MODAL ANALYSIS AND TRANSCRIPTION OF STROKES OF THE MRIDANGAM USING NON-NEGATIVE MATRIX FACTORIZATION

Akshay Anantapadmanabhan[1], *Ashwin Bellur*[2] *and Hema A Murthy*[1] *

[1] Department of Computer Science and Engineering,
[2] Department of Electrical Engineering,
Indian Institute of Technology, Madras, India - 600 036

ABSTRACT

In this paper we use a Non-negative Matrix Factorization (NMF) based approach to analyze the strokes of the mridangam, a South Indian hand drum, in terms of the normal modes of the instrument. Using NMF, a dictionary of spectral basis vectors are first created for each of the modes of the mridangam. The composition of the strokes are then studied by projecting them along the direction of the modes using NMF. We then extend this knowledge of each stroke in terms of its basic modes to transcribe audio recordings. Hidden Markov Models are adopted to learn the modal activations for each of the strokes of the mridangam, yielding up to 88.40% accuracy during transcription.

Index Terms— Modal Analysis, Mridangam, automatic transcription, Non-negative Matrix Factorization, Hidden Markov models

1. INTRODUCTION

The mridangam is the primary percussion accompaniment instrument in carnatic music, a sub-genre of Indian classical music. The mridangam, along with percussive instruments like the tabla and congos, falls under the rare category of pitched percussive instruments. Unlike the western drums which cannot produce harmonics, pitched percussive instruments like the mridangam have significant harmonic properties [1, 2]. There have been numerous efforts to analyze and characterize percussion instruments. However, the focus has been mainly on unpitched percussion instruments, mostly in the context of Western music or for isolated percussion timbre recognition [3, 4, 5, 6].

Recently there have been some efforts to study pitched percussive instruments, especially the tabla [7, 8], a variant of the mridangam. Given that the tabla can produce a number of strokes with unique timbre, both [7] and [8] study the task of automatic transcription of tabla performances. Both works address the task of capturing timbre and building classifiers to transcribe strokes by leveraging timbre.

*Thanks to Charsur Ditigal Workstation for providing studio space along with recording equipment for all experiments and recordings. 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 mridangam is quite different from the tabla in that, the mridangam is a single body instrument with two membranes (one producing treble sounds while the other producing bass sounds) as opposed to the tabla, which consists of two independent bodies. C.V. Raman, in his work on Indian musical drums [1], discusses some of the unique traits of the mridangam as a harmonic percussive instrument. He describes some of the structural similarities between the mridangam and tabla but at the same time highlights some of the major differences in their acoustic properties. He also illustrates the modes of the mridangam using sand figures to reveal the basic physical and acoustic characteristics of the instrument.

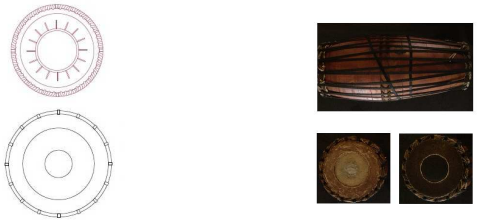
In this paper, we extend Raman's modal analysis to study the strokes produced by the mridangam. We propose a Non-negative Matrix Factorization (NMF) based approach to analyze the strokes of the mridangam in terms of the basic modes as described in [1]. We first create a dictionary of basis vectors representing the fundamental modes of the mridangam using NMF. Then we study the modal activations for each stroke by projecting the strokes along the dictionary basis vectors using NMF. We then extend this knowledge of the decomposition of strokes into basic modes to transcribe audio recordings by a professional mridangam artist. Hidden Markov Models (HMM) are adopted to learn the modal activities for each of the strokes.

The organization of the paper is as follows. Section 2 gives a brief introduction to the mridangam followed by a preliminary description of the strokes that can be produced by the instrument. Section 3 describes the process of building the modal dictionary using NMF. Modal analysis of the strokes is also performed in this section. The relevance of NMF for identifying strokes of the mridangam is illustrated using a few examples. In section 4, the task of transcription using NMF along with HMMs is addressed. Section 5 details the results of the methods proposed for the task of transcription. The paper is summarized and concluded in section 6.

2. INTRODUCTION TO THE MRIDANGAM

The mridangam has been noted in manuscripts dated as far back as 200 B.C. and has evolved over time to be the most prominent percussion instrument used in South Indian classical music [9]. The mridangam has a tube-like structure made from jack fruit tree wood covered on both ends by two differ-

ent membranes. Unlike the western drums which cannot produce harmonics due their uniform circular membranes [10], the mridangam is loaded at the center of the treble membrane (*valanthalai*) resulting in significant harmonic properties with all overtones being almost integer ratios of each other[1, 2]. The bass membrane (*thoppi*) is loaded at the time of performance, increasing the density of the membrane, propagating a bass like sound [1]. Figure 1 depicts a technical draft as well as an image of a mridangam, illustrating the top, bottom and cross-sectional/body views of the instrument.



(a) Technical Draft (b) Three Views

Fig. 1: (a) Technical draft of treble (top) and bass (bottom) membranes of mridangam. (b) Body (top), bass membrane (bottom left), and treble membrane (bottome right) of the mridangam

The two membranes of the mridangam produce many different timbres. Many of these sounds have been named, forming a vocabulary of timbres. Mridangam sounds can be roughly classified into the following three major sound groups:

1. Ringing string-like tones played on the treble membrane. *Dhin*, *cha* and *bheem*[1] are examples. These tones are characterized by a distinct pitch, sharp attack and long sustain.
2. Flat, closed, crisp sounds. *Thi* (also referred to as *ki* or *ka*), *ta* and *num* are played on the treble membrane and *tha* is played on the bass membrane. These tones are characterized by an indiscernable pitch, sharp attack and almost immediate decay.
3. Resonant strokes are also played on the bass membrane (*thom*). This tone is not associated with a specific pitch and is characterized by a sharp attack and a long sustain.

The strokes mentioned above cover all possible single handed syllables that can be played on the mridangam. Composite strokes (played with both hands) are two strokes played simultaneously: *tham* (*num* + *thom*) and *dheem* (*dhin* + *thom*).

3. MODAL ANALYSIS OF MRIDANGAM STROKES USING NMF

In [1], Raman hypothesizes the mridangam to be a *harmonic drum*, whose modes can be excited in isolation, analogous to the harmonic modes of a stretched string which can be excited by plucking at its nodal points. He then describes ways to excite the five modes of treble membrane in isolation by striking the instrument while placing his fingers appropriately at points of nodal chords of the circular membrane. Finally, he validates the modes using sand figures.

Our first experiment was to replicate Raman's approach to the modal analysis of the treble membrane of the mridangam. Recordings were made in a semi-anechoic recording studio using Shure SM-58 microphones and an H4n ZOOM recorder at 44100 Hz. The modes were excited individually by following Raman's finger placements from the sand figures in his work. The fifth mode, which has no image in his paper, was excited by following his written descriptions. Just as it was hard for Raman to obtain good sand figures because of the short duration of the fifth tone, it was also tough for us to obtain a proper recording of that mode. Therefore, the fifth mode is not used in the following analysis. Each recording was cropped to capture only the moment after the drum was struck and the respective mode excited. We hypothesize that these modes are the basic sound units that can define any strokes played on the treble and bass membrane of the mridangam. We assume that strokes played on the bass membrane will excite the treble membrane modes because of the coupling between the two membranes. In order to test these assumptions, we propose a Non Negative Matrix Factorization (NMF) technique.

The methodology to study spectral profiles of harmonically static signals using NMF was first introduced by Smaragdis and Brown [11]. Various adaptations of NMF [12, 13, 14] have then been developed for the purpose of polyphonic music transcription. NMF is a technique using which a non-negative matrix X can be decomposed into two non-negative matrix factors B and Y such that

$$X \approx BY \tag{1}$$

Given X is of dimension $m \times t$, then B and Y are of dimensions $m \times n$ and $n \times t$ respectively. Generally $n < t$. This implies that x_i, the i^{th} column of X, can be represented as a linear combination of basis vectors – the columns of B

$$x_i = \sum_{j=1}^{n} y_{ji} b_j \tag{2}$$

where $y_{j,i}, j = 1, \cdots .n$, representing the i^{th} column of Y, are weights estimated for the linear combination of columns of B. There are numerous algorithms to estimate B and Y, depending on the metric used to quantify the approximation in equation 1. We use the popular euclidean measure and multiplicative update rules proposed in [15] to iteratively estimate B and Y.

In the context of this work, matrix X represents the spectrogram of an audio signal, divided into t frames. x_i represents the magnitude spectrum vector of frame i and $b_j, j = 1, \cdots, n$ are spectral basis vectors that best describe the main components of X. We propose using the NMF technique in a

[1]The name of this stroke varies between different schools of mridangam

two-step procedure to perform modal analysis of the strokes of the mridangam.

1. Creating a dictionary of the modes: Let X_i represent the spectrogram of the i^{th} mode recorded. An FFT of order $m = 2048$ and a hop size of 10ms was used to compute the spectrogram. Matrix factors B_i and Y_i are then iteratively estimated with $n_i = 1$, implying that the number of columns of $B_i = 1$. This implies that we are forcing a single spectral vector B_i to represent the main elements of X_i. A dictionary matrix $D = [B_1\ B_2\ B_3\ B_4]$ of dimension $m \times 4$ is then created. The columns of D concisely represent the 4 recorded modes.
2. Project strokes in the direction of the modes: For this purpose, each stroke on the treble head of the instrument was played in isolation by a professional artist and recorded. Let X now represent the spectrogram for the stroke S (keeping the FFT order and hop size the same). B is initialized by the dictionary D created in step 1. Using multiplicative update rules, Y is iteratively estimated while B is kept constant. If X is of dimension $m \times t$ and given that $B = D$ is of dimension $m \times 4$, Y will be of dimension $4 \times t$. Each of the 4 rows of Y illustrate the strength and temporal structure of the activation of the respective modes from the point of onset until the complete decay of the stroke. We shall refer to B as the basis matrix and Y as the activation matrix.

The waveform, spectrogram and the modal activations for each of the strokes played on the treble head, can be seen in Figure 2. The four rows below the spectrogram are the rows of the activation matrix. The strokes recorded in isolation are displayed adjacent to each other in Figure 2, in order to compare their modal activations.

The stroke *bheem* is an open stroke smartly struck with the pointer finger at the center of the membrane, causing a ringing sound. Hence the first mode which is an open mode gets excited with a long sustain (Figure 2) while the other basis vectors are completely suppressed. The stroke *ta* is played in the same spot but is more heavily struck. It is a closed stroke, therefore the first mode appears damped in Figure 2 for *ta* when compared to the stroke *Bheem*. Since it is a flat sounding stroke with no audible pitch, the excitation is also spread across the other basis vectors.

The stroke *thi* is actually described as one way to excite the first mode in Raman's work. It is claimed in [1] that this approach of modal excitation causes overtones. Figure 2 shows that the first mode is dominant for the stroke *thi* with third harmonic also excited validating Raman's claim about the existence of overtones.

The stroke *Cha* is struck across the nodal diameter of the instrument, and therefore corresponds to the configuration of the second mode of excitement from [1]. Figure 2 confirms this by depicting the second mode as being dominant with the first mode also partially excited. The stroke *dhin* is played with one hand but is similar to the two-handed configuration Raman uses to excite the second mode. As can be seen in Figure 2, the first mode is initially activated but is followed by the activation of the second mode, which has a significantly slower decay rate than the first.

The stroke *num* is a closed, flat stroke struck outside the black circle of instrument. Because the stroke is played across multiple nodal lines, it is logical that the second mode, which vibrates to the left and right of any diameter of the membrane, should definitely be suppressed with other modes active as confirmed by the modal activations under the stroke *num* in Figure 2.

Tha and *thom* are played on the bass membrane of the mridangam. However, because of the coupled nature of the instrument, modes on the treble membrane are activated. This is something Raman does not discuss but is validated by the modal activations in Figure 2. Since the treble membrane is open when the strokes on the bass membrane are played, the first mode (open mode) should definitely be excited for both the strokes as confirmed by Figure 2. Furthermore, since *tha* is a closed note, and *thom* produces a bass sound, *tha* activates the higher harmonics of instrument as shown in the figure.

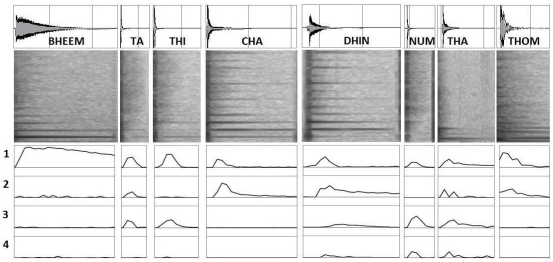


Fig. 2: Waveform (top) and spectral images (middle) of each stroke of the mridangam and their respective modal activations (bottom)

4. TRANSCRIBING STROKES USING NMF AND HMMS

From the analysis in Section 3, it is evident that every stroke, when played in isolation, has distinct modal excitations. This implies that the activations of the modes can be used to determine the identity of the stroke played. To test this hypothesis, two solo performances were recorded by a professional mridangam artist using two instruments tuned to two keys. The modes for each instrument were also recorded as previously described. The solos were then cropped into commonly identifiable phrases (a logical set of strokes) to form two databases for transcription. The instrument tuned to D# consisted of 134 phrases and the instrument tuned to E had 114 phrases. The databases consisted of strokes played individually on both the bass and treble membranes, as well as composite strokes. We also recorded an open mode for the bass membrane by paralleling Raman's configuration for exciting the open mode

Table 1: Stroke Confusion Matrix

Strokes	4 modes										5 modes									
	Bheem	Cha	Dheem	Dhin	Num	Ta	Tha	Tham	Thi	Thom	Bheem	Cha	Dheem	Dhin	Num	Ta	Tha	Tham	Thi	Thom
Bheem	**30**	0	0	0	0	0	0	0	0	1	**31**	0	0	0	0	0	0	0	0	0
Cha	0	**65**	6	2	0	0	3	0	0	0	0	**65**	4	3	0	0	3	1	0	0
Dhin	0	0	**65**	22	0	0	0	1	0	0	0	0	**87**	1	0	0	0	0	0	0
Dheem	0	11	32	**90**	2	0	0	1	0	0	0	1	32	**99**	4	0	0	0	0	0
Num	0	0	10	0	**85**	0	1	32	8	0	0	0	7	15	**91**	0	0	10	13	0
Ta	0	1	0	0	3	**71**	5	20	7	1	0	3	0	4	2	**54**	13	5	27	0
Tha	0	0	1	0	1	5	**99**	15	7	0	2	0	1	0	0	0	**119**	0	6	0
Tham	0	0	3	0	7	2	7	**80**	0	0	0	0	9	1	1	2	0	**83**	2	1
Thi	0	7	3	2	5	31	2	9	**148**	1	0	9	0	2	6	27	6	3	**154**	1
Thom	4	8	8	0	0	1	5	2	3	**129**	0	0	0	0	0	0	3	7	1	**149**

on the treble membrane. This was done to understand the strength of coupling between the bass and treble membranes and if the lack of a bass membrane mode actually reduces transcription accuracy.

Although it was previously confirmed that each stroke in isolation has distinct modal activations, the strength and temporal structure of the activation cannot be assumed to be the same when the strokes are played within a context. The strength and shape (to some extent) of a stroke is affected by the preceding stroke and the tempo of the performance. In order to capture the invariance in the activations for a given stroke played in various contexts, Hidden Markov Models (HMMs) were built. The following procedure was used to automatically extract features and build HMM models for each of the strokes:

- The spectrogram for each phrase in the databases was computed. Using NMF, the activation matrix was estimated by projecting the spectrogram along the direction of the 5 modes (4 modes if the bass mode is not included). In order to determine the onset points, the activation matrix was summed row wise and peaks were picked from the resulting row vector. The number of peaks picked correspond to the number strokes played within the phrase.
- Once the onset points are known, the original activation matrix is divided into sub matrices at the onsets. To prevent activations of succeeding strokes from being included in the feature vector, only eighty percent of the frames from onset to onset was used. This sub matrix can be thought of as a five dimensional feature vector of varying length, each row representing a dimension.
- Feature vectors for each of the ten stroke in the database were then pooled to build, ten continuous-density Hidden Markov Models. After experimenting with Hidden Markov Models with various configurations, three-state, single-mixture HMMs were found to be optimal to represent each stroke.

5. RESULTS AND ANALYSIS

A four-fold cross-validation was performed with 75% of the data used for training and 25% for testing. Models were built using training data as explained in the previous section. Transcription was performed using a 1) four-dimensional feature vector – only projecting onto treble membrane modes and 2) five-dimensional feature vector – including the bass membrane mode as well.

Table 2: Accuracies in %, S - single side stroke, S+C - Single + Composites strokes

Instrument	4 modes		5 modes	
	S	S + C	S	S + C
D#	82.27	72.61	78.26	73.24
E	84.69	74.95	88.40	87.38

Table 2 reports the accuracies for transcription of both instruments in the database. Column S indicates an 8-class classification problem where only the single side strokes are transcribed. Column C indicates a 10-class classification problem with composite strokes also included. As expected, performance drops when composite strokes are included during transcription. As can be seen in the confusion matrix in Table 1 (compiled by consolidating results for the both instruments) there is substantial confusion between strokes *dheem* and *dhin* and also between *num* and *tham*. This is understandable given that *tham* and *dhin* are composed of *num* and *dheem* respectively. Transcription was possible even without the bass mode, although including the bass mode did improve performance. This confirms that there is a coupling effect between the bass and treble membrane that allows for unique treble membrane activations when bass membrane strokes are played.

6. CONCLUSION

We have extended Raman's analysis of the modes of the mridangam by validating the relationship between strokes and modes of the instrument. We have also demonstrated that these modes can be used for transcription. The shortcoming of this approach is that it requires the modes for each instrument – only strokes without a discernible pitch are invariant to instrument modes. In future work we anticipate addressing this issue by applying a transformation to a set of modes to use them for transcribing strokes of the mridangam tuned to any key.

7. REFERENCES

[1] C V Raman, "The Indian musical drums," *in Proc. Ind. Acad. Sci.*, pp. 179–188, 1934.

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

[3] P Herrera, A Yeterian, R Yeterian, and F Gouyon, "Automatic classification of drum sounds: a comparison of feature selection and classification techniques," *in Proc of Second Int. Conf. on Music and Artificial Intelligence*, pp. 69–80, 2002.

[4] P Herrera, A Dehamel, and F Gouyon, "Automatic labeling of unpitched percussion sounds," *in Proc. of the Audio Engineering Society, 114th Convention*, pp. 1378–1391, 2003.

[5] A Tindale, A Kapur, G Tzanetakis, and I Fujinaga, "Retrieval of percussion gestures using timbre classification techniques," *in Proc. of ISMIR*, pp. 541–545, 2004.

[6] V Sandvold, F Gouyon, and P Herrera, "Percussion classification in polyphonic audio recordings using localized sound models," *in Proc. of ISMIR*, pp. 537–540, 2004.

[7] O K Gillet and G Richard, "Automatic labelling of tabla signals," *in Proc. of ISMIR*, pp. 117–124, 2003.

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

[9] S Gopal, *Mridangam - An Indian Classical Percussion Drum*, B.R. Rhythms, 425, Nimri Colony Ashok Vihar, Phase-IV, New Delhi, 2004.

[10] S S Malu and A Siddharthan, "Acoustics of the Indian drum," *arXiv:math-ph/0001030*, 2000.

[11] P Smaragdis and J C Brown, "Non-negative matrix factorization of polyphonic music transcription," *in Proc. IEEE Workshop on Applicaion of Signal Processing to Audio and Acoustics*, pp. 177–180, 2003.

[12] T Virtanen and A Klapuri, "Analysis of polyphonic audio using source-filter model and non-negative matrix factorization," *in Advances in Neural Inf. Process. Syst*, 2006.

[13] E Vincent, N Berlin, and R Badeau, "Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription," *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, pp. 109–112, 2008.

[14] G Grindlay and D P W Ellis, "Multi-voice polyphonic music transcription using eigeninstruments," *in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, WASPAA*, pp. 53–56, 2009.

[15] D D LEE and H S Seung, "Algorithms for nonnegative matrix factorization," *In Neural Inf. Process. Syst*, pp. 556–562, 2001.