Distance Metrics to Aid Rāga-selection in Subjective Studies

Venkata S Viraraghavan1,2, Arpan Pal1, R Aravind2, Hema A Murthy3

1TCS Research and Innovation, Tata Consultancy Services Limited
2Dept of Electrical Engineering, Indian Institute of Technology Madras
3Dept of Computer Science and Engineering, Indian Institute of Technology Madras

Abstract

Recent studies corroborate the beliefs that listening to Indian music is beneficial in terms of mental abilities or mental health. These studies are also akin to those on other genres of music, where similar benefits have been observed. Studies relating to specific rāgas, however, are not as conclusive. The rāga is a key concept in Indian music, and comparable to a mode or scale in Western music. Rāgas are significantly richer than modes/scales due to the deliberate use of continuous pitch variation called gamakas. While most rāgas seem real distinguishable entities, distinctions between some rāgas appear to be due to conditioning. In this paper, we propose a quantitative measure of distance between the histograms of anchor notes in a rāga. Anchor notes were introduced recently to study the nature of gamakas. We consider several distance measures on anchor- and CPN-histograms. For each rāga, other rāgas falling in the lower one-third of the range of distances to that rāga are counted as confused. Out of 1560 possible pairs of confusion, 178 (11%), 434 (28%), and 956 (61%) are confused according to, respectively, the Euclidean, cosine and KL-divergence distances between anchor histograms. CPN-histograms show a higher confusion. The Euclidean and cosine distance metrics between anchor-histograms thus help narrow down the choice of rāgas, such as for the experimental and control groups, in studies relating to the mind.

Index Terms: Carnatic rāgas, anchors, distance metric.

1. Introduction

From ancient times, music is believed to be capable of influencing the human mind in several ways. In the last century, and more so in the last few decades, several scientific studies have focused on the effect of music on the human mind. Improved cognitive ability, stress relief are among the beneficial effects of music [1, 2], while music therapy is also popular [3]. In recent decades, scientific studies on the effect of Indian art music (IAM) on the mind have become prominent [4, 5]. A central concept in IAM is called the rāga, and related studies are, therefore, rāga-based. We survey related work on WM and IAM in Section 1.1.

In most musical traditions of the world, at least twelve notes per octave are used. In IAM, a rāga uses specific subsets of these notes, in a specified order. Rāgas are thus likened to scales or modes in Western music (WM). However, IAM consists of not only segments of music with a fixed-pitch, but also deliberate, continuous pitch variation called gamakas. On a violin, gamakas are played by sliding the finger on the fingerboard with sustained bowing. Gamakas are mandatory in vocal music and must be mimicked faithfully in instrumental music (which is not possible in instruments such as the piano). Carnatic music (CM) and Hindustani music (HM) are two important sub-genres of IAM. Our study focuses on CM, but is relevant to HM too. In this paper, we propose a novel distance-measure between rāgas, based on the recently introduced anchor notes in CM.

1.1. Previous work

We start with a listing of types of studies on the influence of music on the mind. The effect of background music on cognitive abilities was famously reported as the Mozart effect [6], but these effects are yet to be understood [7]. In therapy, musical interventions help patients, e.g., with dementia [8]. Similar therapy based on CM and HM are common in India [9]. In yet another context, music helps in stress relief, and is next only to natural sounds [10]. The emotional impact of music and the emotional perception of music are studied [11]. Further, the impact of specific components of music, such as of scales or of chords in WM [12], and gamakas in CM [13], are also studied. Clearly, the genre and pieces of music have to be carefully chosen for such studies. A key consideration in these choices for studies with CM and HM is the set of rāgas to be used.

Rāga-based approaches continue to dominate studies of HM and CM, despite some evidence to the contrary [14]. For example, the Hindustani rāga Tōḍī has been studied recently [5]. However, since the participants in the control group heard no music, it is not clear whether music, in general, or specifically the rāga Tōḍī has the calming effect. Similarly, the effect of the Carnatic rāga Nīlāmbhari on sleep quality was studied [15], with the rāga Kālāyāni used as a control rāga. As far as we are aware, the rāga for the experimental and control groups are chosen based on their ‘feel’ stated in musicology [16]. However, for scientific studies, a measurable approach is preferred. Quantitative measures of distances between rāgas do exist [17, 18, 19], but the features used are not directly interpretable. Thus, correlation with musicological analyses is limited. Recently, the anchor notes in CM have been defined, and their histograms have been interpreted qualitatively according to musicology [20]. In this paper, we propose a distance measure between rāgas based on the histograms of interpretable anchor notes.

1.1.1. Database

For our experiments in this paper, we use the Carnatic subset of the popular CompMusic dataset. This subset has been used in [17, 21, 22, 20]. We call this subset the ‘database’ in this paper. The database has 40 rāgas with 12 professional-concert renditions in each rāga. The predominant pitch, usually the pitch of vocal music, in each rendition was tracked using the Melodia algorithm [23]. The pitch was measured every \(n/2255\sec \approx 4.4\) ms on 46 ms windows to yield \(f[n]\) in Hz. The range of \(n\) depends on the rendition’s duration. For each rendition in the database, its tonic pitch is also available.
We convert the pitch values in the database, \( \{ f[n] \} \) in Hz, to semitones as \( s[n] = 12 \log_2(f[n]/f_t) \). An octave consists of the pitch-values \( f_t \) to \( 2f_t \), where \( f_t \) is the tonic pitch of the rendition. The tonic pitch corresponds to 0 semitones. If the tonic pitch corresponds to a certain key in a tuned piano, the other keys are at integer semitones. The first key to its right (irrespective of color) is at +1 semitone, the first key to its left is at −1 semitones, the second key to the right is at 2 semitones, and so on. On instruments such as the violin, pitches between integer semitones can be realized. Carnatic musicians deliberately traverse fractional semitones in gamakas.

2. Constant-pitch notes and anchors

The concept of \( \text{r} \text{g} \text{a} \) was introduced in Section 1. A more detailed description can be found in [24, 20], while gentler introductions abound on the Internet. Although CM has a profusion of \textit{gamakas}, the segments of the pitch curve where the pitch is fixed are of great importance too [25]. We call such segments 'constant-pitch notes' or CPNs. Further, we call the pitch curve outside CPNs as transients, which characterize the \textit{gamakas} in CM. It has been found that the direction of the \textit{gamakas} potentially carries \textit{r} \text{g} \text{a} \text{g} \text{i} information [20], which motivates the definition of anchors. Given a pitch curve, the formal definitions of CPN, transients and anchors, adapted from [20], are given below.

CPN: A segment whose pitch-values vary from their mean pitch by at most \( \Delta = 0.35 \) semitones, and lasts for at least \( C_{\text{min}} = 0.08 \) seconds.

Transient: Any pitch curve outside CPNs.

Anchor: A CPN adjacent to a transient

Upward anchor: Anchor whose transient rises in pitch.

Downward anchor: Anchor whose transient falls in pitch.

In this paper, we use the term ‘CPN’ when the directionality of the adjacent transient1 is not relevant. We use the term ‘anchor’ when this directionality matters. For upward and downward anchors, it does not matter if the adjacent transient precedes or succeeds it. It is possible for an anchor to be both upward and downward. The algorithms to extract CPNs and anchors can be found in [25, 22, 20].

A common feature used in the literature is the histogram of pitch-values wrapped around an octave, \( s_n[n] = \text{mod}(s[n], 12) \). A histogram of the wrapped values is computed by counting the number of occurrences in bins of equal size. We use the bin-width of \( \beta = 0.1 \) semitones suggested by [20]. Thus, there are 120 bins per histogram. In this paper, we divide the number of occurrences in every bin by the total number of pitch values to normalize histograms. While histograms are interpretable according to musicology, other multidimensional feature representations, such as in [17], are not interpretable.

The histogram of all pitch-values in a 55-minute rendition in the Carnatic \textit{r} \text{g} \text{a} \text{g} \text{i} \text{K} \text{a} \text{l} \text{y} \text{a} \text{n}i, and histograms of CPNs and transients in the same rendition, are shown in Figure 1. The nature of CM is evident in the histogram. Had it been a piano rendition, the histogram would have had non-zero values only around integer semitones. However, the histogram of all pitch values in the rendition, is observed to be significant even far away from piano-key pitch-values. The same is true about the transients. This behavior is expected because the pitch does vary continuously in CM, and all regions of the octave will be occupied.

Interestingly, the histogram of CPNs shows sharp peaks around piano-keys and is insignificant elsewhere2.

In CM, the characteristics of \textit{r} \text{g} \text{a} \text{s} do not change significantly across music schools. Thus, renditions in a \textit{r} \text{g} \text{a} \text{g} can be combined to increase data for analysis [20]. The normalized histograms of upward and downward anchors for all renditions in the \textit{r} \text{g} \text{a} \text{g} \text{i} \text{K} \text{a} \text{l} \text{y} \text{a} \text{n}i are shown in Figure 2. Again, we observe the sharp peaks near integer semitones. Further, the peaks for upward and downward anchors differ significantly in height. This asymmetry, observed in 31 out of 40 \textit{r} \text{g} \text{a} \text{s} [20], and carries \textit{r} \text{g} \text{a} \text{g}-information. In order to evaluate the distance metrics proposed in Section 3, we split the database into two sets. All renditions in the database are of similar concerts by professional musicians of comparable standing. Further, the order of renditions in the database is arbitrary. Therefore, Renditions 1, 3, 5, 7, 9 and 11 are chosen in Set 1 and Renditions 2, 4, 6, 8, 10, 12 are chosen in Set 2. Figure 3 shows the histograms of upward anchors for the two sets in the \textit{r} \text{g} \text{a} \text{g} \text{i} \text{K} \text{a} \text{l} \text{y} \text{a} \text{n}i. These histograms agree more than the histograms of upward and downward anchors in Figure 2. We seek distance metrics between anchor-histograms where, ideally, inter-\textit{r} \text{g} \text{a} \text{g} distances exceed intra-\textit{r} \text{g} \text{a} \text{g} distances.

3. Proposed distance metrics

We concatenate the histograms of upward-anchors and downward-anchors in each \textit{r} \text{g} \text{a} \text{g} to construct a combined histogram. Given two combined histograms \( h_1[l] \) and \( h_2[l] \), \( 0 \leq l \leq L \).
Figure 3: Upward-anchor histograms in Sets 1 and 2 (Kalyāṇi).

$l < L = 240$, we consider three well-known distance metrics between them. The first is the Euclidean distance metric, $E(h_1, h_2)$, and the second is the cosine distance metric, $C(h_1, h_2)$. The third is the KL-divergence, $KL(h_1, h_2)$, which is a not a metric, but is commonly used to compare probability distributions, such as normalized histograms. They are defined as follows.

$$E(h_1, h_2) = \sqrt{\sum_{l=0}^{l<L} (h_{1}[n] - h_{2}[n])^2} \quad (1)$$

$$C(h_1, h_2) = 1 - \frac{\sum_{l=0}^{l<L} h_{1}[n]h_{2}[n]}{\sqrt{\sum_{l=0}^{l<L} h_{1}^2[n]} \times \sum_{l=0}^{l<L} h_{2}^2[n]} \quad (2)$$

$$KL(h_1, h_2) = \sum_{l=0}^{l<L} h_{1}[n] \log_2 \frac{h_{1}[n]}{h_{2}[n]} \quad (3)$$

In general, we could compute the histograms of upward and downward anchors after combining all available renditions in a rāga[20]. However, to evaluate the proposed metric, we compute the histograms separately for renditions in Set 1 and for renditions in Set 2. For Rāga r, its combined anchor-histograms in Sets 1 and 2 are denoted, respectively, by $h_{r,1}[n]$ and $h_{r,2}[n]$. We construct a $40 \times 40$ matrix $D_E$ with Euclidean distances calculated as:

$$D_E[r, q] = E(h_{r,1}, h_{q,2}) \quad (4)$$

where $r$ and $q$ index the 40 rāgas. The distance matrices $D_{KL}$ and $D_{C}$ are constructed similarly using the KL-divergence and cosine distances, respectively.

Figure 4(a) shows the ‘heat map’ for the Euclidean distance metric. Each square at $(r, q)$, $1 \leq r \leq 40; 1 \leq q \leq 40$, represents the distance between the anchor histograms for Rāga r and Rāga q. The squares on the main diagonal are the intra-rāga distances $D_E(r, r)$ for $1 \leq r \leq 40$. Distances where $r \neq q$ represent the inter-rāga distances. The color of the square corresponds to the distance value as given by the Legend on the right. A distance of 0.0 is shown in dark blue and the maximum is shown in red. For values in between, dark blue to light blue indicate low distances, blue-green to green indicate mid-range distances and yellow to red indicate high distances.

To interpret Figure 4(a), we observe that the main diagonal is dark blue, which shows that the intra-rāga distances are small. The off-diagonal elements show varying colors depending on the closeness of rāgas as measured by the Euclidean distance. A large part of the distance matrix shows mid-range distances, a few areas show large distances. For a square that is red, the rāgas corresponding to that square are far apart. For example, see the pair of rāgas Kalyāṇi and Tōḍī. A pair of rāgas that are close to each other are Sankarābharanam and Bilahari.

Figure 4(b) shows the heat map for cosine distance metric. The heat map is similar to that in Figure 4(a). Most pairs of rāgas that have a low Euclidean distance also have a low cosine distance. However, there are more blue, non-diagonal squares. That is, there are more instances of inter-rāga distances being low than for the Euclidean distance.

The distance matrices are not symmetric by definition. For example, $D_E(r, q)$ is the Euclidean distance between the anchor-histogram for Rāga r in Set 1 and the anchor histogram for Rāga q in Set 2. $D_E(q, r)$ is the Euclidean distance between the anchor-histogram for Rāga q in Set 1 and the anchor histogram for Rāga r in Set 2. In general, $D_E(r, q)$ need not equal $D_E(q, r)$. However, if the rāga’s characteristics do not vary between Sets 1 and 2, a symmetric heat map is expected. The heat maps of Figure 4 are approximately symmetric.

4. Evaluation and results

To choose between the distance metrics, we consider musico-logical and subjective evaluation. Pairs of rāgas that are close to each other according to the heat map should be explainable according to CM practice.

To evaluate how much rāgas are confused according to the chosen metric, we introduce a quantitative measure. For a Rāga r, we first find the maximum value, $M$, and minimum value $m$, among $\{D_E(r, q)\}$. $1 \leq q \leq 40, q \neq r$. We choose a distance threshold $T = \frac{M-m}{40} + m$, and Rāgas q for which $D_E(r, q) < T$ are counted as confused with Rāga r. That is, rāgas in the bottom one-third of the range of distances are counted as confused with Rāga r. Similarly, the number of confused rāgas are also evaluated for the other distance measures.

The number of confused rāga-pairs is given in Table 1, and $40 \times 39 = 1560$ confusions are possible in total. The number of confusions as a percentage relative to 1560 is also shown in brackets in Table 1. These values are shown for the Euclidean, cosine, and KL-divergence distances. The lowest confusion is for the Euclidean metric. Low confusions for one of the metrics suggests that rāgas are real, discernible entities.

Since the KL-divergence distance confuses rāga-pairs most and exhibits the least symmetry (heat map not shown), only the Euclidean and cosine distances are considered for further evaluation. For reference, the confusions based on CPN-histograms with the same metrics are given in the last row of Table 1. Anchors, which consider only the direction of the adjacent continuous pitch movement, contain more information about rāgas than CPNs. This observation may be related to the observation that the direction of a melody is used by listeners to identify tunes [27].

Allied groups of rāgas known to musicologists are found among the confusions based on Euclidean and cosine distances. E.g., (Bhairavi, Mukhari, Hussēnī, (Śrī, Madhyamavati), Some pairs, such as Varālī, Kannavardani, are explained by CM practice. Further, atypical confused pairs such as (Athana, Kamboji), can be explained by anchors, which is not possible with un-interpretable distance measures [17, 18, 19]. The set (Kalyāṇi, Sankarābharanam, Bilahari) is relevant to [15], where the effect of rāga Nīlambāri on sleep was studied. Nīlambāri is believed to induce sleep, but no significant difference was found compared to the effect of rāga Kalyāṇi. Nīlambāri, which is close to Sankarābharanam, is likely to be close to the rāga Kalyāṇi too as per Figure 4. Thus, the choice of rāga Kalyāṇi as a control rāga may need to be reassessed.

8
Table 1: Number (and percentage) of confused rāga-pairs.

<table>
<thead>
<tr>
<th>Type</th>
<th>Euclidean</th>
<th>Cosine</th>
<th>KL-Divergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchors</td>
<td>178 (11%)</td>
<td>434 (28%)</td>
<td>956 (61%)</td>
</tr>
<tr>
<td>CPNs</td>
<td>226 (14%)</td>
<td>535 (34%)</td>
<td>1015 (65%)</td>
</tr>
</tbody>
</table>

The musicological evaluation above is representative of a very small population of experts that are familiar with rāgas. In the context of studies considering the impact of CM on the mind, it is important to consider the perception of people who are not experts in CM.

As far as we are aware, there is (surprisingly!) no study that documents Carnatic rāgas that are confused with each other by non-expert listeners. In this context, we recall an earlier, related experiment [25], where participants had to identify rāgas from ~30-second snippets, some of which were modified digitally. Out of 50 participants, 28 participants correctly identified the rāgas of all original clips. The incorrect choices made by the remaining 22 participants, therefore, are possible pairs of rāgas that are confused. To obtain an initial indication, we re-ran the distance metric for nine confused rāga-pairs that overlapped with our database. Out of seven pairs of rāgas confused by the 22 participants, four pairs are also confused by the cosine distance metric, of which two are confused by the Euclidean distance metric. In addition, four other rāga-pairs are confused by the cosine distance metric, and two others by the Euclidean distance metric, but are not confused by the 22 participants. These results highlight the need for subjective studies that focus on rāga-similarity as perceived by the general population. We observe that initiation into rāga identification in CM is through similarity of songs: “Song $S_1$ sounds like Song $S_2$ in a known rāga $R$; therefore Song $S_1$ is likely to be in rāga $R$.” Thus, the perceived similarity between tunes in different rāgas (and same rāga) may be used in such experiments, without the need for participants to name rāgas.

5. Conclusions

We proposed a novel distance measure between Carnatic rāgas. It is based on the histograms of recently introduced upward and downward anchors. We showed that anchor-histograms, which consider the direction of the adjacent pitch movement, contain significant information about rāgas. The Euclidean distance between histograms showed the lowest number of confused pairs of rāgas. Clearly, further studies on distance metric(s) and detailed subjective evaluations are needed. Currently, the results presented in this paper can help narrow down the choice of rāgas in studies on rāgas and the mind. Rāgas for the main experiment and for the control group could be chosen from pairs whose intersection is not a blue square in the heat maps.

6. Acknowledgements

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).

7. References


