UNIVERSITAT POMPEU FABRA

MUSIC TECHNOLOGY GROUP

MASTER THESIS

Student

Marius MIRON

Automatic Detection of Hindustani Talas

Supervisor: Dr. Xavier SERRA

UNIVERSITAT POMPEU FABRA

Abstract

The thesis aims to develop a system of Hindustani tala automatic recognition which can be trained by building a labeled corpus of Hindustani songs with tabla accompaniment. Most of the research concerning rhythm in the North Indian classical music was developed around monophonic recordings and the scope was just recognizing the tabla strokes or modeling the expressiveness of tabla solos and not the metric cycles in which these strokes usually occur, the talas. The aspects researched were segmentation and stroke recognition in a polyphonic context, as recognizing the talas is a perceptually challenging task and the automatic detection proved to be even more difficult.

CONTENTS

LIST OF FIGURES

Acknowledgments

I would like to thank to my supervisor, Xavier Serra, for his guidance,help and patience throughout the year and for grading me the chance to attend this Master course. I learned lots of things which shaped my knowledge and my future.

I am grateful to the research community at the MTG: Hendrik Purwins and Joan Serra for their advices which really shed light on some difficult subjects and for interest shown in my thesis, Perfecto Herrera for providing me the tabla database and for offering me precious insights on my work and to tall my teachers. Furthermore, I would like to thank Vincent and the researchers working at Freesound for their patience and for accepting me as a part of the developing team and to the interns at the MTG, Ciska, John and Stelios. An enriching experience was sharing thoughts about Indian culture and music with Gopal and collaborating with him. Furthermore, I would like to thank my friends from the Master for supporting me through the whole year and sharing moments that I would never forget.

Nevertheless, I would like to thank to family for encouraging me and for their support.

CHAPTER

ONE

INTRODUCTION

A Hindustani music performance is accompanied by percussion instruments, of which the most common is tabla. The role of tabla in a composition is time keeping and this is accomplished by rhythm cycles called talas. To some extent, we could say that the talas are for Indian music what meter is for western music: to develop a time framework for the melody. However there are many cultural and aesthetic constraints that establish a gap between the two notions. Many of the tabla techniques of playing a tala violate the notions of meter, as we could say that meter is not an appropriate way of describing a specific Hindustani percussion performance, due to its ornamentation, variability in tempo and pulse levels [1].

Unfortunately, most of the technologies created for music are based on the ground truth of the western music theory. Nevertheless, the Music Information Retrieval field is driven by concepts imposed by the current music industry. For the Hindustani music, terms as BPM don't make so much sense as for a DJ. And from what we could see they don't even make sense for some of the western classical music, which can't establish to this extent a steady BPM.

If there is a danger for a form of music that exists for hundreds of years, this is oversimplifying, that is imposed by cultural or industrial, even technological factors. Indian music had been through all this along the past century, if we are to mention the lately lack of tabla percussion in the soundtracks of Bollywood movies or the first recordings of Hindustani music which limited the temporal framework of ragas in order to be accommodated by the recording medium. From being a ritual music, to a music played for the Kings, to nowadays Hindus-

tani music performances for the audience, Indian music transformed, assimilated but preserved its richness and meaningfulness.

The research on how Hindustani talas can be automatically identified proves culturally important and it is a way to see how state of the art MIR technologies can deal with a rich but different music tradition, without diminish its meanings. Furthermore, studying this type of music can give precious hints to solving the current problems in the rhythm or pattern detection.

Tabla in Hindustani Music

Tabla is one percussion acoustic instrument used in Hindustani music(along with Pakhavaj and other drums) which is made of two simple kettle drums: bayan, the left bass drum, usually made of copper or aluminum, and dayan, the right drum, made of wood, which can produce a variety of pitched sounds due to its complex construction.

One of the most important thing in learning a tabla is learning its alphabet. In order to construct the rhythm phrase of the tala, one must learn to speak[1] the language of tabla. Every drum stroke or combination of drum stroke has associated a mnemonic called bol. A bol is the main unit in learning and playing talas. It is very common for a student to practice before the rhythm pattern as spoken phrases, before actually touching the drum. The reason for this could be the fact that the talas were transmitted and learned orally and because the Indian music glorifies the human voice at the point that every instrument tries to imitate the characteristics of the singing soloist.

The bols are obtained by strokes on either the bayan or dayan drum or by stroking the two drums at the same time. The tabla bols are classified as follows:

1. Opened strokes on the right drum, distinguished by a clear sense of pitch, sharp attack, and long sustain. Ta, tin and tun are examples.

2. Resonant bass strokes played on the bass drum, which is actually the stroke ghe. The tabla player modulates the pitch of this stroke by controlling the tension on the skin of the bass drum using the base of his palm.

3. Closed sounds which have sharp attacks and decays and sound damped. Kat played on the bayan, and te, tak, dhe, and re, which are played on the dayan, are examples of this family.

[1] In Hindi to speak is "bolna"

In addition to these, strokes can be played simultaneously forming a compound stroke. Typically, a stroke from the dayan is combined with a bass tone (ge) or a closed tone (ke) on the bayan. The most common compound strokes are dha (na + ghe), dhin (tin + ghe), and dun (tun + ghe).These are conceptualized as distinct bols and not as mixtures.

On the other hand, the physical characteristics place the tabla under the category of drums with harmonic overtones [2]. The properties of dayan were well known long time before by the writings of Raman and emphasize the fact that the tabla could be tuned to the tonic of the song.

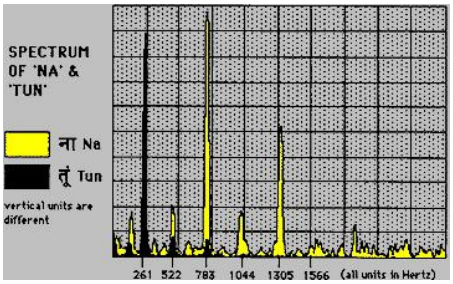

Figure 1.1: Spectrum for the *Na* and *Tun* strokes.

For the tuning the strokes Na or Tin are used. Na is preferred because it has a missing fundamental, compared to Tin who has a strong fundamental, thus the pitch is rather perceived, this thing proving to be very useful when dealing with low quality tablas.

Talas and Rhythm

Once a player masters playing the strokes and the tabla alphabet, he can proceed in learning phrases and the rhythmic cycles, talas. There are hundreds of talas that have been used and are mentioned but nowadays just about ten of them are more common with their known variations.

Talas are fixed patterns of same length made of beats or matras and it is split into sections called vibhags. The whole complete cycle is called avart but the whole tala will always start and end with the first beat called sam, meaning that the first beat it will always add up or should be regarded as adding up at the end of the avart. The clap(tali) and wave(khali) gestures accompany a tala at each vibhag and their scope is complementary to the talas. Basically

they mark accented or unaccented sections and they give precious hints to the soloist when the tala is too difficult to keep track of and when the soloist might be confused [1].

Each tala has associated a pattern of bols called theka. Below is a table with some of the most known talas and their thekas.

Tala	Phrase(Theka)
Dadra	Dha Dhin Na | Na Tin Na
Teora / Tivra	Dha Den Ta | Tete Kata | Gadi Ghene
Rupak	Ti Ti Na | Dhi Na | Dhi Na
Keharwa	Dha Ge Na Ti | Na Ka Dhi Na
Addha	Dha dhin - dha | Dha dhin - dha | Ta tin - ta | Dha dhin - dha
Dhoomali	Dha Dhi | Dha Ti | Tak Dhi | Dhage Tete
Nabam	Dha Den Ta | Tita Kata | Gadi Ghen | Dhage Tete
Jhamp	Dhi Na | Dhi Dhi Na | Ti Na | Dhi Dhi Na
Rudra	Dha | Tat | Dha | Titkit | Dhi | Na | Titkit | Tu | Na | Ka | Tta
Mani	Dha Di Ta | Dhe Tta | Dhage Nadha Ttak | Dhage Nadha Ttak
Chou	Dha Dha | Din Ta | Kit Dha | Din Ta | Tita Kata | Gadi Ghen
Ek	Dhin Dhin | Dhage Tirkit | Thun Na | Kat Ta | Dhage Tirkit | Dhin Dhin
Ras	Dhi | Ttak | Dhi | Na | Tu | Na | Ka | Tta | Dhage | Nadha | Ttak Dhin | Gin
Dhama	Ka Dhi Ta | Dhi Ta Dha - | Ga Di Na | Di Na Ta -
Jhoomra	Dhin -dha Tirkit | Dhin Dhin Dhage Tirkit | Tin -ta Tirkit | Dhin Dhin Dhage Tirkit
Deepchandi	Dha Dhin - | Dha Ge Tin - | Ta Tin - | Dha Dha Dhin -
Teen	Dha Dhin Dhin Dha | Dha Dhin Dhin Dha | Na Tin Tin Na | Tete Dhin Dhin Dha

Figure 1.2: A list of the most common talas phrases

A difference should be made between talas as accompaniment and tabla solo performances. A tabla solo is very often a pre-composed performance where the entire attention is set on the tabla player, which is usually accompanied by a melodic instrument playing a repetitive phrase, having the role of a timekeeper. When a tabla player accompanies another instrument, tabla should keep the time as the performance itself would develop within the tala cycle. Talas are not pre-composed and static. They are based and evolve from the basic pattern, theka, in a semi-improvisatory manner which involves a lot of training and which eventually gives the measure of the virtuosity of the tabla player. Practicing tabla solos compositions could be a way to improve expressiveness as these compositions are usually set in a tala.

The Time in Indian Music

Even it existed before the Mughal's invasion, the tabla gained its importance and developed in the court of the Kings. As explained by Naimpalli [3], the style of playing was constantly developing and under the influence of the historical and cultural events. The traditional style of singing, Dhrupad, which was meant to praise the Hindu gods, was replaced by not so robust Khayal style, which was mainly composed for Kings. Thus required softer and subtle accompaniment, which was not exactly constrained by the rigid rules of ritual singing and would later allow space for improvisation and development. This would reflect later in very personal ways of playing and ornamenting the rhythm cycles. The way authority is distributed in a musical performance also changed, since Taal was granted more importance, in the detriment of the melody, Swar.

It can be mentioned that the talas are almost never static and they involve rhythmic diversity and variation of tempo. The Indian concept of cyclic time is reflected in music as a process of manifestation and dissolution, not bounded by something permanent and unchanging. This can be clearly seen in the evolution of the raga and in the introductory part, the alap, where the raga is becoming, it is searched and it is always there, it returns cyclically, it is gradually becoming. This clearly contrasts with the western approach of a musical piece as a defined structure with a beginning and an end a clear relation between the parts of the structure[1]. There are compositions in Hindustani music as bandis or the tabla solos, but they indicate precisely something that is a priori restricted.

Many things can be brought in order to differentiate between the two approaches – the western based on the logic and progression and the east based on the state and process. For example a tabla performance is regarded as good if it manages to establish the tala.

The theka came under a sufi philosophy influence as way to facilitate improvisation over a rhythmic cycle. The theka gives some hints about learning the tala but it is not the tala itself. It certainly gave space to more cyclical development inside the big time cycle of the raga, but it is not for sure the way to fully understand a performance. With respect to the tala and the basic theka, the tabla player will deviate from the usual pattern, will increase rhythmic density, will swap cycle, permute measures, but everything will happen in the cycle. If in the western music a musician can do everything with respect to the measure, a tabla player can do everything with respect to the cycle.

This way of approaching a performance, different styles of playing and different schools might be the causes of laykari or rhythmic variation. This will encourage the tabla player to embellish the basic theka as a proof of virtuosity (let's not forget the importance of virtuosity for court music) but mainly to establish the tala, in the traditional way of forming it through cyclic variations. If a tabla player manages to keep the tala and to transmit the emotion, this can be a proof of a good performance. However this will make a tala particularly difficult to perceive and recognize.

CHAPTER

TWO

STATE OF THE ART

Background

The current MIR studies on the rhythm in Hindustani music are concerned just with the bol strokes detection in a monophonic context. Gillet and Richard [4] built a labeled database of tabla strokes using a probabilistic approach based on Hidden Markov Models (HMM) to segment and label the specific bols. The logic behind choosing a HMM based design was representing the time dependencies between successive strokes. Real-time transcription of tabla solos with an error of 6.5% lead to the development of an environment called Tablascope. The architecture of the system involves onset detection which segments the signals into strokes, feature extraction - energy in four different frequency bands, learning and classification of the bols - calculating the mean and Gaussian in each of the four frequency bands, modeling of the tabla sequences and finally, transcription. Several classifiers were used with the training data: knearest neighbors (kNN) with k=5, Naive Bayes and kernel Density estimator. However the best results were obtained with a language model - the HMM.

Chordia and Rae [5] developed a system to recognize bols, as part of an automatic tabla-solo accompaniment software, Tabla Gyan. Their research extends the studies of Gillet and Richard and focus more on the tabla solos and the logic of building the improvisation sequences and ornamentation. A larger database was used, comprising recordings of different professional tabla players and different tablas. The descriptors chosen comprised temporal features as well as spectral features(MFCCs). Classification accuracies of 92%,94% and

84% over 10-15 classes were obtained, using different classifiers: Multivariate Gaussian, probabilistic neural network and feed forward neural network.

Further on, Chordia et al [6] described a system which predicts the continuation of tabla compositions, using a variable length n-gram model, to attain an entropy rate of 0.780 in a cross-validation experiment. The n-gram models are extended by adding viewpoints which can improve the actual performance of the system. A multiple viewpoint system tracks variables as pitch classes , notes, onsets times separately, maintaining many predictive models simultaneously, thus offering the advantage of modeling the complex relations between the variables. To prevent that the models will not become too general to be effective, there were considered two types of models: short-term models (STM), which start empty, and long-term models (LTM), built from the compositions in the database. Final results suggested that the strong local patters of the tabla compositions were better captured using STMs, mainly because each composition is based on a specific theme which is progressively varied.

Motivation

As the problem of rhythm detection is not something new, it was applied just to the western music and basically to the popular music. Industrially motivated, most of the research in the MIR field works with just a very small slice from the huge diversity of music created, the western music. The cultural importance of the non-western music could be augmented by adding to this context music from other cultures. Furthermore, this could be a starting point to study the interaction between geographically, historically and culturally different types of music and could provide in the end a broader meaning to the comparative musicology and ethnomusicology studies. A simple example is the Gillet's and Richard's [7] work on tabla strokes influenced their research on indexing and querying drum loops databases by deploying similar mnemonics in their system as the tabla ones. The study of rhythm in Indian music must be made with respect to its meaning in the context of Indian culture. In this way, the research will also have a cultural dimension – to see how the current paradigm of MIR could integrate and describe music from other cultures.

Current MIR research doesn't deal with bol transcription in a polyphonic context. This open topic is similar with the drum transcription one, with the mention that a very small number of classes are used for (single) drum strokes.

This research would be crucial in solving the more ambitious issue of taal detection, a perceptually difficult task which currently can only be performed by musicians and trained audience. The variations introduced by the performer should be modeled in order to understand this kind of improvisation which differs a lot from North India to South India, from school to school, from musician to musician and even from performance to performance. Otherwise, for tala recognition, a system that would understand and integrate all these variations should be build.

CHAPTER

THREE

THE METHODOLOGY

A few tasks needed to be done before detecting the tala. A crucial one was segmenting the target performance into proper segments which would later be fed to an algorithm which would process them and output information which would be used to detect the taal. A tala can also be detected from the pattern of bols. In this case bol transcription in a polyphonic context would be in important issue to study.

The algorithms were implemented in Python 2.6 using the MTG framework Essentia 1.3 offering state of the art MIR extensions. Another useful software was Sonic Annotator using the qm(Queen Mary University) and Aubio Toolbox. For the machine learning part, the Weka platform proved to be very useful, although the Pymix, the Python library for mixture models, was a very solid alternative when using Gaussian Mixture Models. These technologies were successfully tested with western pop music, thus one of the main goals was to evaluate them in the context of Hindustani music.

Database labeling

Indian Classical music has been prodigiously recorded or digitized and nevertheless made available to the audience in different kinds of formats. The indissoluble raga was the subject of medium constraints, when going to compact discs, vinyls or when played on the radio. The constraints were mainly about how the performance was split into tracks or limited as duration.

When building the database, there were considered three type of sources:

compilations of ragas, recorded performances and full albums. The raga compilations as Morning Ragas, Evening Ragas, Night Ragas, The Raga Guide have a wide variety of instruments and styles, always accompanied by the tabla. Furthermore, the introductory, highly improvised part, alap, is missing. On the albums and in the performances, this part was separated from the others. Tabla accompaniment can rarely be found in the alap and usually it is played in a melismatic way [1].

The shorter the performance the highest the chances to find the tala instanced by in a less varied way, by its basic pattern. However, the artists featured on these albums were famous and highly acclaimed Indian classical musicians, which rarely would play the theka as it is. Virtuosity in Hindustani music is appreciated and encouraged. Musicians and especially the best ones would have their special way of playing a tala, differing from performance to performance. For example, depending on the raga, the type of emotion it sets in, the tonic of the song, the instrument, the tala would be played with different types of strokes and with different tablas, tuned in different ways. As usually tabla was tuned to the tonic, in the some recordings (especially with female singers), the tabla was mostly tuned to the fifth.

The school and the tradition in which the musician was educated, could also be a very important factor. Various schools have different ways of interpreting and teaching different talas. We have to consider that recordings of very famous contemporary musicians were used. Nowadays the tabla players include elements from most of the schools, though they would surely mention if some tabla performance is played in some style.

There were labeled 80 albums comprising 450 songs of Hindustani classical music with tabla accompaniment. The tala was annotated from the metadata or album art-work. These albums were labeled in the database with the prefix [TAAL]. Performances that contained other percussion instruments were added the label [OTHER]. If the quality of the recording was low, that album was labeled as [LOFI]. Tabla solo performances were not included in the database, because they are fixed compositions, governed by special rules and different from the talas. The whole database featured 70 musicians.

Furthermore, each file was assigned a class based on the annotated tala.

There are many more instances of the Teentaal then any other rhythm cycle, because this is the most common tala used nowadays in the Hindustani music. The number of measure was also associated as talas with the same number of

Tala	No. Matras (measures)	No. Instances
Dadra	6	9
Rupak	7	8
Keharwa	8	14
Matta	9	6
Jhamp	10	21
Ek	12	30
Deepchandi	14	8
Tilwara	16	5
Addha	16	6
Teen	16	124

Figure 3.1: The tabla bols and their models

measures could be grouped in the same class.

Tuning frequency

Before each raga concert, the soloist and other musicians decide upon the tonic of the song. The other instruments are tuned relatively to the established tonic if not decided otherwise[1]. The same types of rules apply to the tabla. Thus, detecting the tuning frequency is important in detecting the frequency of the Dayan drum. On the other hand, if the strokes in the performance are identified, it can be computed a deviation from the tonic, which could be a measure for how well the tabla is tuned and could bring useful information.

The tuning frequency algorithm as implemented in [8] by Gomez, uses the spectral peaks to detect the tuning frequency of a song. The spectral peaks are computed below 800 Hz using interpolation, sorted by their magnitude and inputed into the Tuning Frequency algorithm. The output frequency is estimated by calculating the frequency deviation of the extracted peaks. The Harmonic Pitch Class Profile (HPCP) is computed for each frame, calculating the relative intensity of each pitch, then a global HPCP vector is computed by averaging the instantaneous values. The FFT analysis parameters were a Hanning window of 2048 size and hope size of 1024.

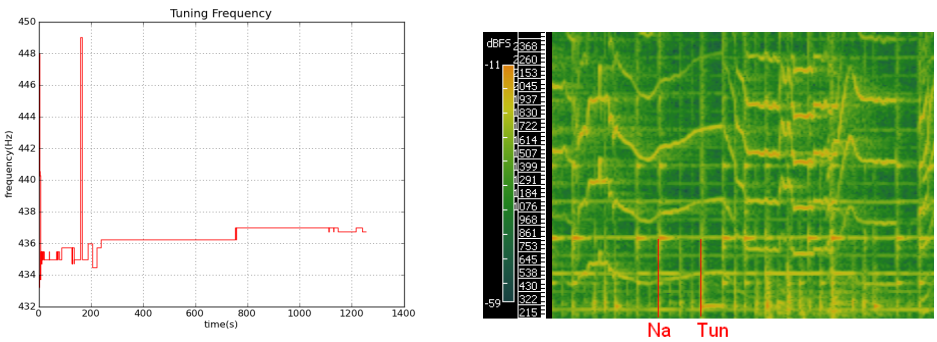

Figure 3.2: Tuning frequency for *Raga Hamir Bahar, Bismillah Khan* and the equivalent tabla tunning for *Na* and *Tun* strokes

Probably a better method which was not implemented in this paper is the joint recognition of raga and tonic, described by Chordia and Senturk[9]. The arbitrary tonic and the micro-pitch structure require the use of tonal representations based on more continuous pitch distributions. Continual tonal representations were found to perform better than pitch class distribution. Best results were obtained with a kernel density pitch distributions, attaining an error of

4.2% for tonic frequency estimation and an error of 10.3% for raga recognition.

Onset detection

Onset detection is an important task, when trying to segment the audio or when transcribing the bols from a tabla performance. The performances of the later tasks would depend on onset detection. Thus, it would be important to evaluate it on Hindustani music.

Onset times were annotated from a 12 minutes long raga performance by the famous singer Girija Devi, "He Mahadev Maheshwar, Khayal madhyalaya,in Raga Bhoop". The tala associated to this concert was the 16 matras teentaal and other instruments accompanying the voice were the tanpura, the sarangi and of course, the tabla.

The High Frequency Content (HFC) of the input spectral frame is efficient at detecting percussive onsets, thus could work better with detecting tabla strokes. As described in [10], the method is based on the fact that the changes due to transients are more noticeable at high frequencies, appearing as a broadband event, producing sharp peaks. A weighted energy measure is computed from the spectrum:

$$E(n) = \frac{1}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} W_k \left| X_k(n) \right|^2$$

where W_k is the frequency depending weighting.

Different HFC implementations were tested: from Essentia, Aubio Toolbox and Queen Mary plugin for Sonic Visualiser. Along with this, a mixed method which weights the energy in six frequency bands along with the high frequency content to obtain the onsets of the tabla.

The detection of the tonic was used to implement a set of band pass filters which would improve onset detection of tabla strokes in a polyphonic context. As described by Raman [11], there are clear ratios between the different frequencies of different vibration modes of the tabla. Using these known frequencies, the filters used could be centered around them.

The energy in these filter bands was used along with the high frequency content to detect the onsets. These functions were normalized and them summed into a global function according to their weights.

Bands	Range	Weight
HFC	HFC	1.5
Energy band 1	60-140 Hz	1.7
Energy band 2	140-210 Hz	1
Energy band 3	centered around the tuning frequency	1
Energy band 4	centered around the first resonating mode	1
Energy band 5	centered around the second resonating mode	1
Energy band 6	centered around the third resonating mode	1

Figure 3.3: The weights of the mixed onset detection function and the relation with the resonating modes described by Raman

Then the silence threshold was applied before finding a series of possible onsets which were later filtered using the alpha and the delay. The following parameters were used: for the FFT, Hanning window of 512 size and a hop size of 256 and for the onsets detection, alpha=0.12, delay=6 and silenceThreshold=0.02. The alpha was used to filter very short onsets and represents the proportion of the mean included to reject smaller peaks. The delay is the size of the onset filter is the number of frames used to compute the threshold. Various thresholds were used to tune the algorithm and filter the onsets. Because of the different contexts and different classes of strokes, this task was difficult to adjust and implement as a general algorithm that would fit many performances. The next thing to do was to set the threshold low and try to filter the onsets by detecting the drum strokes.

The energy based detection follows changes in the energy of a signal by deploying en envelope follower. This method works well when we can find strong percussion events contrasting with the background. Compared to these methods, the complex domain onset detection combines the energy based approaches with the phase based approaches and uses all this information in the complex domain. The function sharps the position of the onsets and smooths everywhere else. Compared to the HFC method, it tends to over-detect the percussive events and it works better with note onsets.

The algorithm implemented to test the onset detection tries to match the best onset candidate C_i for the existing and annotated onset O_k by computing the time difference in seconds between a O_k and every C_i found between O_{k-1}

and O_{k+1} and picking the minimum one. The best C_i is marked as used and the algorithm assigns it to the current index k of O_k, otherwise the onset is marked as missed 0.

```
initalize the onset candidates' index vector with 0
initialize the onset distance vector with 0
search for a possible candidate for the first onset
for each onset candidate C[k]
        while C[k] >  O[i+1]
                increment i
        if the C[k] candidate is closer to O[i]
                check if it's the best candidate for
the O[i] onset
        else check if it's the best candidate for
the O[i+1] onset
```

For this experiment, several parameters of the onset picking algorithm were modified in order to detect all onsets. For the Aubio toolbox the silence threshold was lowered to -110dB whereas the peak picking threshold was lowered at 0.25 from the default of 0.3. Thus it tends to over-detect onsets as it performs very well in detecting all types of strokes in different situations. For the Sonic Visualiser plugin, the global sensitivity is kept at the default of 50%.

Method	Missed Onsets %	Overdetected Onsets %
HFC Essentia	70	18
Complex Essentia	37	39
Mixed Essentia	63	18
Complex Aubio	32	60
HFC Aubio	20	63
Broadband Energy Rise QM	35	67
HFC QM	25	65

Figure 3.4: Results for different onset testing functions

Since the threshold was set really low, all the methods tend to over-detect the onsets, except Essentia HFC which runs with its default parameters. The Aubio implementation of the HFC detection was found to perform slightly better.

Segmentation

Due to the fact that full performances were included in the database, it was necessary to separate the most proper to use segment in order to detect the tala. A proper segment is that part of the performance where the tabla is following as close at it can the basic pattern, where it doesn't improvises or changes the tempo. It is very difficult to find a measure for this kind of variation mainly because the ornaments are introduced very often. These ornaments are more likely to occur at the end of the performance, because the tempo and rithmic variation will vary more and increase to the end.

Three methods were analyzed: the first was to extract descriptors from the audio and use Bayesian Information Criterion to produce useful segments. As various sets of existing features didn't give a measure for the stability of the tala, something related to rhythm was used: using the inter-onsets interval to calculate an onset density and split the audio into sections according to this density. Finally a segmentation based on autocorrelation of inter-onset histogram was implemented.

After performing windowing the sound and performing FFT analysis, several features were extracted from the resulted spectrum. The type of window used was Hanning, of 2048 samples. The hop size was 1024. The features extracted were spectral roll off and spectral complexity. The roll off frequency is defined as the frequency under which some percentage of the energy of the spectrum is contained and it is used to distinguish between harmonic and noisy sounds.

Along with these features, energy in low frequency bands was extracted, considering that the only instrument in that frequency band is the bass drum, the bayan. It was also taken into consideration the fact that the basic pattern is always composed of bayan bass strokes, the part where the ghe stroke is modulated and the energy in low frequency band is higher or there is no energy in low frequency bands and there are lots of ornaments on dayan is the most improvisatory part. A simple segmentation based just on the low frequency band was found to perform better then a mixed segmentation using the spectral roll off or the spectral complexity. Other descriptors considered were loudness and inharmonicity.

Bayesian Information Criterion (BIC) segmentation, used in the AudioSeg project was used to segment the audio into homogeneous portions based on the descriptors above. The algorithm searches segments for which the feature

vectors have the same probability distribution based on the implementation in [12]. The segmentation is done in three phases: coarse segmentation, fine segmentation and segment validation. The first phase uses two sets of parameters to perform BIC segmentation. The second phase uses other two parameters to perform a local search for segmentation around the segmentation done by the first phase. Finally, the validation phase verifies that BIC differentials at segmentation points are positive as well as filters out any segments that are smaller than a minimum length specified which in this case was 7 seconds.

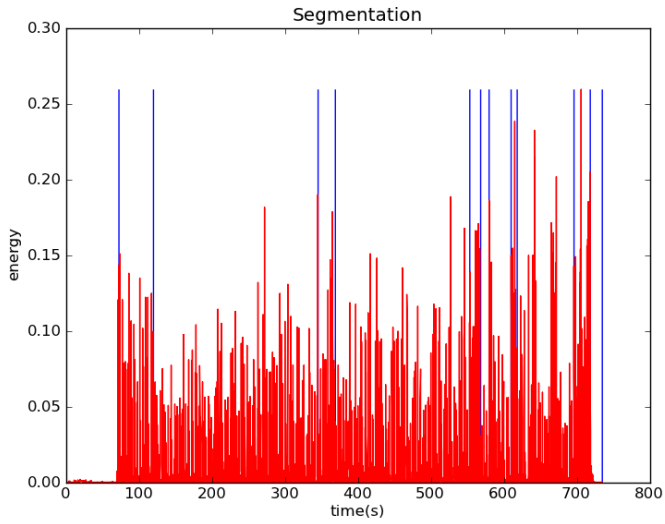

Figure 3.5: BIC segmentation based on the energy in the low frequency band, *Girija Devi - "He Mahadev Maheshwar, Khayal madhyalaya,in Raga Bhoop"*

The example used was, again Girija Devi's "He Mahadev Maheshwar, Khayal madhyalaya,in Raga Bhoop" with the most common 16 beat teentaal as time framework. The most useful segment analyzing a tala is the first after the tabla starts, as we can see from the graph, the interval: [72.4230423:119.76852417]. The tabla actually starts at 70 seconds with short and fast improvisatory ornamentation passage on the dayan drum.

Another option would be to take into consideration the onsets and analyze the segments based on the onset density or the inter-onset intervals.

A first solution would be to compute an onset density for each frame, based on the onsets situated N frames apart and their value, a stronger onset having a stronger weight. An onset density vector is computed for entire audio file and

fed to the segmentation function.

```
initalize the onset density vector with 0
for each detected onset O[i]
        initialize the weight vector w = [0:N/2]*0.1
        for each D[k] with k in [i-N/2:i+N/2]
                D[k] = D[k] + w[k] * V[i]
```

Building the onset detection array involves taking each detected onset O_i and adding its weight to each of the D_k onset density vector, with $k = [i - \frac{N}{2} : i + \frac{N}{2}]$. The weight will be added proportionally, D_i having the highest weight V_i and the elements situated $\frac{N}{2}$ apart, the lowest.

The densities below a certain threshold are assigned a 0 value, in order to prevent the weight added by the over-detected isolated onsets where the tabla doesn't perform.

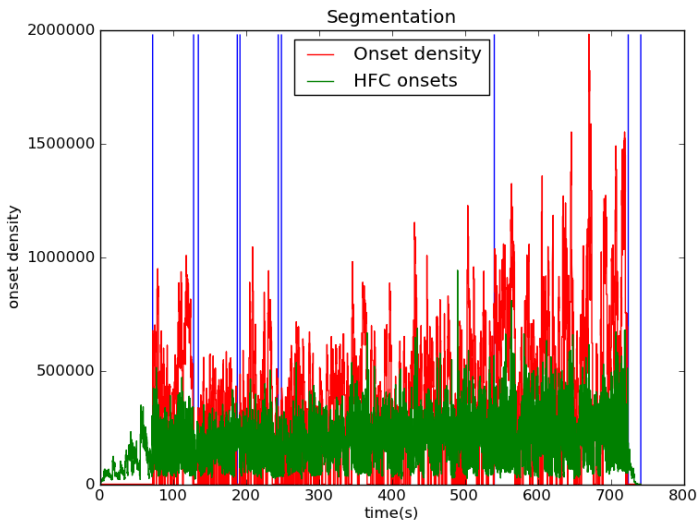

Figure 3.6: BIC segmentation based on onset density, *Girija Devi - "He Mahadev Maheshwar, Khayal madhyalaya,in Raga Bhoop"*

Segments as short as 7 seconds can be detected, which can be useful when segmenting short performances. For a reliable further analysis of the rhythm cycles, segments below a certain threshold should be dropped as tala extends to larger time spans, depending on the tempo(lay) and rhythmic variation(laykari).

The results lead to picking the [71.95864105:128.13931274] interval for the analysis of the tala, and more detailed results on the evolution of the tala, than the low-frequency based method.

A more feasible approach which can separate big variations of the basic pattern from a relative stable tala cycle, is computing the inter-onset histogram and discovering a periodicity over the time spans.The implementation is based on the one presented by Dixon et al[13] which involves accumulating inter-onsets over equally spaced intervals in successive bins and grading these bins a weight. By definition the inter-onset interval(IOI) is the time difference between successive notes $t_i = o_i - o_{i+1}$, or in our case, tabla strokes. However in this case onsets situated as far as 5 seconds apart will contribute to the histogram. Thus, inter-onset intervals will be computed for each pair (o_i, o_k) with k the onset below the 5 seconds threshold. Each of the interval is accumulated in a histogram bin, the number of bins being decided after a specific resolution. In the original implementation, the weight is established not only by the number of the bins but also by their amplitude or deviation from the mean value of the bin. In this case, a simple histogram based only on the number of IOIs is computed over 100ms bins.

```
initialize the histogram vector with 0
for each detected onset O[i]
        for each O[k]<=O[i]+5 seconds
                compute the interval difference O[k]-O[i]
                accumulate interval in correspondent bin
```

A better implementation would be to weight each time the bins with the amplitude of the onsets and their spread.

Another important variable is the sensitivity threshold of the onset detection function, which in this current example, tends to over-detect the onsets, as we could see in the previous section.

After the histogram is obtained, the next step is computing the autocorrelation of the histogram to find the repeating patters through time. The output of the autocorrelation is expressed as lag time and could determine the periodicity of the pulse. Given the IOI histogram x_n with n bins, the discrete autocorrelation R at lag j is defined as:

$$R_{xx}(j) = \sum^{n} x_n \bar{x}_{n-j}$$

After computing the autocorrelation, the next step is finding its characteristics:

- Autocorrelation Frequency - the frequency associated with the highest non-zero-lag peak in the auto-correlation function.
- Autocorrelation Height - The height of the largest non-zero lag peak, an indicator of the peak's strength.

Considering this, the segmentation problem would resume to finding the best time span with the strongest peak. A peak detection algorithm was used to detect the local peaks of the autocorrelation vector and calculate its characteristics. The algorithm follows the positive slopes and outputs a peak when the slope's sign becomes negative and the amplitude of the peak is above the threshold. The peaks are searched in corresponding range of the time span and ordered by their amplitude. The highest peak which does not correspond to the zero lag position is returned along with the autocorrelation frequency and height.

To illustrate the above, analysis prone segments were compared with the ones in which the tala was embellished. The musical piece was split into consecutive 20 seconds segments. Another solution would have been to use the above segmentation just to separate the parts with more percussive events and analyze just those parts.

For Girija Devi's "He Mahadev Maheshwar, Khayal madhyalaya,in Raga Bhoop" in teentaal, for the first segment, where tabla starts, was compared with more stable segments and the was reflected in the periodicity of the IOI histogram. In the first case the algorithm was not able to detect any peaks for the first segment, compared to the results obtain from the segment marked as stable.

Another tala, the 12 beats ektaal was analyzed because its basic theka has short bursts of fast combination of strokes, like the tirikita succession. The algorithm separated two successive segments, a transition followed by a basic pattern.

Because teentaal is a very symmetric tala, its basic theka developing synchronously along the 16 matra cycle, another tala, which is regarded as asym-

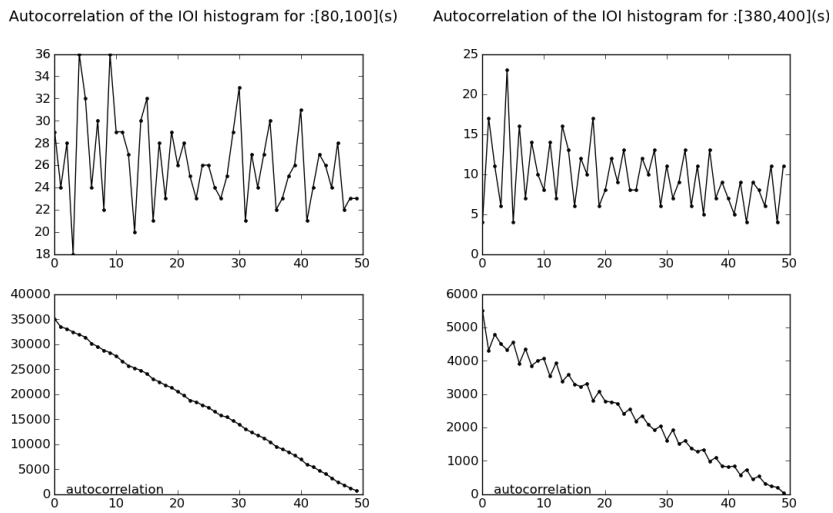

Figure 3.7: Autocorrelation of IOI histograms on the intervals left =[80:100](s) and right =[380:400](s) with $h_{right} = 1.14532$ and $f_{right} = 2$, *Girija Devi - "He Mahadev Maheshwar, Khayal madhyalaya,in Raga Bhoop" - teentaal*

metrical by the tabla players, was analyzed - roopak taal. This tala introduces syncopations even in it's basic theka, because its internal subdivisions are of unequal lengths and it starts with the unaccented beat, the khali, compared to other talas which start with the accented sam.

The second segment in the roopak taal performance is more improvised, with faster strokes on the dayan and this is also reflected in the histogram.

The algorithm is clearly dependent on the performance of the onset detection output. For example, if it over-detect onsets, the segmentation could not work so well when having plucked string instruments - and the autocorrelation could output the periodicity of the inter-onsets of the notes and not the strokes.

Another observation and a further thing to study is that there is a big different between segmenting a 16 matras tala and a 6 matras tala, mainly because over a large cycles, improvisations could happen and the hints would make it easier to keep the time framework.

The time interval chosen, 20 seconds, is not feasible for low tempo talas. It is well known, for example that the slow tempo vilambit teentaal is different from the fast one, drut teentaal, the tabla player deploying a totally different aesthetic in performing at slower tempo. In this case, one 16 beat cycle could

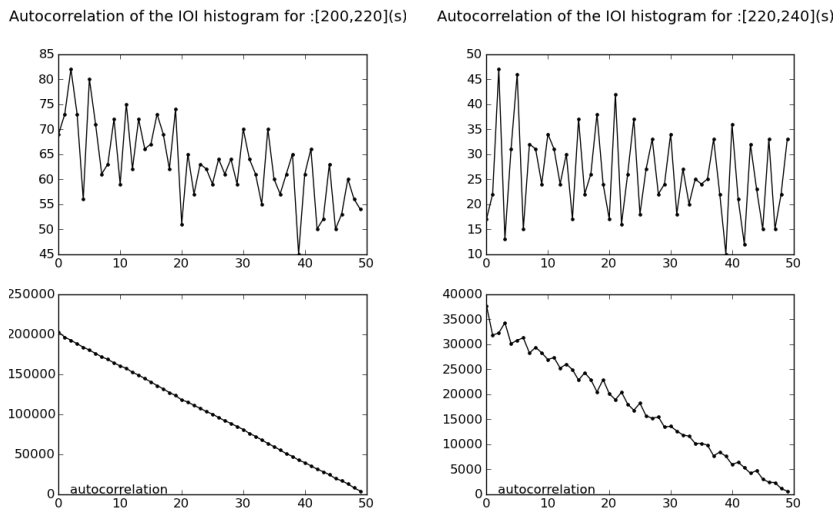

Figure 3.8: Autocorrelation of IOI histograms on the intervals left =[200:220](s) and right =[240:260](s)with $h_{right} = 1.091$ and $f_{right} = 3$, *Hari Prasad Chaurasia - "Madhuvanti" - ektaal*

last longer than 20 seconds.

A better implementation of the segmentation of a tala piece could benefit from how authority is distributed in a raga performance. The tala offers the time framework in order to allow the raga to develop in time cycles. However, when rhythmic variations occur as the point that the basic pattern theka is embellished into something that could look as a virtuous part of a tabla solo, the soloist performs repetitive melodic patterns which draws the focus onto the cyclic time progression and rhythm variation - laykari [1]. In this case, the melody gives time hints for that part of the performance. A future implementation of the segmentation could benefit from the characteristic and implement a joint estimation of the melody and the tabla performance sections.

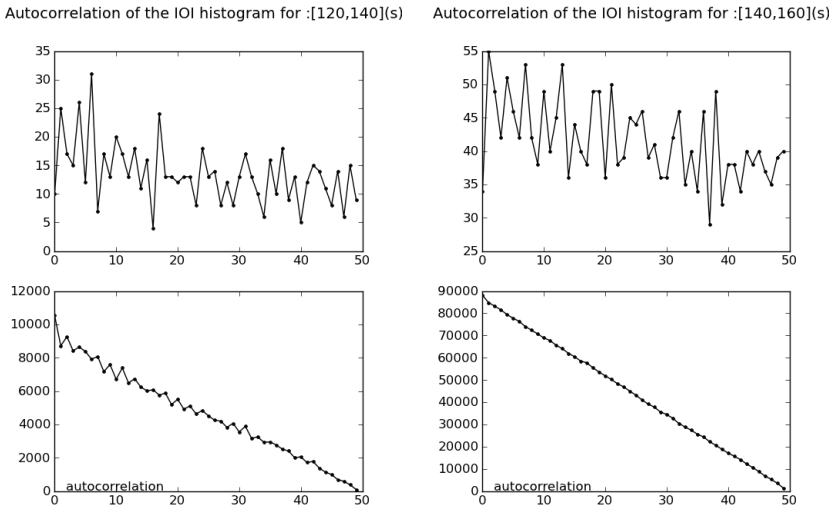

Figure 3.9: Autocorrelation of IOI histograms on the intervals left =[120:140](s) and right =[140:160](s)with $h_{left} = 1.1356$ and $f_{left} = 2$, *Girija Devi - "Jhoola, Aaj do jhool jhoole (in Raga Sindhura-Barwa)" - roopak taal*

Bol transcription using the acoustic properties of tabla

A problem exposed in the sections above is detecting the tabla strokes in a polyphonic context. The method described in this section is using the vibrational modes of the tabla, residing in its acoustic properties as described by Raman [11]. Each stroke determines the drum to vibrate according to a particular mode. A mode could emphasize the fundamental or suppress it. There is a clear relation between the frequency ratios of every stroke and the vibrational modes.

These known relations between the harmonics of each bol could be modeled by mixture of gaussians. The Gaussian Mixture Models are calculated by accumulating the spectrum peaks for each frame. In this way, the spectrum of the bol is regarded as a mixture of a fixed number of gaussians, which depends on the characteristics of each bol.

The univariate gaussian is defined by the probability distribution function:

$$N(x|\mu,\sigma) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where μ is the mean and determines the height of the distribution and σ^2 is the variance and determines the width of the distribution.

Because the number of gaussians is chosen in accordance with the harmonic overtones of the tabla, the number of models depends on the complexity of the bol. For example the Dha bol would be more complex to model than the Na because it represents a combination of two simultaneous strokes(Ghe+Na). The Tun stroke would be easier to model as it has a strong fundamental and it's enough to model it just with only one gaussian.

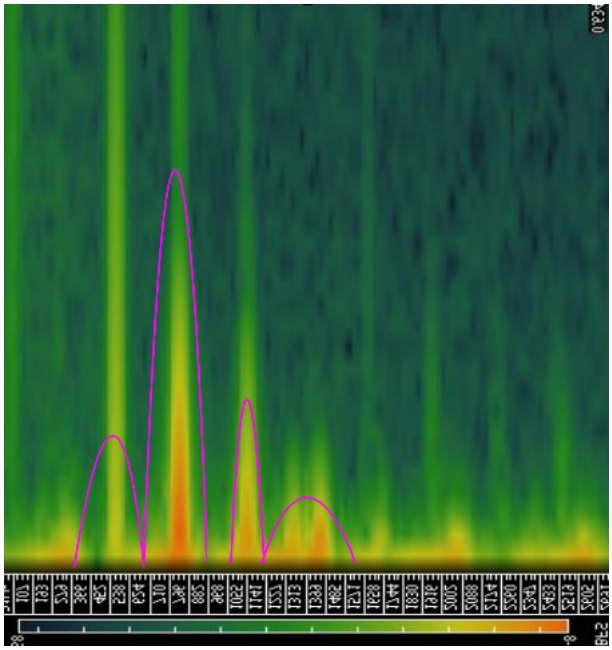

Figure 3.10: Modelling the Na bol with four gaussians

In the figure above we would have to find a pair $\langle \mu_i, \sigma_i \rangle$, where μ is the mean and σ is the variance, for each of the four gaussians. For every bol, several gaussian mixture models need to be built taking into account the tuning of the tabla, which would help identifying the strokes when dealing with tablas of different construction or tuned differently.

Another issue was that the tuning for each stroke was not known as they were only separated into slow and fast. We needed to detect the tuning frequency based on detecting the fundamental frequency and the relations between the spectral peaks. For each bol, we would have to build not just one gaussian mixture model but as many as tunings we will find in the database.

The database used in this experiment comprises more than 8000 labeled monophonic tabla sounds, which were previously used by Parag Chordia in his research regarding the detection of the tabla strokes. The decision of choosing to build a model from samples of tabla was the lack of a solid physical model which would allow to implement a detection algorithm knowing the spectrum for different tunings of tablas and different sizes of a drum.

Bol	Na	Dha	Dhe	Dhin	Te	Tun	Tin	Ke	Ge
No. Models	10	10	1	5	10	5	10	5	5
No. Gauss	4	5	3	3	2	1	3	4	1

Figure 3.11: The tabla bols and the corresponding models

The final purpose of modeling is recognizing the bol in the polyphonic audio. For each frame in target audio, the evaluating system would estimate a probability for each model which would translate into a matrix where the bol chosen would have the best score over a set of frames.

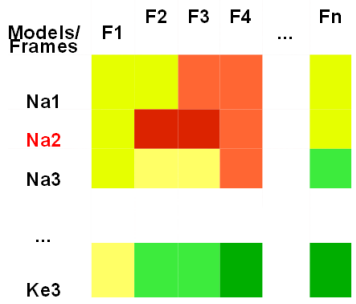

Figure 3.12: Computing the score for each GMM model

As a preprocessing step, an equal loudness filter is applied to the sound, to assure perceptual consistency between the models. Then the signal is low-passed to assure the peaks that will count in building the model would be below 2000 Hz. The values for computing the FFT depend on a Hanning window of 2048 size which allows enough frequency resolution when picking the peaks and a hop size of 256.

Initially, a number of peaks are extracted for each bol, corresponding to the number of gaussians which will model the spectrum. The peaks are sorted and filtered by the magnitude with a threshold of 0.0001. Then for each frame, all

the peaks extracted are accumulated in an array. The threshold assures that only strong peaks are tracked and it relates the magnitude of the peaks for one frame with the whole sample.

```
for each type of stroke
        initalize an empty vector for tuning frequencies
        initialize a empty vector for the peaks
for each stroke
        peaks = detect_spectral_peaks(No_gaussians)
        detect tuning of the stroke
        accumulate peaks in the vector
for each bol class
        compute histogram of the tuning frequencies
        initialize a vector for each tuning
        for each bin in histogram
                accumulate peaks in a each vector
```

On the other hand, the spectrum is summed frame by frame and peaks are computed again in order to detect the tuning frequency which is picked as the strongest peak from the ten peaks which have been extracted. A histogram of N bars, where N is the number of models for each bol, is computed based on all these peaks. Based on the results of the histogram, a bol is added to a tuning category or another. Then, for each tuning category, the univariate gaussian mixture model for the accumulated peaks is computed along with its means, variations and weights.

At this point we would have a number of spectral peaks, $x = [x_1, x_2, ..., x_M]$ where M is the number of peaks or observations from which we would have to find the best gaussian to fit our data. The joint probability of x will be

$$p(x|\mu, \sigma^2) = \prod_{M}^{m=1} N(x_m|\mu, \sigma^2)$$

and we will need to maximize the probability of the data $p_{best}(x|\mu, \sigma^2) = argmax_x[p(x|\mu, \sigma^2)]$, when given the mean and the variance. The problem is we only have the data and we would have to estimate μ and σ^2, finding the likelihood of gaussian which would fit best our data: $p_{best}(x|\mu, \sigma^2) = l(x|\mu, \sigma^2)$.

Expectation maximization is a general method of finding the maximum likelihood estimate of the parameters of a distribution from a given data. Solving

the logarithm of the likelihood would have the advantages that it decomposes the products into a sum of values and it eliminates the exponentials in the original expression, thus we will get:

$$L(\mu, \sigma^2|x) = -\frac{M}{2}(log2\pi + log\sigma^2) - \frac{1}{2\sigma^2}\sum_{M}^{m=1}(x_m - \mu)^2$$

For the expectation maximization algorithm, the first choice was the SciKits package learn, Python module integrating classic machine learning algorithms. The Gaussian Mixture Model is still under developing and this might be one of the causes why the EM algorithm outputted erroneous values for the means and the variances of the models calculated. Then the PyMix module was used with success and easiness. This Mixture package is implemented in python and was previously used in Biology and Genetics. An intermediary function was written in order to initialize each gaussian with a mean and variance. For this purpose, a histogram of the peaks for each model was computed, having as many categories as numbers of gaussians. Then for each category, a mean and a variance was computed and the EM was initialized with these values and equal weights. Because Expectation Maximization only assures convergence for a local maximum and in order to prevent the algorithm to get trapped in a bad local maximum, the EM algorithm was run ten times for each model.

```
compute the spectrum
extract the spectral peaks
for each frame in spectrum
        for each bol model
                calculate loglikelihood with the model
                compute the score
                write the score for the current frame
in the matrix
```

The parameters of the testing algorithm were selected in accordance with the shortest sound sample used, resulting in a window size of 4096 and a hop size of 1024. A number of twenty peaks was extracted for each window, according to their magnitude. For each of this peaks log likelihood was computed in two different ways: the value offered by the PyMix and the fit function which would weight the value of the peak for each gaussian with the peak's magnitude and

the gaussian's weight.

Before testing this model in a polyphonic context, it was tested on a set of monophonic tabla sounds, resulting only 55% accuracy in detecting the specific bols. The causes of this are the lack of a physical model and the fact that the database didn't contained any information about the tuning. Future experiments should either implement a system based on a physical model or build a database of tabla sounds categorized by the tuning and not by their speed. An alternative to this approach would be to model the spectrum of the stroke not with a mixture of gaussian but with a non-negative matrix factorization which would assume separating the bols in the database on tunings and for each category, computing a average bol spectra by decomposing the spectrogram for each instance of the training data into a product of non-negative spectrum and non negative time varying gain and then averaging the non-negative spectrum over all instances [14].

Another issue discovered was that the Parag Chordia's database was used for a real-time system, meant to improvise and recognize strokes even if they are not performed perfectly. It is arguable that this would work perfect on detecting the more accurate strokes of many tabla virtuosos in a polyphonic recording.

Tabla stroke recognition in a polyphonic context

As tabla is known in the Hindustani music as a talking drum, timbre features which were previously used for speech recognition could help in detect the tabla sound in a polyphonic context. These features - MFCCs, spectral centroid, skewness, kurtosis, along with temporal features were also important in detecting tabla in monophonic performances [15] and were used with high efficiency by Herrera et al[16] in classifying the drum sounds. The goal was to train a model for a single performance and use those strokes to annotate this performance.

Descriptors	Temporal features	Spectral features
Herrera et al[16]	logattack time, temporal centroid, zcr	MFCCs,decay strong peak spectral flatness, skewness, kurtosis,energy in 8 custom bands
Chordia[15]	temporal centroid,attack time, zcr	MFCCs,spectral centroid, skewness kurtosis
Used	log-attack time, zcr	MFCCs,spectral flux, centroid, skewness kurtosis, spectral complexity, hfc along with their mean and variation
Selected	-	MFCCs,spectral centroid, spectral complexity, along with their mean and variation

Figure 3.13: Comparison of features used in state of the art methods

In order to segment the tabla bols, a tabla solo by Akram Khan was used. This performance can be found in one of the Tabla Series recordings of famous tabla players. A tabla solo gives more space in the recording to the tabla player and grants him more authority. Furthermore, in a tabla performance, more types of bols are used more frequently than in a normal accompaniment. This was actually a problem when trying to train our system: to find an equal number of strokes for all the classes.

From the first 4 minutes and the last 2 minutes of the first Peshkar part of the tabla solo, the following bols were segmented and labeled:

As there were very few strokes for Tin(23) and Dhin(6), these two classes were not included in the model which was built.

At first the same features used in the state of the art methods were extracted

Bol	Na	Dha	Ge	Te	Tun
No. Instances	121	128	150	128	103

Figure 3.14: The bol classes and the instances used for training

to be refined after testing them with different classifiers.A few temporal features were extracted as log attack, strong decay and the loudness of the sound but they were excluded from the model lately because they didn't offer such a good separation between the classes.

The FFT analysis parameters used were a Hanning window of 1024 and a hop size of 256. Several spectral features were computed along with the mean and the variance: 13 MFCC coefficients and the energy in the corresponding 13 MFCC bands with a high frequency bound set at 7000 Hz, spectral centroid and spectral complexity.

The test data set had strokes from the same audio file but, as the training set was constructed from the begging the performance, for the test we chosen the middle. Then we tested this training set in Weka along with different classifiers.

Bol	Na	Dha	Ge	Te	Tun
No. Instances	263	405	419	188	178

Figure 3.15: The bol classes and the instances used for testing

Suport vector machine(SVM) comprises a set of supervised learning methods which takes as an input a set of data and builds a model that assigns one of two classes for the input. SVM was tested firstly with a polynomial kernel $K(x, y) = < x, y >^p$ and then with a RBF kernel $K(x, y) = e^{-\gamma <x-y, x-y>^2}$.

In the first case the system obtained a 93.53% accuracy, classifying correctly 1359 and incorrectly 94 sounds with a mean absolute error of 0.2432. The parameters used were the complexity parameter C=1, the epsilon for round-off error P=1.0e-12 and the exponent for the polynomial kernel E=1.

There is more confusion between Na and Dha because Dha is basically Na plus a stroke on the bayan, which sometimes was a shorter or longer, more resonating stroke. Also, between Dha and Ge because sometimes the Na wasn't loud enough.

For the RBF kernel 93.59% of instances were classified correctly. The pa-

a	b	c	d	e	<- classified as
229	33	1	0	0	a= NA
1	375	28	0	1	b=DHA
1	10	401	4	3	c=GE
2	0	1	173	2	d=TUN
2	2	1	2	181	e=TE

Figure 3.16: The confusion matrix between the classes with SVM - Poly

rameters used were the complexity parameter C=2, the epsilon for round-off error P=1.0e-12 and $\gamma = 0.1$.

a	b	c	d	e	<- classified as
239	22	1	0	1	a= NA
1	368	36	0	0	b=DHA
1	12	399	5	2	c=GE
1	0	1	172	4	d=TUN
2	0	2	2	182	e=TE

Figure 3.17: The confusion matrix between the classes with SVM - RBF

The descriptors used were evaluated with the SVM attribute evaluator with a complexity parameter C=1. From the 44 used, the mean of the energy in the MFCC bands 5 and 9 were found to separate the best between the classes.

A further evaluation could involve detecting the onsets and detecting the bols on the percussive onsets. This would be a case which would relate more to a polyphonic situation where tabla stroke recognition would be useful for detecting the tala.

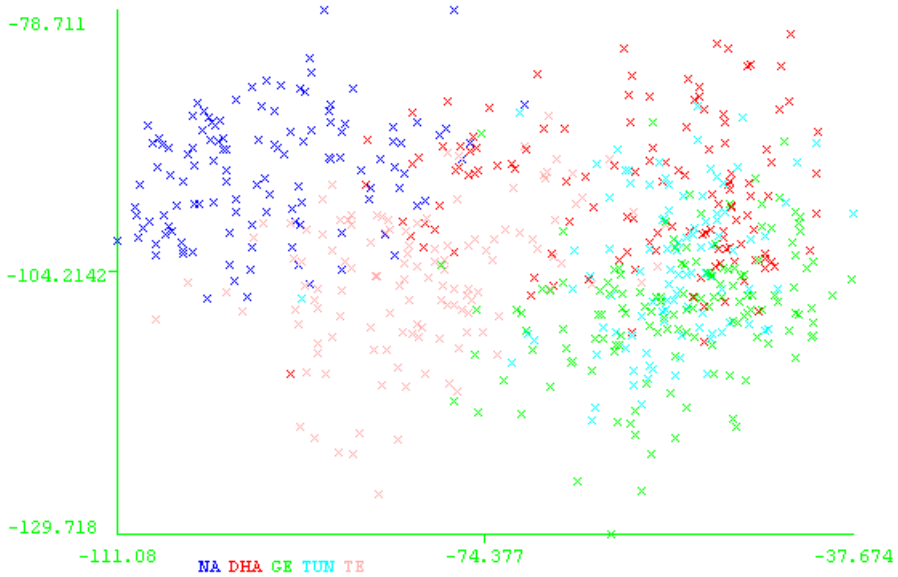

Figure 3.18: The mean in MFCC band 1 vs band 5

CHAPTER

FOUR

CONCLUSIONS AND FUTURE WORK

As the initial goal of building an system which would automatically detect the tala in the Hindustani classical recordings did not succeed, the steps that were done were important if we only mention that no work had been done before concerning tabla in a polyphonic piece of music. Several state of the art issues were confronted with the rich and different tradition of the North Indian classical music. The implicit goal was to find how they could accommodate this tradition: to find if the approach could give answers to the problems that arise.

As a first example, even if at some point during the research a small experiment of BPM detection was conducted over a small set of rhythmic pieces of Hindustani music, it was concluded that the notion of BPM didn't have any meaning for a non-western musician. The annotated surface tempo could not accommodate the more complex notions of lay and laykari. As a matter a fact, it was difficult to find slices of performance with a relatively steady tempo to build our small collection of short excerpts. When dealing with an improvisatory tradition, we should be careful not to analyze it into a limited context, but to find solution which could model it.

A paper from more than a century ago written by the Nobel prize winner C. V. Raman was brought into the context of drum transcription. Detecting the tonic of the raga and the relation established with the tala bols is a characteristic of this music and a chapter of the thesis tries to benefit from this particularity. Even if the method if bol transcription had its flaws and failed, a better implementation that would keep track of the observations made in that chapter could prove successful.

State of the art methods implemented in detecting bols in monophonic recordings were tested and evaluated in a polyphonic context. The same state of the art methods for detecting the tuning frequency, onsets and segmentation were used and tested.

Regarding the database, a big collection of talas was gathered and annotated. This need to be improved because 50% of the collection is made of instances of a single tala. More instances of the not so common talas need to be brought in. Another improvement would be to separate short performances from long performances and talas in vilambit from talas in drut, fast tempos from slow tempos.

A tuning frequency algorithm was used to detect the tonic of the song. The implementation based on HPCPs was yet to be confronted with a method based on Yin fundamental frequency detection, due to the lack of time. The newest method and the more adequate, the joint recognition of raga and tonic should be used instead as it proved to be successful. The relation between the tonic and the tabla need to be studied further as in some performances tabla can be tuned to the fifth.

Different implementations of onset detection algorithm were evaluated with a single piece. This would be needed to be tested further with more musical pieces from different artists, performing with different instruments. A mixed detection method which would combine high frequency content onsets and energy in different frequency bands, established automatically from the tuning of the tabla was tested along. The testing method needs to be reviewed for possible flaws. Onset detection is a crucial thing when dealing with bol transcription or segmentation because the efficiency of these methods would be based on it.

Segmenting the audio in order to find the best part of the performance for a future analyze, the more stable section is similar to the problem of separating short tabla solos from short thekas. Three algorithms of segmentation were implementing: two based on BIC and the last one which proved to be more efficient on autocorrelation of the inter-onset histograms. The efficiency of the three methods needs to be tested against shorter and longer performances, against fast or slow talas and evaluated. The 20 seconds segment length should be re-evaluated in this context. A joint estimation of the melody and the sections could take advantage of the fact that the tala is more varied where the soloist is playing a repeating melody.

The segments obtained by the autocorrelation of the inter-onset histograms,

were analyzed in order to obtain details about the talas. In particular autocorrelation in different frequency bands was computed but no relation between the periodicity of the histogram and the tala could be established. This lead to detection of the bols in a polyphonic context which would detect the tala based on the succession of strokes.

Bol transcription could have been better with the help of a physical model or with a well structured database which would annotate tuning frequency and tabla along with the bols. The number of gaussians to model every bol should be re-thought and the magnitude of a peak should be taken into account along with other timbre features in a multi-variate gaussian mixture model. In this way, the system could discriminate between strokes with harmonic overtones and very noisy strokes. An alternative to this method would be using non-negative matrix factorization or other solutions used in source separation for drums. But for this to be possible we would need, again, a better structured database.

Detecting bols in a monophonic context is already a solved task. The methods used were tested on polyphonic audio, with less classes used than in the monophonic implementation due to the unequal distribution of bols in a performance. The results showed the importance of the timbre descriptors in classifying different types of drum strokes, which were successful even in a polyphonic context. The same tests should be run on other performances, not just tabla solos. Would be preferred that the onsets would be extracted before and the algorithm would be tested against these onsets.

Future work in detecting the tala should use the bol transcription and try to establish relations between the patterns of strokes which would give hints on the corresponding tala. For example, the teen taal is easy to recognize from the successive dhas or nas bols.

Pattern recognition could also be useful when dealing the tala and comparing a basic theka used by teachers with a virtuous performance could be give a measure for the embellishment. Other areas of research like performance modeling would prove useful in establishing the complex relations between the basic pattern and the variations.

BIBLIOGRAPHY

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

[2] C. V. Raman and S. Kumar, "Musical Drums with Harmonic Overtones," *Nature*, vol. 104, pp. 500–500, Jan. 1920.

[3] S. Naimpalli, *Theory and practice of tabla.* Popular Prakashan, 2005.

[4] O. Gillet and G. Richard, "Automatic labelling of tabla signals," in *Proc. of the 4th ISMIR Conf*, Citeseer, 2003.

[5] P. Chordia, T. Malikarjuna, A. Sastry, and A. Albin, "MULTIPLE VIEWPOINTS MODELING OF TABLA SEQUENCES," *Information Retrieval*, no. Ismir, pp. 381–386, 2010.

[6] P. Chordia, A. Sastry, and A. Albin, "Evaluating Multiple Viewpoint Models of Tabla Sequences," *Creativity*, 2010.

[7] O. Gillet and G. Richard, "Drum Loops Retrieval from Spoken Queries," *Journal of Intelligent Information Systems*, vol. 24, pp. 159–177, Mar. 2005.

[8] E. Gómez, "Key Estimation from Polyphonic Audio," *System*, pp. 2–3, 2005.

[9] P. Chordia and S. Sertan, "Joint Recognition of Raag and Tonic in North Indian Music," vol. 30332, no. December, pp. 1–11, 2010.

[10] J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 1035–1047, Sept. 2005.

[11] C. V. Raman, "The Indian musical drums," *Journal of Mathematical Sciences*, vol. 1, no. 3, pp. 179–188, 1934.

[12] M. Cettolo and M. Vescovi, "Efficient Audio Segmentaton Algorithms Based on the BIC," in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*, vol. 6, pp. VI–537, IEEE, 2003.

[13] S. Dixon, E. Pampalk, and G. Widmer, "Classification of dance music by periodicity patterns," in *Proc. International Conference on Music Information Retrieval*, pp. 159–165, Citeseer, 2003.

[14] J. Paulus and T. Virtanen, "Drum transcription with non-negative spectrogram factorisation," in *Proceedings of the 13th European Signal Processing Conference*, pp. 4–8, Citeseer, 2005.

[15] P. Chordia, "Segmentation and recognition of tabla strokes," in *ISMIR*, pp. 107–114, 2005.

[16] P. Herrera, A. Yeterian, and F. Gouyon, "Automatic classification of drum sounds: a comparison of feature selection methods and classification techniques," *Music and Artificial Intelligence*, pp. 69–80, 2002.