Cultural Specificities in Carnatic and Hindustani Music: Commentary on the Saraga Open Dataset

LARA PEARSON[1]
Department of Music, Max Planck Institute for Empirical Aesthetics, Germany

ABSTRACT: This commentary explores features of the "Saraga" article and open dataset, discussing some of the issues arising. I argue that the CompMusic project and this resulting dataset are impressive for their sensitivity to cultural specificities of the Hindustani and Carnatic musical styles; for example, the dataset includes manual annotations based on music theoretical concepts from within the styles, rather than imposing conceptual categories from outside. However, I propose there are aspects of the dataset's manual annotations that require clarification in order for them to be used as ground truths by other researchers. In addition, I raise questions regarding the representativeness of the dataset – an issue that has ethical implications.

WHILE music information retrieval (MIR) is now an extensive research field, there have been relatively few projects systematically addressing issues arising when applying MIR techniques across diverse musical practices – those beyond the usual candidates such as pop, jazz, and Western art music. The Saraga open dataset is just one output from perhaps the most significant example of such a project: the now completed CompMusic project coordinated by Xavier Serra from the Music Technology Group of the Universitat Pompeu Fabra in Barcelona. The dataset in question comprises two collections, both of them drawn from Indian music traditions, namely Hindustani and Carnatic music practices. I approach this commentary as a potential user of the dataset, with my area of expertise lying mainly in the Carnatic style. Although this approach to the commentary means that not all aspects of the dataset are addressed (they are perhaps too numerous to fit into one commentary anyway), I hope it will nevertheless be informative to other musicologists and also to the authors who state that one of their aims is to provide audio and ground truth data for "large-scale data-driven analysis in musicological studies" (Srinivasamurthy et al., 2021). In this commentary, I give an account of the dataset's features that I find to be particularly valuable while also pointing to some of the issues that I believe should be resolved in order for the dataset to achieve its fullest potential.

CULTURAL SPECIFICITIES IN MUSIC

Describing the goals of the CompMusic project, Serra (2014, p. 2) states that, although starting from the idea that there are some universal musical concepts, "we want to emphasise, and show, that many important aspects of a particular music recording can be better understood by considering cultural specificities." In their work on five different music practices (Hindustani, Carnatic, Makam, Jingu, Andalusian), they avoid taking a Eurocentric perspective by accepting that music theoretical units used in mainstream European-originating traditions might not be suitable for the analysis of all musical styles.

As a musicologist with a research focus on the analysis of Carnatic music,[2] I have been aware for some time of potential issues in applying MIR approaches to the style, due to its distinctive features. Firstly, the notation used in the style does not give the naive researcher a clear idea of what is actually performed. Ornaments, known as *gamakas*, are not notated, but must be applied to the theoretical pitch units (*svarasthānas*) according to a musical grammar learned orally/aurally during the years of study involved. In some *rāgas* (melodic types), the majority of notated theoretical pitch units are performed with oscillating *gamakas* that involve moving back and forth between two pitches, at least one of which is not notated. Some

of the oscillating *gamakas* may not even touch on the theoretical (notated) pitch position but instead include pitches that are outside of the theoretical *rāga* "scale" (Krishna and Ishwar 2012, p. 15; Pearson 2016, p. 290). Therefore, analyzing notation will tell researchers little about how the music actually sounds, unless the researchers are themselves Carnatic musicians. So, for example, analyzing prominent pitch distribution in audio recordings and then comparing the findings to theoretical "scales" of *rāgas*, in the hope of identifying them, will lead to a high failure rate.

Secondly, the detection of *rāgas* is best achieved not through identification of *rāga* "scales," but rather through detection of characteristic motifs and phrases; this is, in fact, one of the main ways that Carnatic performers and experienced listeners learn to perform and identify *rāgas* (Krishna and Ishwar, 2012, pp. 17-18). Two or more *rāgas* may have the same theoretical "scale" but differ in the *gamakas* performed on particular pitch positions and in the characteristic phrases present (Viswanathan, 1977, p. 31). Characteristic motifs, phrases, and ornaments are also important in Hindustani *rāg* identification (Rao and Rao, 2014; Ganguli and Rao, 2019).

Because the CompMusic project included Hindustani and Carnatic musicians amongst its researchers and collaborators, the teams were able to adjust their MIR approaches to suit the idiosyncratic features of the styles. One significant branch of their research worked towards *rāga* identification through finding characteristic phrases and motifs, based on analysis of pitch (F0) time-series extracted from audio recordings (Gulati et al., 2016; Ganguli et al., 2017). This research focus was an important achievement of the project and one that is reflected in the article and dataset reviewed in this commentary. The dataset includes not only audio recordings of performances but also pitch (F0) data and manual annotations of characteristic phrases, thus enabling future research projects to analyze the recordings based on phrase units that are meaningful in the style. In addition, manual annotations of other meaningful theoretical units are included, such as the *sama* (start point of the metrical cycle) and named sections of the performances. As in the case of the characteristic phrase annotations, these can be used to ask meaningful questions of the dataset, informed by the annotators' knowledge of the styles.

The project has gone further still, using some of these data (so far focusing on the Hindustani dataset) to create educational applications involving visualizations and game-like interactions. The stated aims of the tools are to develop understanding and appreciation of Hindustani music (Srinivasamurthy et al., 2021). Two platforms are described in the article: the Musical Bridges website, which includes a set of interactive educational tools relating to Hindustani music, and an android mobile app (The Saraga App), which includes a related (although not identical) set of tools. In addition to interactive tools designed to increase understanding of *tāls* (types of rhythmic framework), the platforms include Hindustani *rāg* exploration tools that draw on phrase and section annotations from the Saraga dataset. One of the online *rāg* visualizers includes a moving line that traces the musical pitch of the main melodic line, synchronized to the audio recording and highlighted to indicate when a characteristic phrase is being sung. Meanwhile, other occurrences of the same phrase in the recording are indicated in a timeline along the bottom of the screen. I have found pitch visualizations of this type to be helpful in appreciating unfamiliar musical styles, and there is a long history of ethnomusicologists using pitch tracing, instead of staff notation, to visualize musical sound (Moore, 1974). An early use of this technique can be seen in Hugo Zemp's film on yodeling, titled *Head Voice, Chest Voice* (Zemp, 1987; see also Zemp, 1990). On first watching this film, I remember being struck by how seeing the pitch visualization helped me to "hear" what was being sung – to perceive the vocalization better and understand its features. Ethnomusicologists working on Indian raga performance have often used such an approach in publications (Sanyal and Widdess, 2004; Clayton, 2007), while Rao and van der Meer's AUTRIM project produced an impressive website that displays scrolling visualizations of pitch in dozens of Hindustani *rāg* performances.[3] However, the *rāg* visualizer in the Musical Bridges website goes one step further, allowing the viewer to also easily see where characteristic phrases lie in the performance as they watch the pitch visualizations scroll over time. I suggest this is a powerful tool for *rāg* performance appreciation, informed by a deep understanding of the style.

EXPLORING THE DATASET

In both the Hindustani and Carnatic datasets, for each audio file there are a number of associated files, including some with automatically extracted data (pitch (F0), tonic) and some with manually annotated data (section, *sama*, tempo, melodic phrase), as well as a metadata file that includes information on the *rāga*, *tāla*, main performer, and so on. To my knowledge, there are no comparable Indian music open datasets, and so the team should be applauded for this achievement.

The authors have aimed to provide "easy access to the audio, metadata, and annotations in the collections," and so here, I will discuss my experience using the scripts provided in the Github repository. In the scripts folder, there are four notebooks, two of which can be used to explore the dataset while the other two allow downloading of either the full dataset or user-defined subsections. The download_by_filtering notebook is perhaps the most useful of the four, allowing both exploration of the data and downloading of subsections. Here it is possible to filter the files using a large range of parameters, including the main artist's name, *rāga*, *tāla*, musical form, and file type (mp3, *sama*, phrase, pitch data, and so on). From this filtering, the user can either download the resulting list as a csv file or download the associated files. So, for example, in the Carnatic dataset, I was able to filter and download the phrase annotation files (mphrases-manual) for all of the recordings in *rāga* Bhairavi, as well as the related audio files. As the dataset is not vast (249 Carnatic recordings, 108 Hindustani) and there are a large number of *rāgas* (in this dataset, 96 Carnatic ragas and 61 Hindustani), this particular filtering resulted in only four recordings. But this would still be useful for a case study, and of course many research questions are likely to draw on the entire dataset rather than small subsets.

I did encounter some issues when using the notebooks to download certain sections of the dataset. Using the download_by_filtering script, the multi-track recordings present in the Carnatic dataset failed to download, as did the mphrases-manual annotation files in the Hindustani dataset (although I could gain access to the Hindustani manual annotation files by downloading the entire Github repository). In addition, on examination, I found a few Carnatic phrase annotation files to be empty. Such issues bring me to one potentially problematic aspect of the open dataset, which is its maintenance. With the research project that led to this dataset largely completed, it could be difficult for the authors of this paper to deliver on some of their goals of accessibility and, indeed, growth of the dataset. The article outlines some of the steps through which they hope that the research and music community can contribute to the dataset, either through submission of new audio recordings or through new manual annotations to the existing recordings. Regarding new manual annotations, an acknowledged important step is to have such annotations verified by experts. This type of expert verification can be difficult to secure, as it is laborious. However, reliability of the manual annotations is of utmost importance for the dataset to achieve its potential. I will discuss this issue further in relation to just one part of the dataset – the manual phrase annotations.

CONCEPTUALISING THE "CHARACTERISTIC PHRASE"

While the phrase annotations are one of the most interesting and potentially valuable features of this dataset, I did notice some issues relating to these data. Firstly, as the authors acknowledge, not all recordings are annotated (it is somewhere in the region of fifty percent for both the Carnatic and Hindustani datasets). Secondly, in those that are annotated, a relatively small number of phrase annotations are given. This is particularly the case in the Carnatic dataset, where there are often only a few annotations for recordings that lie between approximately 2 and 60 minutes long. This is not in itself a problem, but clearly the dataset would be more useful if the number of phrase annotations were increased.

However, I suggest there is a more fundamental, conceptual problem with the phrase annotations. The article correctly states that "Every rāga has a set of characteristic melodic phrases that act as building blocks to construct melodies," and that "characteristic melodic phrases are also the most prominent cues used by human listeners for identifying rāgas" (Srinivasamurthy et al., 2021). However, not all phrases in a composition or *rāga* performance can be considered "characteristic phrases," some are grammatically correct but do not point strongly to the particular *rāga*. Furthermore, as the authors acknowledge, there is no closed, definitive list of the characteristic phrases in any given *rāga*. Instead, performers become aware of such phrases through experience, and as a result there may be differences between various experts' assessments of what is, and is not a characteristic phrase. Therefore, even for experts, making these decisions or verifying another musician's annotations is not a straightforward task.

Perhaps in an attempt to manage these conceptual issues, the phrases annotated in the mphrases-manual files have been flagged with a number – 0, 1, or 2 – where, according to the article, 1 indicates a phrase that is representative of the composition, and 2 indicates a phrase representative of the *rāga*. However, the majority of phrases in the files are flagged with 0, which would seem to indicate that the majority of phrases annotated are neither characteristic of the *rāga* nor the composition. This would leave musicologists with even fewer characteristic phrases to work from. In addition, in some cases where the same phrase appears within a file on multiple occasions, on some occasions it is flagged with a 1 or 2 and then on other occasions with a 0. This suggests that there was some confusion regarding the meaning of the flags.

I propose that the dataset should be accompanied by a clearer and more extensive description of the rationale behind the annotators' choices regarding the various phrase categories represented by flag numbers. At present, it is difficult to feel confident in using these particular annotations as ground truths, not because the expertise of the annotators is in doubt – those mentioned as having verified the annotations are highly experienced performers – but due to inconsistencies in the flag numbering and the lack of discussion surrounding the criteria for deciding what is a characteristic phrase.

REPRESENTATIVENESS OF THE DATASET

Serra (2014, p. 2) describes one of the goals of creating a dataset as "putting together corpora that capture the essence of a particular music culture for performing our research work." Therefore, it is reasonable to ask what music culture is this particular dataset representative of. Here, I will restrict my comments largely to the Carnatic dataset, the context of which I am more familiar with, but the same question should be asked of the Hindustani dataset.

It might be assumed that the goal is for the Carnatic dataset to be representative of contemporary Carnatic music practice. But an issue here is that contemporary Carnatic music practice is not homogeneous. As with all musical styles, there are subsets within wider styles, some more dominant than others. One feature of the dataset that immediately stands out is that there are no recordings in which an instrumentalist is the main performer – they are all "vocal concerts", led by one or two vocalists with a violinist and one or more percussionists acting as accompanists. However, both Carnatic and Hindustani contemporary performance practice include concerts where instrumentalists are the lead performers, and where there is no vocalist. It is true that vocal concerts are more popular, particularly in the contemporary Carnatic music scene, but this does not mean that other instruments have ceased to exist and be appreciated by a subset of Carnatic music audiences. The *vīṇā* (plucked lute) and *nāgasvaram* (double-reed aerophone) are two instruments that played particularly important roles in the development of the Carnatic style (Sambamoorthy, 1960), and that are still performed and enjoyed today, although they are admittedly less dominant now than they once were (Pearson, 2018). However, they are completely absent from the dataset.

The article does not address the reasons for this choice, except to state that "vocal music is dominant in both traditions but more in Carnatic music" (Srinivasamurthy et al., 2021). Perhaps a decision was taken to create a certain uniformity amongst the recordings. But as a result of this uniformity, a dataset has been created that is not so much representative of contemporary Carnatic music as a whole, but rather that is typical of a dominant subset: namely vocal performance in a rather "serious" style (lighter and more purely devotional sub-styles are also absent from the Carnatic dataset). The possible repercussions of this are that it perpetuates a rather limited view of what Carnatic music is, and also restricts the sorts of questions that might be asked of the dataset. It would have been fascinating to explore differences between vocal and instrumental performance – for example, in the rendition of *gamakas* where there could be differences arising from the ways in which the human body is able to interact with the particular instrument, the types of sounds that can be produced, and even differences of style lying beyond such factors. Deciding what to include and exclude in a dataset, particularly in an open-access dataset that might be reused for decades to come, is a task that also has ethical implications. Unfortunately, this dataset has mirrored the gradual historical exclusion of *nāgasvaram* and *tavil* (double-headed drum) performance from Carnatic music, which is a social and caste issue (the issue being that of caste exclusion) as well as an issue of musical style (Terada, 1997; Krishna, 2013, pp. 357-360).

Notwithstanding the issues raised in this commentary, I find the dataset to be a significant contribution, which I hope to draw on and/or contribute to in the future. As it is an open dataset supported by a team of knowledgeable and committed researchers, it is clear that additions and improvements can be made. Therefore, it should be possible for the conceptual and categorical issues relating to the manual phrase annotations to be clarified, and for the dataset to become more inclusive of the full range of contemporary Carnatic and Hindustani music performance practices.

ACKNOWLEDGMENTS

NOTES

[1] Correspondence can be addressed to: Dr. Lara Pearson, Max Planck Institute for Empirical Aesthetics, Grüneburgweg 14, 60322 Frankfurt am Main, Germany, lara.pearson@ae.mpg.de.

[2] Carnatic can also be found written as Karnatik and Karnatak. They all arise from the full title for the style, *Karnāṭaka Saṅgīta*.

[3] https://autrimncpa.wordpress.com/

REFERENCES

Clayton, M. (2007). Time, gesture and attention in a *khyāl* performance. *Asian Music*, *38*(2), 71-96. https://doi.org/10.1353/amu.2007.0032

Ganguli, K. K., Lele, A., Pinjani, S., Rao, P., Srinivasamurthy, A., & Gulati, S. (2017). Melodic shape stylization for robust and efficient motif detection in Hindustani vocal music. *Proceedings from the Twenty-third National Conference on Communications (NCC)*, Madras, India. https://doi.org/10.1109/NCC.2017.8077055

Ganguli, K. K., & Rao, P. (2019). On the perception of raga motifs by trained musicians. *The Journal of the Acoustical Society of America*, *145*(4), 2418-2434. https://doi.org/10.1121/1.5097588

Gulati, S., Serra, J., Ishwar, V., Sentürk, S., & Serra, X. (2016). Phrase-based rāga recognition using vector space modeling. In *In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 66-70). https://doi.org/10.1109/ICASSP.2016.7471638

Krishna, T. M. (2013). *A Southern Music: The Karnatik Story*. India: HarperCollins.

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). Barcelona: Universitat Pompeu Fabra. http://hdl.handle.net/10230/20494

Moore, M. (1974). The Seeger Melograph Model C. *Selected Reports in Ethnomusicology*, *2*(1).

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

Pearson, L. (2018). Cultural heritage, sustainability and innovation in South Indian art music. In B. Norton & N. Matsumoto (Eds.), *Music as Heritage: Historical and Ethnographic Perspectives* (pp. 238-257). Abingdon: Routledge. https://doi.org/10.4324/9781315393865-12

Rao, S., & Rao, P. (2014). An overview of Hindustani music in the context of computational musicology. *Journal of New Music Research*, *43*(1), 24-33. https://doi.org/10.1080/09298215.2013.831109

Sambamoorthy, P. (1960). *History of Indian Music*. Madras: Indian Music Publishing House.

Sanyal, R., & Widdess, R. (2004). *Dhrupad: Tradition and Performance in Indian Music*. Aldershot: Ashgate Publishing.

Serra, X. (2014). Creating research corpora for the computational study of music: the case of the Compmusic project. In *Proceedings of the AES 53rd International Conference on Semantic Audio*. London, UK: Audio Engineering Society. http://hdl.handle.net/10230/44221

Srinivasamurthy, A., Gulati, S., Caro Repetto, R., & Serra, X. (2021). Saraga: Open Datasets for Research on Indian Art Music. *Empirical Musicology Review 16*(1), 85-98. https://doi.org/10.18061/emr.v16i1.7641

Terada, Y. (1997). Effects of nostalgia: the discourse of decline in Periya Mēḷam music of South India. *Bulletin of the National Museum of Ethnology*, *21*(4), 921-939.

Viswanathan, T. (1977). The analysis of rāga ālāpana in South Indian music. *Asian Music*, *9*(1), 13-71. https://doi.org/10.2307/833817

Zemp, H. (1987). *Head Voice, Chest Voice* (16mm and videocassette), 23 minutes Meudon, France: Centre National de la Recherche Scientifique.

Zemp, H. (1990). Visualizing music structure through animation: the making of the film *Head Voice, Chest Voice*. *Visual Anthropology*, *3*(1), 65-79. https://doi.org/10.1080/08949468.1990.9966523