Akshara Transcription of Mrudangam Strokes in Carnatic Music

[1]Jom Kuriakose, [2]J Chaitanya Kumar, [1]Padi Sarala, [1]Hema A Murthy and [3]Umayalpuram K Sivaraman

[1]Indian Institute of Technology, Madras
[2]Microsoft India Development Center, Hyderabad
[3]A renowned freelance mrudangam artist, Chennai

***Abstract*—Percussion instruments play a significant role in Carnatic music concerts. The percussion artist enjoys a great degree of freedom in improvising within the defined *tāla* structure of a composition. The objective of this paper is to transcribe the improvisations, treating the percussion strokes as syllables or *akshara*s.**

Onset detection is performed to segment the waveform at each *akshara*. Using the transcriptions from the training data, a three-state Hidden Markov Model is built for each *akshara*. The language model is derived from the training data. Testing is also performed isolated style using onset detection to segment the phrase, and the language model to correct the transcription. Transcription is performed on both concert recordings and studio recordings. This technique yields upto $\approx 96\%$ accuracy on studio recordings and $\approx 76\%$ accuracy for concert recordings.

As the mrudangam[1] is an instrument that is based on tonic; tonic normalised features, namely, Cent Filterbank Cepstral coefficients are used. It is shown that tonic normalisation helps in transcription across different tonics.

I. Introduction

Percussion instruments constitute an integral component across all genres of world music. In addition to keeping track of rhythm, they have been developed to produce individual performances rich in artistic quality. Carnatic music is a genre or system of music associated with the southern part of the Indian subcontinent. Every concert in Carnatic music has a lead performer and accompanying artists. The mrudangam is compulsorily present in every concert. Improvisation is a very important part of Indian classical music and percussion is no exception to this rule. In every concert, a mrudangam artist is given space to perform a solo, referred to as the *taniavartanam*.

The *taniavartanam* is part of the main item in a concert. The artist produces various phrases in a variety of strokes within the framework of the *tāla* (rhythm/beat cycle) of the composition. The duration of the *taniavartanam* can vary from 2 to 20 minutes. The ghatam, kanjira and morsing (also percussion instruments) are used to accompany the main percussion instrument (mrudangam) and play in an almost contrapuntal fashion with the beats [1]. Each of the instruments may be played with the lead artist (usually a vocalist) as a rhythm accompaniment, or individually (*taniavarthanam*). A study of the patterns produced by the percussion instruments can be used in various music information retrieval (MIR) tasks, thus enhancing the listener's experience.

Each stroke produced by the mrudangam is called an *akshara*. A sequence of *aksharas* constitutes a *mātra*, and a sequence of *mātras* constitutes a *tāla* cycle. A number of different *aksharas* can be produced by the mrudangam artist. Further, every school has its own set of *aksharas* ranging from the simplest having ≈ 10 *aksharas* to more complex ones having more than ≈ 40. The objective of this paper is to transcribe mrudangam *taniavarthanam* into a sequence of *aksharas*. We first perform onset detection to segment the waveform into strokes, and build Hidden Markov Models (HMMs) for each *akshara*. During testing the waveform is again segmented into strokes and transcribed into *aksharas*.

The task explored here is automatic stroke transcription for the mrudangam on a large and varied dataset, including concert recordings. Mel Frequency Cepstral co-efficient (MFCC) features are used for this task. Transcription is carried out using an HMM classifier built in two ways - i) isolated style using the onset detection algorithm with a language model, and ii) flat-start based embedded re-estimation model. Mrudangam phrases are split into individual *aksharas* using the onset detection algorithm [2], and transcription accuracy is found to increase by $\approx 20\%$ for isolated style classification. The uni-gram and bi-gram probabilities (language model) are used to correct the output results of the isolated HMM classifier. This results in an absolute improvement in accuracy of $\approx 5\%$ over that of flat-start. The mrudangam has both pitched and non-pitched *aksharas*. Cent filter-bank cepstral co-efficient (CFCC) features are used to recognise *aksharas* using HMMs for a database consisting of databases that consists of different tonics.

The rest of the paper is organised as follows. Section II gives a brief description of mrudangam and Section III given an overview of related work. Section IV describes the datasets used in the study. Section V describes the HMM framework for stroke recognition and the CFCC feature based tonic independent approach to mrudangam stroke transcription. Section VI discusses the results obtained and Section VII concludes the work.

II. Mrudangam

Mrudangam is the main accompanying percussion instrument in Carnatic music. It is a two sided drum whose body is made of hollowed jackfruit wood. The two sides are covered with goat skin which resonate when struck. These are laced together using high tension leather straps. The sides of the mrudangam differ in diameter to produce different bass and treble sounds. The smaller side produces high pitched sounds and the wider side produces low pitched sounds [3].

The mrudangam is tuned before a performance to the base tonic of the lead performer. Various embellishments on

[1]The spelling of mrudangam is spelt as mridangam conventionally. This spelling is based on a discussion with the artist.

the mrudangam enable the production of unique and distinct harmonics. The *akshara*s are categorised based on the side of the mrudangam being played. The left side *akshara*s are tonic independent and some right side *akshara*s are tonic dependent. The sound of the *akshara*s vary based on the position, manner and force with which the membranes are struck. The number of unique *akshara*s are not fixed and varies across schools of mrudangam tutelage. Composite *akshara*s can also be produced by striking the two sides of the mrudangam simultaneously which adds to the complexity of the task of mrudangam transcription.

III. Previous Work

There have been earlier studies on mrudangam by Nobel laureate scientist Sir C V Raman [4], and later by Siddharthan [5], where the harmonics of the *akshara*s are analysed. There have been many previous efforts aimed at recognising *akshara*s in percussion instruments.

In [6], modal analysis and transcription of mrudangam is performed using Non-Negative Matrix Factorization (NMF). Using NMF, a dictionary of spectral basis vectors is first created for each of the modes of the mrudangam. The composition of the *akshara*s are then studied by projecting them along the direction of the modes using NMF. HMMs are used to learn the modal activations for each of the *akshara*s of the mrudangam. In [7], tonic independent stroke transcription is performed on mrudangam *akshara*s. In this paper, the magnitude spectrum of the constant-Q transform (CQT) of the audio signal is used as the feature. Then, using NMF a low-dimensional feature space is obtained. This is used to separate the mrudangam *akshara*s. The resulting feature sequence is made event-synchronous using short-term statistics of feature vectors between onsets. Classification into a predefined set of stroke labels using Support Vector Machines is then performed. In this work, strokes played by a single musician are used. Further, the number of *akshara*s are limited to 12. The tabla stroke recognition in [8], [9] and [10] is very similar to the approach in this paper.

In [8], the tabla *akshara*s are segmented using sudden changes in amplitude and phase. Subsequent recognition is performed using neural networks and tree based classifiers. In [9], variable-length Hidden Markov Models are used for tabla modeling and prediction. In [10], the tabla *akshara*s are segmented into individual *akshara*s and classification is performed using HMMs.

IV. The Dataset

A. Mrudangam

For stroke classification, the mrudangam *tani* dataset (DB-1) [2] (Table-II) along with studio recordings of various tonics is used [3] [6]. A third database which was excised from a standard concert *taniavarthanam* is also used (DB-3). Details of the datasets are given in Table I. The labels of all the *akshara*s have been marked with the help of professional artists. The convention of notation of *akshara*s varies across different schools of mrudangam. Therefore the number of unique *akshara*s in each dataset is different. The dataset from [6] has only 12 strokes, since it has been recorded in a controlled environment. There are many composite *akshara*s in mrudangam, which makes the list of unique *akshara*s very large as shown in Table II and III. The labels associated with *akshara*s also differ across different schools of mrudangam. The list of *akshara*s in DB-1, DB-2 and DB-3 are given respectively in Tables II and III. The mrudangam datasets DB-1 and DB-2 are available online as part of the CompMusic project [12] [13].

Dataset	**Dataset Tonic**	**Duration (min:sec)**	**Unique** *Akshara*
DB-1	C#	15:41	41
	C#	13:10	39
DB-2	B	5:53	12
	C	6:13	12
	D	5:05	11
	E	6:08	12
DB-3	C#	5:44	40

TABLE I: Datasets for *akshara* recognition

*Akshara*s	**Tani-1**	**Tani-2**	*Akshara*	**Tani-1**	**Tani- 2**
Ta	Y	Y	Arai Chapu + THOM	Y	Y
DHI	Y	Y	Chapu + THOM	Y	Y
DHI	Y	Y	DHIN + THOM	Y	Y
Arai Chapu	Y	Y	Long Finger + THOM	Y	Y
Chapu	Y	Y	NAM + THOM	Y	Y
DHIN	Y	Y	Rotating NAM + THOM	Y	Y
DHEEM	Y	Y	DHI+THOM	N	Y
Long Finger	Y	Y	Ta + THOM	N	Y
NAM	Y	Y	DHIN+THOM	N	Y
Rotating NAM	Y	Y	Long Finger + THOM	N	Y
THOM	Y	Y	Tha	Y	Y
DHI + Gumki	Y	N	NAM + THOM + Gumki	Y	Y
DHI + Gumki	Y	N	Rotating NAM + Gumki	Y	Y
DHIN + Gumki	Y	Y	Rotating NAM + THOM + Gumki	Y	N
DHIN + Gumki	Y	Y	Long Finger + Gumki	N	Y
NAM + Gumki	Y	Y	Ta + Tha	Y	N
DHI + Tha	Y	N	DHIN + THOM + Tha	Y	N
DHI + Tha	Y	Y	NAM + Tha	Y	Y
Arai Chapu + Tha	Y	Y	Rotating NAM + Tha	Y	N
Chapu + Tha	Y	Y	Chapu + THOM + Tha	N	Y
DHIN + Tha	Y	Y	DHEEM + Tha	N	Y
Ta + Meettu	Y	Y	Other1	N	Y
DHI + Meettu	Y	Y	Other2	Y	Y
Long Finger + Meettu	Y	N	Other3	Y	N
Other4	Y	N	Other5	N	Y

TABLE II: List of unique *akshara*s - DB-1

V. The Approach

A. Isolated Style Hidden Markov Model

1) Data Preparation: The percussion audio is split into phrases and labeled by trained percussion artists. The phrases are divided based on *tāla* cycles of *akshara*s in the audio. From the labels, the language model (uni-gram and bi-gram probabilities) are learnt. The language model is created using the CMU Statistical Language Modeling Toolkit [14].

2) Onset Detection: Using onset detection, the starting of all *akshara*s are recognized. The onset detection is performed using the group delay algorithm [2] [15], where the *akshara* sequence signal is processed similar to that of syllables in

[2]This database was generated by Vidwan Umayalpuram K Sivaraman Sir (one of the coauthors of the paper). He is an artist who has received Padma Vibhushan, an honorary doctorate from the University of Kerala, and also honored with the Sangita Kalanidhi award by the Madras Music Academy

[3]This database was also specially created by Akshay Ananthapadmanabhan, an upcoming mrudangam artist, especially for the CompMusic [11] project.

Akshara	Akshay				TB	*Akshara*	Akshay				TB
	B	C	D	E			B	C	D	E	
Bheem	Y	Y	Y	Y	Y	Num+Gumki out	N	N	N	N	Y
Bheem+Tha	N	N	N	N	Y	Ta	Y	Y	Y	Y	Y
Cha	Y	Y	Y	Y	Y	Ta+Gumki out	N	N	N	N	Y
Cha+dhom	N	N	N	N	Y	Ta+Tha	N	N	N	N	Y
Cha+Gumki out	N	N	N	N	Y	Ta+thom	N	N	N	N	Y
Cha+Tha	N	N	N	N	Y	Tha	Y	Y	Y	Y	Y
Cha+thom	N	N	N	N	Y	Tha+Ta	N	N	N	N	Y
Dheem	Y	Y	Y	Y	Y	Tha+thi	N	N	N	N	Y
Dhem	N	N	N	N	Y	Thai	N	N	N	N	Y
Dhin	Y	Y	Y	Y	Y	Tham	Y	Y	Y	Y	Y
Dhin+Gumki out	N	N	N	N	Y	Thi	Y	Y	N	Y	Y
Dhin+Gumki in	N	N	N	N	Y	Thi+Gumki in	N	N	N	N	Y
Dhin+Gumki out	N	N	N	N	Y	Thi+Tha	N	N	N	N	Y
Ka	Y	Y	Y	Y	Y	Thom	Y	Y	Y	Y	Y
Ki	Y	Y	Y	Y	N	Gumki in	N	N	N	N	Y
Nam	N	N	N	N	Y	Gumki out	N	N	N	N	Y
Nam+Gumki in	N	N	N	N	Y	Gumki out	N	N	N	N	Y
Nam+Gumki out	N	N	N	N	Y	Dhom	N	N	N	N	Y
Num	Y	Y	Y	Y	Y	Dhi	N	N	N	N	Y
Num+dhom	N	N	N	N	Y	Others+Dhom	N	N	N	N	Y
Num+Gumki in	N	N	N	N	Y	T+Tha	N	N	N	N	Y

TABLE III: List of unique *aksharas* - DB-2 (multiple tonics) and DB-3

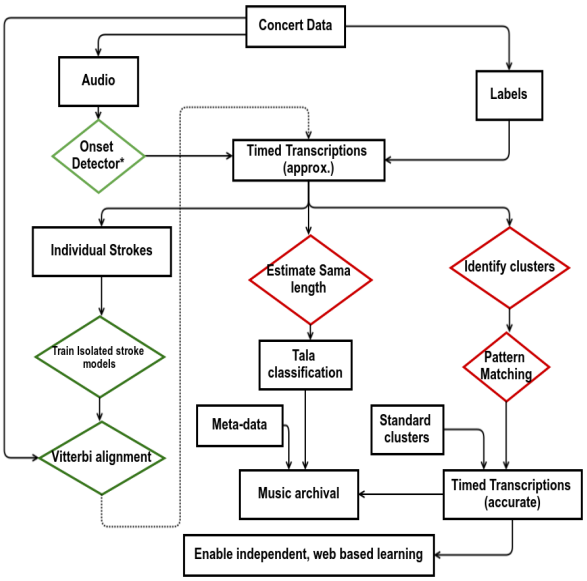

Fig. 1: Flowchart showing the mrudangam transcription details. The green boxes represent the current algo and the red ones represent the future possibilities in this task.

speech. Using the onset detection output, the phrases are divided into individual *strokes*. The individual *aksharas* are labeled based on the *strokes*' name given in the label file. Picture 3 shows the onsets in a phrase of mrudangam.

3) Feature Extraction: Mel-frequency cepstral co-efficients (MFCC) features are extracted using the HTK toolkit[16]. The frame size is 20 msec and the frame shift is 2 msec. The individual *strokes* are short in length and some are shorter than 10 msec, so the frame size is fixed to 20 msec. A 12-dimensional feature vector is extracted along with energy. These 13 features along with their velocity and acceleration co-efficients result in a 39-dimensional vector.

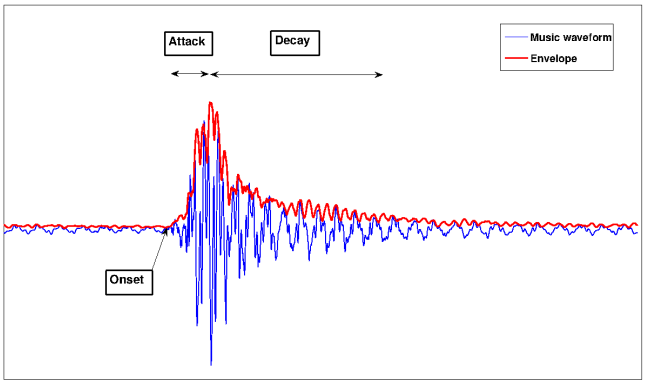

Fig. 2: Onset, attack and decay parts of the stroke

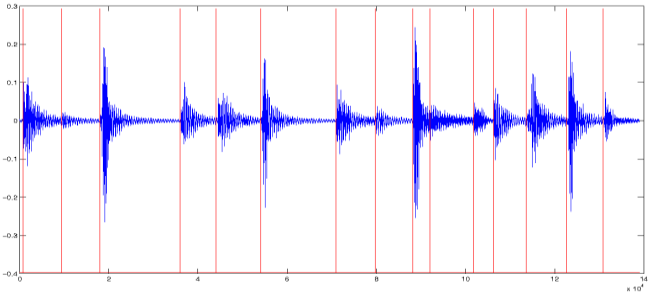

Fig. 3: Onsets marked for mrudangam strokes

4) Akshara Models: Each *akshara* is modeled using a 3-state 1-mixture HMM. The intuition behind taking 3 states is that every *stroke* has 3 parts; onset, attack and decay [17] as shown in figure 2. Each of these part of a stroke is modeled as a single state of the 3-state HMM.

5) Transcription using HMMs: The test phrases of percussion audio are split into individual *strokes* using the onset detector and each individual *stroke* is tested using the trained *akshara* HMM models.

6) Correction of Transcription using Language Model: The probability for individual *strokes* to belong to each of the *akshara* models is computed. These probabilities along with the language model is fed into the algorithm proposed in [18], which is a modification of Viterbi algorithm. It was observed that using the language model to perform stroke recognition improved the accuracy considerably. Figure 4 explains how HMM transcription output is corrected using the language model. Using the language model, probability of all possible paths are computed. Both bi-gram and HMM probabilities are used to find the optimal state sequence.

B. *Flat-start Embedded Re-estimation Hidden Markov Model*

In this method, onset detection is not performed. Here a phrase and its corresponding labels are used to train *akshara* HMM models in an embedded re-estimation framework. Here, the MFCC features are extracted with frame size 100 msec and frame shift 10 msec, unlike in the isolated training method.

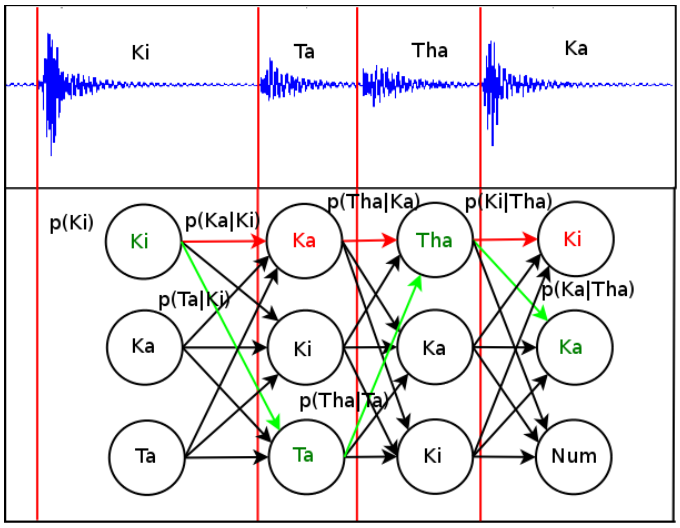

Fig. 4: Red arrows show the HMM transcriptions. Green arrows represent the transcriptions corrected using the language model. Red circles represent wrong transcriptions and the green circles represent the correct transcriptions.

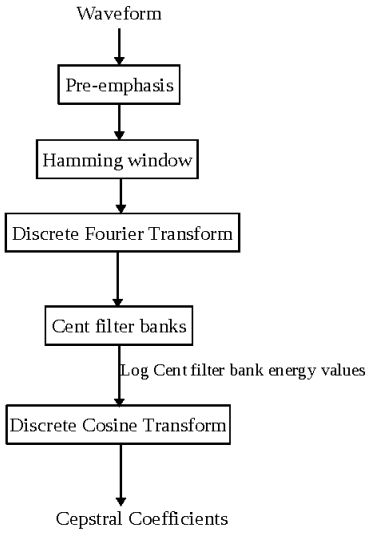

Fig. 5: Flowchart: Computing CFCC

The re-estimation and alignment of labels are performed approximately 10 to 20 times. The final model is used to find the transcriptions on test phrases using Viterbi alignment. The results are given in Table IV below.

C. HMM based Akshara Recognition using CFCC Features

CFCC [19] are tonic normalized features. Notes that make up a melody in Carnatic music are defined with respect to the tonic. The lead performer chooses the tonic, and the accompanying instruments are tuned to the same tonic. The tonic of accompanying instruments depends on the tonic chosen for the concert. For recognition of *aksharas* across different tonics, a tonic independent feature such as CFCC has to be used. CFCC are computed as given below:

1) The audio signal is divided into frames and the short-time discrete Fourier is computed for each frame.
2) The frequency (f) scale is normalised by the tonic. The cent scale is defined as:

$$cent = 1200 \cdot \log_2 \left(\frac{f}{tonic} \right) \quad (1)$$

3) The cent normalised power spectrum is then multiplied by a bank of filters that are spaced uniformly in the linear scale.
4) The filterbank energies are computed for every frame and used as a feature after removing the bias.

In equation (1), *tonic* can be estimated using pitch histograms, as proposed in [20]. The flowchart 5 explains how the CFCC features are extracted from waveforms.

Figure 6 shows the Fast Fourier Transform (FFT) spectral energies of Bheem *akshara* (from DB-2) of different tonics. The values for tonic B and tonic E are not in the same range. However, in the CFCC energy spectra shown in Figure 7, the values are exactly in the same range.

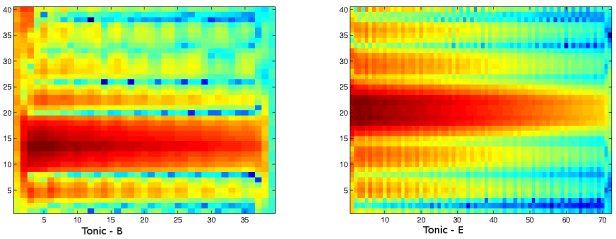

Fig. 6: Spectra for B and E tonics of the *stroke* Bheem using FFT Spectral Energies

VI. Results & Discussion

A. Results

The results of the transcription of *aksharas* on different datasets is given in this section. We show that isolated style recognition gives superior results over that of flat-start based embedded re-estimation method. This is primarily because, in flat-start based embedded reestimation, the duration of the stroke is assumed to be more or less uniform. The significant variation in the duration of a stroke results in poor performance. The accuracy increases significantly when the language model is used for correction of the transcriptions. The results are given in Table IV. Figure 8 is a *tani* phrase from DB-1 transcribed using the algorithm.

Language modeling is done on DB-1 and DB-2 datasets. The result is shown in Table VI. The Table V shows the corrections made by the language model over the isolated style output.

The CFCC features are used to train DB-2 using HMMs. The dataset is taken independent of tonic. The results obtained on DB-2 using tonic independent CFCC features is given in table VII. From the table, it is clear that tonic independent

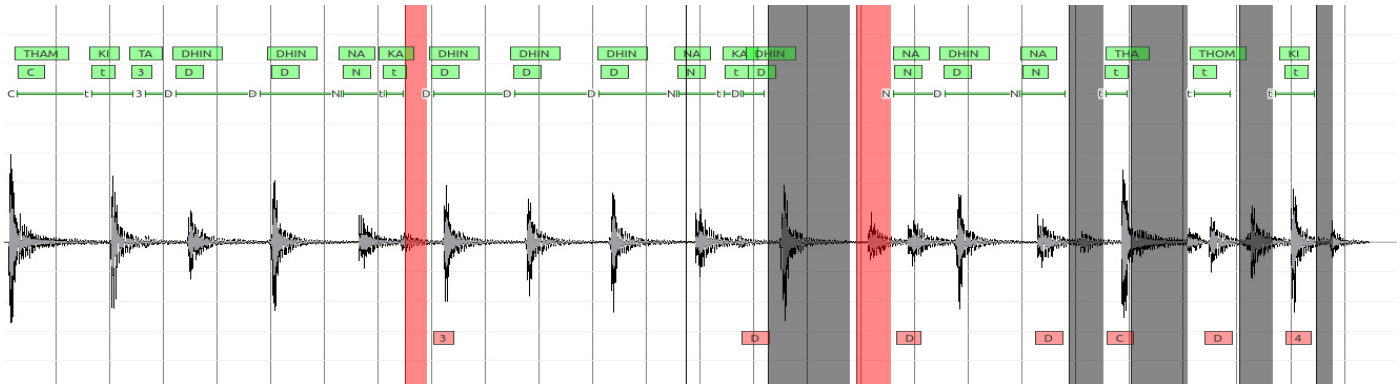

Fig. 8: Mrudangam transcription output. The red segments are insertions and the black segments are substitution. A example phrase from DB-1.

Original Label	Ta	Thom	Ki	Ta	Tha	Ka	Ta	Tha	Cha	Tha	Ki	Ta	Tha	Ka
Isolated Style Recognised Labels	Ka	Thom	Ki	Ta	Tha	Ki	Ki	Tha	Cha	Tha	Ki	Ka	Tha	Ki
Labels Corrected using Language Model	Ta	Thom	Ki	Ta	Tha	Ki	Ta	Tha	Cha	Tha	Ki	Ta	Tha	Ka

TABLE V: Transcriptions corrected using language modeling. The red *aksharas* in second row are recognised wrong in the HMM results. The green colour *aksharas* in the third row represent the corrections obtained using the language model for an example phrase from DB-2.

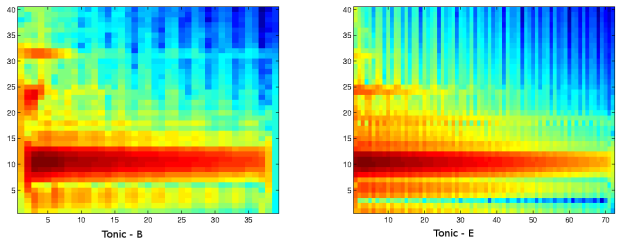

Fig. 7: Spectra for B and E tonics of the *stroke* Bheem using Cent Filterbank Energies

Dataset (Tonic)	**Accuracy**	
	Language Modeling	**Isolated Style**
DB-2 (B)	86.58 %	83.58 %
DB-2 (C)	90.32 %	82.44 %
DB-2 (D)	87.93 %	78.23 %
DB-2 (E)	95.8 %	91.26 %
DB-1 (C#)	75.68 %	72.30 %

TABLE VI: Comparison of results of language modeling and isolated style recognition for DB-1 and DB-2 using MFCC features.

Dataset(Tonic)	**Accuracy**	
	Isolated Style	**Embedded Re-estimation**
DB-2 (B)	83.58%	71.12%
DB-2 (C)	82.44%	55.61%
DB-2 (D)	78.23%	59.30%
DB-2 (E)	91.26%	63.16%
DB-1 (C#)	72.30%	49.32%
DB-3 (C#)	50.36%	39.24%

TABLE IV: Comparison of *stroke* recognition performance in different datasets using MFCC features.

features do improve performance significantly. Constant Q transform (CQT) corresponds to the features used in [7], where the linear spacing in frequency is converted to spaced in $log_2 f$. Although CQT features are also tonic independent, the range of frequencies have to be chosen carefully for every tonic. For CFCC, six octaves ranging from -1200 cents to 6000 cents were chosen across all tonics. The poor performance of CFCC (column 3) is attributed to the fact that some of the strokes are independent of tonic, while the features have been tonic normalised.

B. Discussion

Classification results suggest that HMMs model individual *aksharas* fairly well. The studio-recorded dataset (DB-2) in particular, shows better accuracy when compared to the concert-recorded *taniavarthanams* (DB-1 & DB-3). This can be attributed to the higher tempo and significant variations in

Feature	Trained on all tonics accuracy (%) (Tonic Variant)	Train one tonic, test all tonics(%) (Tonic Invariant)
MFCC	60	50
Chroma	41	40
CQT	74	62
CFCC	77	66

TABLE VII: Accuracy of transcription using CFCC features in DB-2 and comparison with other features.

loudness in *taniavarthanams*. Further, it becomes a challenging task to differentiate between the left and right constituents, or between multiple fingers in composite *stroke*s, which are frequent in the instruments considered. However, a pre-processing step such as onset detection improves the recognition accuracy across all datasets.

In Figure 8, the first red segment shows a insertion of stroke 'DHI'. This corresponds to a composite stroke. The onset detection algorithm distinguishes between the strokes played on the right and left sides of the mrudangam. Further processing is required to eliminate such errors.

The accuracy of detection of *aksharas* at faster tempo is poor as expected. Duration modeling is perhaps required in the HMM, wherein duration information can also be included as part of the context. The language model captures the bigrams probabilities in the labels and corrects the transcriptions. The language model in essence captures the creativity of the musician in question.

VII. Conclusions & Summary

In the current work, mrudangam in Carnatic music is chosen and its stroke characteristics are studied on a fairly diverse database. During the creation of the database, it was observed that the variety and number of types of strokes that can be played varies quite significantly across musicians. Composite strokes are the basic reason for this variety and increased number of strokes in the datasets. New techniques to model these composite strokes using the basic strokes have to be developed, so that a common set of stroke models can be learned for different artists. A unified framework across different musicians and instruments is indeed required for enhancing musical experience in MIR.

VIII. Acknowledgement

This research was partly funded by the European Research Council under the European Unions Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583). The authors are grateful to renowned mrudangam artist Umayalpuram K Sivaraman's disciples, B Sundar, N Hariharan, and S Ramanathan, for providing the dataset DB-1, and for manually transcribing mrudangam *aksharas* for the purpose of stroke recognition. The authors would like to thank Manoj Kumar, Jilt Sebastian, Sridharan Sankaran and Saketh M. for their help and support.

References

[1] "Carnatic music - wikipedia, the free encyclopedia," http://en.wikipedia.org/wiki/Carnatic_music/.

[2] M. Kumar, J. Sebastian, and H. A. Murthy, "Musical onset detection on carnatic percussion instruments," in *National Conference on Communications 2015 (NCC-2015)*, Mumbai, India, Feb. 2015.

[3] "Mridangam - wikipedia, the free encyclopedia," http://en.wikipedia.org/wiki/Mridangam.

[4] C. V. Raman, "The indian musical drums," in *Proceedings of the Indian Academy of Sciences-Section A*, vol. 1, no. 3. Springer, 1934, pp. 179–188.

[5] R. Siddharthan, P. Chatterjee, and V. Tripathi, "A study of harmonic overtones produced in indian drums," in *Physics Education*, 1994, pp. 304–310.

[6] A. Anantapadmanabhan, A. Bellur, and H. Murthy, "Modal analysis and transcription of strokes of the mridangam using non-negative matrix factorization," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013, pp. 181–185.

[7] A. Anantapadmanabhan, J. Bello, R. Krishnan, and H. Murthy, "Tonic-independent stroke transcription of the mridangam," in *Audio Engineering Society Conference: 53rd International Conference: Semantic Audio*, Jan 2014. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=17102

[8] P. Chordia, "Segmentation and recognition of tabla strokes," in *International Conference on Music Information Retrieval (ISMIR)*, Queen Mary, University of London, United Kingdom (UK), 2005.

[9] P. Chordia, A. Sastry, and S. Şentürk, "Predictive tabla modelling using variable-length markov and hidden markov models," *Journal of New Music Research*, vol. 40, no. 2, p. 105–118, 2011.

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

[11] "Compmusic: Computational models for the discovery of the worlds music," http://compmusic.upf.edu/.

[12] "Mridangam tani-avarthanam dataset - compmusic," http://compmusic.upf.edu/node/228/.

[13] "Mridangam stroke dataset - compmusic," http://compmusic.upf.edu/mridangam-stroke-dataset.

[14] "Statistical language modeling toolkit," http://www.speech.cs.cmu.edu/SLM/toolkit.html.

[15] S. A. Shanmugam and H. A. Murthy, "Group delay based phone segmentation for HTS," in *National Conference on Communications 2014 (NCC-2014)*, Kanpur, India, Feb. 2014.

[16] CUED, "HTK Speech Recognition Toolkit," `http://htk.eng.cam.ac.uk`, Cambridge University Engineering Dept., 2002.

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

[18] R. Janakiraman, J. Kumar, and H. Murthy, "Robust syllable segmentation and its application to syllable-centric continuous speech recognition," in *Communications (NCC), 2010 National Conference on*, Jan 2010, pp. 1–5.

[19] P. Sarala and H. A. Murthy, "Cent filter banks and its relevance to identifying the main song in carnatic music," in *International Symposium on Computer Music Multidisciplinary Research*, 2013.

[20] A. Bellur and H. A. Murthy, "A novel application of group delay functions for tonic estimation in carnatic music," in *European Signal Processing Conference (EUSIPCO)*, September 2013, pp. Th–L1.4.