A Matrix Based Approach for Evaluation of Vocal Renditions in Hindustani Classical Music

Kunjal Gajjar[a], **Mukesh Patel**[b]

[a] **Faculty, School of Computer Studies, Ahmedabad University, Ahmedabad, 380009, India, kunjal.gajjar@ahduni.edu.in**

[b] Adjunct Professor, Charotar University of Science and Technology (CHARUSAT), Changa, 388421, India, mukeshpatel.mca@charusat.ac.in

Abstract

Indian or Hindustani Classical Music (ICM and HCM respectively) is based on the Raga system of music. Unlike a Western Classical Music (WCM) composition (usually documented in a written score format), a Raga defines an overall melodic structure to which a vocal or instrumental rendition adheres to. HCM has around 200 Ragas, each of which is specified in term of a subset of notes (that can be part of a composition in that particular Raga) as well as their sequencing (e.g., aroha-avaroha sequence), combinations and permutations. The rules and conventions (here under referred to syntax of Raga) together determine the general melodic structure (often referred to as a mood that the Raga aims to capture) of a rendition. Hence, a computational model for recognition or evaluation of HCM renditions would need to be able to handle a much higher degree of complexity than a similar system for WCM. Here we present a computational model that handles this complexity elegantly and efficiently for 8 Ragas (that are typically learned by novice singers). More specifically we show how a relatively simple note transition matrix based approach incorporating key elements of a Raga's syntax results in highly reliable evaluation of songs and robust error identification (for feedback to novice learners as part of computer based tutoring system).

KEYWORDS

Computational Musicology, Hindustani Classical Music, Indian Classical Music, Raga Evaluation, Transition Matrix, Computational Model, Note, Tutoring System, Novice, Early Stage Learners, Vocal Raga

1. INTRODUCTION

Computational musicology, an interdisciplinary research area is exciting and challenging field which motivates research for music information retrieval, music composition, recognition and analysis of extracted features (Volk et al. 2011). Since 1960's there has been reasonable amount of work done for western classical music (WCM) in this field. However, Indian or Hindustani Classical Music (ICM and HCM) is still relatively unexplored (Pandey et al. 2003) (Agarwal et al. 2013). One main reason for this limited attention is the inherent complexity of HCM resulting from the fact that music is generated based on loosely defined Ragas.

Unlike Western Music, ICM songs or instrumental recitals conform to the structure of a Raga which is determined by rules and conventions about sequencing, combinations and permutations of notes which together determine the overall melodic structure by specifying general boundaries and constraints. It is important to note that these rules and conventions do not determine any specific rendition in a Raga but can be said to define its syntax. Hence, a singer/musician does not compulsorily recite a pre-composed piece of music but renders a version of it; typically by putting more emphasis on certain phrases or repeating some sequence of notes more often than others(Mukerji 2014; Improvisation 2014; Improvisation in Indian Classical music 2012). From the computational point of view such latitude translates into a level of complexity that is relatively difficult to account for in computational models of HCM renditions.

Despite this complexity, attempt to develop computational models of HCM have been made by several researchers. Most such studies have focused on recognition of recitals in various Ragas and automated classification of the music for retrieval systems. The general approach has been to incorporate the rules and conventions as general heuristics. In the model presented here we have extended this approach to a more systemic modelling of the syntax of eight selected Ragas such that it evaluates songs (according to the syntax of the Raga in which it is composed) and identifies any errors in the rendition.

The Raga syntax based computer model of evaluation is part of a computer based tutoring system designed for early stage learners (novices) of vocal HCM which also includes a note transcription module (described in Section 3.1). Typically during the early stage (defined as the first two years) of the learning process, a novice is confined to singing practice songs in the presence of tutor who engages in continuous evaluation of the novice's rendition and providing feedback on wrongly rendered notes/swars or violation of rule such as rendering the aroha sequence wrong. This is obviously a time consuming process (particularly as rarely can it involve more than one learner at a time). Hence, the motivation for exploring the possibility of developing a computer based tutoring system that automates as much of the initial stage of learning process as possible. In particular, it was felt that such a system needs to have robust and reliable process for evaluating whether a song has been rendered correctly, and provide feedback on how to correct any errors with respect to the rules and conventions of the Raga.

In the next section we describe the key elements of HMC that determine the melodic structure of Ragas, and how that translates into essentially non-deterministic and/or unconstrained nature of songs or musical renditions. In Section 2 we present a brief review of relevant previous work, and the scope of the tutoring system with particular emphasis on the evaluation model incorporated in the tutoring system.

1.1 Hindustani Classical Music and Raga Structure

Hindustani classical music (HCM) is one form of Indian Classical Music (ICM) that is primarily practiced in northern and central parts of India. The uniqueness of HCM is in its melodic structure which is determined by a combination of a Raga - musical formalisms and melodic phrases - and beat that determines the tempo (Rabunal 2005). Many of these aspects of Ragas have been codified in detail by Pandit Vishnu Narayan Bhatkhande (Bhatkhande 1970). Here we describe the key musical elements, rules and conventions that collectively determine the structure of each Raga. Vocal and instrumental music (e.g., songs and solo recitals) are usually composed to conform to the structure or framework of specific Ragas.

The tonal material used in compositions known as swars (notes) of HCM scale are; Shadaj (Sa), Rishabh (Re), Gandhar (Ga), Madhyam (Ma), Pancham (Pa), Dhaivat (Dha) and Nishad (Ni). The scale contains these seven natural notes referred to as shuddha (literally pure) swars. However, each swar may also have two different forms which are komal (flat notes) and tivra (sharp notes) except 'Sa' (the tonic) and 'Pa' (the perfect fifth) which are rendered in only one form. Together the following 12 notes make up the scale of HCM Ragas in which music and songs are composed or recited:

Shuddha swar – 7 normal swars: Sa, Re, Ga, Ma, Pa, Dha, Ni

Komal swar – 4 swars, Re, Ga, Dha, Ni can also be rendered in komal form, that is their frequency is lower than as a shuddha swar.

Tivra swar – 1 swars, Ma' can also be rendered in tivra form, that is, its frequency is higher than shuddha Ma.

Collectively these 12 swars are referred to as saptak. Unlike WCM, HCM uses a movable scale which means that there is no fixed pitch for a note. An octave can start from any pitch which is known as tonic frequency (frequency of 'Sa' of middle octave) (Rao and Rao. 2014) (Indian classical music and sikh kirtan 1995). In a rendition a singer/musician can shift across three saptak - Lower (Mandra), Middle (Madhyam) and Upper (Taar). In comparison to madhyam saptak, mandra saptak has lower pitch frequencies and taar saptak has higher pitch frequencies. All the notes in these three saptaks are defined in the relation to their tonic frequency. The frequency ratio with 'Sa' of middle octave (madhyam) is shown in Table 1. For more details see Gajjar and Patel (Gajjar and Patel 2017).

Western Notation	Indian Notation	Notation Used in Proposed Model	Frequency Ratio (natural)	Names
C	Sa	S	1/1	Tonic Sa
C#	Re	R	16/15	Komal Re
D	Re	R	9/8	Shuddha Re
Eb	Ga	G	6/5	Komal Ga
E	Ga	G	5/4	Shuddha Ga
F	Ma	M	4/3	Shuddha Ma
F#	Ma’	M	7/5	Tivra Ma
G	Pa	P	3/2	Perfect Fifth Pa
Ab	Dh	D	8/5	Komal Dh
A	Dh	D	5/3	Shuddha Dh
Bb	Ni	N	9/5	Komal Ni
B	Ni	N	15/8	Shuddha Ni
C’	Sa’ (Next octave)	S_U	2/1	Sa of upper octave

Table 1. WCM and HCM notes with frequency ratio (Rao and Rao. 2014)

The basic structure of each Raga is determined by the selected swars from the saptak (listed in Table 1). Each Raga has minimum of 5 and maximum of 7 swars. It is important to understand that a Raga is not just a sequence of various combination of a set of swars: In addition, a Raga’s structure is determined by rules and conventions that specify do’s and don’ts regarding the sequencing and combination of swars as well as emphasis on certain sub-set (phrases) of notes.
The important conventions are summarized in the list below, that collectively determine the melodic structure (or essence or mood) of a Raga (Rao and Rao 2014)(Dighe et al. 2013)

- Aroha-Avaroha - Aroha is the ascending order sequence of swars that describes how the Raga moves in its ascending order. Avaroha is the descending order sequence of swars that describe how the Raga moves in descending order. For example in Raga Bhimpalasi:

  Aroha sequence is: Sa, Ga, Ma, Pa, Ni, Sa’
  Avaroha sequence is: Sa’, Ni, ***Dha,*** Pa, Ma, Ga, ***Re,*** Sa

As can be seen swars, ‘Re’ and ‘Dha’ can be rendered in descending (avaroha) order, however, they cannot be used in ascending (aroha) order.

- Pakkad – Phrase (or sequence) of notes that are specific to a Raga. Its usage or inclusion in a rendition is regarded as enhancing the quality of song. The pakkad is often used to determine the identity of a Raga, so can be regarded as a Raga’s signature tune. For example, the pakkad “Pa, Ga, Dha, Pa, Ga, Re, Ga, Re, Sa” is specific to Raga Bhupali.

- Vadi swar – the highest frequency swar in a Raga. For example swar ‘Ga’ is Vadi swar of Bhupali. This is reflected in the higher frequency of ‘Ga’ of the Raga’s pakkad, “Pa, Ga, Dha, Pa, Ga, Re, Ga, Re, Sa”.

- Samvadi swar – the second highest frequency swar in Raga. For example ‘Dha’ is Samvadi swar of Raga Bhupali.
- Varjit swar – swars which are regarded as incompatible with the swars of a specified Raga, and therefore to be avoided in compositions of songs/recitals in that Raga. For example swar ‘Ma’ and ‘Ni’ are varjit in Raga Bhupali, which is made up of swars, “Sa, Re, Ga, Pa, Dha”.

- Vivadi swar – Swars which are not part of aroha-avaroha sequence of the Raga but are also not varjit swar are known as vivadi swar. These swars should not be used in the raga hence also referred as “enemy” swar. However experts might sometimes use them to enhance the beauty of raga (Surgyan 2009). Usually these are variations of swars in aroha-avaroha. For example komal swars “Re, Ga, Dha” are vivadi swars for raga Bhupali.

Hence, a set of swars combined with the above conventions (that determine emphasis on specific swars, sequencing of phrase, etc) are the most significant determiners of the structure of a Raga which determines (literally) the mood of a Raga that the singer/musician is supposed to capture in a rendition of a song or instrumental piece in that Raga (and which in turn in evoked in the listener. It is the differences in the conventions applicable to each Raga that enables a listener (or learner) to distinguish them along melodic and/or aesthetic dimensions. So even when two Ragas have the same set of swars and aroha-avaroha sequences: the differences in features such as the vadi and samvadi swars and pakkad renditions in each will result in very different renditions in terms of the moods they invoke. For example, Raga Bhupali and Deshkar both use same set of swars - Sa, Re, Ga, Pa, Dha - yet they sound different because compositions of Bhupali have higher frequency for swars “Sa, Re, Ga”, Whereas in compositions of Deshkar, the swars, “Pa, Dha, Sa” predominate (ITC Sangeet Research Academy 2001).

There are more complicated and esoteric features of HCM, such as, meend (sliding from 1 note to other) and which might even include a varjit (otherwise forbidden) swar for a very short duration; alap, the starting section of a song which is in slow rhythm compared to rest of the song, usually sung before the rendition of the Raga itself (in instrumental or vocal form); or, taan, compositions of Raga sung in very fast tempo. These (and other such conventions) however would typically be used by expert singers/musicians during renditions and as part of complicated improvisation.

Usually early stage learners, apart from learning to render the correct notes of a specific Raga (which includes avoiding the varjit swar), focus on mastering aroha and avaroha sequences and learning when and how to launch into a pakkad. These aspects of a Raga have to be mastered before they can move on to a more advanced stage where the emphasis shifts to improvisation (Improvisation 2014). Hence, the tutoring system presented here has been designed to handle only these basic aspects of songs typically rendered by early stage learners. This approach hence reduces the complexity that needs to be handled by a computational model for automated recognition/evaluation of songs sung by novices

In Section 2.1 (below), we present a brief review of previous research focused on computational musicology models of ICM with an emphasis on findings that are specifically relevant to our research focus. It will be noted that much previous work has focused on exploring and/or modeling the nature of complexity of ICM (as part of a recognition or generation system) (Gajjar and Patel 2017). In Section 2.2 we provide a detailed description of the scope and objective of the proposed system for analysis and evaluation of songs (in one of the eight selected Ragas) typically sung by early stage learners of HCM.

2. PREVIOUS FINDINGS AND SCOPE AND OBJECTIVE OF PROPOSED SYSTEM

2.1 Previous Relevant Research

In this section we review computational musicology models of aspect of ICM that are of relevance to the modeling approach and evaluation algorithms described here. A more comprehensive review and evaluation of computational modeling of ICM can be in found in Gajjar and Patel (Gajjar and Patel 2017).

Bhattacharjee and Srinivasan (Bhattacharjee and Srinivasan 2011) describe a transition probability model for Raga recognition. Their approach is based on the idea that in any Raga the probability for transition between notes in ascending and descending order (ie, aroha-avaroha) is a good indicator of a Raga's structure, and so can be reliably used for Raga identification. The authors carried out manual note transcription of renditions by two singers/musicians in ten Ragas. Based on this note transcription, they hand-crafted a Transition Probability Matrix (TPM) for each Raga. Each TM has 12 columns and 12 rows to represent 144 possible transitions between tuples of 12 swars. The TM has middle octave (madhyan) as the default, so swars rendered in the higher (taar) or lower (mandra) octave are shifted to the middle (madhyam) octave. The probability of transitions for each Raga was determined with a training set of 10 clips of songs, each of 5 seconds duration.

While this approach provides insights in to pattern of transition between notes in each Raga, it is highly constrained with respect to actual renditions. The primary focus on the middle octave (and shifting the notes from higher and lower octave to middle octave) means that critical information is lost as at times aroha-avaroha sequence can change with singer's shift between octaves. For instance, a sequence "Sa (madhyan octave), Ga (taar octave)" is considered aroha, whereas sequence "Sa (madyam octave), Ga (mandra octave)" is considered to be avaroha. Hence the proposed TM would not be able to handle this distinction which can prove to a significant shortcoming for an extended computer based evaluation system for ICM renditions. However, as we will explain in section 3 the TM approach to modelling note transition is a highly promising and efficient approach for an evaluation system for HCM renditions.

Das and Choudhury (Das and Choudhury 2005) focused on generation of HCM, for which they also considered aroha and avaroha features of Ragas. They have designed probabilistic Finite State Machine (FSM) for ascending and descending movements for a particular Raga. This FSM is used to generate a sequence of notes (ie, a rendition of a Raga). However, the authors conclude that the FSM generated renditions were not as good as those rendered by human musicians. The main reason for this being that while the FSM focused on modelling the aroha and avaroha features of a Raga, these two features were not sufficient for generating the complex melodic structure (ie, richness) that is normally evident in renditions by human performers. This finding suggests that conventions such as pakkad, vadi and samvadi swars are critical in determining the essence of a Raga. Hence, an extended FSM that incorporates these additional features is likely to generate music of better quality, however, it would also result in the increased complexity.

Pandey, Mishra and Ipe (Pandey et al. 2003) have proposed an approach for automatic Raga recognition. The note transcription is done using hill peak and note duration heuristics. For Raga identification, the notes identified from transcription module are used with Hidden Markov Model which is trained using Baum-Welch learning algorithm. It relies on the inherent flexibility of a Markov model, which unlike FMS's can handle probability based transition. This approach would be better at handling additional features such as pakkad, vadi and samvadi swars. However, the higher order of Markov model would result in an exponential increase in the number of combinations (of swars). Further, it is highly likely that a model using the same transition matrix will tend to generate music in similar style (McCormack 1996). Overall the conclusion is that this approach would still not be able to generate music that is aesthetically as creative and pleasing as would be by a trained human musician.

Overall computational models for music composition and recognition tend to rely on modelling the past sequence of note to predict next note or to identify patterns in the composition. This approach yields good results for recognition or composition where past sequence of notes can indeed be helpful in predicting the next note(s). However, for HCM at any point in a rendition, subsequent notes are as important as

the preceding notes for deciding the correctness of a song. In order to model that requirement it was decided to implement a transition matrix (using a Markov model approach). This matrix can be used to check the correctness of individual note by considering both, the preceding as well as the subsequent note.
Our approach of designing a tutoring system for novice learners hence is primarily focused on identifying the notes sung and assess their appropriateness with respect to Raga's structure. In the next section we describe scope and architecture of the tutoring system in more detail.

2.2 Scope of the HCM Tutoring System

The primary aim in developing this system was to enable a novice to do singing practice in the absence of a tutor (guru). For many students the necessarily limited duration of one-to-one interaction with a tutor is the biggest reason for the relatively slow progress in mastering singing technique. The tutoring system was therefore designed to address this challenge; more specifically it was focused on automated evaluation of rendered song by an early stage learner of HCM, and to provide feedback on how to correct any mistakes. Ideally such a tutoring system should be designed to "listen" to a song being rendered in real-time and provide feedback in real time as is often the case with a one-to-one learning session with a teacher. However this would have raised the complexity of the system that would be beyond the scope of our research project; particularly so as the key challenge is to have a reliable evaluation system that can identify correctly sung notes (with respect to the Raga in which the song has been composed) and provide list of errors (if any). Hence instead of designing a system robust enough to accept audio recordings of varying quality (depending on the quality of recording equipment) the system was designed to accept audio files in .wav format.
Usually performance by experts of HCM is accompanied by instruments like harmonium/sitar (to provide the tonic scale) and tabla (drums) which is used to set the tempo of the rendered song. However, an early stage learner usually practices in the absence of these instruments, there by rendering song in a scale that he/she is comfortable with and keeping time with hand-claps. Tonic scale and tempo are used for identifying frequency of notes in octaves (lower, middle and upper) and segmentation of audio file respectively in note transcription module (described in detail in section 3.2). Hence, the system requires the learner to select the tempo of song (in beats per minute) and tonic frequency (in hertz).
The tutoring system is designed to evaluate songs in 8 Ragas - Bhupali, Durga, Sarang, Kafi, Bhimpalasi, Khamaj, Bhairavi and Desh - which are generally taught in the initial years of a singers learning process of HCM. Our selection of this subset was based on widely respected two leading proponents of syllabi for novice learners in HCM – first, Gandharv Mahavidyalaya founded by Pt. Vinayakbua Patwardhan in 1932 (Gandharva Mahavidhyalaya 2019); second, Akhil Bharatiya Gandharva Mahavidhyalaya Mandal (Akhil Bhartiya Mahavidhyalaya Mandal 2018) founded

by Pt. Vishnu Digambar Paluskar in 1901. This selection is further supported by a survey of 50 singers of HCM based in Ahmedabad, Gujarat. Of these 92% confirmed that these eight Ragas were the ones that they mastered in the initial phase of learning HCM. Even the remaining 8% agreed on at least 6 six Ragas from this list (while suggesting alternative for the remaining one or two Ragas). In our system the user (novice singer) is required to identify the Raga in which he/she has sung the (pre-recorded) song.

Evaluation of rendered song essentially means determining whether the transition between swars was permissible (with respect to the Raga's syntax). Hence, simple transition matrices were used to encode legitimate transition based on the aroha-avaroha sequence and varjit swar of each Raga. A matrix for the Raga in which the song has been rendered (by the novice) would be used to determine the correctness of a note with respect to both the previous note as well as next one. For instance, in Raga Bhimpalasi, swars 'Dha' and 'Re' are used only in avaroha (descending order). So, in the sequence "Sa, Dha, Pa, Ma, Re, Ga", the (initial) pair 'Sa, Dha' is moving in ascending order (aroha) which is not legitimate. But considering the subsequent (third) note 'Pa' which forms the pair 'Dha, Pa' which is avaroha (descending order) this transition would be considered legitimate. The generalized algorithm used for matrix generation is explained in section 3.3.

Once the notes have been evaluated, users are presented with feedback that highlights the errors (if any) that they have made in the song. This includes indication of where in the song the error occurred with respect to the key aspects (described in Section 1.2 above) that determines a Raga's melodic structure.

However, as explained in section 1.2, though aroha-avaroha and varjit swars are two most important conventions of Ragas, there might be scenarios where other Raga may follow the same convention. Hence to differentiate between Ragas, vadi, samvadi and pakkad are also considered in second stage of evaluation process and which provides further feedback on whether the overall rendering of the song conform to the specified Raga by the novice (or is closer to another Raga's syntax).

In the next section we present the overall architecture of the tutoring system with details of how the song evaluation process has been implemented. This is followed by the results of its performance when presented with 180 songs of 8 Ragas with different types of errors listed in Section 4.

3. ARCHITECTURE AND FUNCTIONALITY OF THE HCM TUTORING SYSTEM

The overall system architecture is illustrated in Fig. 1. The early stage learner would upload the audio file with relevant details of the Raga, tempo, gender and tonic frequency (octave). The input file is then processed for note transcription, the output of which is evaluated for errors. Each part of the system will be described in more detail in the rest of the sections.

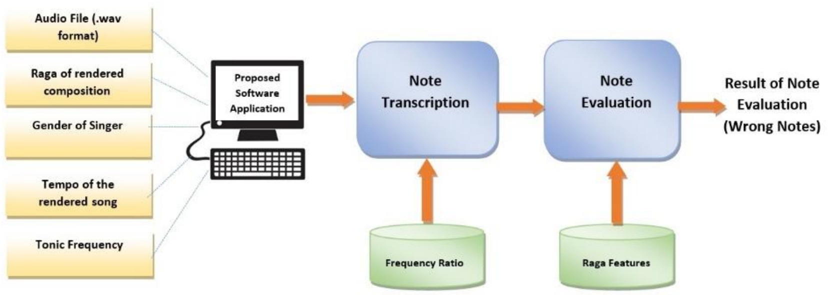

Fig. 1. Architecture of Proposed System

3.1 Note Transcription Process

Fig. 2 shows the sub steps for note transcription and inputs utilized in these steps. The note transcription module considers an audio file uploaded by user and generates list of notes. We have used a third party WavFile class by Andrew Greensted (Greensted 2009) for reading the sample values from uploaded wav file. Next, these samples are processed using algorithm from PRAAT to get the list of frequencies (pitch listing) (Boersma and Weenik 2007) which is further segmented based on the tempo (as provided by early stage learner). The mean value of segmented frequency list is then compared with the note frequency table (generated by using tonic frequency listed in Table 1) to identify each note in the rendition. Fig. 3 illustrated the format of information about notes displayed for the learner.

In general musical note transcription from audio files is a complex process, particularly so for HCM renditions during which sliding between notes is considered a mark of excellence. However, this complexity can be tackled (to the extent required for our purposes) with the note duration heuristics approach described by Pandey, Mishra and Ipe (Pandey et al. 2003). It assumes that the minimum duration of each note is $1/4^{th}$ of the duration of a beat. For example if duration of a beat is 60 seconds, minimum duration of a note is then considered to be 15 seconds.

Fig. 3 is an example of a typical output once a .wav files has been segmented into individual notes. The figure displays identified notes in form of spectrogram and pitch contour. The identified list of notes along with their duration are stored in a separate text file. These files are then accessed by the evaluation module to check for any errors with respect to the Raga structure of the song.

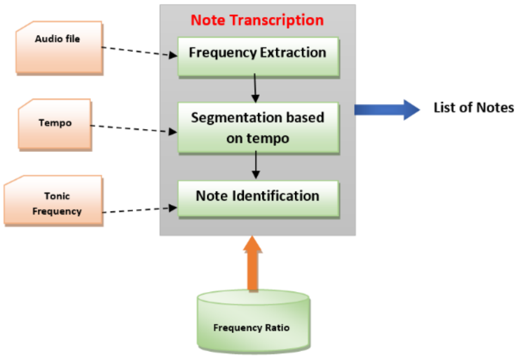

Fig. 2. Note Transcription Module

The note transcription output is then displayed as shown in Fig. 3.

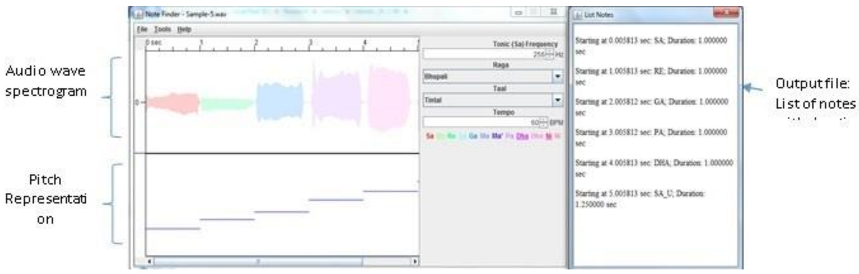

Fig. 3. Notes listed through note identification module

3.2 Evaluation Process Overview

Fig. 4 shows the sub steps for note evaluation process. As explained in Section 2.2, the note evaluation module incorporates a two stage evaluation process. In the first stage the rendered notes are evaluated with respect to the aroha and avaroha sequence of the Raga and varjit swars: rendered notes are compared with a note TM (for the specific Raga of the rendition). The rendition is processed further (second stage) with respect to the vadi, samvadi and pakkad of the Raga. In the next section (3.3) we present details of the algorithm used for generation of TM used in first stage. Sections 3.4 and 3.5 has detailed description of the evaluation process for the first and second stage respectively.

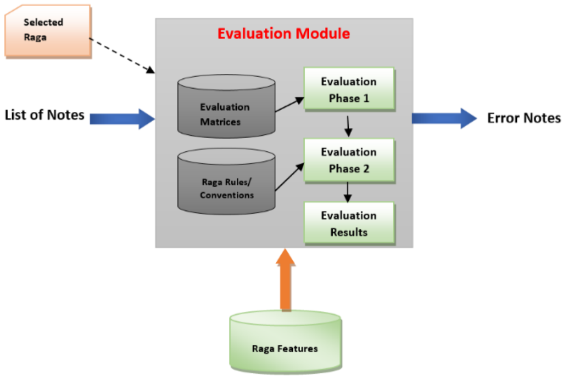

Fig. 4. Note Evaluation Module

3.3 Computation of TM based on Aroha-Avaroha of Raga

For the first step of the evaluation an algorithm based on the aroha-avaroha sequence of each Raga is used for formulation of the TM. Each TM has an 36x36 array, containing 1,296 cells with values '1' or '0' as shown in Table 2, where '1' represents a legitimate transition and '0' a forbidden transition (based on) for that particular Raga. In our implementation there are three sections in each TM; one each of the 3 octaves (mandra, madhyan and taar). The stepwise process for generation of each TM is described below. Once generated the matrices are stored in the system for access during evaluation of songs uploaded by users (early stage learners).

- STEP ONE
  - Let the set of notes of 3 octaves (ie, mandra, madhyan and taar: lower, middle and upper respectively) be denoted as S.
  - $S=\{n_0,n_1,n_2\ldots n_{35}\}$, where n_0=S=Sa of lower octave ,n_1=r=Re of lower octave , n_{12}=S=Sa of middle octave, n_{35}=N=Ni of upper octave
- STEP TWO
  - Let S_A be set of notes in aroha (ascending order), where $S_A \subset S$
  - And S_D be set of notes in avaroha (descending order), where $S_D \subset S$
  - Let i and j represent index of row and column of matrix respectively such that i=0...35 and j=0...3
- STEP THREE
  - For $i \leq j$,

    $V_{ij} =$ 1, if $n_i \in S_A$ and $n_j \in S_A$
    0, if $n_i \notin S_A$ or $n_j \notin S_A$

– For i>j,

$$V_{ij} = \begin{cases} 1, \text{ if } n_i \in S_D \text{ and } n_j \in S_D \\ 0, \text{ if } n_i \notin S_D \text{ or } n_j \notin S_D \end{cases}$$

Considering S_A and S_D, matrix, denoted as M, is populated as

$$M = \left\{ \begin{matrix} V_{00} & V_{01} & \ldots & V_{0n} \\ V_{10} & V_{11} & \ldots & V_{1n} \\ \ldots & & & \\ V_{n0} & V_{n1} & \ldots & V_{nn} \end{matrix} \right\} \quad \text{where } n=35 \text{ and } V_{ij} \in \{0,1\}.$$

The example of TM for Raga Bhupali is shown in Table 2.

Octave		Lower				Middle				Upper			
		S_L	r_L	M_L	N_L	S	R	M	N	S_U	r_U	M_U	N_U
Lower	S_L	1	0	0	0	1	0	0	0	1	0	0	0
	r_L	0	0	0	0	0	0	0	0	0	0	0	0
	M	0	0	0	0	0	0	0	0	0	0	0	0
	N_L	0	0	0	0	0	0	0	0	0	0	0	0
Middle	S	1	0	0	0	1	0	0	1	0	1	0	0
	R	0	0	0	0	0	0	0	0	0	0	0	0
	M	0	0	0	0	0	0	0	0	0	0	0	0
	N	0	0	0	0	0	0	0	0	0	0	0	0
Upper	S_U	1	0	0	0	1	0	0	0	1	0	0	0
	r_U	0	0	0	0	0	0	0	0	0	0	0	0
	M_U	0	0	0	0	0	0	0	0	0	0	0	0
	N_U	0	0	0	0	0	0	0	0	0	0	0	0

Table 2: Example of Transition Matrix for Raga Bhupali

3.4 Evaluation Process - Stage 1

First, the evaluation process compares sequences of notes (generated by the note identification module) with the TM to check for errors with respect to aroha-avaroha sequence of the Raga and varjit swars. Each bi-gram of note sequence is compared with TM in the following step wise process:

Let the sequence of notes be denoted as:

N= $\{N_0, N_1 \ldots N_n\}$, where N_0 is starting note and n represents total number of notes in the sequence.

Transition Matrix (M) used for evaluation (pre-generated) as shown in section 3.3

Apart from matrix, set of varjit swars of a Raga is also used for evaluation. Let it be denoted as V.
Evaluation Algorithm is as follows:

1. For current note N_i and next note N_{i+1}, where $0 \leq i \leq n$ find the value from $M[N_i][N_{i+1}]$.
2. If value of $M[N_i][N_{i+1}]$ is 1, increment i by 1. If value of $M[N_i][N_{i+1}]$ is 0, check if N_i or N_{i+1} belongs to V.
3. If $N_i \in V$, mark N_i as wrong note and increment value of i by 1. Else if $N_{i+1} \in V$, mark N_{i+1} as wrong note and increment value of i by 2.
4. If $N_i \notin V$ and $N_{i+1} \notin V$ and N_{i+1} is not last note in sequence, increment value of i by 1. Check the value of $M[N_i][N_{i+1}]$. If this value is 1, increment i by 2. Else mark N_i as wrong note and increment the value of i by 1.
5. If N_{i+1} is last note and $M[N_i][N_{i+1}]$.=0, mark N_{i+1} as wrong note.
6. Repeat from step 1 until i reaches the last note of sequence N.

Fig. 5 shows a section of the output of first stage of note evaluation of a rendition in Raga Bhupali. The notes in red color are incorrect notes as per the TM based on the aroha-avaroha and varjit swar of that Raga.

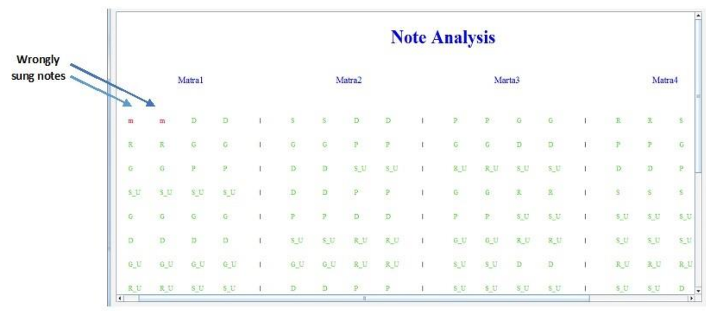

Fig. 5. Note Evaluation Result for Raga Bhupali

3.5 Evaluation Process – Stage 2

The second stage of evaluation is considers the vadi, samvadi and pakkad features of the selected Raga to evaluate the overall extent to which the rendition confirms to the Raga's syntax. This stage is at a two stepped process; first, a simple frequency count of different notes is calculated to check if the vadi and samvadi swars of the Raga are indeed to most and second most used swars in the rendition; and, second a string matching algorithm checks for the occurrences of any/all pakkads in the rendition. This process is implemented as follows:

Consider following sequence of notes which is evaluated for Raga Bhupali:
S_U S_U **D** **D** S_U S_U **D** **D** S_U S_U **D** **D** P P **D** **D** P P **D** **D** P P ***G*** R ***G*** P **D**
P ***G*** P **D** S_U **D** P ***G*** P **D** S_U **D** P ***G*** P **D** **D** **D** ***G*** P **D** S_U **D** P ***G*** ***G*** P **D** S_U
D P ***G*** **D** S **D** P ***G*** **D** S_U **D** P **D** S_U **D** P S_U S_U **D** P S_U S_U **D** P S_U **D**
P S_U **D** P S_U S_U **D** P ***G*** ***G*** P **D** S_U **D** P ***G*** ***G*** P **D** P ***G*** P **D** P ***G*** P ***G*** R S

The above sequence of notes conforms to the aroha-avaroha and varjit swars features of Raga Bhupali, so the first stage of evaluation has indicated not errors as such. However, the second stage evaluation process has highlighted that the swar '**D**' (Dha) occurs more frequently than the swar '***G***' (Ga) which is the actual vadi swar of Raga Bhupali. In addition, the rendition has notably high frequency of the combination of "P(Pa), D(Dha), S_U(Sa_U)" swars which is a pakkad of Raga Deshkar and not Bhupali. Based on such observations, the tutoring system would provide the user qualitative feedback that some features of Raga Bhupali are not captured in the rendition.

4 RESULTS

In this section we present preliminary results of on the performance of the first stage of the evaluation process. Our primary aim is to validate the efficacy of the Transition Matrix approach for evaluation of HCM renditions by early stage learners. For this test we compiled 180 renditions (with average 100 notes) of eight Ragas. The renditions collectively had 380 errors that comprehensively reflected all types of errors that can be typically made by early stage learners. These included:

- Varjit swar as first note, varjit swar as intermediate note in the sequence, 3 consecutive varjit swar in the sequence
- Vivadi swar as first note in sequence, Vivadi swar as intermediate note in the sequence, 3 consecutive vivadi notes in the sequence
- Correct note between 2 varjit swar in the sequence, Correct note between 2 vivadi swar in the sequence
- 5 consecutive wrong notes in the sequence (Combination of varjit and vivadi swars)

- Correct note between sequence of wrong notes
- varjit swar as last note in sequence, varjit swar as last 3 notes in the sequence
- Vivadi swar as last note in the sequence, vivadi swar as last 3 notes in the sequence
- Swar allowed in aroha but not in avaroha used in avaroha sequence, Swar allowed in avaroha but not in aroha used in aroha sequence

Based on the types of errors different testing sequence of notes were generated for 8 Ragas. All the sequences had some combination of errors from the above list. Also, to check for false positives there was at least 1 input sequence with zero errors.

Total Number of Sequences	Average number of notes in each sequence	Total number of notes in all the sequences	Total Number of errors in the sequences	Number of errors identified through the implemented module	False positives
180	100	18,000	380	380	0

Table 3. Test Results for TM based Evaluation (Stage 1)

As can be seen from the results presented in Table 3 the TM based evaluation of HCM performs remarkable well in identifying errors (with respect to aroha-avaroha and/or varjit swar conventions of Ragas), as well as avoiding any false positives. These results confirm that algorithm to generate TM for each of the eight Ragas and their application as part of the evaluation process for HCM renditions is highly efficient and reliable in identifying typical errors while also avoiding false positives.

5 DISCUSSION and CONCLUSION

Our primary objective was to develop a computational model that was capable of evaluating the correctness of vocal renderings (simple songs) in Ragas typically learnt by early stage learners of vocal HCM. The early stage of learning is highly interactive during which the main feedback is on getting the notes and prominent phrases correct. Our analysis and research in the early stage learning/tutoring process (ie, the first two years) highlighted three focus areas of learning (apart from mastering the tonic scale). First, novices learnt how to sing by practicing relatively simple songs (usually in 6-8 Ragas); second, the focus is primarily on rendering the correct notes with respect to aroha-avaroha ,while avoiding the varjit swar ; and third, to gain familiarity with the concept of pakkad. Hence, a computer based tutoring system would need to be able to evaluate renditions with respect to these three aspects of the Ragas typically learnt by novices, and ideally also provide feedback on any errors.

The evaluation module of the tutoring system developed by us is constrained to consider only those elements of Raga's syntax that are expressed in renditions by novices. Given the relative importance of aroha-avaroha, the primary focus of the evaluation process was on assessing the appropriateness of each swar/note with respect to the preceding and subsequent swar/notes. For this assessment transition matrices (for each of eight Ragas) based on aroha-avaroha and varjit swar features were generated. Test results show that this TM based computational model is highly appropriate for assessing the correctness of songs sung by novices.
With respect to the note transcription part of the tutoring system our implementation is a version that incorporates a note duration heuristics method that avoids the need for manual transcription of notes.
The computational model presented here has not considered other features of Raga that determines its overall melodic structure. Further research should focus on extending the evaluation and feedback process for more sophisticated renditions. To do so the TM based model presented here can be extended to consider features such as alap, taan and meend. This would necessarily increase the complexity of the transcription process which may affect performance time. Even further extensions of this approach can attempt top model improvisation which are very much a feature of expert singers/musician of HCM recitals. However, such models would probably need to include non-linear approaches typically used in ML/AI modelling methods. Even so, we hope that the model and results presented here are indicative of the possibilities of developing computational musicology models that reflect and/or account for the richness, sophistication and complexity of HCM.

REFERENCES

Agarwal, P., Karnick, H. and Raj, B., 2013. A Comparative Study Of Indian And Western Music Forms. In ISMIR (pp. 29-34).

Bhatkhande, P.V.N., 1970. A short historical survey of the music of upper India. Journal of the Indian Musicological Society, 1, p.2.

Bhattacharjee, A. and Srinivasan, N., 2011. Hindustani raga representation and identification: a transition probability based approach. IJMBC, 2(1-2), pp.66-91.

Boersma, P. and Weenik, D., 2007. Praat: doing phonetics by computer. Institute of Phonetic Sciences of the University of Amsterdam.

Das, D. and Choudhury, M., 2005. Finite state models for generation of Hindustani classical music. In Proceedings of International Symposium on Frontiers of Research in Speech and Music (pp. 59-64).

Dighe, P., Karnick, H. and Raj, B., 2013, November. Swara Histogram Based Structural Analysis And Identification Of Indian Classical Ragas. In ISMIR (pp. 35-40).

Gajjar, K. and Patel, M., Computational Musicology for Raga Analysis in Indian Classical Music: A Critical Review. International Journal of Computer Applications, 172(9), pp. 42-47

McCormack, J., 1996. Grammar based music composition. Complex systems, 96, pp.321-336.

Mukerji, B., 2014. An analytical study of improvisation in hindustani classical music.

Pandey, G., Mishra, C. and Ipe, P., 2003. TANSEN: A System for Automatic Raga Identification. In IICAI (pp. 1350-1363).

Rabuñal, J.R. ed., 2005. Artificial neural networks in real-life applications. IGI Global.

Rao, S. and Rao, P., 2014. An overview of Hindustani music in the context of computational musicology. Journal of new music research, 43(1), pp.24-33.

Volk, A., Wiering, F. and Kranenburg, V.P., 2011. Unfolding the potential of computational musicology. In Proceedings of the 13th International Conference on Informatics and Semiotics in Organisations.

Akhil Bhartiya Gandharva Mahavidyalaya (2018). http://abgmvm.org/ . Accessed 15 March 2018

Alap. Available at: https://en.wikipedia.org/wiki/Alap. Accessed 22 March 2018

Andrew Greensted(2009), The Lab Book Pages. http://www.labbookpages.co.uk/audio/javaWavFiles.html . Accessed 22March 2018

Audacity. https://www.audacityteam.org/. Accessed 22 March 2018

Gandharva Mahavidyalaya, Pune. https://www.gandharvapune.org/cms/About-Bharatiya-Sangeet-Prasarak-Mandal,-Pune.aspx, Accessed 15 March 2018

Improvisation in Indian Classical Music (2012). Available at: https://auburnsangeet.wordpress.com/2012/08/10/improvisation-in-indian-classical-music/. Accessed 12 December 2018

Improvisation (2014). Available at: https://www.aliakbarkhanlibrary.com/new_to_indian_music/tana.html . Accessed 12 December 2018

Indian Classical Music and Sikh Kirtan (1995). http://fateh.sikhnet.com/sikhnet/gurbani.nsf/d9c75ce4db27be328725639a0063aecc/085885984cfaafcb872565bc004de79f!OpenDocument Accessed 11 February 2018

ITC Sangeet Research Academy (2001). http://www.itcsra.org/Bhupali-Deshkar-Shudh-Kalyan, Accessed 1 March 2018

Shruti. https://en.wikipedia.org/wiki/Shruti_(music), Accessed 11 August 2017

Surgyan: Promoting Indian Classical Music. Available at: http://www.surgyan.com/vivadiswar.htm, Accessed 22 March 2018

Taan. Available at: https://en.wikipedia.org/wiki/Taan(music) . Accessed 22 March 2018

Vishnu Narayan Bhatkhande. https://en.wikipedia.org/wiki/Vishnu_Narayan_Bhatkhande. Accessed 22 March 2018