Proceedings of

DL*f*M 2024

The 11th International Conference on Digital Libraries for Musicology

27th June 2024

In Association with IAML 2024

Programme chair
David M. Weigl, mdw – University of Music and Performing Arts Vienna

ACM International Conference Proceedings Series, ACM, Inc.

ACM ISBN: 979-8-4007-1720-8

ICPS
Published by ACM

The Association for Computing Machinery
2 Penn Plaza, Suite 701
New York New York 10121-0701

ICPS
Published by ACM

ACM ISBN: 979-8-4007-1720-8

Svara-forms and coarticulation in Carnatic music: an investigation using deep clustering

Thomas Nuttall
Universitat Pompeu Fabra
Barcelona, Spain
thomas.nuttall@upf.edu

Xavier Serra
Universitat Pompeu Fabra
Barcelona, Spain
xavier.serra@upf.edu

Lara Pearson
Max Planck Institute for Empirical Aesthetics
Frankfurt am Main, Germany
lara.pearson@ae.mpg.de

ABSTRACT

Across musical genres worldwide, there are many styles where the shortest conceptual units (e.g., notes) are often performed with ornamentation rather than as static pitches. Carnatic music, a style of art music from South India, is one example. In this style, ornamentation can include slides and wide oscillations that hardly rest on the theoretical pitch implied by the *svara* (note) name. The highly ornamented and oscillatory qualities of the style, in which the same svara may be performed in several different ways, means that transcription from audio to symbolic notation is a challenging task. However, according to the grammar of the Carnatic style, there are a limited number of ways that a svara may be realized in a given *rāga* (melodic framework), and these ways depend to some extent on immediate melodic context and svara duration. Therefore, in theory, it should be possible to identify not only svaras but also the various characteristic ways that any given svara is performed - referred to here as 'svara-forms'.

In this paper we present a dataset of 1,530 manually created svara annotations in a single performance of a composition in rāga Bhairavi, performed by the senior Carnatic vocalist Sanjay Subrahmanyan. We train a recurrent neural network and sequence classification model, DeepGRU, on the extracted pitch time series of the predominant vocal melody corresponding to these annotations to learn an embedding that classifies svara label with 87.6% test accuracy. We demonstrate how such embeddings can be used to cluster svaras that have similar forms and hence elucidate the distinct svara-forms that exist in this performance, whilst assisting in their automatic identification. Furthermore, we compare the melodic features of our 54 svara-form clusters to illustrate their unique character and demonstrate the dependency between these cluster allocations and the immediate melodic context in which these svaras are performed.

CCS CONCEPTS

• **Applied computing → Sound and music computing**; • **Computing methodologies → Machine learning approaches**.

DLfM 2024, June 27, 2024, Stellenbosch, South Africa
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1720-8/24/06
https://doi.org/10.1145/3660570.3660580

KEYWORDS

Computational Musicology, Carnatic Music, Indian Art Music, Deep Clustering, Music Analysis, Svara Performance, Gamaka, Coarticulation, Annotation

ACM Reference Format:
Thomas Nuttall, Xavier Serra, and Lara Pearson. 2024. Svara-forms and coarticulation in Carnatic music: an investigation using deep clustering. In *11th International Conference on Digital Libraries for Musicology (DLfM 2024), June 27, 2024, Stellenbosch, South Africa.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3660570.3660580

1 BACKGROUND IN CARNATIC MUSIC

The Carnatic music terms *svara* and *gamaka* are commonly translated into English as note and ornament respectively. However, there are important differences between the way these two sets of terms are conceptualized. Gamaka (ornamentation) performs an integral rather than decorative function in Carnatic music [37]. It is often the specific prescribed gamakas that give a *rāga* (melodic framework) its unique character, and the proper expression of a rāga's character is one of the main goals of the style [37]. Therefore, the performance of particular gamakas (ornaments) on certain svaras (notes) is not optional for the performer, but rather is a necessary part of the performance.

There is a distinction to be made between how gamakas are understood in theory and in practice. Treatises and books of music theory list named and defined gamakas [3], meanwhile in practice, as noted by Carnatic flautist and musicologist T. Viswanathan 'it is virtually impossible to make an exhaustive list of all gamakas, for the many subtle variations of tone or pitch which are characteristic of South Indian vocal style simply defy systematic classification' [37]. Therefore, the gamakas listed in theoretical texts, while indicating broad types of ornamentation, do not provide the musician with sufficient information on precisely how a given svara must be performed in a given melodic context. Instead, this knowledge is acquired by musicians through learning compositions from their teacher, listening to how the teacher plays svaras within the context of each phrase and replicating this.

In any given rāga, there are typically one or more specific ways that each svara may be played in any given melodic context, and this way of performing the svara often involves pitch movement and points of emphasis and deemphasis, as well as rhythms arising from such change. Typical melodic movements on svaras include oscillations, referred to using the gamaka term *kampita*, some of which do not even rest on the theoretical pitch position of the svara [14]. Other common movements include slides (*jāru*) and the same note articulated twice through a brief movement to a lower pitch (*janta*). Two or more forms of melodic movement may be combined on one svara, for example a slide (jāru) may be followed by a brief upwards flick to higher pitch, before continuing to the next svara. Although these movements can be described using theoretical gamaka terms (either singly or in combination), and the study of these terms and their interpretation is valuable for understanding the Carnatic style (for example, [10, 14]), due to the variability in the performance of gamakas in different contexts, and, as described by T. Viswanathan, their tendency to defy systematic classification [37], in this article we will not focus on the gamaka names but rather will take a more practice-based approach, exploring the actual ways that each svara is sung in a single performance of a composition in rāga Bhairavi.

These ways of performing a svara in a given rāga are neither learnt by rote nor listed anywhere, but rather they are picked up through repeatedly playing the phrases within which they lie. There is equally no Carnatic music term to describe the combination of svara together with its correct melodic movement (gamaka), and instead the name of the svara is used. One might say, for example, that the svara *dha*[1] is performed in a particular way in one context, and in another way in a different context. For the purpose of this article, therefore, we will refer to these different ways that any given svara (for example, dha) is performed as 'svara-forms'. In this way we will consider the observable and various forms taken by each svara in rāga Bhairavi, based on a single 20-minute performance by the renowned vocalist Sanjay Subrahmanyan.

2 RELATED WORK

Existing research looking at the ways svaras are actually performed includes work based on computation from audio recordings of pitch curves (f_0 time series of the predominant melody) [12, 15, 16, 31, 32, 38]. Also relevant to this article is an analysis by Robert Morris of the different ways that each svara is performed in two compositions, one in rāga Kalyani and one in rāga Bhairavi, based on manual transcriptions into staff notation [22].

Whilst Morris relies on the manual grouping of svara-forms based on his own expertise, other studies have characterized differences in svara performance using computationally extracted features, broadly separable into two groups: those related to intonation, such as properties derived from pitch probability density functions across the octave (both in Carnatic music and the related North Indian Hindustani music context)[5, 6, 11, 28], and those related to gamaka, such as number of stationary points, or regions of constant pitch, in the extracted pitch curves [34–36]. Such characterization enables a quantitative comparison between svara performance and musically relevant concepts such as rāga and its structural features [5, 6, 28], and can assist with musicologically relevant tasks such as symbolic transcription [36].

More recently, neural network models have proven effective in automatically learning task-relevant features when trained on sufficiently labelled data. These learnt features, whilst often lacking interpretability when compared to handcrafted features, frequently achieve state-of-the-art results when used for tasks such as classification, clustering and dimensionality reduction. In the current context, with the predominant vocal melody described by an f_0 time series, we look to recurrent and convolutional neural networks that are able to encode sequences with consideration for each point in relation to its neighbours. We refer the reader to comprehensive reviews of algorithms for the tasks of time series clustering [18] and classification [9]. One limitation of such highly-parameterized models is the need for large amounts of labelled data for training. Such data can often be expensive and time-consuming to obtain. One model of particular interest is the recurrent, DeepGRU model, which utilizes Gated Recurrent Units (GRUs) for the task of classification, demonstrating comparable performance to LSTM-based models, using fewer parameters, and hence exhibiting faster training time and improved performance on smaller datasets [20].

This article builds on work exploring the influence of melodic context on the performance of svara, which has been examined computationally [28, 32] and also from an embodied music-analytical perspective using the concept of coarticulation [24]. Coarticulation is a phenomenon commonly examined in linguistics and human movement studies in which the performance of a unit, for example the articulation of a phoneme, is influenced by its preceding and succeeding units - that is to say, its context [17]. Pearson applies this concept to Carnatic music, showing how svaras in a violin performance of rāga Todi vary in the way they are realized, influenced by the svaras that precede and follow them (the immediate melodic context), as well as by the overarching oscillatory movements involved in their performance [24]. This coarticulatory tendency is explored in the present article using computational approaches.

Drawing on this existing research, we present a dataset of manually created svara annotations and train a recurrent neural network model, DeepGRU, to learn a fixed-dimension embedding space of svara-relevant melodic features. We demonstrate that these embeddings are valuable for the task of svara similarity comparison and consequently, svara-form clustering. We further explore the influence of melodic context on the clusters. Results are discussed from both a technical and musicological perspective, contributing to a better understanding of the potential uses of the dataset and productive analytical approaches.

3 DATA

We base our analysis on a performance from the Saraga digital library of Carnatic music performances, of the composition, Kamakshi, in rāga Bhairavi, performed by the senior Carnatic vocalist Sanjay Subrahmanyan [30]. From this audio recording we extract a time series of the f_0 corresponding to the predominant vocal melody and also annotate all svara start and end points, as described below. The performance duration is 20 minutes and consists of 1,530 individual svaras.

Rāga Bhairavi is generally considered to be a substantial and musically serious rāga, suitable for extensive exploration in performance. It includes all seven svaras (sa, ri, ga, ma, pa, dha and ni). Its ascending (*ārohaṇa*) and descending (*avarohaṇa*) scales are often written as follows: Ārohaṇa: S G2 R2 G2 M1 P D2 N2 Ṡ; Avarohaṇa: Ṡ N2 D1 P M1 G2 R2 S (the numbers 1 and 2 after the svara letters denote the pitch position (*svarasthāna*) with the octave divided into

[1] The seven svara names used in Carnatic music notation (*sargam*) are sa, ri, ga, ma, pa, dha and ni. Similarly to *solfège* notation, sa can be placed at whichever pitch is comfortable for the performers involved, although this is constrained by certain conventions

12 semitones). Bhairavi therefore includes two forms of dha, D1 and D2, with D1 lying a semitone above pa, and D2 a further semitone above D1. In ascending phrases D2 is used, and in descending phrases D1 is played. Most theoretical texts state that all svaras in Bhairavi apart from sa and pa, are performed with gamakas, but as can be seen from our annotations and clustering results, in practice, even sa and pa are sometimes performed with pitch movement.

3.1 Svara Annotations

The annotations that constitute this svara dataset were created by author 3, who has expertise in the style, with the help of a student assistant who was supported by published composition notation. Annotations were created in Praat [1] and involved identifying the start and end points of svaras and annotating each one with a label corresponding to its svara name (sa, ri, ga, ma, pa, dha or ni). Following this, all annotations made were checked by author 3, both with respect to the svara name assigned, as well as the start and end point of the svara, making changes where necessary. The composition Kamakshi is highly suitable for the purpose of svara annotation and classification, as it is a long composition in a particular format (*svarajati*) where lines, known as *caraṇam*, are first sung using the svara names and then repeated with lyrics. The performance with svara names provides additional certainty that the annotations align with the understanding of both the composer and performer.[2] In addition, some lines in this performance are performed at both slow and faster speeds, which provides further variation in svara-forms produced. Furthermore, the opening *pallavi* section is repeated between each caraṇam, which leads to a great deal of repetition of the same svaras. Through this process, we created a dataset of 1,530 svara annotations, which are provided in the accompanying github repository.

4 METHODOLOGY

We train a DeepGRU model on the time series corresponding to the predominant vocal melody of the individual svara annotations to predict svara label, map each svara time series to the model's embedding space, reduce the dimensionality of each embedding on a svara level, and cluster each svara group to identify unique svara-forms. We compare melodic and melodic-context features of each svara-form group to characterize the difference between these distinct groups.

4.1 Time Series Dataset Creation

For the entire Kamakshi performance we extract a time series corresponding to the predominant vocal melody using a frequency-temporal attention network, FTA-Net trained on a dataset of Carnatic vocal ground truth, implemented as part of the compIAM package [7, 26].

As in previous studies, we interpolate gaps in this pitch curve shorter than or equal to 250ms, such gaps occasionally occur within ornaments due to glottal stops or rapid vocal movements [8, 23]. We then apply a 2nd order Savitzky-Golay filter with window length of 145ms for smoothing. The smoothing window size and interpolation length is set according to manual inspection of the pitch curve quality when compared to the performance audio.

The pitch curve is segmented into individual svara observations based on the manual annotations presented in Section 3.1 and trailing or leading silence is removed. Those that contain silences or pitch jumps of over 1000 cents, owing to errors in the pitch extraction process, are removed. Finally, svara pitch curves for which more than 80% of values exist outside of the middle octave of the vocalist's range are transposed by 1200 cents, to lie within that octave.

An example of 4 observations of ni is shown in Figure 1. The forms seen here are just a few of the various ways in which ni can be performed in rāga Bhairavi.

4.2 Model Training and Deep Embeddings

We train a DeepGRU model on our time series dataset to predict the 7 svara class labels. The model is trained on 70% of the data and evaluated using classification accuracy on the remaining 30%. The input data to the network is z-score normalized using the training set, the initial learning rate, 0.001 and the batch size, 64. For regularization we use batch normalization and a dropout of 0.5. The number of features in the hidden state is 256. These parameter decisions are made based on repeated experiment with a range of sensible values, selecting those which maximize train accuracy.

Each svara observation in our time series dataset is embedded using the trained model and the hidden state for each used as features for within-svara clustering.

4.3 Clustering

We perform a SHAP analysis on the svara embeddings to understand how each affects the final prediction [19], the results of which are not included in this document but available in the accompanying github repository. We observe that although all latent features do contribute to the final prediction, not every feature is relevant for every svara. Since our goal is to cluster these embeddings for each svara, i.e. the subset of embedddings corresponding to each svara are clustered separately, we first reduce the dimensionality of each subset using Uniform Manifold Approximation and Projection (UMAP), in an attempt to reduce complexity whilst maintaining only the information relevant to each svara. [21]. The embeddings are first z-score normalized and mapped to 5 components using UMAP.

We apply k-means clustering on the 5-dimensional UMAP embeddings for each svara, 7 clusterings in total. The number of clusters, K, is determined individually for each svara using the elbow method on the inertia criterion, where inertia is defined as the sum of squared distances of samples to their closest cluster center [27].

Generally, inertia decreases monotonically with increasing K, a widely accepted method of selecting the 'correct' value of K is by computing the inertia for multiple values of K and selecting the point that corresponds to a sudden marked flattening of the curve, the point beyond which any further increase in K provides diminishing returns in terms of inertia value improvement [33]. We experiment with values of $5 \leq K \leq 30$ and select the optimum

[2] Regardless of the care taken with the annotations, the transcription of audio into svara notation involves a degree of subjectivity, largely arising from different decisions made regarding the level of detail to include in the transcription[25]. In this transcription we have aimed to follow the level of detail in the published notation and uttered svara names.

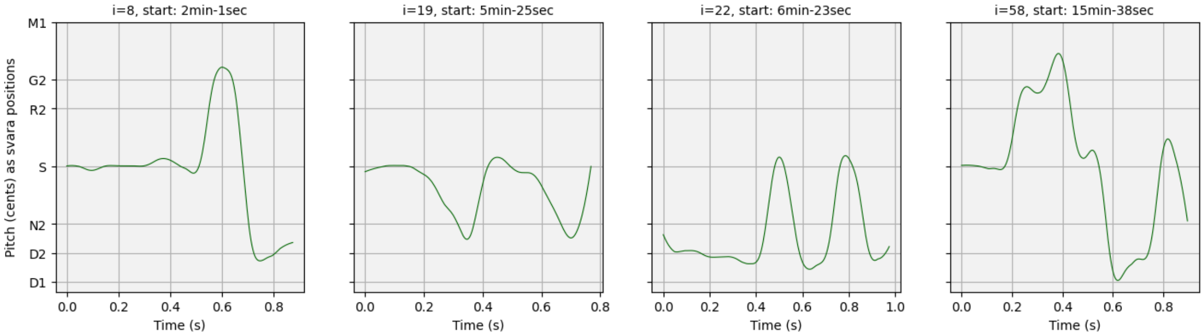


Figure 1: 4 pitch curves of the svara, ni

number of clusters using this method, resulting in 54 distinct svara-form clusters, detailed further in Section 5.

4.4 Post Processing

In an effort to prune clusters of incorrect or anomalous observations we apply an outlier detection to each. For each observation we compute the average euclidean distance between itself and all others in the cluster, $D(o_{ik})$ for observation i in cluster k.

$$D(o_{ik}) = \left(\sum_{\substack{j=1\\ i\neq j}}^{j=N} d(y_i, y_j) \right) / N \tag{1}$$

Where N is the total number of observations in cluster, k, y_i is the UMAP embedding vector for observation, i and $d(y_i, y_j)$ is the euclidean distance, $\sum_{v=1}^{V}(y_{iv} - y_{jv})$ between observation y_i and y_j. Observations are removed from a cluster if $D(o_{ik}) > Q3_k + 1.5IQR_k$, or $D(o_{ik}) < Q1_k + 1.5 * IQR_k$. Where $Q1_k$ is the first quartile , $Q3_k$ is the third quartile, and IQR_k is the interquartile range, of $D(o_{ik})$ values in cluster k. These observations are excluded from our analysis but are available in the accompanying results.

4.5 Cluster Characterization

To assist in the characterization of each svara-form cluster we extract certain melodic and melodic context features for every svara observation, averaging these values over each cluster.

4.6 Numerical Melodic Features

For each svara observation, we compute the following melodic features:

min_pitch - Minimum pitch value in cents

max_pitch - Maximum pitch value in cents

pitch_range - *max_pitch* - *min_pitch*

av_pitch - Mean pitch value in cents

av_start_pitch - Mean pitch value in cents in the first 10% of the pitch curve

av_end_pitch - Mean pitch value in cents in the final 10% of the pitch curve

num_change_points - Number of peaks or troughs in the pitch curve. To discount small peaks and troughs owing to subtle vibrato, while retaining the wider kampita gamaka and other musically meaningful pitch fluctuations, we discard those with a topographic prominence of below 70 cents.

4.7 Categorical Melodic Context Features

The comprehensive nature of our svara annotations provides us with melodic contextual information about each svara in the dataset, i.e for each svara observation, we know which svara (if any) precedes or succeeds it. This information is useful in studying the relationship between svara performance and its immediate melodic context. Each svara observation is labeled with the following melodic context features:

Preceding Svara - Svara label of the svara performed immediately before the current one. ‘silence’ if no svara precedes it.

Succeeding Svara - Svara label of the svara performed immediately after the current one. ‘silence’ if no svara succeeds it.

Preceding-Succeeding Svara - Unique combination of *preceding svara* and *succeeding svara*, e.g. ‘ni-pa’

Preceding Direction - Either ‘ascending’ or ‘descending’ depending on whether the current svara is above or below the preceding one. ‘from silence’ if the current svara follows silence.

Succeeding Direction - Either ‘descending’ or ‘ascending’ depending on whether the current svara is above or below the succeeding one. ‘to silence’ if there is no succeeding svara.

Preceding-Succeeding Direction - Either ‘ascending’ or ‘descending’ depending on whether the succeeding svara is above or below the preceding one. ‘to silence’ if there is no succeeding svara. ‘from silence’ if there is no preceding svara.

4.7.1 Normalized Mutual Information. To quantify the relationship between cluster allocation and immediate melodic context, we compute the Normalized Mutual Information (NMI) between the categorical melodic features in Section 4.7 and the categorical ‘feature’ of cluster label. The NMI of two variables, A and B is defined as:

$$NMI(A, B) = \frac{2 \times I(A, B)}{H(A) + H(B)} \tag{2}$$

Where $I(A, B)$ is the mutual information between A and B and $H(X)$ is the Shannon entropy of the variable X [2]. An NMI value of 0 indicates that the two features are completely independent and a value of 1 indicates that they are perfectly correlated. An

	Sa	Ri	Ga	Ma	Pa	Dha	Ni
# Clusters	6	9	8	8	7	9	7

Table 1: Number of svara-form clusters identified for each svara

investigation into the relationship between correlation coefficients and mutual information as a measure of dependence can be found in [29].

5 RESULTS

All results, pitch plots and code to reproduce this analysis can be found in the accompanying github repository.[3]

5.1 Embeddings Model

Our model converges after around 100 training epochs. We evaluate the model using 2-fold cross validation, reporting a test prediction accuracy of 87.6%. Figure 2 shows the confusion matrix for predictions on the test set. Whilst overall the model predicts svara very well, it has most difficulty in classifying dha, particularly in distinguishing it from ni. This might be influenced by the fact that the two svaras are neighbours, and are performed in rāga Bhairavi with a wide range of gamakas, some of which result in similarly shaped pitch curves for ni and dha across a highly similar pitch space.

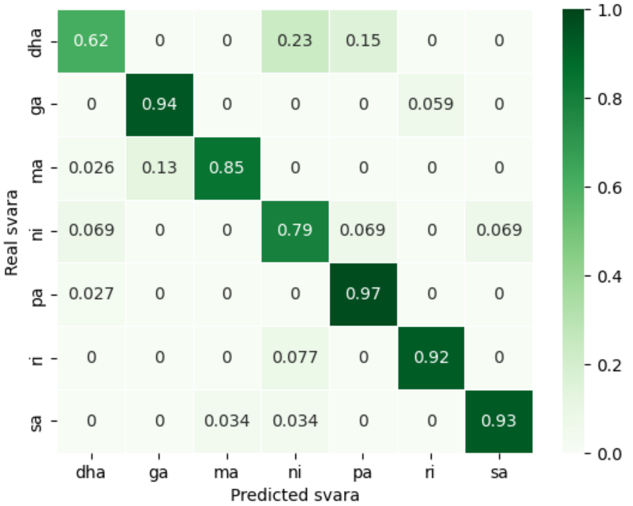


Figure 2: Confusion matrix of test predictions from DeepGRU model. Values represent the proportion of <y-axis svara> predicted as <x-axis-svara>

5.2 Clustering

In total we identify 54 svara-form clusters, Table 1 displays the number of clusters for each svara.

	Normalized Mutual Information (NMI)					
	P	S	PS	Pdir	Sdir	PSdir
Sa	0.224	0.237	**0.350**	0.180	0.145	0.177
Ri	0.324	0.179	**0.370**	0.268	0.148	0.145
Ga	0.340	0.217	**0.380**	0.261	0.181	0.251
Ma	0.385	0.117	0.403	**0.405**	0.117	0.391
Pa	0.248	0.239	**0.398**	0.144	0.057	0.172
Dha	0.326	0.271	**0.496**	0.326	0.148	0.145
Ni	0.509	0.374	**0.609**	0.327	0.274	0.311

Table 2: Normalized mutual information (NMI) values between cluster prediction and melodic context. Values correspond to the NMI between the categorical variables, cluster allocation and melodic context, introduced in Section 4.7. Where P is the preceding svara, S is the succeeding svara, PS is the unique combination of preceding and succeeding svara, Pdir is the preceding direction, Sdir is the succeeding direction, and PSdir the preceding-succeeding direction.

For 45 of these 54 clusters, at least one of the melodic context features introduced in Section 4.7 corresponds to at least 70% of svaras in that cluster. Table 2 further clarifies this relationship with the NMI values for each combination of svara and melodic context feature. We see that in all cases the melodic context feature with the strongest relationship to cluster allocation is the combination of preceding and succeeding svara: the full melodic context. This is particularly notable for ni and dha, as can also be seen when exploring the clustering results. We note that in almost all cases, melodic direction is not as important as the exact preceding and succeeding svara in determining cluster allocation, though naturally, these two features are closely related.

5.3 Interactive Svara Explorer

So as to provide a method of exploring the results of this analysis to non-technical readers, we present our clustering results in an interactive web application, available via the project Github repository. The DeepGRU embeddings of each svara are mapped to 2 dimensions using UMAP and visualized as points in a 2d scatter plot, the cluster and svara labels corresponding to each observation are indicated with distinct colors. Theoretically, proximity of points (svaras) in this plot correspond to similarity between embedding vectors. It is important to note that these 2-dimensional UMAP embedddings are not those that are used for clustering in Section 4.3, for which we use 5 components.

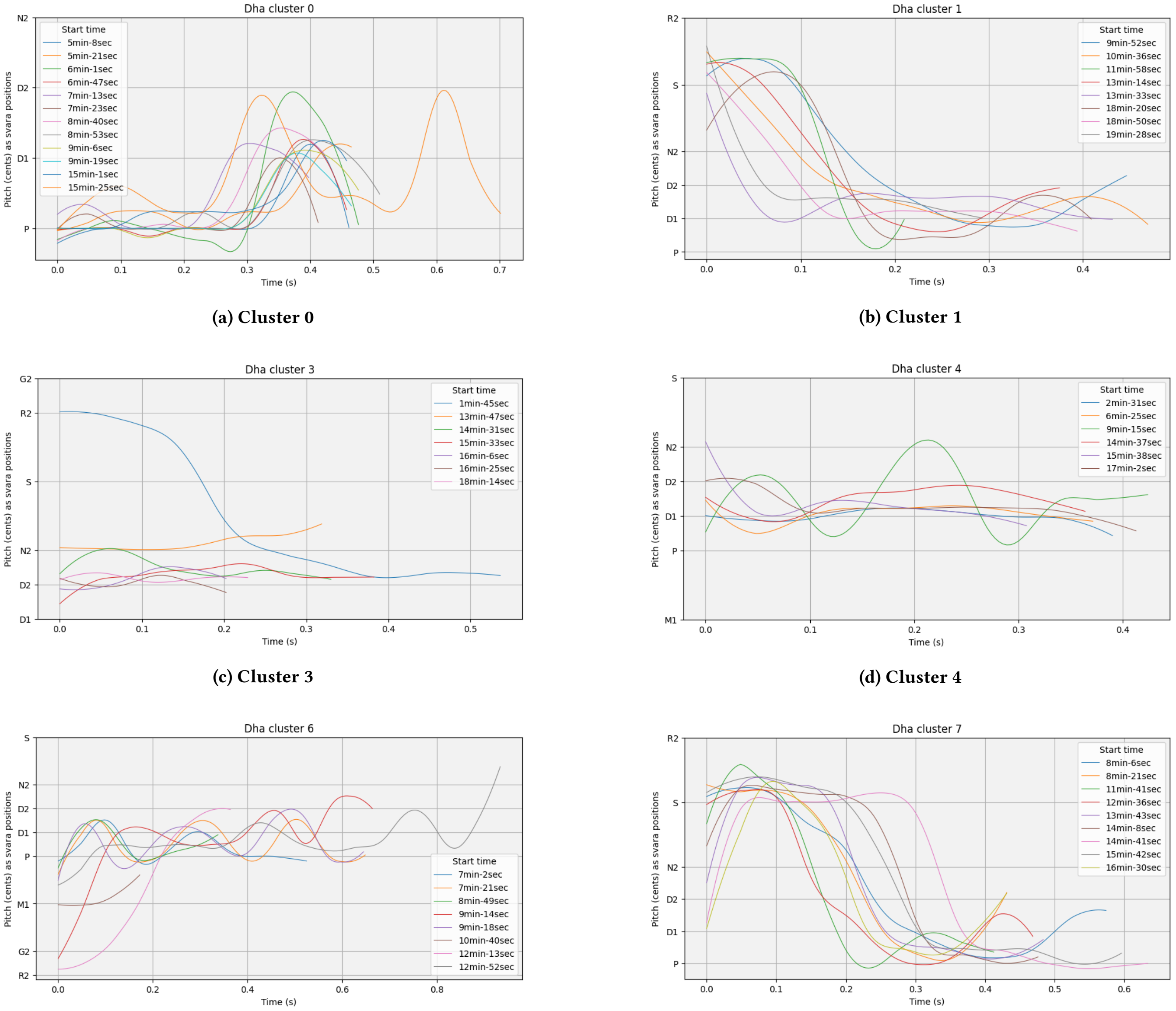
Users can select a svara, explore the individual observations, view their pitch curves, listen to the corresponding audio, visualize their melodic and melodic context features, and view a feature summary averaged across the svara-form cluster.

6 DISCUSSION

In this section, we reflect on the musicological relevance of the computational results, based on a combination of qualitative analysis and knowledge of the style.[4] The analysis focuses on the clusters of

[3] https://github.com/MTG/svaraforms-coarticulation

[4] This is founded on the experience of author 3, who spent several years in South India learning to play the style, followed by ten years of research in collaboration with Carnatic musicians.



(a) Cluster 0

(b) Cluster 1

(c) Cluster 3

(d) Cluster 4

(e) Cluster 6

(f) Cluster 7

Figure 3: Pitch curves corresponding to 6 dha clusters

dha, and the method used was to first listen to all examples of the svara in our dataset without clustering, followed by identifying and describing perceptually distinct svara-forms from amongst these. Following this, the clustered forms resulting from the computational analysis were examined to consider how well they matched with the qualitatively identified forms, and also to explore the musical features, common to each cluster.[5]

Through listening to the svara audio files in each cluster and examining their respective pitch curves, it becomes apparent that many of the clusters consist of highly similar svara-forms, while there are some clusters that are more mixed. In cluster 0 (see Figure 3a) nearly all of the examples involve a pitch movement that starts at pa and ends with a brief touch of dha. The one outlier in the cluster is an oscillation between pa and dha, which nevertheless is similar to the others in parts of its pitch curve. Overall, this can be considered a highly consistent group, nicely highlighting a clear svara-form of dha in rāga Bhairavi.

Cluster 1 also shows a good degree of consistency, with all pitch curves starting at sa and ending at dha (see Figure 3b). There is however a slight difference in the shape of the pitch curves. Some traverse sa-pa-dha in an elegantly curved sliding motion

[5] For the sake of visual clarity, three pitch curves in total that did not accurately reflect the audio were removed from the figures accompanying this musicological discussion, but these remain present in the dataset and overall results.

(*plots/dha/ni-82-pa.png* in the github repository), while others comprise a simple brief slide from sa to dha (*plots/dha/ni-52-pa.png* in the github repository). In this case, although these two slightly differing variants both appear in one cluster, the clustering process also helped reveal a form of the svara that was not initially identified by the analyst based on listening alone (the brief slide from sa to dha), demonstrating the way in which the clustering may be useful for music performance analysis, particularly when combined with existing knowledge of the style. Cluster 7, meanwhile, displays another set of the sa-pa-dha svara form (Figure 3f), but includes four examples that remain on pa rather than rising to dha at the end. Therefore, cluster 1 and cluster 7 present variations on the same overall melodic movement, all sliding down from sa, with one cluster more focused on the movement to dha, and the other more consistently reaching pa.

Cluster 2 comprises 13 examples, all of which dwell largely on pa, some with a slight movement up towards dha at the end (but not to the extent of those in cluster 0). This cluster therefore nicely represents the dha svara-form that consists largely of pa. Meanwhile clusters 3 and 4 comprise a mostly steady D2 and D1 respectively (Figures 3c and 3d). It should be noted that the description 'steady' is relative, as there will always be some slight pitch fluctuation in vocal pitch curves. We therefore use the term steady to refer to svaras that can be perceived as a single pitch, although perhaps including slight vibrato or other subtle pitch movement.

A particularly characteristic form of svara dha in Bhairavi is an oscillatory movement between pa and dha, which can be described as a kampita gamaka. In Bhairavi, this is often performed on dha when followed by ma, as in the phrase "ma pa ni dha ma", although, as seen in this cluster, a similar oscillation can also occur on dha followed by pa. The majority of instances of this oscillatory form of dha fall in cluster 6, six of which have this oscillatory form (Figure 3e). Also appearing in this cluster however are three renditions of dha that have no oscillation at all, demonstrating that although the current clustering is helpful in highlighting the various svara-forms, it cannot be used as the sole source of information in a musicological analysis, but rather must be combined with knowledge of the style.

Although the different forms of each svara are not learnt by rote but rather through learning phrases, Carnatic musicians are of course aware that the performance of svaras vary in ways that are influenced by immediate melodic context. In research literature, such influence has been explored from both computational and music analytical perspectives [24, 28, 32]. The Normalized Mutual Information (NMI) results from this paper indicate an influence of melodic context on clustering. As the clustering itself does not perfectly match the various svara-forms identified qualitatively, the current NMI results do not directly relate to these svara-forms. However, as noted above, the clusters are indicative of many qualitatively identified svara-forms, with highly consistent results on numerous clusters. Therefore, it can be argued that the NMI results relate not only to clusters but also, to some extent, to the qualitatively identified svara-forms. Focusing on those clusters that are consistent in their svara-form we can see concomitantly consistent immediate melodic context. For example, in cluster 0, every dha is preceded by pa, and is succeeded by either pa or ma. Therefore, all of these renditions open with an ascending motion to dha and are followed by a descending motion, either to pa or ma. Cluster 1 is even more consistent, with every example preceded by ni and succeeded by pa. Clusters that are similarly consistent in both svara-form and immediate melodic context can be found across the dataset.

The contextual consistency is lower for those clusters of dha that consist of largely steady svaras, which is perhaps indicative of the fact that such steady svaras do not include contextual information in their form; they are not coarticulated with previous and following svaras, and, for example, they are not subsumed by an overarching oscillation that begins in the preceding svara and continues through to the succeeding svara, as sometimes can be the case [24]. As a result, such steady pitch svara-forms can be sung in a variety of contexts and, therefore, the immediate melodic context in the clusters is more varied. Furthermore, the contextual consistency appears lower in those clusters that have been qualitatively identified as having a lower degree of svara-form consistency. In conclusion, the combined evidence points towards an effect of melodic context, which would be expected based on the coarticulated melodic structure of Carnatic music.

As noted in Section 1, written Carnatic music theory has not attempted to define the specific ways that each svara can be played in a given rāga [37]. Instead this knowledge exists as part of the oral tradition, distributed across expert performers of the style. Evaluation of the svara-form clusters presented in this article can therefore only be made through expert assessment and/or annotation. Such expert assessment of a phenomenon without well-defined categories will be subject to significant variation, depending on the degree of detail considered and the individual annotator's perception of the borders between categories - between similarity and difference [25]. For example, in cluster 1 of svara dha, should there be two categories or only one? Either option is plausible. Therefore, it is necessary to accept that categorization of svara-forms will involve several plausible truths rather than a single ground truth. A potentially fruitful avenue for further research would be to employ multiple expert annotators (experienced Carnatic musicians) in order to assess the degree of inter-rater agreement on categorization of svara-forms and any 'subjectivity ceiling' that might exist [4, 13].

7 CONCLUSIONS

In this study we show that deep embeddings of pitch time series for Carnatic svaras are useful for the task of svara-form clustering. For a single performance in rāga Bhairavi, we automatically identify and present 54 unique svara groups that are broadly indicative of forms that are likely to be recognized by Carnatic musicians and musicologists. We further demonstrate how these forms relate to immediate melodic context, explicating knowledge that is implicitly understood by practitioners of the tradition and showing that the style can be considered coarticulatory, in the sense that immediate melodic context is an important determinant of svara-form. Alongside our analysis we publish a dataset of comprehensive svara annotations for the performance and an interactive web application to explore the svaras and their clusterings.

We hope that our open dataset and clusterings will stimulate further research on the identification and exploration of svara-forms across this and other Carnatic rāgas. Meanwhile the embeddings model, svara representations, and annotations should prove valuable for other computational music analysis tasks such as automatic symbolic transcription. Further work could include the improvement of pitch extraction and/or manual removal of inaccurate pitch curves from the dataset. In addition, svara-form categorizations made by multiple Carnatic musicians, would allow us to learn more about the extent and areas of agreement amongst experts.

ACKNOWLEDGMENTS

The authors would like to acknowledge Lluís Solsona for his assistance in creating the svara annotations.

This research was carried out as part of the "IA y Música: Cátedra en Inteligencia Artificial y Música" (TSI-100929-2023-1), funded by the Secretaría de Estado de Digitalización e Inteligencia Artificial, and the European Union-Next Generation EU, under the program Cátedras ENIA 2022 para la creación de cátedras universidad-empresa en IA.

REFERENCES

[1] Paul Boersma and David Weenink. 2019. Praat: Doing phonetics by computer [Computer application]. https://www.praat.org

[2] Thomas M Cover. 1999. *Elements of information theory.* John Wiley & Sons.

[3] Subbarama Diksitar. 2008. *Sangita Sampradaya Pradarsini (English Web Edition).* http://ibiblio.org/guruguha/ssp.htm.

[4] Arthur Flexer and Thomas Grill. 2016. The Problem of Limited Inter-rater Agreement in Modelling Music Similarity. *Journal of New Music Research* 45, 3 (July 2016), 239–251. https://doi.org/10.1080/09298215.2016.1200631

[5] Kaustuv Kanti Ganguli, Sankalp Gulati, Xavier Serra, and Preeti Rao. 2016. Data-driven exploration of melodic structure in Hindustani music. In *Devaney J, Mandel MI, Turnbull D, Tzanetakis G, editors. ISMIR 2016. Proceedings of the 17th International Society for Music Information Retrieval Conference; 2016 Aug 7-11; New York City (NY).[Canada]: ISMIR; 2016.* 605–611.

[6] Kaustuv Kanti Ganguli and Preeti Rao. 2017. Towards Computational Modeling of the Ungrammatical in a Raga Performance. In *Hu X, Cunningham SJ, Turnbull D, Duan Z. ISMIR 2017 Proceedings of the 18th International Society for Music Information Retrieval Conference; 2017 Oct 23-27; Suzhou, China.[Suzhou]: ISMIR; 2017.* 39–45.

[7] Genís Plaja-Roglans and Thomas Nuttall and Xavier Serra. 2023. *compIAM.* https://mtg.github.io/compIAM/

[8] Sankalp Gulati, Joan Serrà Julià, Kaustuv Kanti Ganguli, Sertan Sentürk, and Xavier Serra. 2016. Time-delayed melody surfaces for rāga recognition. In *Devaney J, Mandel MI, Turnbull D, Tzanetakis G, editors. ISMIR 2016. Proceedings of the 17th International Society for Music Information Retrieval Conference; 2016 Aug 7-11; New York City (NY).[Canada]: ISMIR; 2016.* 751–757.

[9] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. 2019. Deep learning for time series classification: a review. *Data mining and knowledge discovery* 33, 4 (2019), 917–963.

[10] R. S Jayalakshmi. 2002. *Subbarama Dikshitarin Sangita-sampradaya-pradarsiniyil gamakangal.* PhD. University of Madras. https://musicresearchlibrary.net/omeka/items/show/2483

[11] Gopala Krishna Koduri, Joan Serrà Julià, and Xavier Serra. 2012. Characterization of intonation in carnatic music by parametrizing pitch histograms. In *Gouyon F, Herrera P, Martins LG, Müller M. ISMIR 2012: Proceedings of the 13th International Society for Music Information Retrieval Conference; 2012 Oct 8-12; Porto, Portugal. Porto: FEUP Ediçoes, 2012.*

[12] Madhu Mohan Komaragiri. 2013. *Pitch analysis in South Indian music: with a critical examination of the theory of 22 śruti-s.* Munshiram Manoharlal Publishers, New Delhi.

[13] Hendrik Vincent Koops, W. Bas de Haas, John Ashley Burgoyne, Jeroen Bransen, Anna Kent-Muller, and Anja Volk. 2019. Annotator subjectivity in harmony annotations of popular music. *Journal of New Music Research* 48, 3 (May 2019), 232–252. https://doi.org/10.1080/09298215.2019.1613436

[14] T. M. Krishna and Vignesh Ishwar. 2012. Carnatic music: Svara, gamaka, motif and raga identity. In *Proceedings of the 2nd CompMusic Workshop*, X. Serra, P. Rao, H. Murthy, and B Bozkurt (Eds.). Universitat Pompeu Fabra, Barcelona, 12–18.

[15] Arvindh Krishnaswamy. 2003. Application of pitch tracking to south indian classical music, Vol. 5. IEEE, 557–60. https://doi.org/10.1109/ICASSP.2003.1200030

[16] Arvindh Krishnaswamy. 2004. Melodic Atoms for Transcribing Carnatic Music. In *International Society for Music Information Retrieval Conference.* https://api.semanticscholar.org/CorpusID:1349104

[17] Barbara Kühnert and Francis Nolan. 1999. The origin of coarticulation. In *Coarticulation* (1 ed.), William J. Hardcastle and Nigel Hewlett (Eds.). Cambridge University Press, 7–30. https://doi.org/10.1017/CBO9780511486395.002

[18] Baptiste Lafabregue, Jonathan Weber, Pierre Gançarski, and Germain Forestier. 2022. End-to-end deep representation learning for time series clustering: a comparative study. *Data Mining and Knowledge Discovery* 36, 1 (2022), 29–81.

[19] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

[20] Mehran Maghoumi and Joseph J LaViola. 2019. DeepGRU: Deep gesture recognition utility. In *Advances in Visual Computing: 14th International Symposium on Visual Computing, ISVC 2019, Lake Tahoe, NV, USA, October 7–9, 2019, Proceedings, Part I 14.* Springer, 16–31.

[21] Leland McInnes, John Healy, and James Melville. 2020. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426 [stat.ML]

[22] Robert Morris. 2011. Tana Varnam-s: An Entry into Rāga Delineation in Carnatic Music. *Analytical Approaches to World Music* 1, 1 (2011), 1–27. https://journal.iftawm.org/wp-content/uploads/2022/02/Morris_AAWM_Vol_1_1.pdf

[23] Thomas Nuttall, Genís Plaja-Roglans, Lara Pearson, and Xavier Serra. 2022. In search of Sañcāras: tradition-informed repeated melodic pattern recognition in carnatic music. In *Rao P, Murthy H, Srinivasamurthy A, Bittner R, Caro Repetto R, Goto M, Serra X, Miron M, editors. Proceedings of the 23nd International Society for Music Information Retrieval Conference (ISMIR 2022); 2022 Dec 4-8; Bengaluru, India.[Canada]: ISMIR; 2022.* 337–344.

[24] Lara Pearson. 2016. Coarticulation and gesture: an analysis of melodic movement in South Indian raga performance. *Music Analysis* 35, 3 (2016), 280–313. https://doi.org/10.1111/musa.12071

[25] Lara Pearson and Brindha Manickavasakan. 2023. Annotating Karnataka Music: Encounters Between a Musical Tradition and Computational Tools. In *Second Symposium of the ICTM Study Group on Sound, Movement, and the Sciences (SoMoS)*, Filippo Bonini Baraldi (Ed.). Barcelona, Spain, 23–27. https://zenodo.org/records/10423805

[26] Genís Plaja-Roglans, Thomas Nuttall, Lara Pearson, Xavier Serra, and Marius Miron. 2023. Repertoire-Specific Vocal Pitch Data Generation for Improved Melodic Analysis of Carnatic Music. *Transactions of the International Society for Music Information Retrieval* (Jun 2023). https://doi.org/10.5334/tismir.137

[27] Andrei Rykov, Renato Amorim, Vladimir Makarenkov, and Boris Mirkin. 2024. Inertia-Based Indices to Determine the Number of Clusters in K-Means: An Experimental Evaluation. *IEEE Access* PP (01 2024), 1–1. https://doi.org/10.1109/ACCESS.2024.3350791

[28] Sertan Sentürk, Gopala Krishna Koduri, and Xavier Serra. 2016. A score-informed computational description of svaras using a statistical model. (2016).

[29] Lin Song, Peter Langfelder, and Steve Horvath. 2012. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC bioinformatics* 13, 1 (2012), 1–21.

[30] Ajay Srinivasamurthy, Sankalp Gulati, Rafael Caro Repetto, and Xavier Serra. 2021. Saraga: Open Datasets for Research on Indian Art Music. *Empirical Musicology Review* 16, 1 (2021), 85–98.

[31] M Subramanian. 2002. Analysis of gamakams of Carnatic music using the computer. *Sangeet natak* 37, 1 (2002), 26–47.

[32] M Subramanian. 2007. Carnatic ragam thodi–Pitch analysis of notes and gamakams. *Journal of the Sangeet Natak Akademi* 41, 1 (2007), 3–28.

[33] RL Thorndike. 1953. Who belongs in the family? Pyschometrika 18 (4): 267–276.

[34] Venkata Subramanian Viraraghavan, Rangarajan Aravind, and Hema A Murthy. 2017. A Statistical Analysis of Gamakas in Carnatic Music. In *Hu X, Cunningham SJ, Turnbull D, Duan Z. ISMIR 2017 Proceedings of the 18th International Society for Music Information Retrieval Conference; 2017 Oct 23-27; Suzhou, China.[Suzhou]: ISMIR; 2017.* 243–249.

[35] Venkata Subramanian Viraraghavan, Rangarajan Aravind, and Hema A Murthy. 2018. Precision of Sung Notes in Carnatic Music. In *In: Gómez E, Hu X, Humphrey E, Benetos E. Proceedings of the 19th ISMIR Conference; 2018 Sep 23-27; Paris, France.[Canada]: ISMIR; 2018.* 499–505.

[36] Venkata Subramanian Viraraghavan, Arpan Pal, Hema Murthy, and Rangarajan Aravind. 2020. State-Based Transcription of Components of Carnatic Music. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 811–815.

[37] Tanjore Viswanathan. 1977. The Analysis of Rāga Ālāpana in South Indian Music. *Asian Music* 9, 1 (1977), 13–71.

[38] Harsh M. Vyas, Suma S. M., Shashidhar G. Koolagudi, and Guruprasad K. R. 2015. Identifying gamakas in Carnatic music. In *2015 Eighth International Conference on Contemporary Computing (IC3).* 106–110. https://doi.org/10.1109/IC3.2015.7346662