CARNATIC MUSIC ANALYSIS: *Shadja, Swara* IDENTIFICATION AND *rAga* VERIFICATION IN *AlApana* USING STOCHASTIC MODELS

Ranjani, H. G.[1] *, Arthi, S.*[2] *and T. V. Sreenivas*[3]
Departments of EE and ECE, Indian Institute of Science, Bangalore, India
[1]ranjanihg@gmail.com, [2]arthis@ece.iisc.ernet.in, [3]tvsree@ece.iisc.ernet.in

ABSTRACT

We analyze the *AlApana* of a Carnatic music piece without the prior knowledge of the singer or the *rAga*. *AlApana* is a means to communicate to the audience, the flavor or the *bhAva* of the *rAga* through the permitted notes and its phrases. The input to our analysis is a recording of the vocal *AlApana* along with the accompanying instrument. The *AdhAra shadja*(base note) of the singer for that *AlApana* is estimated through a stochastic model of note frequencies. Based on the *shadja*, we identify the notes (*swaras*) used in the *AlApana* using a semi-continuous GMM. Using the probabilities of each note interval, we recognize *swaras* of the *AlApana*. For *sampurNa rAgas*, we can identify the possible *rAga*, based on the *swaras*. We have been able to achieve correct *shadja* identification, which is crucial to all further steps, in 88.8% of 55 *AlApana*s. Among them (48 *AlApanas* of 7 *rAgas*), we get 91.5% correct *swara* identification and 62.13% correct $\Re$ (*rAga*) accuracy.

Index Terms— Note transcription, *rAga* verification, SC-GMM, *sampurNa rAga*, *AlApana*, *Shadja*

1. INTRODUCTION

A typical rendering of Carnatic music by a professional vocalist will comprise of a melodic accompaniment (violin), a rhythm accompaniment (*mridanga*) and a *tambura* which provides a drone throughout the performance. The *AlApana*, is mainly a solo exposition of the *rAga* and its *bhAva* (see Table 1 for terminology), without the rhythm accompaniment or any influence of the *sAhitya*, i.e., the song or the musical text. There are a few rules which govern the way in which *AlApana* of each *rAga* is rendered, but mostly it is extempore, left to the creativity of the artist. The artist would choose different note phrases of the *rAga* to bring out its *bhAva*. So, there is no one pattern to attempt a pattern matching for the identification of a *rAga* from its *AlApana*. But, yet most music enthusiasts recognize a *rAga* based on short piece of *AlApana* and enjoy its *bhAva*, challenging us to find an analytical reason. Further, automatic annotation and identification of *rAga* will be of much use for understanding the creativity in Carnatic music by a student as well as a general listener.

Unlike western music, where the note frequencies are fixed (scale), Indian music gives room for the singer to choose a suitable base frequency for the scale. This is called *AdhAra shadja*. The *AdhAra shadja* can vary even for a particular singer, for different renderings of even the same *rAga*, as illustrated in Fig 2. All notes get fixed with respect to the base frequency. Hence, identification of *AdhAra shadja* itself becomes a fundamental problem. In Carnatic music, the whole *rAga* structure is later built by the artist to suit the *sAhitya* and *bhAva*. Table 2 gives the names, position and notation for the *swaras* in Carnatic music. *Swaras* at the same position will not occur in the same *rAga* together.

The notes in an octave of western music are equally tempered, while in Indian music they are not. Though there is a controversy about the number of notes in an octave (12 or 22), the 12-note octave is widely used. Table 2 gives the relative ratios of *swaras* with respect to *AdhAra shadja* in Carnatic music for a 12-note octave convention [7].

Once the *shadja* is identified, the notes are relative to the base note *(shadja)*. The *AdhAra shadja*, S_0, its 3/2 note (*panchama*, P_0) and its octaves are called *achala swaras* and would have least variance while singing. Rest of the notes would have higher variance depending on the singer style and the *gamakas* which are characteristic of the *rAga*, but would be more or less fixed around a mean. Past efforts in computational analysis of Indian music [1], [2] are based on identifying and matching musical phrases. More recently, HMMs are used for *rAga* classification[3]. In the work by Chordia [4]-[6] for *rAga* classification and recognition, pitch class distribution and pitch class dyadic distribution are used as features. In most of the papers, manual identification of *AdhAra shadja* is used for further analysis. Here, we are attempting a fully automatic scheme.

We propose a method to determine the *shadja* from the pitch values of the input *AlApana*. Since the notes are not fixed but vary around a mean ratio, we explore semi-continuous GMM (SC-GMM) approach to note modeling. Fig 1 gives the block diagram for the proposed analysis.

RAgas with missing notes would pose additional difficulties for *shadja* identification. So we confine to the *sampurNa rAgas* and estimate the 7 most probable notes using a stochastic model approach. Thus, determining the labels for each pitch interval also gives us a transcription of the *AlApana*.

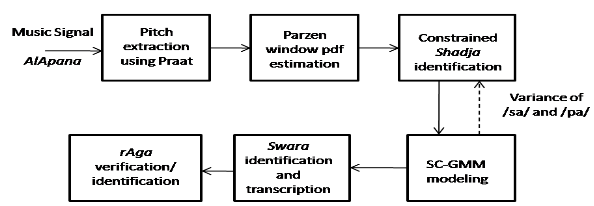

Figure 1: *Block diagram of the proposed AlApana analysis*

2. ESTIMATING *ADHARA SHADJA*

For determining S_0, the frame-wise (10 ms) pitch frequencies for the entire *AlApana* are obtained using Praat software. An *AlApana* will typically last for 3-10 mins. A non-parametric density estimate of the pitch using the Parzen window method (with a Gaussian window of variance 0.7) is determined. Typically, S_0 for different singers ranges over 100-200 Hz. It has been observed that the pitch for the S or P has a high probability in a typical *AlApana* extract, as

All words in italics refer to Indian language words transliterated in Roman alphabet

Carnatic	Western music equivalent (approximately)
swara	Note of 7 types (/sa/, /ri/, /ga/, /ma/, /pa/, /dha/, /ni/) (or 12 positions)
AdhAra shadja	*Swara* /sa/ corresponding to base octave
rAga	Musical composition
sampurNa rAga	All 7 *swaras* in the scale
janya rAga	*RAgas* derived from *sampurNa rAga*
shadja	Defining note of the scale, or the tonic
gamaka	Similar to vibratos
AlApana	Melodic improvisation that introduces and develops a rAga
bhAva	The emotive state induced by the *rAga* and *sAhitya*
lakshaNa	Characteristic
sAhitya	Text of the song
ArOhaNa	Ascending scale of notes in a rAga
avarOhaNa	Descending scale of notes in a rAga

Table 1: *Carnatic music names and its nearest western equivalent*

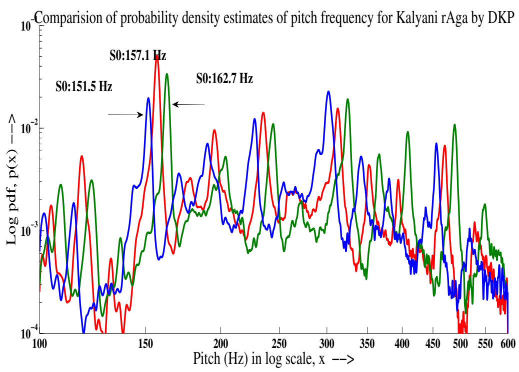

Figure 2: *Probability density estimate of 3 different AlApanas of Kalyani rAga, rendered by Smt D. K. Pattammal*

Pos	*Swara* name	Label	f.r.(C)	f.r(W)
1	*Shadja*	S	1.000	1.000
2	*Shuddha rishaba*	R1	1.067 (16/15)	1.059
3	*Chatushruthi rishaba*	R2	1.125 (9/8)	1.122
4	*Shatshruthi rishaba*	R3	1.200 (6/5)	1.189
3	*Shuddha gAndhara*	G1	1.125 (9/8)	1.122
4	*ShAdhArana gAndhara*	G2	1.200 (6/5)	1.189
5	*Anthara gAndhara*	G3	1.250 (5/4)	1.259
6	*Shuddha madhyama*	M1	1.333 (4/3)	1.334
7	*Prati madhyama*	M2	1.416 (17/12)	1.414
8	*Panchama*	P	1.500 (3/2)	1.498
9	*Shuddha daivatha*	D1	1.600 (8/5)	1.587
10	*Chatushruthi daivatha*	D2	1.667 (5/3)	1.681
11	*Shatshruthi daivatha*	D3	1.800 (9/5)	1.781
10	*Shuddha nishAdha*	N1	1.667 (5/3)	1.681
11	*Kaisika nishAdha*	N2	1.800 (9/5)	1.781
12	*KAkali nishAdha*	N3	1.875 (15/8)	1.887

Table 2: *The position (Pos), names and label of swaras in Carnatic music. Also shown is the ratios of notes in Carnatic scale, (f.r.(C)) [7] against Pythogorian equitempered scale, f.r(W) =* $2^{(Pos-1)/12}$

can be seen in Fig 2. But a simple peak picking in the PDF will not solve as shown in Fig 3. This is because *rAga Kharaharapriya* has its characteristic *swara* $R2$, which has a higher probability than S or P. In the backdrop of such possible cases, we propose a method-

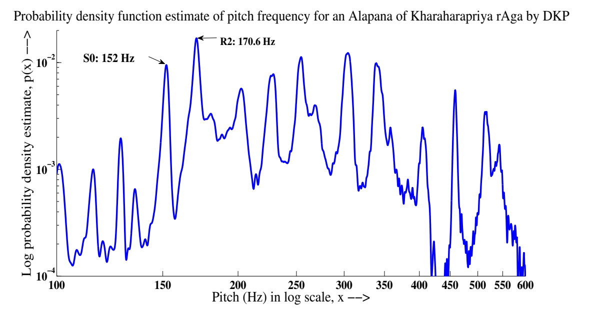

Figure 3: *Probability density estimate of AlApana of Kharaharapriya rAga rendered by Smt D K Pattammal*[1]*. It can be seen that R2 peaks up more than S in all octaves with low variance in both R2 and D2, revealing possibility of erroneous choice of shadja*

ology for estimating S_0 as follows.

Let M_k be the note frequencies which correspond to largest local peaks in the density estimate, $p(x)$ where $x = log_2(y(t))$ and $y(t)$ are the pitch frequencies at time t, obtained using Praat. Let,

$$M_k \in [100\,Hz : 600\,Hz], \quad 1 \leq k \leq 10 \tag{1}$$

Let probable *shadja* candidates, $S_0(j)$ be a subset of M_k such that,

$$S_0(j) \subseteq M_k \ \ and$$
$$S_0(j) \in [100\,Hz : 200\,Hz], \ \ 1 \leq j \leq J < 10 \tag{2}$$

The number of candidates for S_0, J is found to vary across songs and across *rAgas*. One of the key attributes of S_0 and P_0 is their least variance across all octaves. To identify variance and weights of each *swara*, we fit a SC-GMM (semi-continuous Gaussian mixture model) to $p(x)$ as described in Section 2.1.

2.1. Semi-continuous GMM

In Indian classical music, the same *swarasthAna* of base octave are used in other octaves also, i.e., lower *shadja* and upper *shadja*. Similarly, the phrases(note sequences) of the *rAga*, are not altered in different octaves. Considering that a singer spans 3 octaves, the relative note ratios of all the 12 notes in each octave, with respect to S_0, are known. Hence, the density estimates, $p(x)$ can be constrained using Gaussian mixtures spanning these intervals. Since the means of the Gaussian mixtures will correspond to the set of discrete ratios, we need only to find the variances and their weights and hence we use the semi-continuous GMM formulation [8]. The density estimate of log pitch frequencies, $p(x)$ is modeled using 36 mixture GMMs, 12 GMM per octave, covering all the 3 octaves. The means, μ_i of the GMM are fixed and chosen from the ratios f.r.(C) given in Table 2. The means of lower octave $[S_-$ to $S_0)$ is half of that in the middle octave $[S_0$ to $S_+)$, while that of higher octave $[S_+$ to $S_{++})$ is twice. The SC-GMM parameters are initialized as in (3)

$$\sigma_i = \sum_{j=1}^{N_i}(x_j - \mu_i)^2; \alpha_i = \frac{N_i}{\sum_{i=1}^{36} N_i}$$
$$p(x) = \sum_{i=1}^{36} \alpha_i \mathcal{N}(x; \mu_i, \sigma_i) \tag{3}$$

DKP: Sangeetha Kalanidhi Damal Krishnaswamy Pattammal (March 28, 1919 July 16, 2009) was a prominent Carnatic musician. She, along with her contemporaries M. S. Subbulakshmi and M. L.Vasanthakumari are popularly referred to as the Female trinity of Carnatic Music. She has been awarded **PadmaBhushan and Padma Vibhushan** by Govt of India for her contributions to music.

where N_i is the number of pitch frequencies associated with i^{th} note-interval. Here, α_i is weight and σ_i is variance of each *swara* determined using EM algorithm. This SC-GMM is used for both *shadja* identification and *swara* classification. Fig 4(b) gives an example of the SC-GMM for a particular *AlApana* pitch estimate. To estimate S_0, we estimate a separate SC-GMM for each $S_o(j)$ estimated. Let $\alpha_{S0}, \alpha_{P0}, \alpha_{S+}$ and $\sigma_{S0}, \sigma_{P0}, \sigma_{S+}$ correspond to weights and variance of S_0, P_0, S_+ respectively. Let θ be the estimated *shadja* from $S_0(j)$. We define 5 estimators of *shadja* as:

$$\theta_a = argmin_{s_0(j)}\{\sigma_{S_0}|s_0(j)\}; \qquad j \in [1, J] \quad (4)$$

$$\theta_b = argmin_{s_0(j)}\{\sigma_{S_0} + \sigma_{P_0} + \sigma_{S_+}|s_0(j)\}; \qquad j \in [1, J] \quad (5)$$

$$\theta_c = argmin_{s_0(j)}\{\frac{\sigma_{S_0}}{\alpha_{S_0}}|s_0(j)\}; \qquad j \in [1, J] \quad (6)$$

$$\theta_d = argmin_{s_0(j)}\{\frac{\sigma_{S_0}}{\alpha_{S_0}} + \frac{\sigma_{P_0}}{\alpha_{P_0}} + \frac{\sigma_{S_+}}{\alpha_{S_+}}|s_0(j)\}; \qquad j \in [1, J] \quad (7)$$

$$\theta_e = argmin_{s_0(j)}\{\frac{\sigma_{S_0} + \sigma_{P_0} + \sigma_{S_+}}{\alpha_{S_0} + \alpha_{P_0} + \alpha_{S_+}}|s_0(j)\}; \qquad j \in [1, J] \quad (8)$$

All three, S_0, S_+, P_0 have higher probability in the considered *rAgas*. A KL-Divergence based estimation of *shadja* fails because S_0 and P_0 give similar SC-GMMs. Hence, we explore the following 5 different estimators to disambiguate the *shadja*. We find that the accuracy of *shadja* estimation is *rAga* dependent. Different estimators gave different performances. Table 3 gives a comparison for a set of *rAgas* for the above 5 estimators. For *rAgas Kalyani* and *Panthuvarali*, S_0 is 100% identified. In the case of *Mayamalavagowla* there is confusability of S_0 with P_0. In case of *Kharaharapriya*, S_0 is confused with $R2$. We see that θ_b, gives 100% performance in case of *Kharaharapriya* while θ_c gives better performance for *Harikambhoji* , and θ_e for *Shanmukhapriya*. Over all across *rAgas*, we find that θ_c is giving a good average performance.

3. *SWARA* IDENTIFICATION AND *RAGA* VERIFICATION

In SC-GMM, we determine the weights for each of the 36 notes based on their probability. We recognize that the singer, although traverses different octaves, s/he has to retain the ratio of note frequencies in all octaves. For the goal of *swara* identification, i.e., which 7 of the 12 *swaras* are used in the *AlApana*, it is logical to fold over the weights of the GMMs for the 3 octaves, (modulo 12). For instance,

$$\alpha_{R1} = \alpha_{R1_-} + \alpha_{R1_0} + \alpha_{R1_+} \quad (9)$$

Similarly for all 12 notes of the *AlApana*. Since any *sampurNa rAga* must have all the 7 *swaras*, the most probable R, G, M, D and N among their variants are chosen based on the largest of the alternate mixture weights of the folded SC-GMM as in (10). This we refer to as unigram based identification, $\hat{\xi}_u$.

$$\begin{aligned} \hat{\xi}_u(R) &= argmax\{\alpha_{R1}, \alpha_{R2}, \alpha_{R3}|s_0\} \\ \hat{\xi}_u(G) &= argmax\{\alpha_{G1}, \alpha_{G2}, \alpha_{G3}|s_0\} \\ \hat{\xi}_u(M) &= argmax\{\alpha_{M1}, \alpha_{M2}|s_0\} \qquad (10) \\ \hat{\xi}_u(D) &= argmax\{\alpha_{D1}, \alpha_{D2}, \alpha_{D3}|s_0\} \\ \hat{\xi}_u(N) &= argmax\{\alpha_{N1}, \alpha_{N2}, \alpha_{N3}|s_0\} \end{aligned}$$

Hk: Harikambhoji; Pv: Panthuvarali; Mk: Mechakalyani or Kalyani; Sb: Shankarabharana; Kh: Kharaharapriya, Sn: Shanmukhapriya, Mm: Mayamalavagowla

$\Re$	Q	% θ_a	% θ_b	% θ_c	% θ_d	% θ_e
Hk	6 (3)	50.00	66.67	100	83.33	83.33
Mk	6 (3)	100.0	66.66	100.0	100.0	100.0
Kh	9 (2)	44.44	100.0	66.67	77.78	88.89
Mm	9 (3)	77.78	55.56	77.78	33.33	55.56
Pv	7 (3)	71.43	71.43	100.0	100.0	100.0
Sb	10 (1)	60.0	60.00	90.0	90.0	80.0
Sn	8 (3)	37.50	62.50	87.50	87.50	100.0
Avg	55	63.02	68.97	88.85	81.71	86.83

Table 3: *Performance of S_0 estimators, as percent correct of # of shadja using different estimators for 7 melakartha rAgas*[2], $\Re$. *Second column gives # of songs (# of singers), Q*

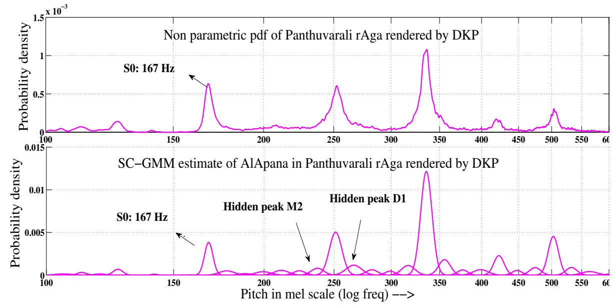

Figure 4: *Semi-continuous Gaussian mixture model estimate for AlApana of Panthuvarali rAga by DKP.*

From the designated $\{R,G,M,D,N\}$ for a *rAga*, if all the *swaras* are correctly recognized, we decide that the *rAga* is verified. Table 4 shows that we are able to recognize the *swaras* 86.3% correctly using unigram model, among 44 *AlApanas* of 7 *rAgas* rendered by different singers. We next examine Markov model based bigram and trigram models for the identifying the *swaras*. It is expected that Markov models will bring on the temporal structure (*swara* sequence) which is the first step to identify characteristics of *rAga*. We assign a note label U_t for every pitch frame(10 ms duration) value $y(t)$ according to the probability of the note as given by SC-GMM model. The bigram note probabilities are obtained as:

$$P(U_t = j|U_{t-1} = i) = \beta_{ij} \qquad i, j \in 1:36 \quad (11)$$

This gives the probability of transition from note-i to note-j for all time. We use a median filter to reduce short labeling errors while not sacrificing the true transitions from one *swara* to the next. The note labels are integer numbered to aid ordering and have been median filtered with a filter of order 5, removing label errors of ≤ 2. The output is down-sampled by 5. The state occupancy probabilities are wrapped around the 3 octaves as: $\beta_{R1} = \beta_{R1_-R1_-} + \beta_{R1_0R1_0} + \beta_{R1_+R1_+}$; similarly for all 12 *swaras*. The self transitions are used to identify the notes, $\hat{\xi}_b$ as in (12).

$$\hat{\xi}_b(R) = argmax\{\beta_{R1R1}, \beta_{R2R2}, \beta_{R3R3}|s_0\} \quad (12)$$

Similarly bigram estimates $\hat{\xi}_b(G), \hat{\xi}_b(M), \hat{\xi}_b(D), \hat{\xi}_b(N)$ are obtained for the corresponding notes G, M, D & N using β_{ii} analogous to (10). It is observed that note identification accuracy increases with note bigram probabilities (β_{ii}) as against unigram probabilities, as seen in Fig 5(b) and 5(c). This is because β_{ii} correlates with duration of state occupancy for any note-i. β_{ii} is consistently higher for valid notes in a *rAga*. Similarly, trigram note probability can be obtained as:

$$P(U_t = k|U_{t-1} = i, U_{t-2} = j) = \gamma_{ijk} \qquad i, j, k \in 1:36 \quad (13)$$

We use γ_{iii} for finding *swaras*, $\hat{\xi}_t$ similar to bigram.

A change in note label will denote the end of one note and transition to another. Thus, we also get the approximate annotation of the *AlApana* along with timing information as shown in Fig 5.

4. EXPERIMENTS AND RESULTS

All the analysis above is for *AlApana* and *swarakalpanas*. The audio files have been chosen from [9]. The sampling frequency of the audio is 22 kHz. The test procedure:

1. Input.mp3 file is processed by audio software, Praat. The pitch frequencies are estimated for every 10ms.
2. Pitch contour is converted to a density estimate to estimate the *shadja* S_0 using Parzen window method
3. SC-GMMs give probability measures for all 36 notes and *shadja* is estimated.
4. Notes are identified and *rAga* verification is done using the 7 maximum probable notes. We have considered only two variations of R $\{R1, R2\}$, G $\{G2, G3\}$, D $\{D1, D2\}$, N $\{N2, N3\}$ in our experiments.

The results of identifying the notes of the *rAga* and the *rAga* itself are shown in Tables 4 & 5

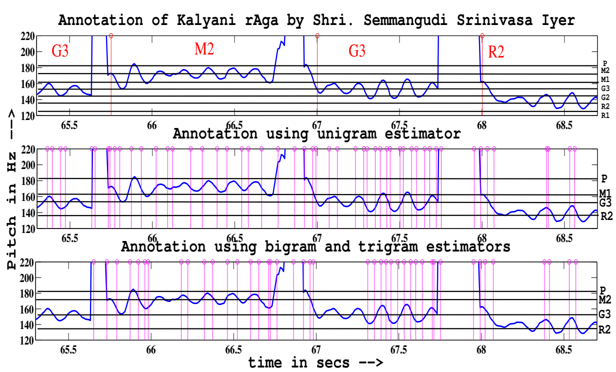

Figure 5: *Segmentation of a clip of Kalyani rAga containing gamakas (rendered by SSI) using unigram, bigram and trigram estimators. The bigram and trigram estimators giving same annotation. It can be noted that the unigram estimator is giving wrong swara, M1 while bigram and trigram give correct swara M2.*

5. CONCLUSIONS

This is a first effort towards fully automated *AlApana* analysis. By musicology, a *rAga* is not just the specific notes of *ArOhaNa* and *avarOhaNa*. It is additionally characterized by certain phrases, important notes and their transitions (*rAga lakshanas*) which need to be further explored. In *shadja* identification, we observe that estimation errors are more for low S_0. In *swara* identification, bigram and trigram perform better than unigram. Trigram has lesser accuracy than bigram because, for bigram the pitch is required to remain around the mean value for 100 ms only. But for trigram, it is expected to remain for 150 ms, which may not happen in the rendering of the *AlApanas*. Error in *rAga* verification is more due to pitch doubling errors, which get concentrated around 600 Hz, near S_{++} and N_+. However, use of stochastic model is promising and also required, because of enormous variability associated with a *rAga*. We have confined to a set of *sampurNa rAgas* and more experiments need to be done to understand and model the *janya rAgas*.

$\Re$	# of $\mathcal{A}$	$\hat{\xi}$% correct			Wrong notes	$\Re$ % for $\hat{\xi}_b$% correct
		$\hat{\xi}_u$	$\hat{\xi}_b$	$\hat{\xi}_t$		
Hk	6	92.0	86.7	93.3	$N3, M2$	50.0
Mk	6	80.0	100	100	-	100.0
Kh	6	96.7	90.0	93.3	$N3, G3$	50.0
Mm	7	85.7	88.6	77.1	$D2, N2$	42.9
Pv	7	80.0	100	97.1	-	100.0
Sb	9	80.0	95.5	93.3	$D1, N1$	77.8
Sn	7	88.6	80.0	74.3	$N3, D2$	14.3
Avg	**48**	86.3	**91.5**	89.8		62.13

Table 4: *Accuracy of different rAgas, $\Re$ and their swaras using θ_c shadja estimator for shadja rightly identified AlApanas, $\mathcal{A}$.*

$\Re$	# of $\mathcal{A}$	Confused *ragas* (# of misclassification, mistaken *swaras*)		
Hk	6	*Sb* (1, $N3$)	*Vachaspati* (1, $M2$)	*Mk* (1,$M2$,$N3$)
Kh	6	*Gaurimanohari* (2, $N3$)	*Hk* (1,$G3$)	-
Mm	7	*Suryakantha* (3,$D2$)	*Gayakapriya* (1,$N2$)	-
Sb	9	*Sarasangi* (1,$D1$)	Hk (1,$N1$)	-
Sn	7	*Simhendra-madhyama* (1, $N3$)	*Dharmavati* (1, $D2$, $N3$)	*Hemavati* (4, $D2$)

Table 5: *Confused rAga $\Re$, in case of misclassification using bigram estimators*

6. REFERENCES

[1] H. V. Sahasrabuddhe, Searching for a Common Language of RAgas, Proc. Indian Music and Computers: Can Mindware and Software Meet?: August 1994

[2] R. Upadhye and H. V. Sahasrabuddhe, On the Computational Model of Raag Music of India, Workshop on AI and Music: 10th European Conference on AI, Vienna: 1992.

[3] G. Pandey et al, Tansen : A system for automatic rAga identification, Proceedings of the 1st Indian International Conference on Artificial Intelligence, 2003

[4] P. Chordia, J Jayaprakash and A. Rae, Automatic Carnatic Raag Classification, Journal of the Sangeet Research Academy (Ninaad), 2009.

[5] P. Chordia, A. Rae, Automatic Raag Classification Using Pitch-class and Pitch-class Dyad Distribution, Proc. 7th International Conference on Music Information Retrieval (ISMIR), 2007.

[6] P. Chordia, Automatic rag classification using spectrally derived tone profiles, Proc International Computer Music Conference (ICMC), 2004.

[7] A. Krishnaswamy, On twelve basic intervals in south Indian classical music, 115th AES Convention, New York, 2003.

[8] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. Redmond, Washington: Prentice Hall, 2001

[9] `http://www.shivkumar.org/music/manodharma/index.html`