Computational analyses of coarticulatory processes in Karnatak music

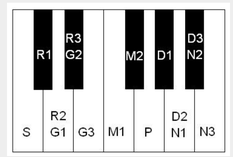
Thomas Nuttall
Universitat Pompeu Fabra, Barcelona
thomas.nuttall@upf.edu
Lara Pearson
University of Cologne, Germany

Karnatak music is a South Indian tradition with a distinctively ornamented style. Ornaments, known as gamakas, are not notated and are learnt within phrases. Musicians are aware that the performance of gamaka depends on melodic context, and this has recently been theorized using the concept of coarticulation. In this study we explore coarticulatory processes in Karnatak music through a combination of distance-based clustering, codependency analysis, and deep embeddings.

Karnatak Music | Svara and Gamaka

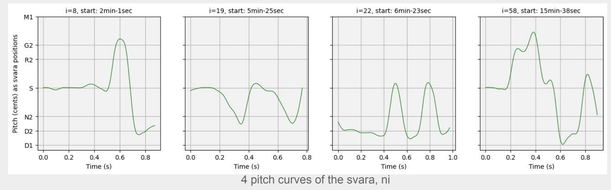
Svara | Note

S - Shadja (Sa)
R - Rishabha (Ri)
G - Gaandhaara (Ga)
M - Madhyama (Ma)
P - Panchama (Pa)
D - Dhaivata (Dha)
N - Nishadha (Ni)



Source: http://www.learnkeyboardonline.com/p/blog-page.html

Gamaka | Ornamentation

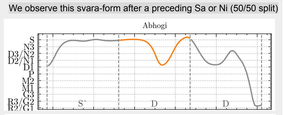
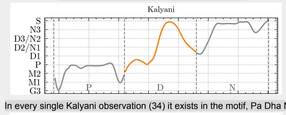
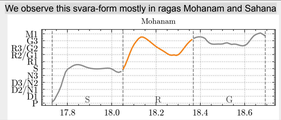
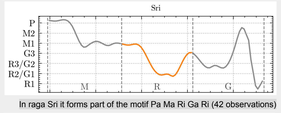
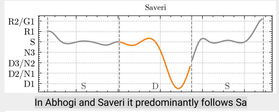
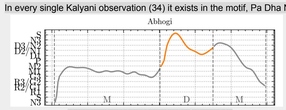
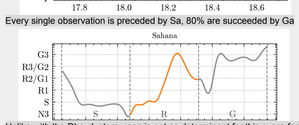
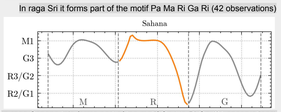
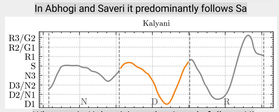
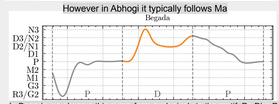
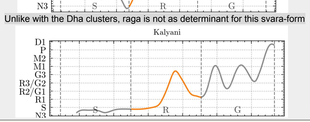
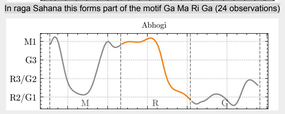


4 pitch curves of the svara, ni

- Svaras are note-length units with theoretical pitch positions
- Referred to using *sargam* syllables: **sa ri ga ma pa dha** and **ni**
- Sargam is used for written notation, and also in certain musical formats when the svara names are sung
- In practice, most svaras are performed with gamakas (ornamentation) that can have a strong impact on the pitch movement[1]
- Gamaka performs an integral rather than purely decorative function in Karnatak music - use of certain gamakas in certain contexts is not optional for the performer, but they are not notated

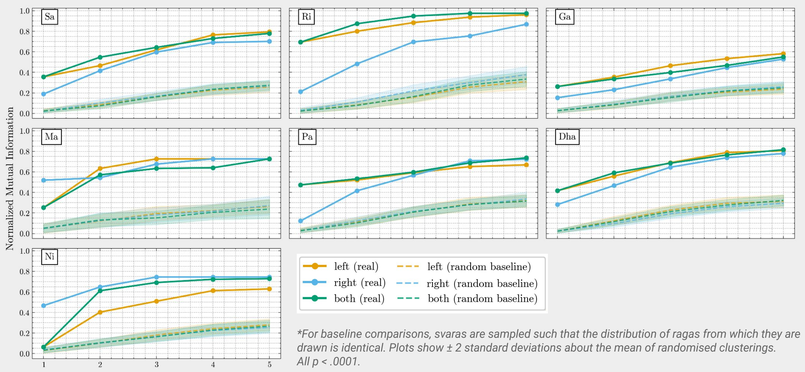
Coarticulation Hypothesis: The performance of each svara is influenced by its immediate melodic context.[2]
In phonetics, coarticulation is defined as the variation that a speech sound undergoes when influenced by neighbouring sounds.[3]

Svara-Forms | Unique Svara-Gamaka Clusters

Svara:	Dha		Ri	
Svara-form:	sds (66)	psd (66)	sgr (50)	mr (100)
	We observe this svara-form after a preceding Sa or Ni (50/50 split)	72% of the time, this svara-form follows Pa	We observe this svara-form mostly in ragas Mohanam and Sahana	80% of observations for this svara-form exist within the motif Ma Ri Ga
				
	In Abhogi and Saveri it predominantly follows Sa	In every single Kalyani observation (34) it exists in the motif, Pa Dha Ni	Every single observation is preceded by Sa, 80% are succeeded by Ga	In raga Sri it forms part of the motif Pa Ma Ri Ga Ri (42 observations)
				
	Whereas in Kalyani, it predominantly follows Ni	In Begada we observe this svara-form exclusively in the motif, Pa Dha Pa	Unlike with the Dha clusters, raga is not as determinant for this svara-form	In raga Sahana this forms part of the motif Ga Ma Ri Ga (24 observations)
				

Cooarticulation | Dependency on Melodic Context

Normalised Mutual Information between Svara-form group and Melodic Context



For baseline comparisons, svaras are sampled such that the distribution of ragas from which they are drawn is identical. Plots show ± 2 standard deviations about the mean of randomised clusterings. All p < .0001.

Key Findings

- For this analysis we work with 21 distinct svara-forms (1460 svaras total). Many of these are common to different ragas and performers.
- Melodic context is a key determinant of how a svara is performed. With the neighbouring 2 or 3 svaras on both sides typically being most correlated.
- Within single ragas all examples of a svara-form tend to have the same melodic context, but across different ragas they often have different contexts
- For this dataset, certain svara-forms have a one to one mapping with the 5 contextual neighbouring svaras, i.e., they exist almost exclusively within one motif
- Ultimately svara-forms depend on a combination of svara context and raga
- Our findings corroborate the findings of our previous svara-forms analyses[4,5]

Dataset | Svara Annotations of Karnatak Varnams

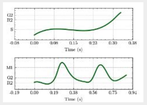
Carnatic Varnam Dataset[6]

- 7 varnams in 7 ragas sung by 5 professional singers
- 28 performances total
- The recordings are annotated with tala cycles (beat annotations)
- And accompanied by written sargam notation

Semi Automatic Alignment

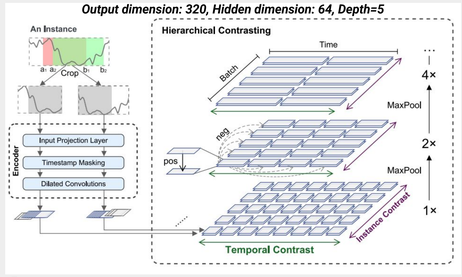
- Tala annotations are combined with written notation to produce svara-level transcriptions across all performances
- Incorrect transcriptions are removed upon manual inspection

Dataset of ~6000 individual svara transcriptions



Predominant pitch time series for each svara are extracted using FTA-NET

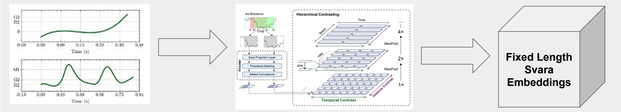
Embeddings | Deep Representations of Svara



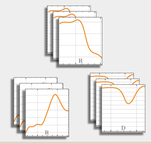
- TS2Vec[7] is a self-supervised time series representation model
- We train TS2Vec on all pitch time series of our svara annotations
- Learning a universal embeddings space that maximizes distance between augmented views of the same sequence whilst separating dissimilar sequences.

Clustering | Identifying Svara-Forms

All svara time series are embedded into the learnt space



And clustered with HDBSCAN to identify svara-form "**prototypes**"



By minimizing within cluster DTW - the clustering is optimized so as to prioritise very tight, small clusters - candidates for svara-forms

Finally the entire dataset of svaras is classified to a prototype using DTW kNN

Svaras are assigned if their normalised DTW distance to a prototype is below 20 cents

Analysis | Investigating Coarticulation

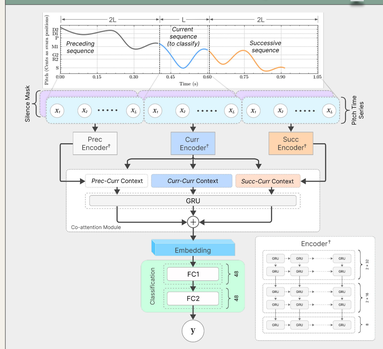
- Each clustered svara is labeled with a token corresponding to the k-length melodic context that precedes (**LEFT**), succeeds (**RIGHT**), or surrounds (**BOTH**) it.
- For each svara we compute the normalized mutual information (NMI) between the context tokens and it's unique svara-form group allocation.
- We do this for k in [1,5] for **LEFT**, **RIGHT**, and **BOTH** contexts.

For svara **R** in the sequence, DMPGM**R**GRSRG

k	LEFT context	RIGHT context	BOTH context*
1	M	G	M
2	GM	GR	M G
3	PGM	GRS	GM G
4	MPGM	GRSR	GM GR
5	DMPGM	GRSRG	PGM GR

BOTH contexts are built by iteratively taking from each side*

Use Case | Improving Svara Representation Models



- We adapted the existing DeepGRU model to encode both svara and its surrounding melodic context, before combining them via a co-attention mechanism prior to classification.[8]
- We train the model on one raga only (bhairavi) and include a context length equal to two neighbouring svaras (informed by analysis findings)
- The inclusion of melodic context leads to a 6.6% absolute increase in F1 score for svara label classification (from 78.3% to 84.9%), and an 7.8% absolute increase (from 59.9% to 67.7%) for classification of svara-form

Melodic Context	Target	F1 mean	F1 std
No	Svara	0.783	0.026
Yes	Svara	**0.849**	0.025
No	Svara-form	0.599	0.006
Yes	Svara-form	**0.677**	0.033

By incorporating melodic context information into svara representation models. We can significantly improve svara representation learning. Useful for computational tasks like **transcription**, **pattern recognition**, and **synthesis**.

[1]Krishna, T. M., & Ishwar, V. (2012). Carnatic music: Svara, gamaka, motif and raga identity. In X. Serra, P. Rao, H. Murthy, & B. Bozkurt (Eds.), *Proceedings of the 2nd CompMusic Workshop* (pp. 12–18). Universitat Pompeu Fabra.
[2]Pearson, L. (2016). Coarticulation and gesture: An analysis of melodic movement in South Indian raga performance. *Music Analysis, 35*(3), 280–313. https://doi.org/10.1111/musa.12071
[3]Hardcastle, W. J., & Hewlett, N. (1999). *Coarticulation: Theory, data and techniques*. Cambridge University Press.
[4]Nuttall, T., Serra, X., & Pearson, L. (2025). Svara-Forms in Carnatic Music: Contextual Influences on the Performance of Svara. In *Workshop on Indian Music Analysis and Generative Applications (WIMAGA)*, a satellite workshop of ICASSP
[5]Nuttall, T., Serra, X., & Pearson, L. (2024). Svara-forms and Coarticulation in Carnatic Music: An Investigation Using Deep Clustering. In *Proceedings of the 11th International Conference on Digital Libraries for Musicology (DLfM 2024)* (pp. 15–22). ACM.
[6]CompMusic. (n.d.). *Carnatic varnam dataset*. Repositori Digital de la UPF. Retrieved from http://hdl.handle.net/10230/27956
[7]Yue, Z., et al. (2022). Ts2vec: Towards universal representation of time series. In *Proceedings of the AAAI Conference on Artificial Intelligence, 36*(8).
[8]Nuttall, T., Vijayan, V., Serra, X., & Pearson, L. (accepted). Leveraging Melodic Context for Improved Svara Representation. In *Proceedings of the 17th International Symposium on Computer Music Multidisciplinary Research (CMMR 2025).*