COMPUTATIONAL ANALYSIS OF MELODIC MODE SWITCHING IN RAGA PERFORMANCE

Nithya Shikarpur[1] **Asawari Keskar**[2] **Preeti Rao**[1]
[1]Department of Electrical Engineering, [2]Department of Physics
Indian Institute of Technology Bombay, India
{nithyas, prao}@ee.iitb.ac.in

ABSTRACT

Melodic mode shifting is a construct used occasionally by skilled artists in a raga performance to enhance it by bringing in temporarily shades of a different raga. In this work, we study a specific North Indian Khyal concert structure known as the Jasrangi jugalbandi where a male and female singer co-perform different ragas in an interactive fashion. The mode-shifted ragas with their relatively displaced assumed tonics comprise the identical set of scale intervals and therefore can be easily confused when performed together. With an annotated dataset based on available concerts by well-known artists, we present an analysis of the performance in terms of the raga characteristics as they are manifested through the interactive engagement. We analyse both the aspects of modal music forms, viz. the pitch distribution, representing tonal hierarchy, and the melodic phrases, across the sequence of singing turns by the two artists with reference to representative individual performances of the corresponding ragas.

1. INTRODUCTION

Ragas are melodic modes that underpin all performances of Indian art music across both the North Indian (Hindustani) and South Indian (Carnatic) traditions. There are dozens of ragas in common practice, distinguished by their salient melodic properties which include the choice of tonal material, the hierarchy of notes (*svara*), their intonation and typical phrasal contexts and, finally, the association with a particular mood. A drone sounds the tonic throughout the concert making the relative intervals of all the notes clearly apparent to the listener providing, thus, strong cues to the raga identity in terms of the corresponding tonal material and hierarchy, and the melodic phrases. Automatic raga identification from computed pitch class histograms, normalised by the concert tonic, have worked well, especially when ragas with different tonal material appear in the dataset [1, 2]. Ragas with the identical scale notes relative to the tonic such as allied ragas have also been successfully differentiated with pitch distributions that exploit the differences in intonation and emphasis across the common set of svara [3]. A particular raga performance construct that has not received much attention computationally is melodic mode switching. As in other modal music, it is possible to develop a large number of scales by means of the modal shift by using any note of a raga as the tonic and building a new set of scales by maintaining the intervals or ratios between the notes and the tonic [4]. Termed '*murchana*', this is a subtle form of modulation (similar to key modulation from a major to its relative minor) and is used by skilled musicians to temporarily bring in shades of another raga during a performance.

In this work, we specifically consider a recent but widely acclaimed development in the North Indian Khyal classical music scenario of mode-shifted ragas performed together in concert by a pair of singers. Motivated by a desire to create a space where a male and a female singer can perform together in spite of the ½ octave difference typical of their vocal pitch ranges, the two ragas are chosen such that the M (the lower octave fourth) of the higher-pitched voice serves as the S (tonic) for the lower-pitched voice with all actual note values (i.e. in terms of fundamental frequency or MIDI number) being identical in the two [5]. Figure 1 shows the scales of the pentatonic raga pairs considered in this paper. With the concert drone typically tuned to the Sa of the female voice, the singers strive to maintain the character of their respective ragas in the '*jugalbandi*' performance that interleaves the two singers' voices in equally weighted roles, as in what may be considered a 'call and response' musical format. What makes it particularly interesting is that the call and the response, both improvised, are drawn from different ragas. With a common set of notes, the challenge lies in meaningfully linking the phrases during the interaction while also carefully preserving the individual raga-specific characteristics. There are occasional episodes of singing together.

The above discussed form, known as the "*Jasrangi jugalbandi*" after proponent Pt. Jasraj, has been performed over the past decade by a handful of well-known Hindustani vocalists drawing from a limited number of raga pairs [6]. The chosen raga pairs presumably satisfy the music theory and aesthetic requirements for the simultaneous presentation in the jugalbandi concert format. In the present work, we carry out the computational analy-

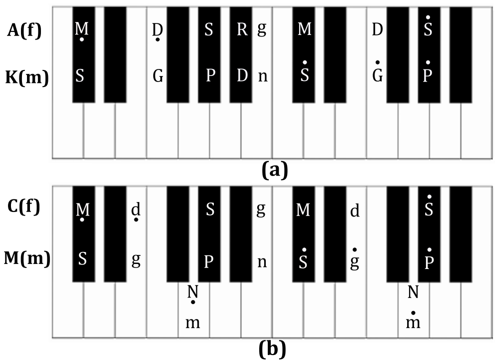

Figure 1. The solfege of each raga superposed on the standard keyboard. 'S' denotes the tonic note of the raga. We see that the same set of keyboard notes is shared across ragas in each pair. (a) Abhogi (A) and Kalavati (K); (b) Chandrakauns (C) and Madhukauns (M). Ragas A and C are sung by the female (f) artist of the corresponding pair, ragas K and M by the male (m).

ses of the melodic content of available Jasrangi jugalbandi (shortened to "JJ" in this paper) concerts in two different raga pairs. Specifically, we seek to answer questions around the JJ concerts with reference to the corresponding normally performed individual ragas that can possibly be addressed with computed melodic representations. An important question is the extent to which raga characteristics are preserved when performed in the context of a JJ song. Considering that the interaction between singers is the main highlight of a JJ performance, we are also interested in analysing the melodic relationships of the call and response format. The larger goal for this empirical study is to derive the structure or schema of the JJ concert in terms of the individual raga presentations and the dynamics of the interaction between the two artists. Apart from its value to music appreciation and pedagogy, the insights obtained could serve to identify new raga pairs that potentially fit the format for future JJ performance. In the next section, we present our dataset with descriptions of the represented raga pairs. The manual annotation and the observations directly derivable from this are provided. The melodic representations considered in this work are discussed next. Our experimental analyses are presented with a discussion of the observations and the implications of the outcomes.

2. DATASET & AUDIO PROCESSING

There have been about 7 distinct raga pairs in publicly performed JJ concerts so far. We select, for our work, the two raga pairs that are best represented in the publicly accessible JJ concert recordings. These are the pentatonic raga pairs, *Abhogi-Kalavati* (A-K henceforth) and *Chandrakauns-Madhukauns* (C-M). We also identify a number of the corresponding raga-specific (i.e. individual) concert recordings.

A concert may comprise more than one song (*bandish*), complete with improvisation and chosen composition, and we segment the concert audio accordingly. Our

Feature	Abhogi	Kalavati	Chandrakauns	Madhukauns
Aaroh	SRgMDṠ	SGPDnDṠ	SgMdNṠ	SgmPnṠ
Avaroh	ṠDMgMgRS	ṠnDPGPGS	ṠNdMgMgSṆS	ṠnPmgS
Vadi, Samvadi	S, M	P, S	M, S	P, S
Nyas	S, R, M, D	S, P, D	S, M, N	S, m, P
Char. Phrases	ḌSRg, MRS, RḌSRg, MgRS, MDṠ	SGPD, PDnD, GPDṠ, GPDP, GPGS	SgMgS, ṆS, gMdNṠ, NdṠ, NdMgMgS	Sgm, PmgmP, mPnṠ, ṠnPmg S

Table 1. Raga grammar details; dots over and under indicate upper and lower octave pitches respectively

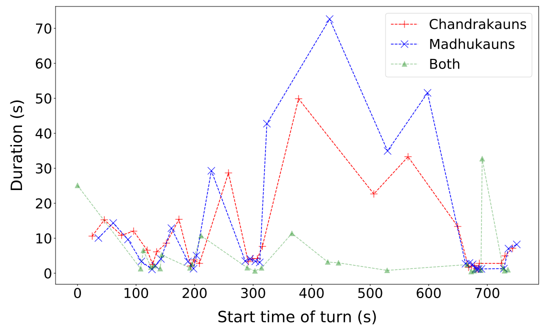

Figure 2. Turn duration versus start time plotted separately for each singer and for the simultaneous (both) singing for a C-M JJ song from our dataset

dataset, summarised in Table 2, comprises recordings from the Hindustani music corpus Dunya compiled as part of the CompMusic project [7–9] supplemented by available YouTube audio content by well-known Khyal artists, especially for the relatively more scarce JJ concerts. The chosen recordings are vocal performances accompanied by the tanpura (drone), tabla and harmonium. The JJ concerts are manually segmented into male, female and simultaneous-singing episodes or turns Figure 2 illustrates the temporal sequence of singing turn durations across the three labels for one of our C-M JJ songs. We note the dominance of solo singing with roughly equal durations across the two artists, an important characteristic of the JJ format. The simultaneous singing episodes are relatively short and are observed to correspond to the refrain (mukhda) of the song and sometimes the long held notes.

Our melodic analysis is based on vocal pitch contours automatically extracted from the audio recordings. The pre-trained 4-stems Spleeter model [10] is used to obtain the vocals component from the stereo mixtures. While not trained on Indian art music, the model produces vocals with either absent, or low enough, levels of the accompanying melodic instruments for the ensuing vocal pitch detection step to work reliably. The regions of the separated vocals, corresponding to the previously labeled solo singing regions, are processed for F0 detection at intervals of 10 ms using an autocorrelation function based method followed by temporal smoothing with search range restricted to the anticipated two octave range of the singer [11, 12]. The resulting pitch time series were checked for accuracy via listening to the resynthesis for any extended duration errors that could be corrected with suitable adjustments of

Raga	Number of songs (minutes)		
Pair	Raga 1	Raga 2	JJ
A-K	12 (185)	12 (227)	5 (89)
C-M	13 (214)	14 (171)	4 (69)

Table 2. Number of songs and total audio duration for each raga and JJ raga-pair in our dataset

the pitch analysis settings. The concert tonic is obtained for each song using an automatic tonic detection method that exploits the presence of the drone in a multipitch detection framework and further manually verified [13, 14].

3. MELODIC REPRESENTATIONS

In Western music, the psychological perception of "key" has been linked to distributional and structural cues present in the music [15]. Likewise, the bare essential theoretical description of a raga lists the allowed pitch classes, the dominant (*vadi*) and sub-dominant (*samvadi*) notes, the resting notes (*nyas*) and the raga motifs or characteristic phrases. Table 1 presents the same for the ragas in our study. The distribution of pitch classes in terms of either duration or frequency of occurrence in scores of Western music compositions has been found to correspond with the perceived tonal hierarchies in different keys [16–19]. That is, the emphasis given to the different notes in a composition is indicative of the underlying musical scale.

First-order, octave-folded pitch distributions (or pitch profiles) computed from the concert recording, and normalised by the concert tonic F0, have been widely used in raga identification. Given the pitch continuous nature of the Indian traditions, different kinds of representations and distance measures have been experimentally evaluated in different contexts [1]. Derived from the continuous pitch contour, the instantaneous pitch samples can be binned as such to obtain 'continuous pitch histograms'. The bin width is an important analysis parameter choice in this case with finer bin widths more fully capturing precise note intonation and note transitions that are possibly distinctive of the raga. Bin widths of up to 25 cents (i.e. 48 intervals per octave) were found to obtain perfect separation in the clustering of allied raga concerts, with performance degrading at larger bin widths [20].

Alternately, a stage of segmentation and quantization can be applied to the continuous pitch time series to obtain 'stable note' regions. To account for non-standard note intonations, the underlying scale interval locations or svara are estimated from the most prominent peaks of the finely-binned long-term tonic-normalized continuous pitch histogram across the concert. Following [3], segments of the melodic contour of duration greater than 250 ms that display a deviation of within +/-35 cents from a svara location, with gaps upto 100 ms discounted, are labeled as stable notes corresponding to the particular svara.

Octave-folded histograms are computed for each raga-specific song by accumulating the pitch values across the song audio. The continuous pitch (CP) distribution with 25 cent bin width has a vector dimension of 48. The stable notes histogram has a dimension equal to the number of raga notes or pitch classes. We consider two distinct interpretations for the strength of a pitch class in the song audio, viz. its total duration and the number of occurrences (or count) to obtain two types of stable-note (SN) histogram representations. The previous studies on raga recognition using first-order distributions have considered a variety of distance measures to quantify the similarity between two histograms. For this work, we implement the two measures that have been found to perform best, viz. correlation and Bhattacharya distance [3,9]. We further extend the study to unfolded pitch distributions to investigate how the octave dependence of the realised notes contributes to the discrimination of the mode-shifted ragas in performance.

Apart from pitch profiles, the melodic character of a raga lies in sequential representations including motifs, as is true for other forms of modal music [21, 22]. Similarity of phrase shapes has been exploited in raga recognition using sequence matching techniques [8, 23, 24]. Different ways to deal with the challenges from melodic variation inherent to oral traditions and the imperfect correspondence of any automatically quantized pitch contour to the underlying note sequence give rise to a variety of melodic representations and distance computation methods [25]. For the current task, we employ the continuous pitch contour as well as the extracted stable note (SN) sequence to derive various features that potentially capture the salient characteristics of the JJ song call and response phrases.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Our experiments are targeted towards (i) examining how closely the singers adhere to the raga-specific characteristics of their respective parts in the JJ song while drawing from the identical sets of notes in terms of the standard keyboard as in Figure 1, and (ii) describing the interaction between the two singers to model the phenomenon of 'call and response'. Raga-specific characteristics are modeled from the individual raga concerts in our dataset. We implement the distinct melodic representations, presented in the previous section, for each raga-specific song as also for the separated raga components of each JJ song. In order to simulate the JJ concert scenario of a single overall concert tonic, we pitch transpose all the concert audios so that the female-sung raga songs (i.e. Abhogi and Chandrakauns) assume a tonic (i.e. the svara S) of 207 Hz (G#3 on the standard keyboard) and their complementary raga songs assume C#3 for the tonic, corresponding to the depiction in Figure 1. This ensures that all the represented note pitches in terms of the standard keyboard notes are drawn from the same set for all concerts within a raga pair, making the raga discrimination ambiguous to that extent (just as it would be to the listener of the JJ performance). The experiments, presented next, are organised based on the specific melodic representation employed in the comparisons.

4.1 Pitch distributions

We consider three distinct first-order pitch distributions: the 25-cent bin width continuous pitch (CP) histogram and the stable-note histograms based on duration, SN-d, and count, SN-c. We evaluate both octave-folded and unfolded histogram representations of each type. The fixed tonic normalization gives rise to histograms that are aligned across the two ragas of a pair as in Figure 1. The similarity between two songs is then computed by the distance between their corresponding distributions. Based on the assumption that the individually performed ragas adhere closely to the raga grammar, we use the estimated 'goodness of clustering' for the raga-specific songs (i.e. the individual raga concerts only) to obtain a model for the subsequent investigation of the JJ songs. A popular measure of cluster quality is the silhoutte coefficient [26]. For the 2-class problem, it is computed as a normalised difference between the mean distance of a point to points in the opposite cluster and the mean distance to other points in its own cluster [27]. It can take on values in [-1.0, 1.0] with higher positive values indicating superior clustering, i.e. the points are better matched to their own cluster members. The silhoutte coefficient is computed for each raga-specific song with respect to the two clusters, viz. other songs in the same raga and all songs of the complementary raga.. Finally, an average silhoutte coefficient is obtained for each raga pair across all the individually performed songs in either raga, as presented in Table 3.

4.1.1 Discriminating raga-specific songs

All the considered combinations of pitch distribution and histogram distance measure appear in Table 3. A number of observations are apparent. We obtain positive valued coefficients in all cases implying that the tonal hierarchy is sufficiently differentiated within a pair even with identical keyboard notes and fixed tonic. For both raga pairs, the unfolded representations yield superior clusters, with the distinctions between other variations being somewhat less marked. The unfolded distribution is, of course, helped by the pitch transposition effected between the ragas, as reviewed earlier in this section. The distance measures, on the other hand, exhibit differences. As for the octave-folded representations, only the A-K ragas demonstrate good separation with the stable-note duration (SN-d) representation taking on the highest values for both the distance measures. A possible explanation for reduced separation in C-M is the overlapping nyas (rest notes) across the 2 ragas. We can conclude that of the considered methods of comparing pitch distributions, the Bhattacharya distance between SN-d representations overall best separates the individual songs across the two ragas of the complementary raga pairs with the octave unfolded representation doing better. This therefore forms our computational model of melodic similarity to be applied to the investigation of raga components of the JJ songs, as described next.

4.1.2 Discriminating ragas performed in JJ songs

The silhoutte coefficient computed using the above model on a JJ song raga component (i.e. the song part rendered by a specific singer) with respect to the corresponding two raga-specific song clusters can tell us about its "faithfulness" to its own raga characteristics in the JJ context. We obtained positive valued silhoutte coefficients across the set of JJ song components for all the representations in Table 3, but report here only the values computed with the unfolded SN-d and Bhattacharyya distance measure, viz. 0.75 (A-K) and 0.77 (C-M).

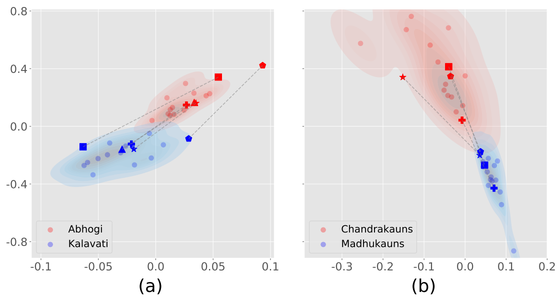

Figure 3. MDS scatter plots for A-K and C-M raga pairs with unfolded SN-d representations and Bhattacharyya distance. All raga-specific songs appear as uniform filled circles with the corresponding kernel density estimates superposed. The JJ songs are indicated with distinct symbols of shared shape but different raga-specific colours for the two raga components of the same song, also connected with a dashed line.

Further, a more visual rendition of the similarities between JJ song components and their own class of raga-specific songs is possible with multi-dimensional scaling (MDS) [27, 28]. The chosen representation (unfolded SN-d) and distance measure (Bhattacharyya) are used to obtain all inter-song distances in the corresponding multi-dimensional space of the representation which is then projected to a two-dimensional space where the similarity between items is preserved in the visual distances using MDS. Figure 3 presents the MDS plots for each raga pair separately. The different colours in each plot separate the two ragas of the corresponding pair. The JJ song components are depicted with special symbols that distinguish them from the raga-specific songs but also serve to identify the same-song components. We note that the raga-specific songs as well as the JJ component songs of the A-K are well separated in the 2 dimensional space. The two raga components of the same JJ song tend to be at least as well separated as the most closely spaced raga-specific songs drawn from opposite ragas. The similar observations hold for the C-M dataset indicating the preservation of the distributional characteristics of the raga in the JJ songs.

4.2 Melodic phrases

Our goal is to study the interaction between the two JJ singers in terms of the nature of 'call' and 'response' phrases. We isolate the individual singing turns from the JJ song pitch time-series and create pairs out of consecutive turns with one each from the female and male singers. The pairing is achieved from the manual annotation of the singers' parts as described in Section 2 and starts from the first singing turn in the song. The interaction, if any, is

Raga Pair	Metric	Folded			Unfolded		
		CP (48)	SN-d (5)	SN-c (5)	CP (128)	SN-d (14)	SN-c (14)
A-K	Correlation	0.47	0.54	0.35	0.55	0.55	0.61
	Bhattacharyya	0.36	0.53	0.31	0.82	0.78	0.79
C-M	Correlation	0.16	0.19	0.03	0.71	0.68	0.68
	Bhattacharyya	0.11	0.18	0.08	0.87	0.84	0.83

Table 3. Average silhoutte coefficient for raga-specific songs in each raga pair for different representations and distance measures. (CP: continuous pitch; SN-d and SN-c: stable notes weighted with duration and count respectively. The corresponding vector dimension appears in parentheses.)

expected to be more evident with shorter duration turns. While any JJ song has singer turns with a wide range of durations, short turns (10 s or lower) are relatively more frequent in the faster tempo (*drut laya*) songs. We therefore restrict this part of the study to the drut songs in our dataset, i.e. 4 A-K songs (83 turn pairs in total) and 3 C-M songs (80 turn pairs in total), omitting altogether only one JJ song from each of the raga pairs in our dataset. Figure 2 shows the variation of the duration of a turn with start time of the turn in a drut song. The presence of rapid pitch modulations (taans) in the drut songs leads to fewer detected stable notes. We therefore relax the duration threshold used in their extraction from the continuous pitch contour to 150 ms.

We begin by examining the local pitch distributions at the turn level. This allows us to compare the tonal content across the two turns in a call and response pair. The two singing turns in a pair are assigned the same ‘index’. Figure 4 presents 2 JJ song examples, one from each of our raga pairs. We see the time-varying distribution of the stable notes in terms of their durations in a given turn. An immediate observation is the fairly consistent vertical displacement in the range of notes covered in going from one raga turn to the same-index turn in the complementary raga. Further, it appears that the shape of the distribution of notes per turn, as it changes with turn index, is roughly matched across the 2 singers with a visually somewhat closer match in the case of the C-M song.

4.2.1 *Melodic shape features*

In order to obtain a more quantitative analysis comparing call and response across turn pairs, we define a few meaningful scalar features that can be reliably computed from the melodic pitch contours. The previously paired turns are assumed to comprise a call and its response. We compute the linear correlation between the corresponding features of the two turns in a pair across all the pairs in the A-K drut songs; similarly, the C-M drut songs. The selected features are the turn duration (in seconds), the number of notes in the SN sequence representation of the turn and the pitch range spanned by the turn as computed from the corresponding continuous pitch contour. The estimated correlations obtained for each raga pair appear in Table 4 as contrasted with the corresponding correlations between randomly paired turns averaged over 50 shuffles of the set. We note moderately high positive correlations for all the features, suggestive of the similarity between the melodic contours of the paired turns.

Raga pair	No. of turn pairs	Duration (s)	No. of notes	Pitch range (cents)
A-K	83	0.52 (0.02)	0.61 (0.02)	0.47 (0.04)
C-M	80	0.81 (0.02)	0.64 (0.03)	0.55 (0)

Table 4. Correlations between matched-index turns from the 2 ragas for turn duration, number of notes and pitch range, averaged over the turns in each raga pair. Values in parentheses are the corresponding correlations across randomly paired turns, serving as a baseline.

4.2.2 *Transposition interval*

With the turn-level pitch distributions of Figure 4 suggestive of a fixed pitch interval shift between turns in a call and response pair, we attempt to estimate the transposition interval. In order to determine a suitable computational model for this, we manually examined the continuous pitch contours of a few call and response pairs where the singers used the raga solfege as lyrics, providing us with a ready transcription of the phrase in terms of the raga notes.

Figure 5 shows instances of two distinct kinds of call and response interactions observed in a C-M JJ song. We have (a), where the turns correspond to the identical keyboard pitch classes, i.e. the singers are actually in unison (or one octave apart). In such a case, the solfege transcriptions are quite different across the phrases, as expected from different ragas and different assumed tonics. While the notes uttered are valid for each raga given its assumed tonic, the realised phrase is not necessarily a raga characteristic phrase. The more common pattern, however, is (b), where the turns have similar solfege notation, i.e. the singers utter (mimic) the same solfege notes (svara) as far as possible. When the svara is not available to the responding singer, they draw from the closest available svara of their raga (verifiable from the raga grammar of Table 1). The case (b) is expected to correspond to an exact tranposition by a fifth (700 cents), a desirable state of harmony (*samvaad*), in the case of svara common to the two ragas but only approximately so otherwise. We also occasionally observe geometric transformations such as melodic contour inversion or reversal across call and response.

Given the known challenges in transcribing the continuous melodic contours to solfege notation (evident also in Figure 5), we use the simplified, even if crude, measure of the mean pitch (in Hz) of a turn to estimate the pitch offset between the corresponding phrases. Figure 6 presents a histogram computed for each raga-pair showing the distri-

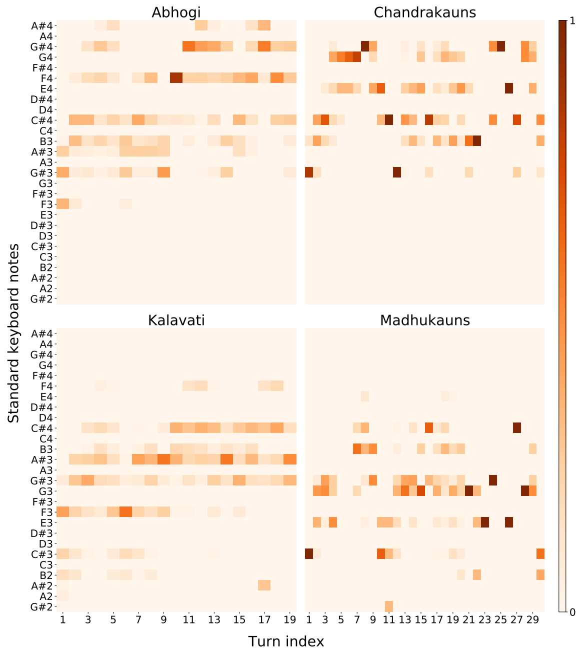

Figure 4. Distribution of stable notes in a turn versus turn index, separately (top and below) for each singer, in an A-K JJ song (left) and a C-M JJ song (right). The song notes are provided in terms of standard keyboard notes with normalizing F0 selected so that the assumed tonic of the female component corresponds to G#3, and therefore that of the male voice to C#3.

bution of the estimated intervals between the mean pitches of turns in a call and response pair. We note that both raga pairs show distributions concentrated around the interval of a fifth. While the C-M distribution is centered around 700 cents, A-K is more skewed to lower intervals. To explain the relatively high A-K histogram peak at (a relatively dissonant) 600 cents, we examined turn pairs that exhibited this specific mean pitch difference. A prominent example of this turned out to be phrases that matched 'g' in Abhogi with 'G' in Kalavati. Another case was the upward going 'D-Ṡ' transition in Abhogi being matched with a 'D-n-Ṡ' transition in Kalavati. In any case, given that we do not have precise within turn alignments of the svara rendered in call and response, we can only infer that the phrases in a pair are largely centered at matching pitch intervals relative to their respective assumed tonics. With the two tonics offset by a fifth, we see the transposition by 700 cents. The occurrence of mean pitch differences of values below and above 700 cents in Figure 6 may be explained by our observation that the automatic pairing of call and response turns is occasionally incorrect with singers switching roles at times.

5. CONCLUSION

We have presented here what is probably the first empirical analysis of mode-shifting (murchana) in Hindustani raga performance using MIR tools. The specific form considered was the Jasrangi jugalbandi where two artists collaborate, each singing a different raga with the same set of pitches shared across the two. Computed song-level pitch distribution similarity revealed that the JJ singers adhere to raga characteristic tonal hierarchy to the same extent as in the corresponding individually performed ragas. Simple melodic features of the singing turns during the artists' interaction revealed interesting insights about the melodic shape transformations and transposition interval in the course of the call and response. Overcoming the limitations posed by data scarcity would facilitate the investigation of more complex models of melodic similarity that also take note sequences into account.

Supplementary material available at this link.

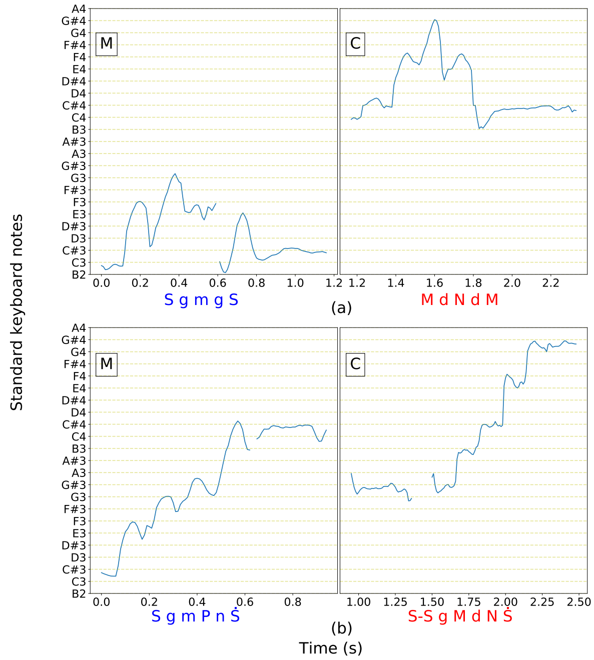

Figure 5. Two examples of call and response patterns from a CM JJ song. (a) replicating the keyboard notes of the call phrase, but an octave apart; (b) approximating the solfege of the call phrase relative to own assumed tonic.

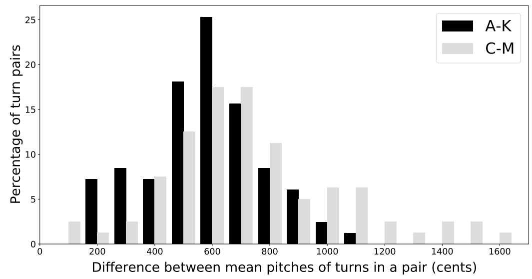

Figure 6. Histograms of the differences between the mean pitches between turns in a pair, computed for all turn pairs corresponding to A-K (83 turn pairs) and C-M (80 turn pairs) drut songs.

6. ACKNOWLEDGEMENTS

We thank Ankita Joshi, disciple of Pandit Jasraj, for her guidance and useful discussions.

7. REFERENCES

[1] G. K. Koduri, S. Gulati, P. Rao, and X. Serra, "Rāga recognition based on pitch distribution methods," *Journal of New Music Research*, vol. 41 (4), pp. 337–350, 2012.

[2] P. Chordia and A. Rae, "Raag recognition using pitch-class and pitch-class dyad distributions," in *Proc. of the 8th Int. Soc. for Music Information Retrieval Conference*, Vienna, Austria, 2007.

[3] K. K. Ganguli and P. Rao, "On the distributional representation of ragas: Experiments with allied raga pairs," *Transactions of the Int. Soc. for Music Information Retrieval*, vol. 1 (1), pp. 79–95, 2018.

[4] S. Bagchee, *Nad: Understanding Raga Music*. Business Publications Inc., 1998.

[5] H. Hirlekar, *Nuances of Hindustani Classical Music*. Unicorn Books, 2010.

[6] M. Desai, "Jasrangi jugalbandi: A wondrous show of musical acumen," https://creativeyatra.com/reviews/jasrangi-jugalbandi-a-wondrous-show-of-musical-acumen/, 2018, accessed: May-2021.

[7] X. Serra, "A multicultural approach in music information research," in *Proc. of the 12th Int. Soc. for Music Information Retrieval Conference*, Miami, USA, 2011.

[8] S. Gulati, J. Serrà, V. Ishwar, S. Şentürk, and X. Serra, "Phrase-based rāga recognition using vector space modeling," in *Proc. of the 41st IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Shanghai, China, 2016.

[9] S. Gulati, J. Serrà, K. K. Ganguli, S. Şentürk, and X. Serra, "Time-delayed melody surfaces for rāga recognition," in *Proc. of the 17th Int. Soc. for Music Information Retrieval Conference*, New York City, USA, 2016.

[10] R. Hennequin, A. Khlif, F. Voituret, and M. Moussallam, "Spleeter: a fast and efficient music source separation tool with pre-trained models," *Journal of Open Source Software*, vol. 5, p. 2154, 2020.

[11] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing parselmouth: A python interface to praat," *Journal of Phonetics*, vol. 71, pp. 1–15, 2018.

[12] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]," http://www.praat.org/, 2021, version 6.1.38, retrieved May 2021.

[13] J. Salamon, S. Gulati, and X. Serra, "A multipitch approach to tonic identification in indian classical music," in *Proc. of the 13th Int. Soc. for Music Information Retrieval Conference*, Porto, Portugal, 2012.

[14] D. Bogdanov, N. Wack, E. Gómez, S. Gulati, P. Herrera, O. Mayor *et al.*, "Essentia: an audio analysis library for music information retrieval," in *Proc. of the 14th Int. Soc. for Music Information Retrieval Conference*, Curitiba, Brazil, 2013.

[15] D. Huron, *Sweet Anticipation: Music and the Psychology of Expectation*. MIT Press, 2006.

[16] C. L. Krumhansl and L. L. Cuddy, *Music Perception*. New York: Springer, 2010, ch. A Theory of Tonal Hierarchies in Music, pp. 51–87.

[17] N. A. Smith and M. A. Schmuckler, "Pitchdistributional effects on the perception of tonality," in *Proc. of the 6th Int. Conf. on Music Perception and Cognition*, Staffordshire, England, 2000.

[18] R. Raman and W. J. Dowling, "Real-time probing of modulations in south indian classical (carnatic) music by indian and western musicians," *Music Perception: An Interdisciplinary Journal*, vol. 33 (3), pp. 367–393, 2016.

[19] D. Temperley, "What's key for key? the krumhansl-schmuckler key-finding algorithm reconsidered," *Music Perception*, vol. 17 (1), pp. 65–100, 1999.

[20] K. K. Ganguli and P. Rao, "Towards computational modeling of the ungrammatical in a raga performance," in *Proc. of the 17th Int. Soc. for Music Information Retrieval Conference*, Suzhou, China, 2017.

[21] B. Cornelissen, W. Zuidema, and J. A. Burgoyne, "Mode classification and natural units in plainchant," in *Proc. of the 21st Int. Soc. for Music Information Retrieval Conference*, Monrèal, Canada, 2020.

[22] S. Sertan and P. Chordia, "Modeling melodic improvisation in turkish folk music using variable-length markov models," in *Proc. of the 12th Int. Soc. for Music Information Retrieval Conference*, Miami, USA, 2011.

[23] S. Dutta, K. S. PV, and H. A. Murthy, "Raga verification in carnatic music using longest common segment set," in *Proc. of the 16th Int. Soc. for Music Information Retrieval Conference*, Malaga, Spain, 2015.

[24] S. T. Madhusudhan and G. Chowdhary, "Deepsrgm - sequence classification and ranking in indian classical music with deep learning," in *Proc. of the 20th Int. Soc. for Music Information Retrieval Conference*, Delft, The Netherlands, 2019.

[25] N. Kroher, A. Pikrakis, and J.-M. Díaz-Báñez, "Discovery of repeated melodic phrases in folk singing recordings," *IEEE Transactions on Multimedia*, vol. 20 (6), pp. 1291–1304, 2018.

[26] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[28] I. Borg and G. P., *Modern Multidimensional Scaling - Theory and Applications*. Springer, 1997.