Reliable Tempo Detection for Structural Segmentation in Sarod Concerts

T. P. Vinutha **Suryanarayana Sankagiri** **Preeti Rao**
Department of Electrical Engineering
Indian Institute of Technology Bombay, Mumbai 400076, India.
Email: {vinutha, prao}@ee.iitb.ac.in, surya.narayana@iitb.ac.in

***Abstract*—Tempo, or perceived speed, is an important characteristic of any music piece. In Hindustani instrumental music, different sections of a concert are characterised by the absence or presence of a regular rhythm and the speed of the music. In this paper, we analyse the tempo of concerts of a particular plucked-string instrument, the *sarod*. We first describe and evaluate different methods to detect plucks of *sarod* strings, as the timing pattern of the plucks, or note onsets, decides the tempo. We then use one such accurate method to track the tempo of a concert by analyzing the short-term periodicity of the onsets. Simultaneously, we also generate a measure of the salience of the periodicity to distinguish between rhythmic and non-rhythmic sections. We show that the tempo and salience vary significantly across sections in a concert, and thus, are valuable discriminative features in the task of automatic segmentation of *sarod* concerts.**

I. Introduction

Instrumental music holds an important position in the tradition of Indian classical music. Not only are a wide variety of instruments used, a single instrument can produce an astounding variety of sounds, making an instrumental concert a musically rich experience. The notes and the sequence in which they are played elaborate the *raga*, or melody, of the musical piece. In parallel, the temporal pattern or rhythm in which the notes are played also plays an important role in the musical experience. In Hindustani instrumental music, concerts follow the structure that originated in the ancient *Drupad* style and have the following sections-*alap*, *jod*, *jhala* and *gat*, in the order mentioned [1]. The first three segments are improvised where the instrumentalist elaborates upon the raga and plays solo. In the *gat*, the main instrument is accompanied by a percussion instrument, which distinguishes it from the others sections. The *alap* is not rhythmic; the intervals between notes are irregular and can vary greatly. A short string of notes played in quick succession is often followed by a long pause. The *jod* marks the beginning of the rhythmic section. As this section proceeds, the tempo gradually increases. The improvised section ends with the *jhala*, where the artist plays at an extremely rapid speed. Often, a sudden change in tempo marks the boundary between the *jod* and the *jhala* [1]. The rhythmic structure, in particular the tempo, is an indicator of the underlying segments in Hindustani instrumental concerts. Thus, the reliable detection of the time-varying tempo from the audio signal is crucial in segmenting the concert into the above sections [2].

In this paper, we focus on estimating the tempo in *sarod* concerts, which is a popular instrument used in Hindustani music. The *sarod*, like the *sitar*, is a plucked-string instrument with an elaborate structure. A sarod's strings can broadly be classified into two types-the playing strings, which are actually plucked, and sympathetic strings, which lie below the playing strings and resonate with the plucked strings to reverberate the sound. Among the playing strings, there are 4-5 melody strings (that are pressed at one end in order to produce different notes) as well as a couple of drone strings (tuned to low frequency notes), and chikari strings (tuned to high frequency notes) that are played as open strings. Unlike the *sitar*, the *sarod* is fretless, and thus an artist can produce a continuous glide in a note [3].

In a *sarod* recital, the tempo is set by the rate at which the strings are plucked. It is therefore essential to accurately detect the plucks (onsets of notes) of a string in order to estimate the tempo. A robust onset detection method is one that correctly detects and localises most of the onsets. Many of the existing onset detection techniques are based on simple musical intuition [4]–[6]. However, the challenge here is to choose a method that works across the variety of onsets observed in *sarod* playing. In our work, we highlight how certain characteristic features of the *sarod* play a role in selecting an onset detection method.

Having identified a good onset detection technique, the next task is to extract the tempo from the detected onsets. Tempo detection in music is a well-studied area [7]–[9]. In our work, we address some of the challenges in tempo detection that arise in *sarod* concerts. Firstly, there is considerable departure in local behaviour from exact periodicity in any real performance. Secondly, the tempo varies greatly over a concert (by a factor of three or more). The tempo detection algorithm must accurately match the perceived tempo, despite these challenges. We also present a measure of the periodicity present in the music signal in a small interval. This measure plays an important role in segmenting out the non-rhythmic portion (the *alap*) from the rhythmic portion.

The rest of the paper is organised as follows. In Section II, we present our approach to the task of tempo detection, that allows us to focus on the onset detection and tempo estimation individually. We then describe the onset detection and tempo estimation algorithms that are used in this paper. In Section III, we discuss different types of plucks in a *sarod* concert and analyse the effectiveness of different onset detection methods in detecting these plucks. We then perform a comprehensive evaluation of these methods by testing their accuracy on short concert segments. Section IV goes on to describe the effectiveness of different tempo detection schemes in tracking the tempo in a concert. We also analyse the validity of our salience measure. In Section V, we present the results of our analysis on a database of 15 concerts, illustrating the point that the tempo and the salience of periodicity are indeed distinguishing features of the different segments of a sarod concert. In Section VI, we discuss the results and suggest possible directions for future work.

II. TEMPO DETECTION IN MUSIC

In this work, we follow the generic framework of tempo detection as presented in [7]. The task of tempo detection is split into two discrete stages (Fig 2.1). The first stage involves computing an onset detection function (ODF) which provides peaks localised at time instants corresponding to onsets of musical events such as notes in the music signal. A good ODF should have prominent peaks coinciding with note onsets, and at no other time instants. Having computed the ODF, the task of tempo detection reduces to finding the periodicity in the peaks of the ODF. A local tempo estimate is obtained based on the short-time periodicity of the ODF, and is computed over the concert duration. Significant changes in the local tempo would indicate the underlying structural boundaries in the audio.

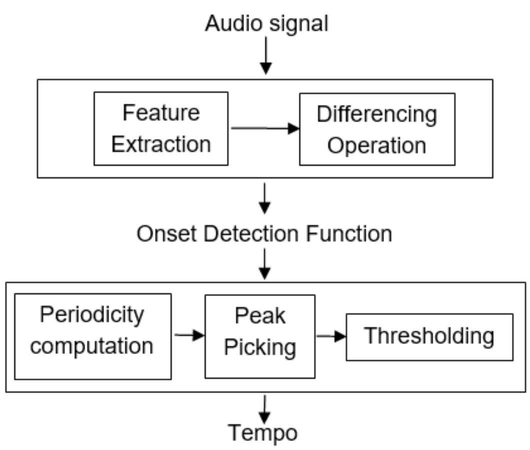

Fig 2.1: Block-diagram of tempo detection system

A. Onset Detection Function

Onsets of musical events are typically marked by an abrupt change in one or more acoustic parameters of the signal. A time-frequency representation such as the spectrogram can help to achieve the time localisation of the possibly frequency dependent changes. Hence, the first step in deriving an onset detection function is to compute the music signal's short-time Fourier transform, $X[n,k]$ and obtain the spectrogram [5]. While other methods of onset detection exist, as mentioned in [4], [6], the spectrogram offers a powerful representation to develop appropriate onset detection methods. In our work, we choose a window length of 40 ms and a hop of 10 ms between successive windows. For simplicity, we deal only with the magnitude of the spectrum, although there are methods that take into account the phase as well that will be explored in future work [4], [5].

The task now is to detect a 'novelty' (clear change) in a certain musical feature, such as the amplitude or pitch, that occurs at the onset of a pluck. Dixon [5] elaborates on onset detection functions that are based on the amplitude of the signal, while Hainsworth [6] describes methods that captures harmonic changes. We first describe the different methods to calculate the ODFs that we have used, before analysing their relevance to *sarod* concerts in Section III.

1) Spectral flux method: In plucked string instruments (indeed, for many others), the amplitude of the music signal increases whenever there is a note onset. This is natural, as by playing a note, the player imparts some energy into the music signal, which otherwise is produced by the natural oscillations of the string. An increase in amplitude of the music signal would lead to an increase in the magnitude of the spectral components. We define the spectral amplitude function (SA) as the sum of the spectral bins at each instant in time, and look for points where it rises sharply. To obtain peaks localised at these points, it is natural to take the derivative of the SA. One option is to consider the spectral flux, which is the first-order difference in the spectral amplitude, as the onset detection function [4]. However, random fluctuations in the SA would introduce many spurious peaks in the ODF. In order to mask impulsive disturbances, we take the derivative that involves smoothing via a bi-phasic filter $h[n]$, identical to the one used in [10], [11].

$$SA[n] = \sum_{k=0}^{N/2-1} |X[n,k]| \quad ; \quad ODF_{SF}[n] = SA[n] * h[n]$$

2) KL divergence method: In this method, the magnitude spectrum at each time slice is normalised and treated as a probability distribution over frequency. To measure the change of the spectrum with time, we compute the Kullback-Leibler (KL) divergence between successive time slices [6]. The onset detection function is simply the KL divergence measure varied with time. A change in note, cause by a pluck, is accompanied by a sudden change in the spectrum shape, resulting in a spike in the KL divergence. After the pluck, the normalised spectrum remains steady and the KL divergence is close to zero. In contrast to the spectral flux method, the KL divergence is sensitive to the distribution of the spectrum across frequency bins (and therefore, the pitch), and is less sensitive to overall amplitude changes of the signal. To smoothen out short range fluctuations, we take the average of five adjacent time slices in calculating the probability vectors [6]. This method is concisely captured in the following equations:

$$\widetilde{X}[n,k] = \sum_{i=n}^{n+4} |X[i,k]| \quad ; \quad \widetilde{X}[n,k] = \widetilde{X}[n,k] / \sum_{k=0}^{N/2-1} \widetilde{X}[n,k]$$

$$ODF_{KL}[n] = \sum_{k=0}^{N/2-1} \widetilde{X}[n+1,k] log \left(\frac{\widetilde{X}[n+1,k]}{\widetilde{X}[n,k]} \right)$$

3) Euclidean distance method: Another method of computing the onset detection function involves treating the magnitude spectrum at each time slice as a vector in an $N/2$-dimensional vector space. To detect the change of the spectral features with time, the Euclidean distance between adjacent time frames is calculated [6]. A large distance indicates some change in the spectrum, whether it be in the amplitude or pitch. In the case of *sarod*, the Euclidean distance function rises sharply at the onset of a pluck and remains large for a considerable time until steady state is reached, as successive frames are quite different from each other. Therefore it has broad peaks. In order to refine the peak location, we compute its derivative, which peaks only at the location where it rises sharply. This is done using the bi-phasic filter, as before:

$$ED[n] = \left(\sum_{k=0}^{N/2-1} (|X[n+1,k]| - |X[n,k]|)^2 \right)^{0.5}$$

$$ODF_{ED}[n] = ED[n] * h[n]$$

B. Tempo Extraction

One way of estimating the local tempo is to explicitly mark note onsets from the onset detection function, and compute and inter-onset interval histogram [12]. However, this would involve peak-picking and thresholding [4], and is prone to errors if there are peaks of different magnitudes in the ODF. A robust way of capturing the local periodicity within the ODF is by computing its auto-correlation function (ACF) [13]. This is because a periodic function has significant overlap with a time-shifted version of itself when shifted by any multiple of its time period. A second method to extract the local tempo is to compute the DFT of the ODF [13]. Both the methods are particularly well-suited for tempo detection in *sarod* as they bring out the tempo even if there are small pauses that interrupt the regular plucking of the strings.

In both the computations, the ODF is analysed only in a short window, in order to extract the local tempo. The choice of window duration is a crucial decision to be made in any specific problem context. The trade-off involved is between sufficient averaging across the underlying repeating cycles and the ability to track genuine temporal changes in rhythm structure. In our case, we find a window length of 10 s to be optimal. By shifting the analysis window by 1 s each time, we obtain the tempo as a function of time, which we prescribe in the unit beats-per-minute (bpm). Below, we specify the algorithms we use to evaluate the tempo from the ACF and the DFT.

1) Combined DFT-Frequency mapped (FM) ACF method: Ideally, the peaks of the ACF lie at multiples of the tempo period of the ODF, while the peaks of the DFT lie at multiples of the tempo frequency. Thus both methods are vulnerable to octave errors in the estimated tempo. To take advantage of the different natures of octave errors of the two methods, Peeters [7] suggests an effective combination. If we were to plot the ACF on a frequency scale, the peaks of the ACF would lie at fractions of the tempo frequency. By overlapping the frequency mapped ODF with the DFT, the peak at the tempo frequency becomes most prominent and provides a relatively octave-error free estimate of the tempo.

2) GCD of ACF peaks method: Although it is difficult to find an appropriate global threshold to identify the first peak of the ACF, we observed that one could set a threshold above which most peaks of the normalised ACF would lie. The inter-peak intervals would be a multiple of the tempo-period. Computing the tempo-period is thus a simple matter of computing the GCD of the inter-peak intervals. We obtained the intervals between the first peaks (above a certain threshold) from each ACF vector, and computed the best estimate of the GCD of these numbers. The restriction to five peaks was done to maintain uniformity in tempo computation across different sections with varying tempo and therefore with different number of peaks in the ACF within the same time lag.

C. Salience

As the *alap* section is devoid of regular pulsation (that characterises all the other sections), it is unreasonable to assign a tempo to it. Therefore, the tempo does not serve as an indicator of the *alap-jod* boundary. Rather, a measure of the degree of periodicity, or salience of periodicity, would serve to discriminate the *alap*. A good salience measure would be one that remains low for the *alap* and high for the rhythmic sections. In this work, we develop our own salience measure. First, we make the observation that a periodic ODF would result in a periodic ACF with prominent peaks. To measure the strength of the periodicity, we add five different time-compressed versions of the ACF and choose the value of the largest peak in this function as the salience. This procedure is similar to that of Harmonic Product Spectrum used for pitch detection [14]. The salience obtained in this manner would be high only if the ACF were periodic.

III. Detecting Onsets of Notes in a sarod

In this section, we present an analysis of the three different onset detection methods, reviewed in Section 2, when applied to *sarod* concerts. We found that the spectral flux method can be successfully adopted for detecting onsets on the *sarod*, as string plucks are inevitably accompanied by a rise in amplitude. The KL divergence method, one that is sensitive to changes in the spectral distribution, also showed peaks at onsets. Fig 3.1 shows the ODFs obtained by these methods for a six second excerpt of the *alap* of *sarod* concert. The ground-truth onsets (those perceived by the authors) are marked in green. The spectrogram of the section is shown to illustrate the varying acoustics of the sarod plucks.

The spectral flux method, although a very common method for onset detection, does not perform well in some scenarios for reasons specific to the *sarod*'s playing style. The sympathetic strings in a *sarod* make it possible for the sound to reverberate after a pluck, making the amplitude rise well after the onset of the pluck. In such cases (observed often in the *alap* section), this method would yield a spurious onset (e.g. at time 4.5 s in Fig 3.1). Also, in expressive playing, the strings are plucked gently, and the spectral flux ODF might not highlight such onsets. Often, the artist alternates between a strong melody pluck and a mild *chikari* pluck in order to maintain the tempo as well as to provide a tonal base. It is therefore crucial to detect mild plucks well in order to estimate the tempo. A comparison of the two methods shows that mild onsets, such as the first one in Fig 3.1, are picked more prominently by the KL divergence ODF than the spectral flux ODF. Moreover, the KL divergence method is not sensitive to amplitude changes that may occur after a single pluck, a desirable feature.

It is evident from Fig 3.1 that the KL divergence method too has certain pitfalls. Firstly, small variations in the pitch following a single pluck can lead to redistribution of weights among neighbouring frequency bins, making the KL divergence yield many spurious peaks. As the *sarod* is a fretless

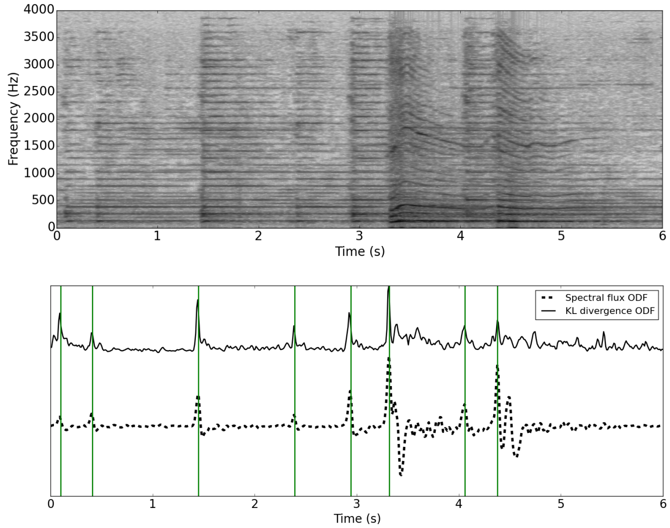

Fig 3.1 Spectrogram (top) and ODFs (bottom) of a *sarod* *alap* excerpt. Vertical lines indicate true onsets

instrument, an artist can vary the pitch of a struck note by varying the length of the playing string. Such an event can also cause spurious peaks. This is seen in the region around 3.5 s in Fig 3.1.

Secondly, in a concert, the artist may play the same note in succession, multiple times. This is often the norm in the *jhala* section of a concert, where the player continuously strums the *chikari* strings to keep the rhythm, in between melodic phrases (where different notes, corresponding to the *raga*, are played). In such a scenario, it is the amplitude, not the pitch, that is the main indicator of an onset and the KL divergence fails to detect the onsets accurately. The Euclidean distance method, which takes into account the amplitude as well as the pitch, turns out to be a better method in this scenario.

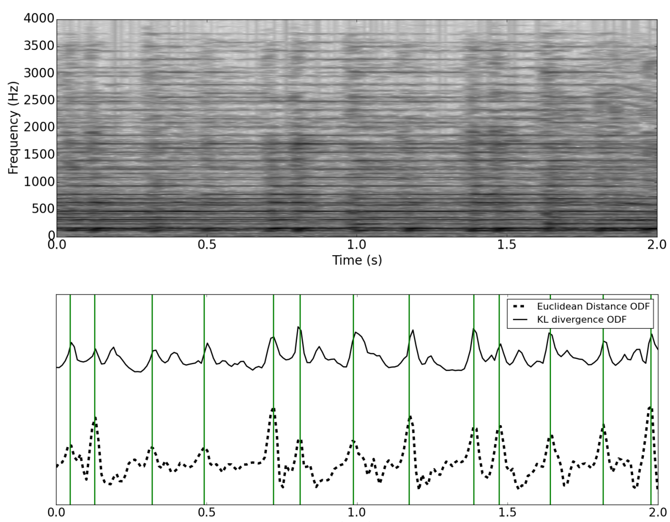

Fig 3.2 Spectrogram (top) and ODFs (bottom) of a *sarod* *jhala* excerpt. Vertical lines indicate true onsets

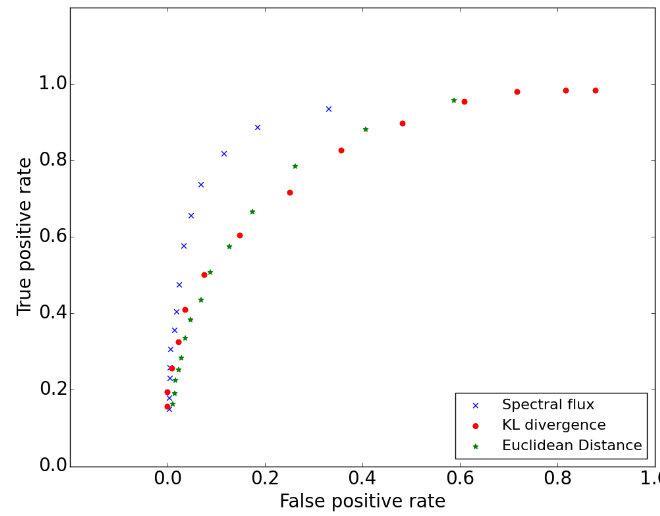

Fig 3.3: ROCs of different onset detection methods

This point is illustrated in Fig 3.2, which shows the ODFs of these two methods (again, normalized for comparison) from a 2 second excerpt from a *jhala* section. The ODF from the KL divergence method has peaks corresponding to onsets which are not particularly prominent compared to the other random variations. In comparison, the Euclidean distance ODF shows more prominent peaks at the desired locations. The Euclidean distance method too is not perfect. It yields spurious peaks when the amplitude rises well after an onset (like the spectral amplitude method), or when the pitch changes, unaccompanied by an onset (like the KL divergence method).

In order to test the suitability of the different methods, each of them were evaluated on small clips of *sarod* concerts. We chose a section each of the *alap*, the *jod* and the *jhala* from four concerts by different artists. The true onsets were manually marked by the authors, with the aid of the spectrogram. More than 400 onsets were marked across all segments. The onsets captured in these clips are of a diverse variety, and include the types of onsets described above.

To identify the detected onsets, each onset detection function was first normalised and thresholded. For different threshold values, the locations of peaks of the ODF lying above the threshold were compared with the true onsets. Detected onsets lying within a margin of 50 ms from a true onset were labeled as correct onsets, and the rest were labeled as false onsets. For each ODF, a higher threshold gave both a lower number of correct onsets as well as a lower number of false onsets. The percentage of correct onsets among existing onsets was plotted against the percentage of false onsets among detected onsets. The data across all twelve segments is shown in Fig 3.3, in an ROC plot, as done in [4].

The plot clearly shows that the spectral flux method works best overall. This is primarily because the ODFs obtained from the other methods have a much larger number of spurious peaks. Also, since we use the ACF to compute the tempo, the effect of a few missed mild onsets is not felt. Therefore, we choose the spectral flux method for detecting onsets, and use it to compute the tempo.

IV. Periodicity Detection and Tempo Determination

To test the accuracy of the tempo calculated by different algorithms, we need a reference value of the tempo for comparison. Moreover, we also need to know how well our measure of salience distinguishes the rhythmic and the non-rhythmic portions. The rhythmogram, which is a spectrogram-like representation of the ACF with time on the horizontal axis and time-lag on the vertical axis, is a powerful visual tool that captures the variation of the tempo with time. Fig 4.1 shows the rhythmogram, tempo and the salience of a complete *sarod* concert with the boundaries of segments manually marked.

In analysing many *sarod* concerts, we noticed that the *jod* section often has a sudden change in tempo. As we are interested in tracking the tempo and identifying boundaries based on tempo changes, we split the *jod* section into two parts, *jod-1* and *jod-2*, by placing a boundary wherever there is an abrupt and significant change in the tempo. Therefore, Fig 4.1 shows four sections. This split allows us to track the tempo variation across sections better, as shown in Section V.

The tempo and the salience can be compared with the visual cues available in the rhythmogram to obtain some measure of the accuracy of the tempo and salience measures. The striations of the rhythmogram mark the peaks of the ACF, and the distance between them is the period of the music signal.

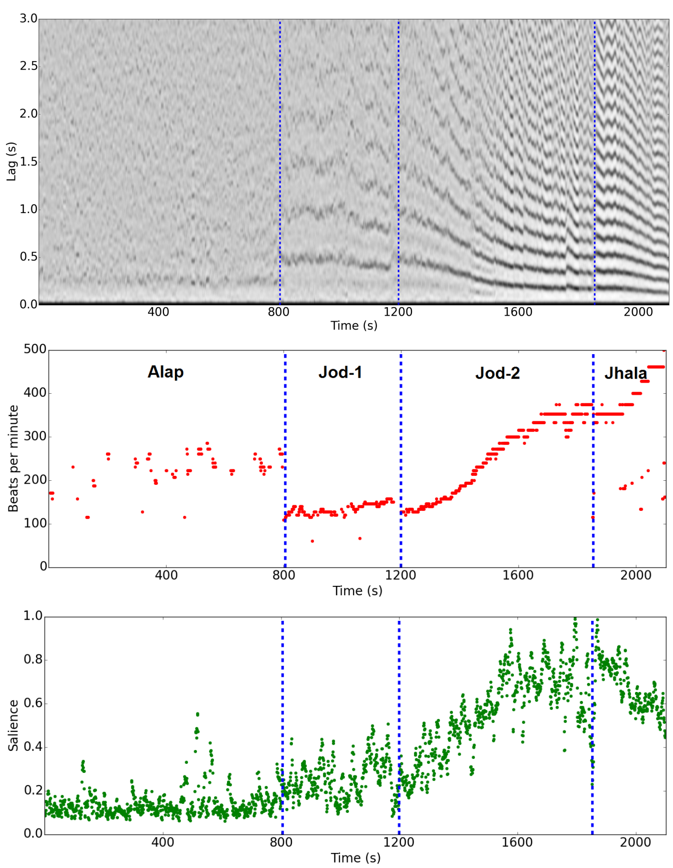

Fig 4.1 Rhythmogram, tempo and salience of a typical *sarod* concert. Tempo obtained from DFT-FM ACF method

The increasing tempo of the concert is seen by the continuous reduction in the distance between the striations. Moreover, the non-rhythmic portion is evident by the absence of striations. We find that the tempo frequency (in bpm), obtained from the DFT-FM ACF method as described in Section II, follows the inverse trend of the rhythmogram striations. Having calculated the tempo, we also evaluate the value of the ACF with the tempo lag as the argument, and ascribe that as a confidence measure of the tempo. Tempo values of low confidence, which occur when the tempo doesn't coincide with any ACF peak (ACF value < 0.2), are omitted from consideration. The tempo trend obtained from the other method is extremely similar to that shown here, and is therefore not plotted. We also find that the salience remains low for the entire duration of the *alap*, and rises quickly once the rhythmic section begins. Thus, both the tempo detection algorithms and the salience measure appear to be reliable.

V. Dataset Description and Experimental Evaluation

To verify the accuracy of our tempo and salience detection algorithms, we check their consistency with our expectations. In this section, we present the results of analysing several *sarod* concerts, each with a *alap-jod-jhala* structure. Our database comprised of 15 concerts by 6 different *sarod* maestros, with different playing styles. The average length of each piece was about 30 minutes, and the total time of all the concerts was 432 minutes.

Each concert was manually segmented into four sections, *alap-jod1-jod2-jhala*, based on obvious changes in the perceived tempo. In order to combine data for each section across the concerts, the tempo values within each concert are normalised to the (0,1) range by using concert minimum and maximum tempi (determined manually). Because of inaccuracies in automatic tempo estimation, some tempo values lie outside this range too. The distribution of normalised tempo values of the sections *jod-1*, *jod-2* and *jhala* from all the concerts are concisely represented in Fig 5.1. The tempo calculations are from the DFT-FM ACF method (method 1), and the GCD of ACF peaks method (method 2). Both the boxplots show an increasing trend in the tempo, from *jod-1* to *jod-2* to the *jhala*. The normalised salience across all concerts is also plotted as a boxplot in Fig 5.2.

A good tempo detection method would result in a small spread of data-points within a section, and a clear separation between different sections, and can be seen from the boxplots. A one-way ANOVA test was conducted to compare the effectiveness of the algorithm in segregating the returned tempo values to sections of the audio. The analysis confirmed that the distinction in tempo across sections is statistically significant, as F-values for the *jod-1*−*jod-2* boundary from methods 1 and 2 are 3100 and 2048 respectively, while those for the *jod-2*−*jhala* boundary are 250 and 407 respectively ($p < 0.0001$ in all cases). Thus, both the methods give us accurate tempo values which allow us to clearly notice the section boundaries. We also see that method 1 works better for a slower tempo, and method 2 is superior for a faster tempo. True to our expectations, we find that the salience in the *alap* section is much lower than the other sections, which are rhythmic (F-value = 14779, p = 0).

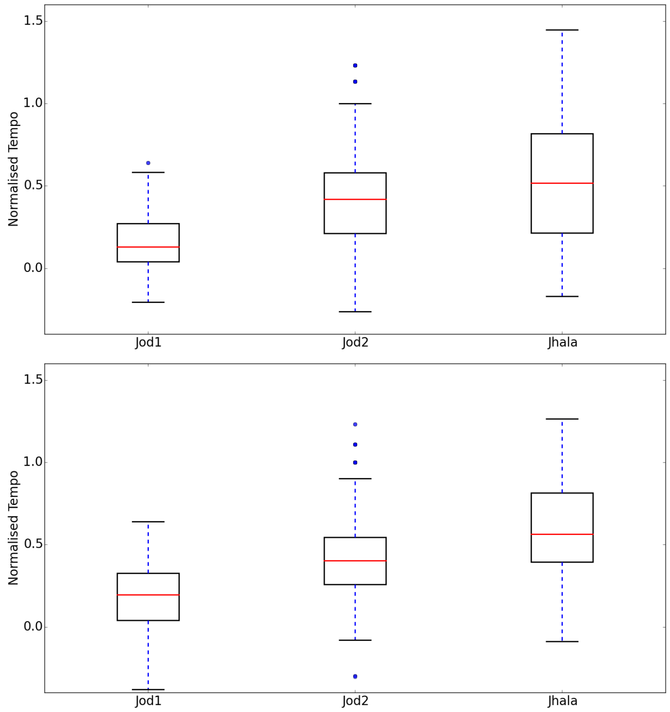

Fig 5.1 Tempo distribution obtained from method 1 [top] method 2 [bottom] for different rhythmic sections

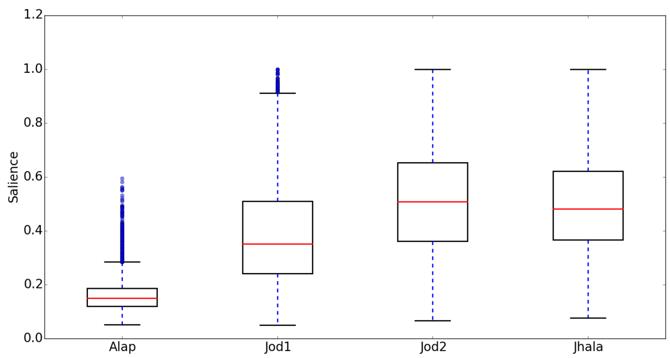

Fig 5.2 Salience distribution for different sections

VI. Conclusion and Future Work

In this work, we aimed to come up with an accurate method of detecting the tempo in *sarod* concerts. We chose the approach of first computing an onset detection function from the spectral changes effected by string plucks, and subsequently extracting the tempo, based on the periodicity of onsets from this function. We analysed three onset detection techniques that rely on the magnitude spectrum and understood the merits and demerits of each, with regard to detecting onsets in *sarod* concerts. In future, by including the information from the phase as well, we hope to improve the accuracy of detecting the variety of pluck onsets encountered in *sarod* concerts. In principle, one can combine multiple methods, e.g., spurious peaks in the spectral flux ODF can be ignored if they are unaccompanied by a sudden change in phase. We expect that a similar approach will be applicable to other plucked string instruments as well. However, we need to study and optimise the parameters involved in the feature extraction step of onset detection, as these are instrument specific.

Our method of tempo detection involved computing the ACF of the onset detection function, without any processing. We observed that our method often fails to detect the correct tempo in the *jhala* section. In this section, many mild strokes are interspersed with strong strokes. Therefore, even though they are regularly spaced, the onsets do not contribute equally to the ACF. If the pattern of mild strokes are and strong strokes repeat at some period, the ACF turns out to be periodic with this false larger period. This results in a low tempo calculation. By detecting peaks and replacing them with uniformly sized peaks, one can hope to get a better ACF. Finally, we need to incorporate further features indicative of other characteristics of the concert section, followed by statistical methods for segment boundary detection, as done in [2].

Acknowledgment

This work received partial funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 267583 (CompMusic).

References

[1] M. Clayton, *Time in Indian Music: Rhythm, Metre, and Form in North Indian Rag Performance*. Oxford University Press, 2001, ch. 7: Performance practice and rhythm.

[2] P. Verma, T. P. Vinutha, P. Pandit, and P. Rao, "Structural segmentation of hindustani concert audio with posterior features," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 136–140.

[3] B. C. Wade, *Music in India: The classical traditions*. Manohar Publishers, 2001, ch. 4: Melodic Instruments.

[4] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1035–1047, 2005.

[5] S. Dixon, "Onset detection revisited," in *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, 2006, pp. 133–137.

[6] S. Hainsworth and M. Macleod, "Onset detection in musical audio signals," in *Proc. Int. Computer Music Conference*, 2003, pp. 163–6.

[7] G. Peeters, "Template-based estimation of time-varying tempo," *EURASIP Journal on Applied Signal Processing*, vol. 2007, no. 1, pp. 158–171.

[8] K. Jensen, J. Xu, and M. Zachariasen, "Rhythm-based segmentation of popular chinese music," in *The International Society of Music Information Retrieval*, 2005, pp. 374–380.

[9] S. Gulati and P. Rao, "Rhythm pattern representations for tempo detection in music," in *Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia*, 2010, pp. 241–244.

[10] P. Pradeep Kumar, P. Rao, and S. D. Roy, "Note onset detection in natural humming," in *Conference on Computational Intelligence and Multimedia Applications, 2007*, vol. 4, 2007, pp. 176–180.

[11] D. J. Hermes, "Vowel-onset detection," *The Journal of the Acoustical Society of America*, vol. 87, no. 2, pp. 866–873, 1990.

[12] S. Dixon, "Automatic extraction of tempo and beat from expressive performances," *Journal of New Music Research*, vol. 30, no. 1, pp. 39–58, 2001.

[13] P. Grosche and M. Müller, "Extracting predominant local pulse information from music recordings," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1688–1701, 2011.

[14] D. J. Hermes, "Measurement of pitch by subharmonic summation," *The journal of the Acoustical Society of America*, vol. 83, no. 1, pp. 257–264, 1988.