To appear in Proc. of Frontiers of Research on Speech and Music (FRSM)
15-16 December 2017, NIT Rourkela, Orissa, India.

DISCRIMINATION OF SITAR AND TABLA STROKES IN INSTRUMENTAL CONCERTS USING SPECTRAL FEATURES

Dhanvini Gudi, Vinutha T.P. and Preeti Rao
Department of Electrical Engineering
Indian Institute of Technology Bombay, Mumbai 400076, India
{vinutha, prao}@ee.iitb.ac.in

Abstract

A Hindustani instrumental gat is characterized by the interplay between the solo melodic and the percussive instrument. Previous work has shown that the structural segmentation and resulting visualization of the concert in terms of musically meaningful sections is possible by the estimation of the stroke densities of the individual instruments, as these densities represent the inter-related tempi that evolve with the progression of the concert. Motivated by the need for the estimation of individual instrument tempo, we propose spectral features to discriminate between sitar and tabla strokes in performance recordings. The resulting classification method is tested on a concert dataset and the performance is discussed. Finally, the usefulness of the methods is demonstrated via the analysis of a complete sitar gat where the different concert sections are observed to be clearly distinguished via the rhythmic patterns computed on the segregated strokes.

1. Introduction

A typical Hindustani classical music concert revolves around a main vocal or instrumental lead, with an accompanying percussion instrument. The focus of this paper is the discrimination of sitar and tabla strokes to facilitate application of signal processing and classification techniques to reveal the structure of Hindustani instrumental concerts. This can be attempted by segmentation of typical concert into sections, on the basis of identification of strokes of each instrument played in a typical polyphonic concert, which in the case of sitar or sarod concerts, calls for a method to distinguish between the melodic (sitar/sarod) strokes and the percussive or tabla strokes.

Previous work [1] involves identification of the onsets of all strokes (sitar and tabla) and picking up the tabla strokes based on the onset signal characteristics to obtain rhythmic patterns in the form of rhythmograms to visually demarcate segments of typical sitar gat into sections such as *alap*, *vistar*, *layakari* and *tabla-solos*. Traditional stroke discrimination techniques employ feature extraction and classification [2]. Clustering approach was used to categorize the percussive onsets and then the rhythmic pattern of the stream of onsets was computed to extract the rhythmic features, tempo and the rhythmic density. Rhythmic analysis of Indian and Turkish music has been explored for vocal and instrumental audios [3]. State of the art MIR methodologies were evaluated for beat tracking, meter estimation and downbeat detection tasks.

This paper discusses the motivation behind the consideration of spectral features for the purpose of classification of the strokes. The section below describes the initial analysis of monophony sitar and tabla strokes (ignoring feeble strokes of tanpura playing in the background) and extraction of discriminatory features from the observation of these strokes. The motivation for extending this study of monophony strokes or isolated strokes to polyphony strokes or strokes in ensemble (both sitar and tabla playing) obtained from sitar concert clippings are discussed. Following this, an application of the study, in order to test the method on rhythmograms, to be able to visually demarcate the segments of a test sitar concert is discussed. The next section presents the results obtained from the above study, plots and accuracies of classification. Following this section, inferences and discussions based on the obtained results are presented.

2. Feature Extraction

2.1 Acoustic Signal Differences between Tabla and Sitar Strokes

Sitar is one of the oldest North Indian classical instruments. A sitar can have 18 - 21 strings. Six or seven of these are the main strings, which are placed over frets. The rest of the strings are sympathetic strings that resonate with the main strings. The frets are movable, which allows for fine tuning. A hollow gourd is used as a resonating chamber to produce melodious notes on plucking the strings [4, 5]. Tabla is an ancient Indian percussion instrument which is composed of two hollow cylindrical structures covered by a thick hide of different layers and materials. The hollow cylindrical structures, typically constructed from either gourd or metal, acts as a resonating chamber to produce the percussion beats. Music is produced by striking the fingers or the hand on the thick hide layers covering the hollow cylindrical structures, to produce sharp and resonant beats [4].

The structural differences of the two instruments and the style of playing produce two entirely different sounding strokes. These differences are sought to be captured using spectral flux and spectral slice representations at the point of time the stroke has been played, and at small intervals after the stroke, to be able to extract discriminatory features for the two kinds of strokes. Sitar strokes, having been obtained from plucking of strings at different frets, over a resonant body, create sounds that are harmonic in nature. These vary from tabla strokes, which are not as harmonic as compared to sitar strokes. Additionally, there is a sharp increase in energy at all frequencies at the onset of a tabla stroke, whereas, in case of a sitar stroke, there is a predominant increase of energy at the harmonic frequencies of the note that is being played on it.

2.2 Discriminatory Spectral Features of Tabla and Sitar strokes

In order to capture the differences between the two kinds of strokes, small windows of varying time width were placed around the time stamps of manually annotated sitar or tabla strokes. For the time frames present in this time window, spectral features such as spectral centroid, spread and other such features were plotted with respect to time. The above features do not give much information to differentiate between the tabla and sitar strokes, hence spectral flux was considered, as this feature provides a description about the energy variation of a stroke along time. To obtain an understanding about the time intervals of the sitar and tabla stroke onset, and to study the duration of most of the strokes, histogram of inter-onset intervals (IOI) versus time in seconds is plotted as shown in Figure 2.2a. This plot indicates our choice of keeping 0.3s as the duration of the stroke and this led to 0.3s to be considered as a window around a stroke to consider it as an independent stroke without the influence of the other strokes while analysing the spectral properties. Hence, a time interval of 0.3s was considered after a given onset and the spectral flux was calculated with a window size of 30ms and a hop size of 5ms.

Some of the observations were that the sitar stroke has a much more gradual slope reduction when compared to the tabla stroke after an onset. The peak of a tabla onset is also relatively higher than the sitar onset. To improve the understanding of the transition and minimize chances of missing the true onset, the spectral flux plots of the strokes were considered between the time intervals of 0.05s prior to

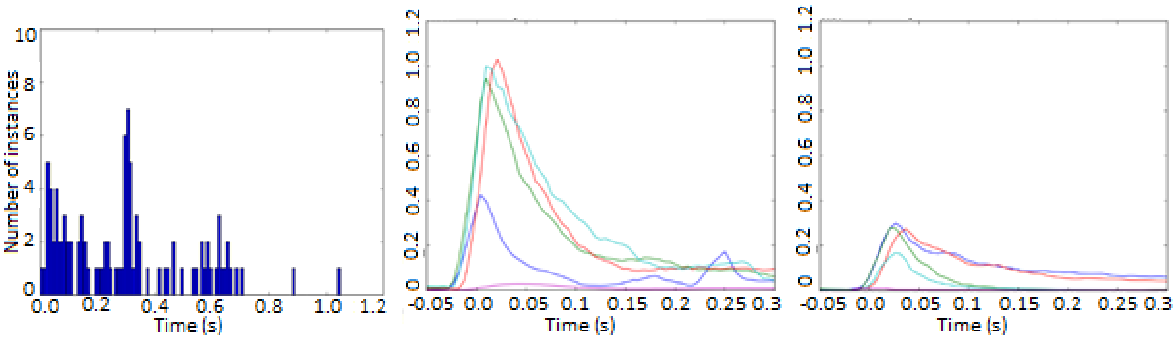


Figure 2.2: a) Histogram of inter onset intervals for both sitar and tabla strokes b) Representative spectral flux plots for tabla strokes. c) Representative spectral flux plots for sitar strokes

the onset to 0.3s after the onset, keeping the window size as 30ms with a hop size of 5ms. Different statistical measures were considered for the spectral flux plots of all the above strokes as features for classification [5].

In addition to spectral flux derived features, spectrogram for short polyphonic clips were computed, and the spectral slices at the time of the onset of the stroke, 10ms after the onset, and 20ms after the onset are considered. The change in the statistical properties during the decay was sought to be analyzed through this process. The spectral slice at these three time instants was computed and harmonic peaks were found out by selecting the peaks that were above 20% of the maximum value of the spectral slice. Some of the features extracted from the spectral slice are, number of peaks of the spectral slice, number of peaks after 2kHz, inter-peak distance and statistics related to the inter-harmonic peak distance. These features were added to the features obtained from the spectral flux curve for the discrimination of the two kinds of strokes using different classification techniques, which are elaborated in the next section. From the spectral flux and spectral slice plots, differentiating features were observed in some of the statistical measures, which were used as features for classification. In spectral flux plots of characteristic strokes as shown in the Figure 2.2.b) and c), it was observed that typically, sitar strokes have a lower energy at the onset when compared to the tabla strokes, and they also decay relatively slower than the tabla strokes. Sitar strokes also showed a lower ratio of the maximum height to minimum height after decay, when compared to tabla. The sitar spectral flux plots also appeared more skewed after the onset, when compared to the corresponding tabla plots. The set of features extracted for classification have been listed below.

A) Spectral slice derived features:
- i) Number of peaks observed at the onset of the spectral slice
- ii) Number of peaks over 2 kHz observed on onset of spectral slice
- iii) Ratio of maximum to minimum value of spectral slice on the onset
- iv) Mean of inter harmonic peak interval distance on onset
- v) Range of inter harmonic peak interval distance on onset
- vi) Standard deviation of inter harmonic peak interval distance on onset

All the above features are also computed at 0.1s after the onset and 0.2s after the onset, so that we have 18 spectral slice derived features.

B) Spectral flux derived features:
- i) Average value of plot
- ii) Peakiness of spectral flux plot
- iii) Skewness of spectral flux plot
- iv) Decay rate of spectral flux plot from the instant of onset to 0.3s after stroke
- v) Ratio of the maximum to minimum value of spectral flux plot.

The above features were computed using the audio feature extraction library Essentia. The spectral flux is obtained for time windows of 30ms, with a 5ms hop over a duration of 0.05s before the onset of the stroke to 0.3s after the onset. Peakiness and Skewness of the spectral flux waveform were also obtained via the Essentia library [7].

3. Classification Experiments

On analyzing all the features discussed above, a set of 23 features was selected based on the observed visual differences in the spectral flux and spectral slice plots of some of the representative tabla and sitar strokes. The selected features were computed for a dataset of 225 sitar and 295 tabla monophony strokes, which were manually annotated by observing the spectrogram and obtaining the points of highest energy to mark the onsets. Table 3.1 shows the segregation of sitar strokes picked from different sections of alap for monophony strokes, collected from isolated tabla theka samples corroborated to study the acoustics of tabla strokes. The polyphony strokes of sitar and tabla are picked from gat sections of the concerts of musicians, Pt. Nayan Ghosh, Pt. Niladri Kumar, Pt. Nikhil Banerjee and U. Shahid Parvez.

Monophony strokes				Polyphony strokes	
Alap	**Jod**	**Jhala**	**Tabla strokes**	**Sitar**	**Tabla**
69	224	154	295	340	265

Table 3.1: Dataset for monophony and polyphony stroke samples

Different methods of classification were explored to be able to effectively discriminate between the sitar and tabla strokes based on the features discussed in the previous sections. Classifiers such as Simple K-means and CART were attempted, however, due to the large number of features in consideration, a sequential minimal optimization (SMO) SVM classifier performed better[8]. The implementation of this classifier was performed on WEKA, an open source machine learning platform.

The experiment involved computation of the features for all the data strokes which satisfied the threshold conditions, and were placed wide enough to satisfy the minimum inter onset time interval considered. These features were stored as feature vectors, and passed through the classifier to obtain accuracy of classification with a tenfold cross validation. These preliminary studies on the monophony strokes were extended to the analysis of polyphony sitar and tabla strokes played in the same concert, as the real usefulness of this discrimination lies in the ability to do so in polyphonic Hindustani concerts. After satisfying the same conditions of thresholding and inter onset time intervals, a set of 5170 polyphony samples (3786 sitar and 1384 tabla) were obtained, after resampling, i.e. an instance based increase of the dataset to obtain a large enough dataset for the analysis, and the features for these samples were computed and stored as feature vectors. The original number of 517 strokes (381 sitar strokes and 136 tabla strokes) were repeated to obtain a dataset of ten times the original number, while keeping the ratio of sitar to tabla strokes, nearly constant. A larger number of sitar strokes were taken for training to account for the large variation in the flux and spectral slice plots of sitar strokes.

To test the usefulness of this method in the segmentation of Hindustani sitar concerts, a complete gat audio in Rag Kamod, by Ustad Shahid Parvez, is given as a test input to the algorithm. The algorithm picked up the sitar and tabla strokes according to the set rules, and the corresponding features were obtained, to be able to classify these strokes as tabla or sitar. The classification returned labels is utilized to derive the tabla specific ODF and the sitar specific ODF from the all onsets ODF generated by the proposed method in [1]. Tabla specific ODF is derived from all onsets ODF, with masking of the entire signal, except 0.3s before and after the detected tabla strokes for the gat audio. Similarly, sitar specific ODF is obtained by masking of the all onsets ODF signal, except 0.3s before and after the detected sitar strokes in the entire audio clip [9]. From these classifier derived sitar specific or tabla specific ODFs, rhythmograms were obtained [10]. The resulting plots have been presented in the next section.

4. Results

Classification results for monophony strokes in shown in the Table 4.1a. An accuracy of 99.4% was obtained for the monophony data set, with three strokes misclassified out of 520. On examination of the misclassified strokes, it was observed that two types of sitar strokes and one type of tabla strokes were wrongly classified. The misclassified tabla stroke was played by striking both the cylindrical drums that created a highly resonant sound, while the misclassified sitar strokes were short and low energy strokes, which were followed by longer and more resonant sitar strokes.

On classification of the polyphony strokes, shown in Table 4.1b, it was observed that an accuracy of 77.5% was achieved under the same classification conditions as mentioned above. Examination of the misclassified strokes here comprised of feeble sitar strokes, which had short decay periods along with a comparatively larger number of tabla strokes with the background tanpura louder than those strokes.

Annotated / Predicted	Sitar	Tabla
Sitar	224	2
Tabla	1	293

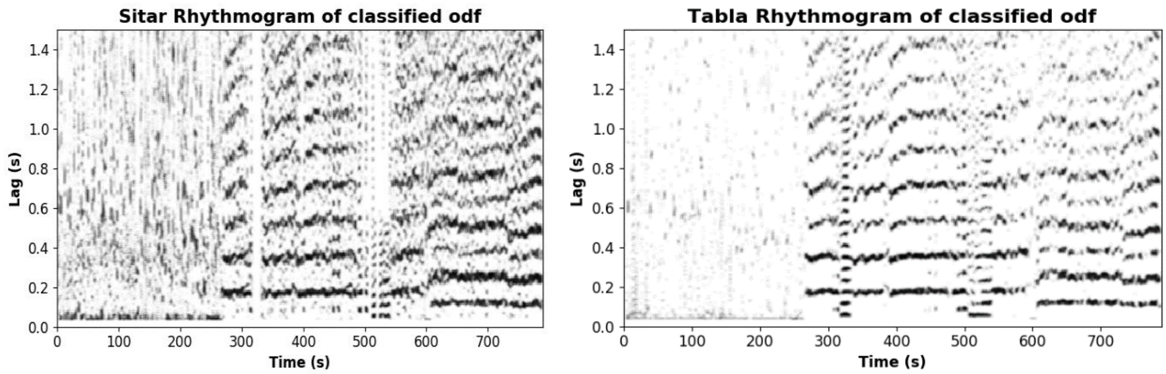
(a)

Annotated / Predicted	Sitar	Tabla
Sitar	3692	1067
Tabla	94	317

(b)

Table 4.1: Confusion matrix for stroke classification (a) Monophony strokes (b) Polyphony strokes

The sitar and tabla specific rhythmograms obtained by the classification approach discussed in this paper have been plotted below in Figure 4.1. From the figures, it is observed that the sections of the *tabla-solos* (314s to334s and 488s to 541s) are enhanced in the tabla specific rhythmogram obtained by classification, while the *layakari* segment appear to be subdued. In the sitar specific rhythmogram obtained by classification, while the *layakari* segment (442s to 463s) appear prominent, the *tabla-solo* segments are very faint. The tabla specific rhythmogram obtained by classification is similar to the one obtained in [1]. When compared to the surface rhythmogram, the tabla and sitar specific rhythmograms obtained by classification appear to complement each other.



(a) (b)

Figure 4.1: Rhythmograms by classification and masking a) sitar specific rhythmogram b) tabla specific rhythmogram.

5. Inference and Conclusions

Classification approach using spectral features has discriminated the sitar and tabla strokes fairly well. A higher accuracy was obtained for monophony stroke discrimination over polyphony stroke discrimination. This could be due to the introduction of the tanpura, along with sustained sitar sounds during some of the feeble tabla strokes in polyphonic music. The misclassified polyphony strokes were largely tabla, which can be explained by the constant presence of tanpura and trailing sitar strokes in the background of the polyphonic music. Further analysis of the training data used for this method and larger training dataset may help improve the accuracy of classification. Modifications can be made to the training data by identifying and incorporating more representative strokes of sitar and tabla, along with experimentation on simulated polyphony stroke samples to analyze where the method of discrimination can be improved upon.

On examining the rhythmograms, it is observed that the sitar and tabla rhythmograms obtained by classification approach highlights the prominent sitar and tabla dominant segments respectively. This result shows promise in the application of this method of discrimination of strokes to segmentation of sitar concerts. Future work includes testing the method on several different kinds of sitar and sarod concerts, and examining the cases where this method gives a good differentiation between the different segments, and where it fails.

6. References

[1] T.P. Vinutha, S. Suryanarayana, K. K. Ganguli and P. Rao, " Structural segmentation and visualization of Sitar and Sarod concert audio ", Proc. of the 17th International Society for Music Information Retrieval Conference (ISMIR), Aug 2016, New York, USA.

[2] A. Elowsson and A. Friberg, “Modeling the perception of tempo”, *The Journal of the Acoustical Society of America*, *137*(6), pp.3163-3177, 2015.

[3] A. Srinivasamurthy, A. Holzapfel, and X. Serra, “In search of automatic rhythm analysis methods for Turkish and Indian art music”, Journal of New Music Research, 43(1):94–114, 2014.

[4] S. Bagchee. Nad: Understanding Raga Music. Business Publications Inc., India, 1998.

[5] B.C. Wade, “*Music in India – The Classical Traditions, Chapter 4: Melody Instruments”*, Manohar Publishers, New Delhi, India, 2008

[6] Kumari M., Kumar P., Solanki S. S., “Classification of North Indian musical instruments using spectral features.” Computer Science & Telecommunications. 2010, Vol. 29 Issue 6, p11-24

[7] Essentia: Algorithms reference, http://essentia.upf.edu/documentation/algorithms_reference.html, last accesses: 29/8/2017

[8] Sequential Minimal Optimization: A fast algorithm for training support vector machines, https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines/, last accessed: 29/08/2017.

[9] Bello, J.P, Daudet, L, Abdallah.S, Duxbury.C, Davies.M, and Sandler.M.B., “A tutorial on onset detection in music signals. IEEE Transactions on Speech and Audio Processing”, 13(5):1035–1047, 2005

[10] K. Jensen, J. Xu, and M. Zachariasen, “Rhythm-based segmentation of popular Chinese music,” in Proc.of Int. Conf. Music Inf. Retrieval (ISMIR), London, U.K., 2005.