

MELODIC AND METRICAL ELEMENTS OF EXPRESSIVENESS IN HINDUSTANI VOCAL MUSIC

Yash Bhake Ankit Anand Preeti Rao

Department of Electrical Engineering, Indian Institute of Technology Bombay, India

yash.bhake@iitb.ac.in, ankit0.anand@gmail.com, prao@ee.iitb.ac.in

ABSTRACT

This paper presents an attempt to study the aesthetics of North Indian *Khayal* music with reference to the flexibility exercised by artists in performing popular compositions. We study expressive timing and pitch variations of the given lyrical content within and across performances and propose computational representations that can discriminate between different performances of the same song in terms of expression. We present the necessary audio processing and annotation procedures, and discuss our observations and insights from the analysis of a dataset of two songs in two *ragas* each rendered by ten prominent artists.

1. INTRODUCTION

A *Khayal* performance consists of a composition (the chosen *bandish* or song with lyrics) that also forms the base for improvisation [1]. That is, while the first occurrence of a *bandish* line is typically rendered in its canonical form from memory, the accomplished artist renders successive repetitions of the line with pitch and timing variations in specific ways to create an aesthetic experience appreciated by listeners trained in the genre. A motivation for this work is that suitable computational representations can help obtain insights or verify hypotheses regarding this performance practice. They can also inform tools to generate realistic audio for artistic and educational contexts [2].

Performance expression, according to Juslin [3], concerns "the small and large variations in timing, dynamics, timbre, and pitch that form the microstructure of a performance and differentiate it from another performance of the same music". While the structure of the songs themselves carries specific emotion – this certainly holds for *bandish* given the *raga* that underlies the melody, and the semantics of the lyrics, - expressive variations contribute to the emotional impact of the performance. We present a review of some of the work of musicologists who have studied *Khayal* music from the viewpoint of composition and improvisation.

In a scholarly work based on several vocal performances of a single *bandish* by different artists, Morris [4] used his transcriptions of the recorded music (well aware of the ambiguities inherent in the manual process that involves discrete symbols) to compare an artist's rendition with their self-declared notation for the composition. He noted that some portions of *bandish* were reproduced identically, others were more flexibly produced, and proceeded to investigate the extent and nature of the differences. Another example of such work is the analysis of a single live performance by van der Meer [5], focusing on *raga*-related vocal pitch inflections in the *raga* improvisation section (*alap*) that triggered emotion as detected from the audience's spontaneous audible reactions. He relates these expressive moments to the occurrence/execution of specific ornaments. Finally, a close parallel of our problem is that of tracking *sangatis*, which are lineage-dependent variations of the same lyric lines in Carnatic compositions [6].

More recently, the similar questions using slightly larger, curated data sets have been facilitated by the availability of computational methods. Previous computational studies have examined the variability of *raga* phrases in the course of *Khayal* improvisation with respect to underlying tempo and metrical location [7]. The context of singing *bandish* lines, on the other hand, entails additional constraints given the special importance accorded to the lyrics in the context of *bandish*. Recently, a small comparative study of two performances of the same *bandish*, with reference to temporal deviations of note events from the canonical metrical positions [8], served as a preliminary validation for methodology using automatically detected onsets for the audio-level and text-level alignment of manual and reference syllable-level transcription. Their visual representation of the distribution, across each performance audio, of the timing offset corresponding to a specific metrical position showed the musicologically anticipated behaviour, viz. reduced deviation closer to the *sam* or downbeat of the rhythm cycle. It also helped visualise the difference between the two performances in terms of the use of timing expressiveness. In contrast, the present work (i) uses a significantly larger dataset for a timing expressiveness study while also adding new attributes; (ii) includes a new study of pitch-based expressiveness; (iii) proposes computational measures for artist-level expression on each of the studied dimensions and validates their potential for discriminating performances based on expressiveness.

In the next section we provide some of the music back-



ground needed to appreciate the specific questions we set up for our work and our overall approach. Following this, we present our dataset, audio and text processing methods and the devised computational measures. Finally, we discuss our observations and attempts to draw insights.

2. BACKGROUND

Bandish are lyrical compositions that have been handed down across generations by oral transmission and serve as a sort of dictionary for the associated *raga* in terms of the overall pitch movement and characteristic phrases [9]. The two verses of the *bandish*, termed *sthai* and *antara*, typically sung early on in any concert, are defined by their lyrics and tune. In the early 20th century, music scholar Pt. V.N. Bhatkhande launched on a mission to collect and notate traditional compositions from across the country. He devised his own ways of notating the melodies that so far lay within the framework of oral transmission and largely within hereditary musician families. The notation, considered a "schematic" form, requires an understanding of the intricacies of the *raga* and *tala* system in order to interpret in performance. Now widely respected, his monumental work helped preserve the traditional compositions for posterity [10]. Although commonly featured in learning contexts, it is held that the notation corresponding to a given performed *bandish* is dependent to an extent on the lineage (*gharana*) of the artist. Text differences exist as well, but these are mainly due to spelling and dialects or verb forms, while maintaining the recognizability of the lyrics.

Bhatkhande's book [10] typically provides a single notation sequence for every unique line of the song with a *raga swar* (note), or a short sequence of *swar*, assigned to every syllable of the lyrics. Figure 1 shows the canonical notation of a *bandish*, as available

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x				2				o				3			
						^m P	-	^m g	-	R	S	R	ⁿ n	S	-
						Jaal	-	Jaal2	-	Re	-	A	Pa	Ne	-
^s S	-	^s m	m	m	-	ⁿ n	P	^m g	R	S	R	n	S	-	
Man	-	di1	Ra	Waa	-	Jaal	-	Jaal2	-	Re	-	A	Pa	Ne	-
^s S	-	^s m	m	m	-	ⁿ n	P	^m g	R	S	R	n	S	-	
Man	-	di1	Ra	Waa	-	Sul	Na1	Paa	-	We	-	Gii	-	Saa	-
R'	n	S'	P	g	-	P	-								
Sa	Na2	Na3	di2	Yaa1	-	Jaal	-								
								P	P	P	m	P	-	^m g	m
								Su2	Na4	Ho	Sa	daa	-	Ran	Ga1
P	P	^s n	^s n	S'	S'	S'	-	^s n	^s n	S'	g'	R'	R'	S'	-
tu1	Ma1	Ko1	Chaa	Ha1	ta	Hain	-	Kyaa	-	tu2	Ma2	Ha2	Ma3	Ko2	-
^s n	n	S'	S'	P	m	n	P								
Chha	Ga2	Na5	di3	Yaa2	-	Jaal	-								

Figure 1. The composition *Ja Ja Re* from Bhatkhande [10] as a machine-readable CSV file for the two verses of 2 lines each.

in the book and the created machine-readable format, retaining the note label (both pitch and lyric syllable) and timing information as indicated by the labels of the 16-beat *tala* cycle. In this study, all *bandish* are set in *teentaal* (16-beat cycle), with salient *matra* (beats) like the *sam* (down-beat, 'x') and *khali* (9th beat, 'o'), coinciding with the start of the second half cycle) explicitly marked. The top row indicates beat number, the row below it indicates *vibhag* symbols, marking the start of each quarter cycle. Each line

of lyrics of the *bandish* spans 16 beats, with each beat containing either a syllable (or, seldom, multiple syllables), a rest (empty string), or a continuation of the previous note ('s'). Each cycle is represented using two aligned rows: (1) Notation row (black), which displays the main note (*sargam*) with any ornamentation indicated as a superscript preceding the main note (such as a grace note or a glide); and (2) Lyrics row (red), which contains the corresponding *bandish* lyric syllables aligned to the beat positions and associated notes.

In performance, singers typically repeat each line (and sometimes its component phrases) several times with variations introduced after the first utterance is rendered in the canonical form. The successive repetitions are characterised by a certain fluidity in the notes and timing of where a given syllable falls [4]. These variations are normally extempore in concert settings but could show similarities across the artists from the same musical lineage (*gharana*). Viewing these as expressive variations of the *bandish* line, as specified by its lyrics, we investigate the range and nature of the expressive gestures as a function of the specific word (via its component syllables) across its multiple occurrences in the singing. That is, we hope to obtain insights about the structural moments of the song where the expressive gestures are added and precisely which acoustic parameters are varied to realise this.

3. DATASET AND PROCESSING

Our dataset¹ is designed to serve the objectives of our study, namely, to observe the variations across repeated utterances of a given *bandish* line across performances of the same and different artists. We consider two specific popular compositions, each with multiple performances by several prominent artists in the genre, obtained from across commercial and free internet sources. The sung *bandish* lines are manually segmented from full concerts (which also typically include several other sections, including free improvisation in the chosen *raga*). The details of the dataset used in this study appear in Table 1. An immediate observation is the skewed distribution across the 4 *bandish* lines with Line 1 being sung far more frequently than any other. This is common in the concert setting, given that the *bandish* line 1 is known as the *mukhda* or refrain of the song, analogous to the *pallavi* in Carnatic style of Indian classical music.

We carry out vocals separation on the concert audio segments to eliminate the accompanying instruments. We use a pretrained model (OpenAI/Whisper) for speech to text conversion with a prompt comprising the words of the song lyrics in Devanagiri script, together with expected pronunciation variations. The obtained word sequence is then aligned at the phone level to the audio using forced alignment with a Kaldi TDNN Hindi speech trained acoustic model [11]. Next, the sequence of phones is segmented into the syllables of the lyrics words. Each resulting syllable's onset instant is then identified as the frame corre-

¹ More details appear in the supplementary material

sponding to the consonant-vowel (CV) transition.

The resulting alignments are checked and manually corrected for the occasional errors that arise mainly due to the presence of long vowels in singing, poor enunciation at times, as well as the occurrence of significant pitch inflections. We observe the manually corrected onset locations—i.e. the syllable onsets as realised by the artist—marked at the top of the waveform in the example of one full *tala* cycle of the *bandish* in Figure 2. To obtain the beat locations, we annotate the tabla stroke onsets using the source-separated accompaniment file; we manually mark the salient beats (downbeat ‘x’ *sam* and 9th beat ‘o’ *khali*) in 1 cycle. Assuming a consistent local tempo, we divide each half cycle (interval between one downbeat and the adjacent tth beat into 8 equal parts, thereby obtaining all the estimated beat instants across the rendition.

Next, we mark the canonical locations of the matching syllables, positioning the syllables according to the Bhatkhande notation (as in Figure 1). The position mapping of the realised syllable onsets to the corresponding canonical syllable was implemented following the method proposed in previous work [8]. We observe from Figure 2 how the realised onsets lag the canonical locations most of the time.

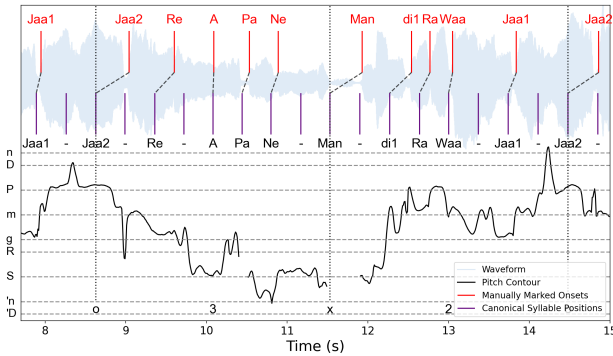


Figure 2. Pitch contour (bottom) and sung syllable alignment (red) with canonical beat positions of the same syllable (black) for an excerpt of *Ja Ja Re* by ABD.

Finally the vocal pitch is extracted at 10 ms intervals using an autocorrelation based method for fundamental frequency and voicing [12]. Brief pauses and unvoiced regions are linearly interpolated to obtain a continuous pitch contour for each sung syllable region. The pitch contour is converted to cents by normalisation with the known performance’s tonic. Eventually, we obtain for each performance in our dataset, the segmented audio of each sung line annotated at the syllable level with syllable name, boundaries and the pitch (cents) at 10 ms intervals. We use these low-level features to define quantities that capture the singing variations across repetitions of a *bandish* line within and across singers. The reference for the comparison is the syllable identity (i.e. its name and metrical location) as defined in the canonical notation as presented in Figure 1.

Raga	Bhimpalasi	Yaman
Bandish	Ja Ja Re	Yeri Aali
Tala	Teentaal	Teentaal
Swar	S, R, g, m, P, D, n, S	S, R, G, M, P, D, N, S
# Concerts	15	13
# Artists	15	12
# Repetitions (L1, L2, L3, L4)	167, 39, 47, 47	94, 32, 35, 23
Matra per min range	138–200	111–203

Table 1. Summary of our dataset of concert recordings across *ragas*, *bandish*, and artists. *Swar* notation details are in the supplementary.

4. MEASURING EXPRESSIVENESS

We wish to quantify and compare the variability observed in the acoustic realisation of a given syllable, from a specific line of the *bandish*, across (i) repeated utterances within an artist’s performance, and (ii) utterances of the same syllable across different performances/artists. The acoustic parameters that we detect are: (i) the onset time of the syllable, (ii) the syllable duration (as the time interval between the current syllable’s onset and either the onset of the next syllable or the start of the following silence segment, whichever occurs first. and (iii) the pitch contour shape across the syllable interval. We illustrate the process by providing examples of the processing and analyses of the audio rendering of a chosen line by one artist.

4.1 Timing expression

The deviation of the detected onset from its reference assigned beat index in the canonical notation gives us an estimate of the lag/lead of the singer for the syllable in question. We represent the deviation in terms of fraction of the local beat interval; this normalization facilitates the comparison across instances and concerts. We can view the thus measured timing offsets as evidence of expressive timing, especially if this quantity shows variability across repetitions of the syllable within the concert.

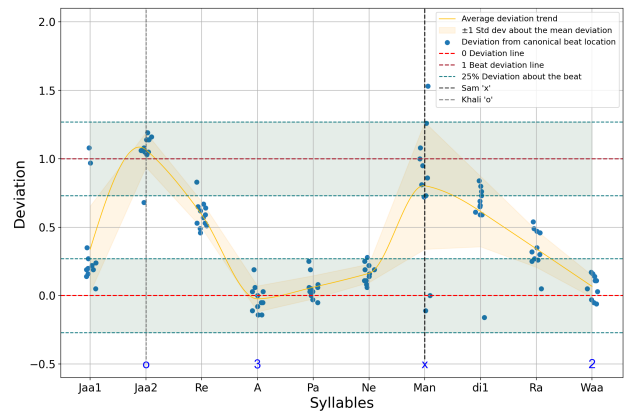


Figure 3. Deviation of the sung syllable onsets from the canonical locations measured in the units of beat duration for *Ja Ja Re* Line 1 by ABD for multiple repetitions of the line.

Figure 3 captures the onsets of syllables in *bandish* 1, Line 1 as rendered by singer ABD. The syllable names

are shown with their canonical *matra* locations at the bottom. We note that some syllables occupy 2 beats and others 1 beat in the canonical form. We observe, for example, that "Jaa2" and "Man" (both 2-beat syllables) show variations with a mean lag of about one beat. "Man" instances however are much more dispersed. While a uniform offset could potentially indicate a structural difference between the artist's version of the *bandish* and that of the Bhatkhande book, a high standard deviation (like in "Man") points to the artist's in-the-moment expressive variations. On the other hand, the 3 syllables preceding "Man" show near-zero offsets across repetitions.

4.2 Pitch expression

Analysing the rendered song for pitch-based expressiveness is a rather involved task. There are many ways in which the artist injects expressiveness in their performance via pitch variation. As we want to analyse the different ways of realising the same line of a *bandish*, we are interested in the variability in the pitch contour (PC) shape of a syllable across the multiple repetitions of that line. A greater diversity of PC shapes can then be interpreted as higher expressiveness in an artist's repetitions of the line.

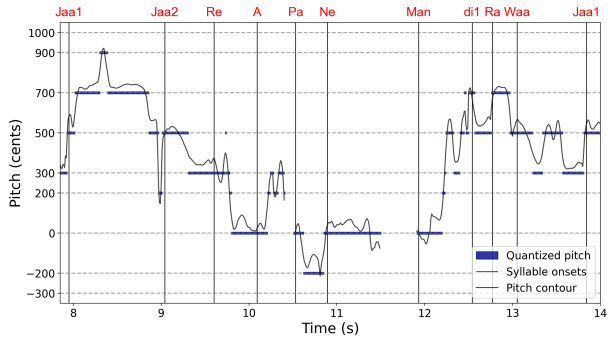


Figure 4. PC quantisation to the nearest *raga* note for one instance of *Ja Ja Re* *bandish* line 1 rendered by artist ABD

For each syllable, the associated PC spans the duration of that syllable; hence, this is not a fixed-length time series. We represent this variable dimension PC for a given syllable with a lower and fixed-dimensional vector. We first implement a piece-wise aggregate approximation (PAA) over the syllable PCs. PAA is a time-series representation that has been used widely in data-mining tasks [13]. This is implemented as follows.

The syllable PC values are each first quantised to the nearest *raga swar* (note), Figure 4 shows this process. Next, the quantised PC values for a syllable across repetitions, is aggregated by dividing the quantised PC into a fixed number of uniform intervals and assigning the mode of the values to each interval. The number of fixed intervals is set empirically to 10 intervals per beats allotted to the syllable (treating the syllable extensions indicated by '-' in Figure 1 to be a part of the previous syllable, thereby adding to its allotted beats). The choice of 10 equal segments per beat interval is based on the tempo range of our dataset (110-200 BPM or 300 ms to 545 ms

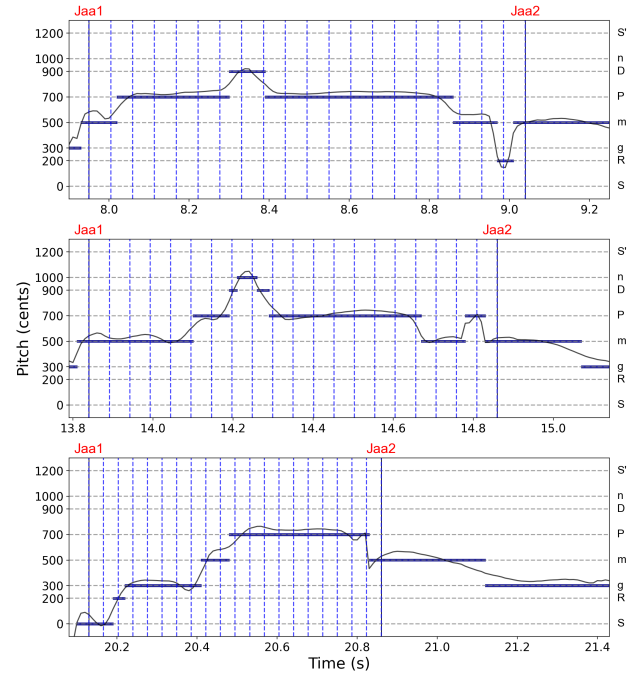


Figure 5. Three distinct renditions of the syllable "Jaa1" by artist ABD, each represented by a fixed number of uniform time intervals. Each interval is mapped based on its modal pitch to the nearest *raga* note, giving us the PAA string representation for the syllable's pitch shape. Figure 6 describes this process and the PAA strings for the above PCs.

per beat) and sampling period (10 ms/sample) of the pitch contour. This results in segments, each represented by a short sequence of samples of the pitch contour. This balance allows capturing dynamic pitch fluctuations which are prevalent in Hindustani classical music (HCM), without over-quantisation. Now, each PAA interval within a syllable is assigned a discrete symbol, where the symbols are drawn from a suitable alphabet which comprises notes (*swar*) across the relevant octave ranges. The resulting string of note values (one per PAA interval) then represents coarsely the realised pitch shape of the syllable. Figure 6 shows this process. The 3 string sequences in Figure 6 correspond to the 3 PCs in Figure 5. This type of aggregation presents a tradeoff of generality vs specificity.

Like in the case of syllable timing, we are interested in the variation, if any, in pitch shape of a given syllable across repetitions. We achieve this by computing the similarity of the syllable PCs for pairs drawn from the set of repetitions in a single concert. The Levenshtein edit distance [14] between the PAA strings provides us with the number of note substitutions. We evaluate the Normalised Levenshtein Substitution Score (NLSS) for each pair as a measure of the dissimilarity. Figure 7 shows a matrix representation (heat map) of NLSS values for a chosen syllable as rendered by one artist across 14 repetitions.

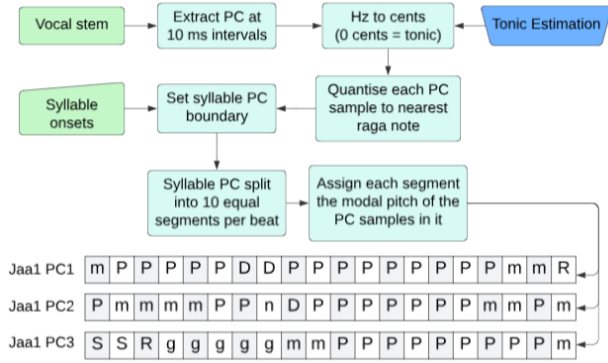


Figure 6. Processing pipeline for generating PAA string representations for the pitch contour for different repetitions of a syllable. The strings correspond to the PCs of the syllable repetitions given in Figure 5 spanning 2 beats.

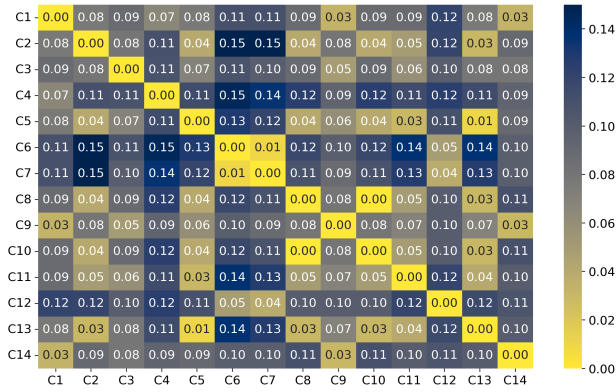


Figure 7. Heat map showing NLSS for each pair of "Jaa1" syllable PCs drawn from the set of repetitions of line 1 of *bandish Ja Ja Re* rendered by artist ABD

5. OBSERVATIONS AND DISCUSSION

We are interested in the within-artist variation for a given *bandish* line and its syllables. This can facilitate potentially valuable insights about (i) the preferred locations (in terms of chosen syllables) for expressive gestures by individual artists and across artists and (ii) the extent and nature of expressive gestures for a given artist. Due to space limitations, we present the analysis results for line 1 of one *bandish*, with the other *bandish* presented in the supplementary.

Figure 8 presents for each artist and syllable, the standard deviation (s.d.) of the timing deviation (as captured in the example for artist ABD in Figure 3). The s.d. helps us focus on the *variability* of offsets rather than on actual offset values (which might be attributed to structural differences between the artist's version and Bhatkhande's version of the *bandish*, rather than expression-related).

In Figure 8, we note the dominance of the first 3 syllables for most artists. The full range of behaviours, however, includes IN at one end with minimal variations to DG and RK, who introduce new variations on practically all syllables. That IN does not exercise any flexibility is not

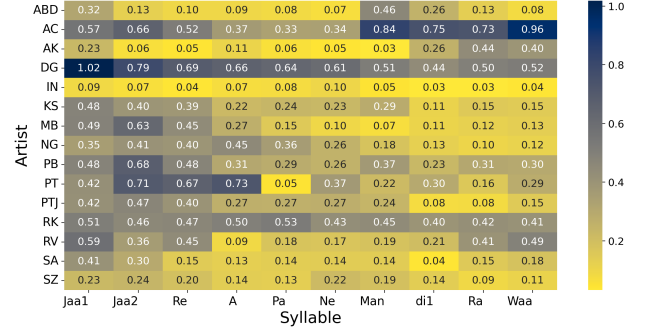


Figure 8. Standard deviation of the distribution of the fractional timing deviation for every syllable over multiple repetitions in one rendition, across artists.

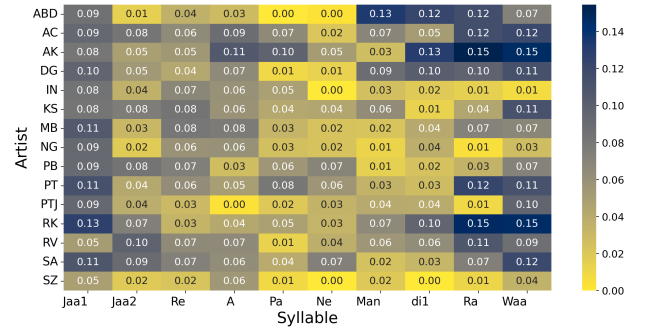


Figure 9. Mean NLSS over all pairs of repetitions of each syllable in line 1 of *bandish Ja Ja Re*, computed per artist. serving as a dissimilarity measure across different repetitions of a syllable by an artist.

surprising given that his performances were explicitly created to closely follow the prescribed Bhatkhande notation, as discussed here [15]. RK, on the other hand, is considered a virtuoso musician.

Aggregating across the rows, we obtain the per syllable behaviour across artists in Figure 10. The mean values show that the first 3 syllables carry the most expressive timing, with "Jaa2" also showing the most spread across artists (consistent with Figure 8). We see, for example, that PT shows a large range in per syllable SD, again agreeing with Figure 8. In the case of the artist ABD, we can observe that the temporal deviation is spread relatively evenly, while peaking for a particular syllable "Man", which is the down-beat. This can be easily appreciated in listening to the audio, which can be accessed in the supplementary material.

To assess pitch variability, we calculate the average number of pitch substitutions by taking the mean of the NLSS for all pairs of repetitions per artist and per syllable, (normalised by the string length), calculated across all possible pairs of the given syllable utterances within a concert. Figure 9 shows the values per artist and per syllable of Line 1. We can observe in Figure 8 and Figure 9 that both the pitch and temporal variation across repetitions and across different artists is more prominent at the beginning of the line. Near the end of the line, the pitch variation increases, which can be attributed to the emphasis on a semantically

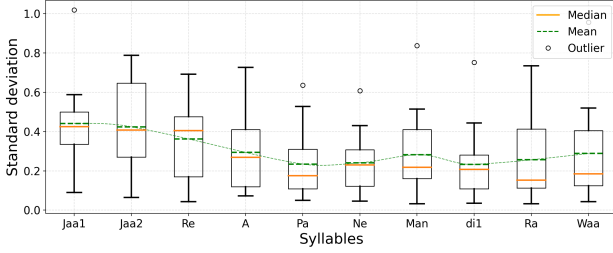


Figure 10. Box-plot of s.d. of timing deviation parameter per syllable of *Ja Ja Re* Line 1 as aggregated across concerts and artists. The *tala*-cycle ends at the syllable "Ne", and a new cycle starts from downbeat at the syllable "Man".

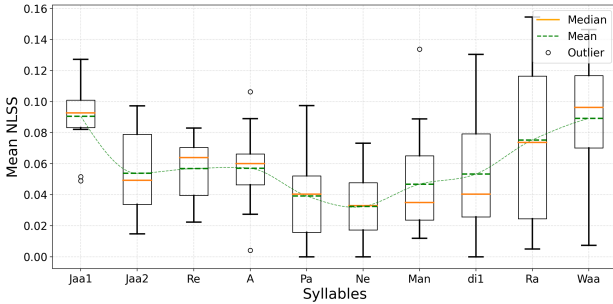


Figure 11. Box-plot of the mean of the averaged NLSS per syllable of *Ja Ja Re* Line 1 as aggregated across concerts and artists.

important word- "Mandirava".

The syllables "Pa" and "Ne" are near the *tala*-cycle boundary, and we observe a minimum pitch and temporal variation; this is consistent with the observations of [8]. As the performer heads towards the *tala*-cycle boundary, their overall tendency of variation and improvisation decreases, aiming rather towards resolving the melodic and temporal expression they have come up with in the particular *tala*-cycle.

Aggregating across rows of Figure 9, we obtain the per syllable behaviour over all artists in Figure 11. The dotted line joining the means indicates that the pitch variation decreases as one approaches the *tala*-cycle boundary at the syllable "Ne", while the pitch expressiveness is higher in the start and end of the line, which falls on the 7th and 5th beats of the *tala*-cycle respectively, which is far from the cycle boundary, hence providing more scope for timing and pitch expressiveness. We see, for example, that AC and AK show high pitch variation across the repetitions of the same line.

Hierarchical clustering of all the pairs from the set of repetitions of a syllable by an artist provides us with information about the variation clusters. A threshold can be defined that decides if a variation belongs to a cluster or not. More diverse variations would indicate more number of clusters, indicating higher expressiveness. Dendrograms are excellent for visualising such clusters. The realisation of the previous syllable has an influence on which cluster the following syllable variation would belong to.

Comparing Figure 8 and Figure 9, we note that expressive gestures that utilise pitch are not necessarily at the same locations that exhibit timing deviation in terms of the preferred syllable. It is rather interesting to look at the least amount of expressiveness in both pitch and timing lie with the syllables that are near the *tala*-cycle boundary. An analysis at the individual audio level would provide a more accurate picture of the correlations, if any, and is left to future work.

Our observations, reported here on the Line 1 of one *bandish*, largely hold with the second *bandish*. The underlying reasons for the choice of specific syllables exhibiting larger variability are similar to those discussed by Morris [4]. These include larger variations at line or phrase ending syllables due to the effect of previous and next contexts, and the choice of syllables belonging to more emotionally loaded words in the lyrics.

6. CONCLUSION

In this paper, we articulated the problem of modeling expressive variations in the context of performance of Hindustani traditional compositions by established artists of the genre. A well-known *bandish* in the chosen *raga* is always sung at the beginning of a concert with multiple repetitions of the lines, marked by variations in the location and type of the expressive gestures. Based on our proposed methodology, we showed that it is possible to arrive at systematic patterns across artists by treating the syllables of the lyrics as reference points for a study of the range of variation. This also helped us discuss interesting correlations between the roles of melody and rhythm in expressiveness.

We presented a dataset that was annotated with a combination of manual and automatic tools to obtain a rich repository of distinct realizations of the lines of two popular traditional compositions. While much further exploration remains possible, this work demonstrates the potential of computational models for improvisation in the context of compositions in the *Khayal* genre. This work lays the foundation for generative applications by capturing high-level performance features that reflect an artist's distinctive style. A preliminary experiment was performed to generate the temporal deviations discussed above, where the distributions formed by all the deviations of a syllable from its canonical location for an artist were used to sample out new points for each syllable. A sine-tone based audio was synthesized from the generated pitch contour. The reference (Bhatkhande canonical form) and generated tracks are available in the supplementary. This lets us create infinite possibilities for rendering the same line, while at the same time capturing some hint of the artist's style. Extending this approach to other acoustic dimensions for expression such as timbre and dynamics can enable the generation of classical music that embodies the unique identity of individual artists.

7. ACKNOWLEDGMENTS

We take this opportunity to acknowledge and express our gratitude to all those who supported and guided us during this research work. We are thankful to Madhumitha S., whose thesis work was a crucial base for our project. We also thank Mr. Himanshu Sati, and Mrs. Hemala Ranade for their scholarly musicological insights, which gave direction to our work.

8. REFERENCES

- [1] B. Wade, “Music in India: The classical traditions,” Manohar Press, 2001.
- [2] C. E. Cancino-Chacón, M. Grachten, W. Goebel, and G. Widmer, “Computational models of expressive music performance: A comprehensive and critical review,” *Frontiers in Digital Humanities*, vol. 5, p. 25, 2018.
- [3] P. N. Juslin, “Cue utilization in communication of emotion in music performance: Relating performance to perception,” *Journal of Experimental Psychology: Human perception and performance*, vol. 26, no. 6, p. 1797, 2000.
- [4] A. Morris, *Transmission and performance of Khayal compositions in the Gwalior gharana of Indian vocal music*. PhD thesis, Univ. Of London, S.O.A.S., 2004.
- [5] W. Van der Meer, “Audience response and expressive pitch inflections in a live recording of legendary singer kesar bai kerkar,” in *Expressiveness in music performance: Empirical approaches across styles and cultures*, D. Fabian, R. Timmers, and E. Schubert, Eds. Oxford University Press (UK), 2014, pp. 170–184.
- [6] S. Sankaran, P. V. K. Sekhar, and A. M. Hema, “Automatic segmentation of composition in carnatic music using time-frequency cfcc templates,” in *Proceedings of 11th International Symposium on Computer Music Multidisciplinary Research (CMMR)*, 2015.
- [7] K. K. Ganguli and P. Rao, “A study of variability in raga motifs in performance contexts,” *Journal of New Music Research*, vol. 50, no. 1, pp. 102–116, 2021.
- [8] Y. Bhake and P. Rao, “Expressive timing in Hindustani vocal music,” in *Proc. of ICASSP 2025 Workshop on Indian Music Analysis and Generative Applications (WIMAGA)*, Hyderabad, India, 2025, accessed at:link.
- [9] S. Rao and P. Rao, “An overview of Hindustani music in the context of computational musicology,” *Journal of New Music Research*, vol. 43, no. 1, 2014.
- [10] V. Bhatkhande, *Kramik Pustaka Malika*. Sangeet Karyalaya Hathras, India, 2013.
- [11] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, “The kaldi speech recognition toolkit,” in *IEEE 2011 workshop on automatic speech recognition and understanding*. IEEE Signal Processing Society, 2011.
- [12] Y. Jadoul, B. Thompson, and B. De Boer, “Introducing parselmouth: A python interface to praat,” *Journal of Phonetics*, vol. 71, pp. 1–15, 2018.
- [13] J. Lin, E. Keogh, L. Wei, and S. Lonardi, “Experiencing SAX: a novel symbolic representation of time series,” *Data Mining and knowledge discovery*, vol. 31, no. B, pp. 107–144, April 2007.
- [14] W. J. Heeringa, “Measuring dialect pronunciation differences using Levenshtein distance,” 2004.
- [15] “2000 classic compositions from Bhatkhande on cd,” <https://scroll.in/article/726180/>, accessed: 2024-04-10.