of half rhythm cycle duration are more prominent than the ones at tatum duration. This prominence is also manifested in the comb filter output as high peak at half cycle duration, which for this song is within the search range for tatum duration (relatively fast song). Thus, we end up estimating incorrect tatum duration. If we make this kind of octave error in the tatum estimation, naturally we will get prominent peaks in meter vector at the multiples of 4, 8, 16 of this duration (which corresponds to integer multiples of rhythm cycle duration which is highly periodic.) and finally a wrong meter value for the song.

For the double meter songs also we encounter errors in tatum estimation, but for most of these cases even if we incorrectly estimate tatum duration as half or quarter rhythm cycle, we get correct meter value because the prominent strokes are indeed at the multiples of 4,8,16 of this duration (coinciding with rhythm cycle boundaries).

In the case of septuple meter, we found that the tatum estimation errors were very few. One reason for this is that these songs are typically of low tempo values, so less possibility of encountering half cycle duration within tatum search range. Secondly, as the rhythm cycle for the songs in septuple meter is such that it doesn't have any intra cycle repetition of sections (even in term of beat locations if not timbre), we don't get a prominent peak in comb filter output corresponding to any subsection duration of the rhythm cycle. Figure 4 shows the spectrogram of a song in septuple meter, we can notice that no intra cycle repetition exist (which is expected as 7 is a prime number).

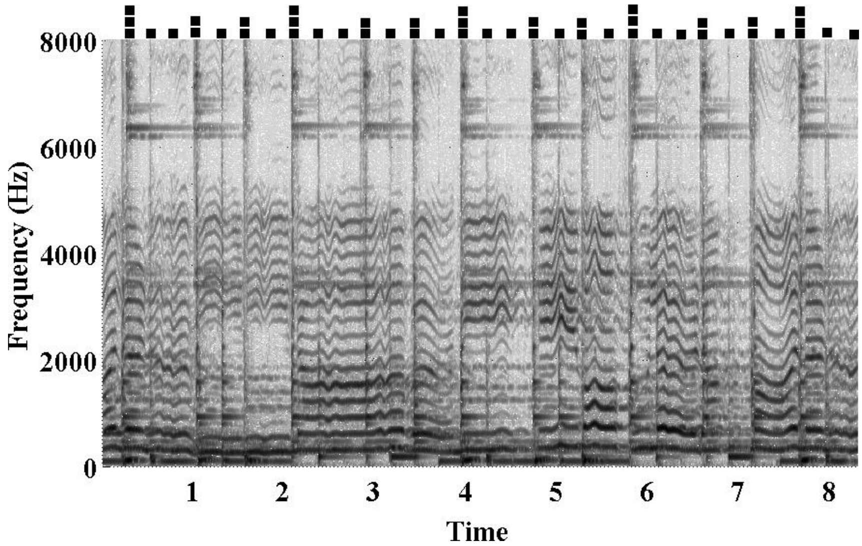

Fig. 4. Spectrogram of a song in septuple meter. Rhythm cycle boundaries are indicated by three boxes, sectional divisions by two boxes and tatum by single box. This rhythm structure is of type 3+2+2.

To verify that incorrect tatum estimation is the reason for erroneous meter estimation in most of the triple meter songs, we consider the same triple meter song as in Figure 3, for which meter was incorrectly estimated as double meter and manually correct the tatum duration value to the ground truth value. Figure 5.a. and Figure 5.b. display the meter vectors computed from an incorrectly estimated tatum duration value (half-cycle duration) and the true tatum duration respectively. Clearly the salience of the double meter (S2) is high for figure 5.a. and that of the triple meter (S3) is high for Figure 5.b.

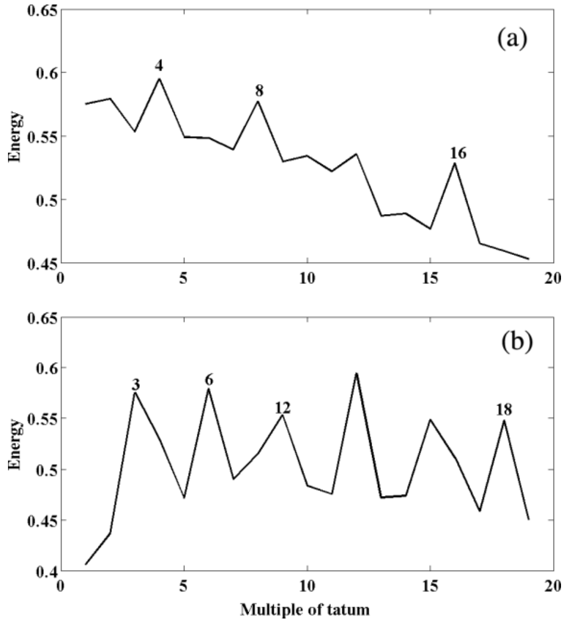

Fig. 5.Meter vector of a song from Indian music dataset in triple meterfrom (a) incorrectly estimated tatum (b) true tatum duration

4 Conclusion and future work

The high overall accuracy 87 % in meter classification reveals that this approach for accent extraction from audio is suitable for the automatic detection of meter for culturally distinct datasets, viz. Western ballroom dance music and Indian music. The approach, originally proposed for simple meters like double and triple, was successfully extended to determining complex meters like septuple. However confusions between triple and double meters are found to occur resulting from incorrect tatum duration estimation. Further work is needed to reduce these confusions. We also intend to test the above system on other complex meters like 5/8 and 9/8 and extend its application to the task of tal detection in Indian music where time signature information can be a useful input. The automatic segmentation of music by rhythmic structure is also worthy of future investigation.

References

[1] P. Klapuri, A. J. Eronen, and J. T. Astola, “Analysis of the meter of acoustic musical signals,” IEEE Transactions on Acoustics Speech and Signal Processing, 14(1):342–355, 2006.

[2] A. Holzapfel and Y. Stylianou “Rhythmic similarity in traditional turkish music,” in Proceedings of International Conference on Music Information Retrieval (ISMIR), 2009.

[3] S. Gulati, P. Rao, “Rhythm Pattern Representation for Tempo Detection in Music”, in proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia, Dec, 2010, Allahabad, India.

[4] S. Dixon, “Onset Detection Revisited,” in Proceedings of the International Conference on Digital Audio Effects (DAFx’06), Montreal, Canada, 2006.

[5] A. Klapuri, “Sound Onset Detection by Applying Psychoacoustic Knowledge,” in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, March 1999.

[6] S. Dixon, “Automatic extraction of tempo and beat from expressive performances,” J. New Music Res., vol. 30, no. 1, pp. 39–58, 2001.

[7] D. P. Ellis, “Beat tracking by dynamic programming” J. New Music Res., vol. 36, no. 1, pp. 51–60, 2007.

[8] M. Goto and Y. Muraoka, “Music Understanding At The Beat Level Real-time Beat Tracking For Audio Signals,” in Proceedings of IJCAI-95 Workshop on Computational Auditory Scene Analysis, page 6875, 1995.

[9] F. Gouyon and P. Herrera, “Determination of the meter of musical audio signals: Seeking recurrences in beat segment descriptors,” 114th Audio Engineering Society Convention, March 2003.

[10] M. Clayton: Time in Indian music: rhythm, metre, and form in North Indian rãg performance, Oxford University press Inc., New York, 2000.

[11] B. Schuller, F. Eyben, and G. Rigoll, “Fast and robust meter and tempo recognition for the automatic discrimination of ballroom dance styles,” in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’07), pp. 217–220, Honolulu, Hawaii, USA, April 2007.

[12] F. Gouyon, “A computational approach to rhythm description – Audio features for the computation of rhythm periodicity functions and their use in tempo induction and music content processing,” PhD dissertation, Music Technology Group, PompeuFabra University, 2005

[13] Scheirer, E. 1998. Tempo and beat analysis of acoustic musical signals. J.Acoust. Soc. Amer., vol. 103, no. 1, pp. 588–601.

[14] Gouyon, F., Dixon, S., Pampalk, E., Widmer, G. 2004. Evaluating rhythmic descriptors for musical genre classification, In Proceedings. AES 25th Int. Conf., New York, 2004, pp. 196–204.