International Journal of Current Engineering and Technology
ISSN 2277 – 4106

Available at http://inpressco.com/category/ijcet

Research Article

Music Structure Analysis of Hindustani Music for Transcription

T. P. Vinutha[a*]

[a]Shah & Anchor Kutchchi Engg College, Mumbai 400088

Abstract

Music is structured audio comprising the harmonious blend of melody and the rhythmic elements. Transcription of an audio signal involves the essential task of melodic structure analysis at the instances marked by the rhythmic events. Signal processing methods can be used to reveal the music structure of the audio signal using the descriptors of melody and rhythm chosen to describe the particular music tradition. In this work, challenges in developing required algorithms in the analysis of melodic and rhythmic structures of Hindustani music have been explored.

Keywords: *Melodic structure, rhythmic structure, pitch histogram, onset detection.*

1. Introduction

Music structure analysis represents one of the crucial decision making steps of the transcription system. Music transcription refers to the analysis of an acoustic music signal so as to represent the pitch, onset time and source of the signal. Music notation is any system that represents aurally perceived music, through the use of written symbols. In this work, the concepts of Hindustani music that have to be relied on to explain the perception of the underlying pitch contour of the melodic segments and the intricacies of the rhythm in a performance have been investigated.

The generally-accepted musical phrases (*pakad* and *chalan*) carry the flavour of the raga along with other attributes like *vadi-samvadi* (most frequently occurring *swara*) and *aaroh-avaroh* (S. Rao, et al, 1999). It is interesting to note that these attributes are clearly revealed by the underlying melodic structure even though two performances of the same raga or two performances of a raga by the same artiste in general differ in their appeal. A rich symbolic body of data exists for each raga of Hindustani music in the form of collection of notated vocal compositions called *bandishes*. Rhythmic pattern of any composition in Indian classical music is described by the *Taal* system. In a broader sense rhythm refers to all aspects of musical time patterns such as the way syllables of the lyrics are sung, the way the strokes of the instrumental music are played and the inherent tempo of the melodic piece. Understanding the rhythmic structure underlying the melody is essential in the tasks involving music structure analysis like various Music Information Retrieval (MIR) applications (M. Clayton, et al, 2000).

The music analysis work has been started with movie songs that are based on Hindustani classical music and then extended to popular *bandishes* (raga specific compositions). Melodic structures have been analysed on signal processing platform to validate the raga rules pertaining to the intonation of phrases. In some cases, *swaras* that are particularly tailored in the framework of *tala* have been analysed. Owing to the complexity of polyphonic audio consisting simultaneous presence of more than one musical sound source and the variability inherent in the Hindustani music performance, the structure analysis task is quite challenging. This work has been validated in consultation with persons having sound musical knowledge.

2. Musical structure of Hindustani classical performance

In a Hindustani classical performance, an artiste chooses a particular raga fitting the occasion and the audience and presents the raga within the framework of the *bandish* with the melodic improvisation adhering to the raga phraseology. Melodic mode of Indian music is embedded in the raga of an audio. Intonation of phrases and some semitones (*swaras*) in particular, along with the strength and duration of notes also form the attributes of raga in addition to *aaroh* or *avaroh*, *pakads* and *chalans*. Thus phrases can be considered as essential building blocks of any melody and most ragas can be adequately represented using up to 8 phrases.

The *pakad* or the characteristic phrase of a raga contains the smallest combination of notes by which a raga can be identified and *chalan* is the expansion of the *pakad*. The title phrase, *mukhda* of the *bandish* carries the flavour of the raga and occurs frequently in the performance. *Mukhda* is characterized by its text, as well as the melodic content and hence can be considered as the signature phrase of the raga rendition. It is interesting to note that all mukhda phrases occur at the *sam*, the first beat of the

*Correspoding author: **T. P. Vinutha**

rhythmic cycle or *tala* (J. Cheri Ross, et al, 2012). Rhythm in Hindustani music comprises *chand*, *tal* and *laya*. Chand refers to the syllabic pattern of the lyrics in vocal music, *tal* is the metrical pattern imposed by the accompanying percussion instrument and *laya* refers to the tempo or the rate of rhythmic events (rhythmic density) and under some contexts *laya* can mean the ratio of rhythmic density to tempo. Indian music has traditionally three main tempos or *laya*; *vilambit* (slow), *madhya* (medium) and *drut* (fast). In a performance, the listener perceives the surface rhythm but within the context of *tal* framework as the reference matrix, as the musician fills *matras* in *vilambit* and skips them in *drut laya* (M. Clayton, et al, 2000).

3. Signal processing in music structure analysis & implementation

Melody and rhythm form the basic elements of music representation. Signal processing methods can be applied to the audio to derive the features revealing the desired musical aspect of melody and rhythm. Melody of a song can be defined as the monophonic pitch sequence that a listener might reproduce while humming a segment of polyphonic music. The melodic pitch sequence is usually represented as the fundamental frequency, (F0) contour of the voice in the monophonic music or of the lead musical instrument in the polyphonic mixture (V. Rao, et al, 2009). Rhythm of an audio is inherently captured in the beat locations and the patterns of beats exhibited in the periodicities of onsets.

3.1 Pitch histogram based analysis of melody

Depending on the sequence of notes in the scale of the raga in Indian music, the artist chooses a certain position for the note to ensure consonance with the previous or next note. The hypothesis is that, the ragas sharing the same scale would differ to certain extent in different intonations of certain swaras of the raga to achieve their uniqueness. Thus along with the probability of occurrence of a swara, position of the pitch of each *swara* can also be explored for applications like raga recognition (S. Belle, et al, 2009). Pitch histogram is a statistical representation of the frequency of occurrence of each note in a musical piece. Features calculated from the histogram can be used in applications like raga recognition. In that sense, pitch histogram should be able to capture the unique melodic features of a particular raga (G. Tzanetakis, et al, 2003).

A. Analysis of melodic features of raag Marwa and raag Puriya

For the purpose of melodic analysis, vocal performances in *raag Marwa* and *raag Puriya* that share the same scale are selected. Both of these performances are in *tintal* having 16 *matras* distributed evenly in 4 cycles and rendered in *drut laya.*

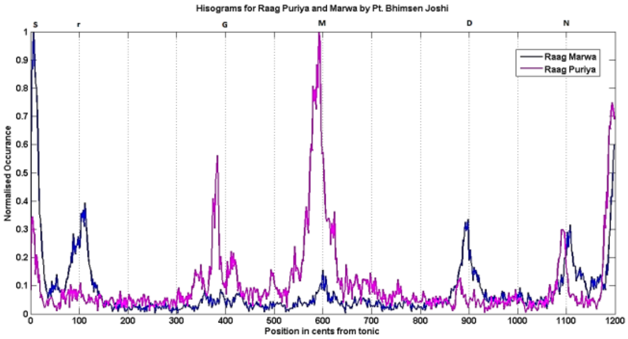

Figure 1 Pitch frequency histogram of a bandish in Raag Marwa and Raag Puriya

Selected audio segment is converted to mono and down sampled to 8 kHz. By identifying the vocal segments and employing the predominant F0 extraction algorithm, pitches were extracted every 10 ms. Re-synthesized pitch contour is listened for validation and any pitch tracking errors were corrected for the specific segments, by readjusting the parameters of the melody extractor. As pitch is the perceived frequency, it is expressed in logarithmic units. However, for musically relevant representation of pitch time series, logarithmic frequency scale is expressed in cents (1200 equal divisions of an octave).

Pitch Class Distribution (PCD), a data representation of pitches with 12 bins that represent the probability of occurrence of the 12 *swaras* over one octave is computed. Unlike western music that uses 12 bins per octave, Hindustani music requires fine binning of 10 cents leading to 120 bins per octave. This is the requirement even in Turkish music (A. C.Gedik, et al, 2010). Fig. 1 shows the PCD of performances in *Marwa* and *Puriya ragas.*

B. Theoretical concepts validated through pitch histogram

In *raag Marwa*, most frequently occurring *vadi swar* is r and *samvadi* swar is D. So, r should have highest occurrence than D. This can be seen in the plot for *Marwa*. Intonation of *swara* 'r', *komal* re is a feature by which a performance in *Marwa* can be distinguished from that of *Puriya*. Being a *vadi swar*, r is passed through frequently and also its duration can be longer in *Marwa*. However in *Puriya*, r should be *ati-komal* and it comes quickly with S. In the pitch histogram of Fig. 1, we can observe higher occurrence of r for *Marwa* than for *Puriya*. Also, the spread of r in *Puriya* is more towards S, unlike the spread of r in *Marwa,* which is distinct from that of S. The key phrase of *Puriya* is 'MGrs', hence G will have a tendency to shift from the mean position towards M. Here, this shift is 16 cents for the peak of G and hence can be justified as a valid shift.

3.2Melodic motifs of Hindustani music

Key phrases or the *pakads* of a *raga*, which give out the cue of the *raga* can be considered as motifs of the particular raga. *Pakads* of the raga and the intonation of the *swaras* of the *pakad* are strong indicators of the *raga*. Fig. 2 shows manually annotated pitch contour of the key phrase 'DMGr' of *raag Marwa*.

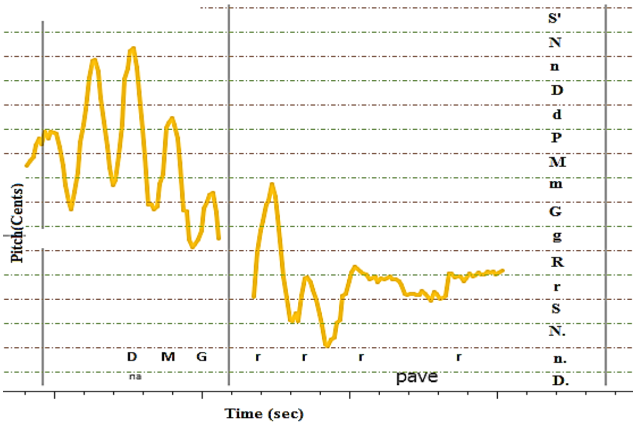

Figure 2 'DMGr' phrase in Gurubina, Marwa

Horizontal lines represent the *swara* locations of *raag Marwa* with respect to the tonic of the singer. This phrase could not be located in the entire audio of *raag Puriya*, even though they share the same scale. Fig. 3 shows the pitch contour of 'DMGM' phrase appearing in the clip of *raag Puriya*, which clearly indicates the different intonation for DMG.

3.3 Percussion Onset Detection

Hindustani classical music performances are accompanied by the percussion instrument *tabla*. Music is an event based phenomenon and beat location is represented by the onset of the *tabla* stroke. Percussion onset detection is an essential step in a number of musical analysis and musical information retrieval applications. Musical level annotation of one cycle of '*Gurubina'* is shown in the spectrogram of Fig. 4. Vertical lines in this spectrogram represent the *tabla* strokes, which are wideband events.
The standard procedure adopted by majority of the onset detection algorithms is to capture the beat location in the detection function, an intermediate signal that reflects in a simplified form, the local rhythmic structure of the original signal (J. P. Bello, et al, 2005).

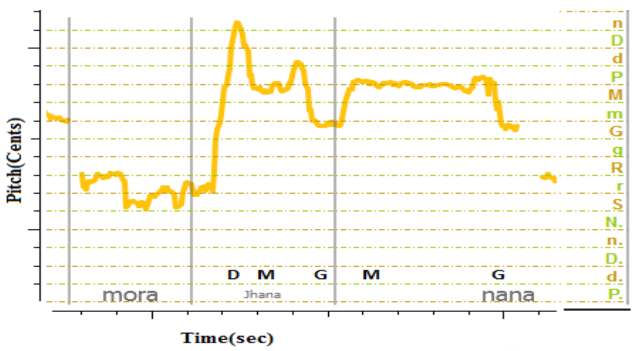

Figure 3DMGM' phrase in raag Puriya

Acoustic analysis is carried out by Short-term spectral analysis using 30 ms hamming window with a hop size of 15ms. The short-term DFT of the input sampled signal x[n] is represented by X[n,k], where the index k ranges from 1 to K with K representing fs/2 (half the sampling frequency).

A. Sub-band Spectral flux as detection function

Sub band spectral flux is a detection function that captures fairly well the *tabla* onsets and it represents the change in magnitude in the frequency bins, restricted to (5000, 8000Hz) range. Detection function, SF(n) represents the rectified spectral flux that is summed for all the bins.

$$SF(n) = \sum_{k=1}^{K} H\left[\left(|X(n,k)| - |X(n-1,k)|\right)W(k)\right] \qquad (1)$$

where, $H(x) = \frac{x+|x|}{2}$ is the half-wave rectifier function.

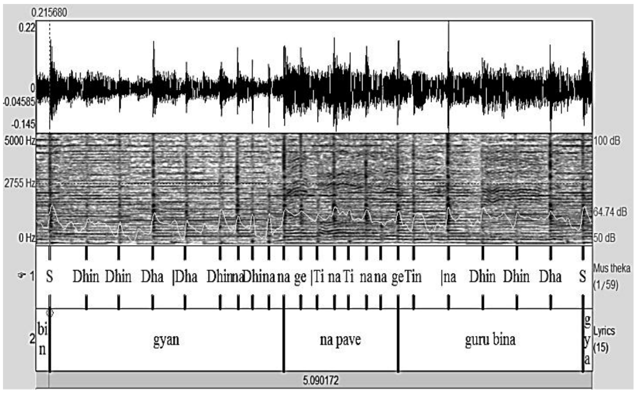

Figure. 4: Musical level annotation of 1 cycle of Gurubina bandish

Rectification has the effect of counting only the onsets that represent the increase in energy rather than offsets, and *W(k)* is the band limiting filter response with unity gain in the frequency region of 5000Hz to 8000 Hz, and falling off linearly to zero gain over a frequency region of 100 Hz in the transition band (S. Dixon, et al, 2006). Detection function, *d[n]* is normalized by subtracting the mean so that the average will be zero and dividing by the maximum absolute deviation so that the function will be in the interval [-1, 1]. A well defined threshold effectively separates the event-related and non event-related onsets. In fixed thresholding method, peaks where the detection function exceeds the threshold *δ(n),* a positive constant are considered as onsets. However, it does not consider the great dynamics common in a musical signal.

Hence, an adaptive threshold based on the local mean is implemented. Results are improved by peak picking, wherein detection function values which are local peaks are picked as final onsets (S. Dixon, et al, 2012). Fig. 5 shows the results of the algorithm for 3 cycles of Gurubina bandish. Musical annotation of one cycle of the audio clip has been marked in green and is treated as the ground truth. Here, 59 strokes have been marked as ground truth and the algorithm has resulted in 48 strokes as true positives, 18 as false peaks and 8 as false negatives.

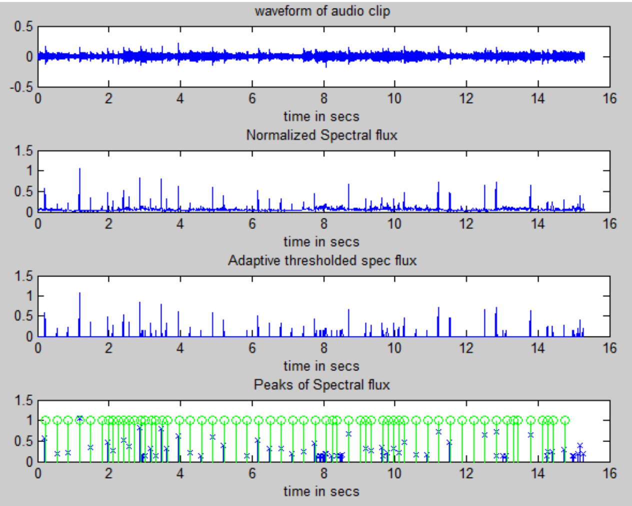

Figure. 5: Percussion onsets of 'Gurubina' in Marwa

4. Conclusion

Through the position of peaks and the spread of distribution corresponding to the swara in the pitch histogram, melodic features of ragas sharing the same scale could be revealed, so as to categorize the clip to a particular raga. Melodic contours of key phrases represented on the grid of swara locations clearly exhibit the intonation of swaras in particular and phrases in general, confirming to the raga rules. Sub-band spectral flux is a good representation of percussion onsets and these onsets can characterize the rhythmic pattern of an audio clip fairly well. Harmonious blending of melody and rhythm in Hindustani music is always very well exhibited in the mukhda phrase appearing on the sam.

Acknowledgment

The author wishes to thank Prof. PreetiRao, IITB, Mumbai, Dr. Uma Rao, SAKEC, Mumbai and Dr.Vishwas Kulkarni, B.A.R.C.

References

S. Rao, W. van der Meer and J. Harvey (1999), The Raga Guide: A Survey of 74 Hindustani Ragas, *Nimbus Records with the Rotterdam Conservatory of Music*.

M. Clayton (2000), Time in Indian Music : Rhythm , Metre and Form in North Indian Rag Performance, *Oxford University Press*.

J. Cheri Ross, Vinutha T.P. and P.Rao (2012),, Detecting melodic motifs from audio for Hindustani classical music, *Proc. of ISMIR* .

V. Rao, S. Pant, M. Bhaskar and P. Rao (Sept 2009), Applications of semi-automatic melody extraction interface for Indian music, in Proc. 12th *Intl. Conf. Digital Audio Effects* , Como, Italy.

S. Belle, R. Joshi and P. Rao (Dec 2009), Raga Identification by using Swara Intonation, *Journal of ITC Sangeet Research Academy*, vol. 23.

G. Tzanetakis, A. Ermolinskyi and P. Cook (2003), Pitch Histograms in Audio and Symbolic Music Information Retrieval, Journal of New Music Research Vol. 32, No. 2, pp. 143–152

A. C.Gedik , B. Bozkurt (2010), Pitch-frequency histogram-based music information retrieval for Turkish music, *Signal Processing* 9, 1049–1063.

J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler (2005) A Tutorial on Onset Detection in Music Signals, *IEEE transactions on Speech and Audio*, vol. 13, no. 5, pp. 1035–1047.

S. Dixon (2006) Onset Detection Revisited, In Proceedings of the International Conference on Digital Audio Effects (DAFx'06), Montreal, Canada.

C. Ros˜ao, R. Ribeiro, and D. Martins de Matos (2012) Influence of Peak selection methods on Onset detection, 13th *International Society for Music Information Retrieval Conference* , ISMIR.