

# Saraga: Open Datasets for Research on Indian Art Music

AJAY SRINIVASAMURTHY[1,2]

*Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain*

SANKALP GULATI[3]

*Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain*

RAFAEL CARO REPETTO[4]

*Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain*

XAVIER SERRA[1]

*Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain*

**ABSTRACT:** We introduce two large open data collections of Indian Art Music, both its Carnatic and Hindustani traditions, comprising audio from vocal concerts, editorial metadata, and time-aligned melody, rhythm, and structure annotations. Shared under Creative Commons licenses, they currently form the largest annotated data collections available for computational analysis of Indian Art Music. The collections are intended to provide audio and ground truth for several music information research tasks and large-scale data-driven analysis in musicological studies. A part of the Saraga Carnatic collection also has multitrack recordings, making it a valuable collection for research on melody extraction, source separation, automatic mixing, and performance analysis. We describe the tenets and the process of collection, annotation, and organization of the data. We provide easy access to the audio, metadata, and the annotations in the collections through an API, along with a companion website that has example scripts to facilitate access and use of the data. To sustain and grow the collections, we provide a mechanism for both the research and music community to contribute additional data and annotations to the collections. We also present applications with the collections for music education, understanding, exploration, and discovery.

Submitted 2020 April 30; accepted 2020 December 22.

Published 2021 December 10; <https://doi.org/10.18061/emr.v16i1.7641>

**KEYWORDS:** *Open annotated datasets, Indian Art Music, computational analysis*

## INTRODUCTION

CAREFULLY curated open datasets of music are indispensable resources for research. In various information retrieval domains (such as computer vision), significant advances were possible after the release of large open datasets. Such resources promote the reproducibility of research and ensure quality control (hence improvement). They also improve access by enabling researchers without the means to collect large datasets to contribute to research in these fields.

Digital artifacts of music accompanied with metadata and expert annotations is a basic need for data-driven approaches used in Music Information Retrieval (MIR) and computational musicology (Serra, 2017). There is a scarcity of open music recording collections for music research, primarily due to copyright issues. Some of the common ways the music research community has been tackling the problem of open datasets have been through the sharing of automatically computed audio features instead of the original recordings (Bertin-Mahieux, 2011; Porter et al., 2015) or providing researchers access to machines that can run analyses on local servers where data is stored. These limited sharing opportunities provide important means for reproducibility while facilitating collaborative work. However, the full potential (as observed in domains like computer vision) is far from being achieved due to the inaccessibility of the actual data itself, severely



limiting research to problems that utilize available features. In addition, recent developments in deep learning technologies have led to the development of end-to-end systems for music applications, where the input is the audio recording.

Serra (2014) has defined the principles to be taken into consideration for the creation of new research corpora, which can also serve for evaluating the goodness of the data collection for a particular research task. According to such principles, the purpose of the corpus has to be explicit, the corpora must have good data coverage for the phenomenon under study, the data must be as complete as possible and of good quality, and finally, corpora should be reusable, implying that the data should be available for other researchers. Wilkinson et al. (2016) recently formulated a set of guidelines to improve the findability, accessibility, interoperability, and reuse (FAIR) of digital assets.[5] While Serra (2014) emphasized the quality and coverage of data, the FAIR principles emphasize machine-actionability because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data. We consider the tenets and guiding principles by Serra (2014) to build research corpora with good quality and coverage while emphasizing FAIR principles to store, navigate and reuse data. In addition, we wish to provide open access to audio, metadata, and annotations, while allowing for the datasets to grow both in size and quality through community-driven efforts.

Tzanetakis (2014) observed an increasing interest in expanding data-driven research in MIR to music traditions around the world not commonly researched in this discipline, which poses hitherto unknown challenges and interesting research directions to improve the current state of the art. A major effort in that direction was the CompMusic [6] project focused on developing culture-specific analysis methods for various music cultures (Serra, 2011). The project provided an impetus to data-driven MIR in those music cultures, opening up new research directions, problems, and challenges. A primary effort of the project was to build research corpora and datasets for MIR in those music cultures. These corpora [7] have been created to study particular music traditions and include audio recordings plus complementary information that describes the recordings.

Srinivasamurthy et al. (2014) presented various datasets [8] for Indian Art Music (IAM), collected within the CompMusic project. Porter et al. (2013) initially presented a system named Dunya [9] for browsing audio music collections in the music cultures studied in CompMusic and an Application Programming Interface (API) for accessing metadata and audio files. Dunya has evolved to comprise the music corpora and related software tools that have been developed as part of the CompMusic project. The set of open-source software tools PyCompMusic [10] can be used to organize, access, and analyze these corpora. Each corpus has specific characteristics, and the developed software tools allow us to process the available information to study and explore the characteristics of each musical repertoire.

While metadata and annotations from the CompMusic datasets could be shared openly, we do not have the rights to share all the audio recordings since the music is copyrighted. Although the recordings are publicly accessible (through commercial albums), copyright still restricts many forms of usage that might be necessary for computational research. In addition, annotations could be performed only for parts of the audio, which limited their potential to be used in data-driven musicological studies.

In this paper, we present Saraga Indian Art music (Carnatic and Hindustani) open data collections, consisting of audio recordings and corresponding time-aligned melody, rhythm, and structural annotations for the complete duration of the recordings. The audio in Saraga collections comprises a subset of Dunya collections curated within the CompMusic project. The Carnatic collection includes recordings of live concerts, with a part of the collection available in multitrack. The Hindustani collection contains compilations by several renowned artists. The audio data spans about 96 hours of music accompanied by editorial metadata and time-aligned structural, melodic, and rhythmic annotations. We have obtained permission to share these audio recordings for research purposes under Creative Commons (CC) licenses [11], and the annotations are publicly available. The audio, metadata, and annotations are accessible through an API. A companion website [12] for the Saraga collection provides Python notebooks and example code snippets to access and explore the collections.

The datasets are intended to be useful for different MIR tasks in Indian art music and statistical computational musicology. The audio recordings and concerts were chosen to reflect the current state of music and performance practices. The availability of data, annotations, and code under open licenses helps create standardized open datasets for different MIR tasks in Indian art music, such as predominant melody extraction, meter tracking, structural segmentation, pattern similarity, pattern detection, and pattern discovery. An API to store/access data and annotation allows for effective indexing, sharing, and reusing the datasets.

The audio with rich metadata and annotations can help bootstrap datasets for relevant research problems posed in the analysis of Indian art music. These annotated collections are also useful to explore supervised machine learning methods for analysis. The annotations act as ground truth to build and improve automatic extractors for different melodic and rhythmic descriptors. Multitrack audio recordings are rarely available in MIR datasets. Hence, they are valuable resources when performing additional tasks that might be difficult or impossible in stereo mixes. For example, predominant melody extraction is much easier when a monophonic vocal track is available, as opposed to a stereo mix with another melodic instrument, and onset detection of percussion strokes will likely work better on drum tracks. The availability of multitrack audio in the Saraga Carnatic music collection is the first of its kind in IAM and will enable novel research directions such as melody extraction, source separation, automatic mixing, and performance analysis. The well-curated editorial metadata in the collections are useful for semantic analysis of different entities in editorial metadata, such as artists, compositions, forms, and relationships between them. The size of the collections allows the inference of statistically valid insights into music and aid in computational musicology. In addition, the collections can be useful to music listeners by providing an enriched music listening experience.

The Saraga collections were conceived to be a living collection that continues to evolve in current socio-cultural contexts and grow in quantity, quality, and comprehensiveness of metadata and annotations. To facilitate this goal, we provide a mechanism for the research/music community to contribute to the collections with more audio, manual annotations, or algorithmic extractors for automatic annotations. It is envisioned that Saraga datasets are adopted by the community at large for MIR with Indian Art Music.

## INDIAN ART MUSIC

In the context of this paper, IAM refers to two art music traditions of the Indian subcontinent: Carnatic music, widespread in the southern regions of the Indian subcontinent (South India and Sri Lanka), and Hindustani music (or North Indian Classical music), prominent in the northern and central regions of India, Pakistan, Nepal, Afghanistan, and Bangladesh. IAM today is a confluence resulting from cultural interactions between the Persian, Greek, Arabic, Iranian and Indian cultures (Saraf, 2011).

While Hindustani music has a higher impact from influences external to the Indian subcontinent, Carnatic music is relatively insulated from those influences (Subramaniam, 1999). Both traditions are well studied with sophisticated and grounded music theory. IAM has a large audience, continues to evolve in the current socio-cultural context, and has attracted high interest from music scholarship. Different aspects of IAM history, theory, practice, and performance have been studied and documented in musicological and popular literature in both English and Indian languages (e.g., works by Bhatkhande, 1984, 1990; Sambamoorthy, 1998; Viswanathan et al., 2004; Trivedi, 2008; Raja, 2012; Ramanathan, 1999; Meer et al., 1980; Jha, 2001; Deshpande, 1989; Danielou, 2010; Krishna, 2013 provide a non-exhaustive sampling of current and past literature). The presence of a large, dedicated audience and research literature forms a solid basis for studying these music cultures from both musicological and computational perspectives.

Both Hindustani and Carnatic music is performance-oriented, heterophonic, with the main melody being sung or played by the lead artist and strongly improvisatory in nature. Vocal music is dominant in both traditions but more in Carnatic music. A typical arrangement in a performance of IAM consists of a lead performer, a rhythm accompaniment (typically mridangam and tabla in Carnatic and Hindustani music, respectively), a constant sounding drone in the background and an instrumental melodic accompaniment. In both traditions, rāga is the melodic framework, and tāla is the rhythmic framework. Rāga (melodic framework) and tāla (rhythmic framework) concepts have been largely discussed in a multitude of studies and will not be reviewed in detail here (Bagchee, 1998; Sambamoorthy, 1998), but it is emphasized that these concepts are fundamental in the analysis of Indian art music.

IAM is taught orally through a lineage of teachers and students. While a skeletal prescriptive notation based on an Indian solfège (called the sargam) exists, the notation is not standardized and mainly serves as a mnemonic aid for melodies and lyrics and is not directly used during a concert performance. The improvisatory nature and the absence of descriptive music scores are significant contributing factors to the recognizable stylistic differences that exist between different interpretations of a composition, melody, or a musical concept.

Ganguli and Rao (2018) summarize the rāga as being somewhere between a scale and tune, providing a grammar that specifies tonal material, tonal hierarchy, and a set of characteristic melodic phrases. Every rāga has a set of characteristic melodic phrases that act as building blocks to construct melodies and provide a base for artists to express their creativity through improvisation within the rāga grammar. A typical

concert starts with the rendition of characteristic phrases of the rāga being performed. Musicians and musicologists often consider that a rāga can only be learned by getting familiar with landmark compositions and hence, the typical phrases in the rāga. Characteristic melodic phrases are also the most prominent cues used by human listeners for identifying rāgas (Krishna & Ishwar, 2012). Annotated characteristic melodic phrases in an audio recording of IAM is hence an important descriptor for melodic analysis.

Rhythmic organization in IAM relies on the tāla framework, which comprises fixed-length hierarchical time cycles. The instant of the beginning of each tāla cycle (also the end of the previous) is referred to as sama (or sam in Hindustani music), and it is highly significant structurally, marking time boundaries of important melodic and rhythmic events. The sam frequently marks the coming together of the rhythmic streams of soloist and accompanist and the resolution point for rhythmic tension (Clayton 2008, p. 81). Sama annotations are hence valuable for rhythmic and metrical analysis of IAM recordings.

## DATA COLLECTION AND CONTENT

Many types of musically relevant digital artifacts are used in MIR. The most commonly used ones are music scores, digital audio recordings of musical performances, and music-related texts. Given that an audio recording of a music piece is the central element of music consumption, distribution, and description, the Saraga collections are organized in the same fashion. The meta-information for the audio recording is often as important as the recording itself since it helps to build interlinking relationships between the recording, musical concepts, and musicians. The collections hence comprise albums/concerts with audio recordings, along with the accompanying metadata and annotations for each recording. Music-related texts or scores are not included.

Hindustani music is spread over a wider geographical region. As a result, it is much more diverse and heterogeneous than Carnatic music. Several forms and styles exist, among which the khyal form is the most frequently performed. The Saraga dataset includes Hindustani recordings in the khyal form. In Carnatic music, a concert is the natural unit of music performance (which is also the preferred unit for organization and digital distribution). Carnatic music concerts typically comprise around ten music pieces involving performances based on compositions. The Carnatic collection of the Saraga dataset is also organized in a similar fashion. While both the music traditions have a rich instrumental repertoire, Saraga collections focus on and comprise only vocal performances.

Saraga collections have been designed under the principles outlined by Serra (2014). Throughout the CompMusic project, several challenges and opportunities were identified in MIR for Indian art music, which provided indicators of necessary metadata and annotations useful for computational research. Some of the most relevant annotations have been carried out with the purpose of developing datasets for a variety of well-defined melodic and rhythmic analysis tasks targeted in the CompMusic project. While the Saraga collections are envisioned to grow in both quality and quantity through community contribution, the current state of the collection includes carefully curated content described in the following subsections.

### Audio Data

Audio recordings for Hindustani music were collected directly either from the lead artists/vocalists or, in some cases, from the institutes that reserve the rights for the distribution of the recorded material. In Hindustani music, in the majority of the cases, an audio release comprises a compilation of different rāga renditions by an artist. The recordings are a combination of live performances and studio recordings. In the case of Carnatic music, the audio recordings were procured from Arkay Convention Center[13], an institute that hosts and records professional concerts and reserves distribution rights for the audio recordings. None of the audio recordings in the collection were recorded for the sole purpose of using them in the Saraga collections.

The artists voluntarily contributed their recordings to the Saraga collection, and no additional compensation was provided beyond what they had already received for their concerts or recordings. Due permissions were taken from artists and the licensees for all the audio recordings to store, process, and distribute them for research purposes. The artists in the Carnatic music collections are contemporary professional musicians, while the Hindustani collection comprises contemporary artists and recordings of yesteryear maestros. The audio collection contains mp3 files (128 kbps stereo mixes sampled at 44.1 kHz) for ease of transmission and storage.

A part of the Carnatic collection was provided as raw multitrack audio, which was later processed by an audio professional to obtain the stereo mix versions. When present, the multitrack recordings (also stored as 128 kbps mp3 files sampled at 44.1 kHz) are also a part of the collection and available for use in addition to the stereo versions. The number of tracks in each multitrack concert varies with the number of artists in the concert. Most of the multitrack recordings in the collection comprise four tracks - vocal, violin, mridangam left, mridangam right, with occasional additional tracks comprising a second vocal track and other accompanying instruments. Since some music pieces in a concert use only a subset of instruments (e.g., only the lead vocal and violin tracks would be active during a melodic improvisation piece), the active tracks in multitrack audio for a particular piece of music has to be inferred from the instrument relationship metadata.

## Editorial Metadata

Relevant editorial metadata for each release is obtained either from its cover-art if present or by consulting the artists themselves. The metadata comprises the name, composition, composer, names of the artists involved in the recordings and their roles, information about the rāga, tāla, and form.

Metadata for each recording in the collection is stored in MusicBrainz [14]. Storing metadata in MusicBrainz allows us to uniquely identify a recording (with a MusicBrainz Identifier - an MBID) and each piece of metadata associated with the recording while allowing for ontological exploration (Koduri, 2016) of the metadata in the collections. All the metadata is accessible through the MusicBrainz API and can be edited and supplemented on the MusicBrainz website, thus allowing for community contribution and improvement.

## Annotations

Annotations are musically relevant labels particularly useful for computational methodologies targeted towards analyzing a specific musical concept. The annotations are useful in various ways. Many musicological studies focus on a specific aspect of a performance and study the relationships between the observed characteristics and the music theory concepts or cultural context. Such studies often rely on annotations of the performance in terms of relevant descriptors. For example, Clayton (2007) studied the entrainment phenomena in IAM, making use of statistical analysis of beat annotations together with observational analysis of the video of the performance.

Music exploration is an emerging music technology field that targets facilitating users' access to cultural content in an engaging way. Annotations are particularly useful for developing exploration tools that enrich the listening experience via visual representation of annotations. Such tools facilitate relating musical concepts (such as rhythmic cycles or sections of a form) with the actual performance and the option of browsing using musical concepts.

Data-driven information extraction from digital artifacts of a culture is a research methodology gaining high popularity, thanks to advances in computing, machine learning, and an increase in machine-readable data resources. One of the standard approaches in that field is the application of computational methods to model (manually or automatically extracted) features of large amounts of data. High-quality manual annotations serve as ground truth for supervised automatic analysis tasks and enable us to apply machine learning techniques. One example is our previous work (Srinivasamurthy et al., 2017), focusing on meter analysis of IAM. For such studies, large sets of annotated data are a basic requirement.

Manual annotation by experts is difficult, time-consuming, and hence costly. In the CompMusic project, several datasets for IAM containing manual annotations have been collected and reviewed in recent theses (Gulati, 2016; Koduri, 2016; Srinivasamurthy, 2016). For multiple MIR tasks and music concepts, there exist methods for automatic analysis that can provide reliable automatic annotations for a music piece. The Saraga collections contain both manual and automatic annotations for melody, rhythm, and structure. We further describe the annotations in two subsections: manual annotations by experts and automatically extracted descriptors computed using state-of-the-art analysis tools developed by the CompMusic research team.

## Manual Annotations

Many of the recordings in the Saraga collections are accompanied by various manual annotations done and verified by experts. Professional musicians and collaborators of the CompMusic project Vignesh Ishwar [15]

and Kaustuv Kanti Ganguli [16] verified the manual annotations in Carnatic and Hindustani music collections, respectively.

#### SECTION ANNOTATIONS

Every musical performance has an overall structure that is an organized sequence of sections. Each section defines some uniformity of the rhythmic, melodic, lyrical, or harmonic organization depending on the musical context. Characteristic variations of specific aspects in a performance (for example, tempo changes) are often synchronized with the sections. Hence, segmenting a performance into sections is a prior step in a large number of analytical studies.

The sections in IAM define the underlying rhythmic, melodic, or lyrical organization of a performance (Clayton 2008). Hindustani music pieces are lengthy and comprise sections that are performed often in different tempo and rhythmic structures. Since most Carnatic music pieces are performances of compositions (with elaborate improvisations), the sections are a combination of lyrical sections of the corresponding composition and the improvised sections of the corresponding form. The Saraga collections include time-aligned manual annotations of sections stored as start and end timestamps, together with the section's name for each audio recording. Further, when section boundaries are accompanied by rhythmic and melodic transitions, the timestamps of these events are aligned accurately and named with the section (e.g., the name of the *tāl* of a section in Hindustani music).

#### SAMA AND TEMPO ANNOTATIONS

The sama annotations are particularly useful for studies focusing on cycle level rhythm and meter analysis, as done by Srinivasamurthy et al. (2016, 2017). They can be used as a pre-processing step for further structural and melodic analysis of the music piece. The time-aligned sama annotations in the Saraga collections were obtained in a semi-automatic method. State-of-the-art meter analysis algorithms for IAM (Srinivasamurthy, 2016) were used to extract sama candidates for the music pieces, which were then manually corrected and verified by professional Hindustani and Carnatic musicians using Sonic Visualizer [17] (Cannam, 2010). The sama annotations are stored as a sequence of timestamps aligned with the audio recording.

With the *tāla* information (and hence the rhythmic structure of the piece) from metadata, we can derive the tempo of the piece from the sama annotations. We extracted a time-varying average beat interval by dividing the time interval between consecutive sama instants by the number of beats in the cycle. Dividing 60 by the beat interval, we obtain the more familiar tempo measure in beats per minute (BPM). The median tempo of each section of the piece is computed and stored, along with *tāla* related metadata for each piece. To summarize, the tempo annotations accompanying an audio recording have the average inter sama interval (cycle length), tempo in BPM, and length of the cycle in beats for each section of the music piece, identified by start and end timestamps.

#### CHARACTERISTIC MELODIC PHRASE ANNOTATIONS

Gulati (2016) summarizes a melodic phrase as “a unit of melody that encapsulates an idea or a musical thought by an artist” (p. 15). Phrase annotations can serve as an excellent resource for musicologists studying melodic patterns in the *rāga* framework and modeling melodic ornaments, tuning, and style in rendering melodic phrases.

The phrase annotations for both the music traditions in Saraga collections were done by professional musicians using Sonic Visualizer. They listened to the entire audio recordings. While listening, whenever they identified a melodic phrase characteristic of either the composition in the recording or of the underlying *rāga*, they marked the timestamps of the phrase and labeled it using Indian solfège symbols (Gulati, 2016, p. 241). The phrase annotations include a number of phrases per recording with the timestamps, annotation of the phrase, and a flag that indicates if the phrase is representative of the composition or of the *rāga* (1: representative phrase of the composition in the recording, 2: representative phrase of a *rāga*, which also implies a representative phrase of the recording).

It is to be noted that there is no closed formal inventory of characteristic phrases of a *rāga*, and hence, the annotated phrases in the collections are not comprehensive. Furthermore, due to the laborious nature of phrase annotations, a relatively small proportion of a characteristic phrase in a recording was annotated.

However, the goal was to get sufficient manually annotated ground truth phrases available to aid in meaningful analysis or to bootstrap an automatic melodic phrase extractor.

### **Automatic Annotations**

Computational analysis applied to music makes use of some low-level acoustic features/descriptors automatically extracted from audio signals. It is a common practice to extract frequently used features and include them in data collections. This practice facilitates the implementation of algorithms aimed at higher-level modeling using these features. The following automatic annotations are obtained using the extractors in PyCompMusic tools, which internally use the open-source Essentia library[18].

#### **PREDOMINANT MELODY**

One of the most common descriptors in music analysis is the fundamental frequency (or pitch) of the lead voice, estimated on short frames of audio signals and largely used in melodic analysis tasks. For example, a melody in the context of IAM can be considered as a time series of pitch values corresponding to the lead artist in an audio recording.

We have included pitch series information estimated using the Melodia algorithm, a state-of-the-art melody extraction method proposed by Salamon and Gómez (2012), stored as a sequence of timestamps and pitch values in Hertz (Hz). When a vocal-only track is available as a part of a multitrack release in Carnatic music, a pitch series extracted from the vocal track is a more reliable estimate of predominant melody, and hence, is also stored.

#### **TONIC**

The tonic pitch is the base frequency that serves as a reference and foundation for melodic rendition and integration throughout a performance. In an IAM performance, tonic pitch is constantly reinforced by the drone sound in the background that is typically generated by the tanpura. In contrast to music cultures applying standardized tuning (with say, A4=440 Hz), tonic frequencies in Carnatic and Hindustani music recordings are set by the vocalists without any specific restriction. In IAM, tonic hence refers to a particular pitch value (in Hz) rather than a pitch-class.

Melodic analysis of IAM often involves a step of normalization with respect to the tonic pitch of the recording since all the tones of a melody are produced in reference to the tonic pitch. Thus, identifying the tonic pitch in the recording is a crucial first step in the melodic analysis of IAM. To facilitate melodic analysis, we automatically extracted tonic information using the algorithm by Gulati et al. (2014) and stored it as a part of the collection.

### **Content**

The audio recordings in the Saraga Carnatic and Hindustani collections are a part of the CompMusic Dunya collections. Compared to other CompMusic collections [19], the main differences are the open sharing license and the accompanying annotations. The Carnatic [20] and the Hindustani [21] CC collections in MusicBrainz provide the complete list of recordings in the Saraga collections. However, not all recordings are accompanied by all the possible metadata and manual annotations. The current content in the collections is summarized in Table 1 (and Table A1 in the Appendix). Statistics about the total size, coverage, and completeness of the collection after any future additions, will be updated on the companion website of Saraga collections. For the Carnatic collection, 36.3 hours (168 recordings, 19 concerts) of audio are available in multitrack, corresponding to about 69% of the total duration of the collection.

## **ORGANIZATION AND ACCESS**

The organization of the Saraga collections is based on the tenets of easy access for both human and machine consumption, open availability of data and code, reproducibility, easy adoption by research, musician and music listener community, and enabling contribution by the community to grow the dataset in quality, quantity, coverage, and completeness.

**Table 1.** The content and coverage statistics of the Saraga collections

<b>Content</b>	<b>Hindustani collection</b>	<b>Carnatic collection</b>
Total releases	36	26
Total recordings	108	249
Total artists (lead and accompanying)	36	64
Compositions	113	202
Unique rāga	61	96
Unique tāla	9	10
Total duration	43.6 hours	52.7 hours

The contents of the collections can be accessed through an open-source API for machine consumption. The API allows access to audio, metadata stored in MusicBrainz, and manual/automatic annotations. Alternatively, a publicly available repository with a complete dump of metadata and annotations is also available to enable versioning and community contribution. The extractors used to derive automatic annotations are available through PyCompMusic tools.

## Organization

The Saraga collections are grouped by music culture into Carnatic and Hindustani collections. Each collection comprises several albums/concerts with multiple recordings. Each recording is uniquely identified using the MBID of the recording, which can be used to obtain the editorial metadata of the recording from the MusicBrainz database (e.g., artist, release, year, lead instrument, rāga, tāla). Metadata is stored and accessed in the JSON format[22], which can be easily parsed (or directly read using standard decoders).

Each recording is associated with multiple annotations stored as files and identified by a file type identifier (slug) that can be used to access the file. The annotation files are classified into source files and derived files. Source files refer to typically audio (including multitrack) and manual annotations, while derived files refer to the automatic annotations obtained from automatic extractors. Currently, the collection comprises audio, multitrack audio, melodic phrases, sama, section, and tempo stored as source files for each recording, while pitch and tonic are stored as derived files. The format and description of each currently available source and derived files, along with the number of recordings with those accompanying files is detailed in Table A1 and A2 in the Appendix.

## Access: The PyCompMusic API

The PyCompMusic tools built as part of the CompMusic project comprise an API that can be used to access the content in the Saraga collections. The API is organized around an audio recording uniquely identified by its MBID and can be used to fetch metadata, source files, and derived files of a recording. The API also allows fetching some or all of the associated metadata and files in a collection. To further facilitate researchers' access to the database and preclude the need for a deep understanding of the PyCompMusic API, we share sample python scripts, notebooks, and other resources in the companion website [12] of the Saraga collections.

## Access: Github repository

In addition to the API access to the Saraga collections, we provide a version controlled publicly available git repository[23] with a complete dump of metadata and annotations. The git repository serves to provide a complete and up-to-date overview of the content of the dataset. In addition, it serves to create versioned

snapshots and subsets of the Saraga collections to build task-specific datasets for research experiments that could be shared within the community for reproducibility.

The repository is also a mechanism for the community to contribute additional manual annotations or improve upon the existing ones (e.g., add more characteristic melodic phrase annotations). Since audio and pitch annotations are large files not suitable for a git repository, we provide a tarball of audio and pitch annotations. However, the repository acts as a version control for those files by storing the MD5 checksum of the audio and pitch annotation files, which can be verified against the original files.

### **Contributing to the collection**

The research and music community are welcome and encouraged to contribute to the Saraga collections. Audio data contributions of high quality that can be released under CC licenses will be considered for addition to the collections.

The editorial metadata for the audio recordings needs to be added to MusicBrainz so that we generate unique MBIDs for each recording, which can then be used to add manual or automatic annotations to the recording. New manual annotations or updates to the existing ones could be submitted through a pull request to the git repository, which will be merged after due verification by community experts. Automatic annotations can be added to the recordings in the collection by writing an extractor in PyCompMusic tools, which will then extract automatic annotations to be added as derived files. Contributors to collections will be duly credited and acknowledged in the collections' documentation. Additional information for contributors can be found on the companion website.

## **APPLICATIONS**

In addition to the MIR and computational musicology research problems described in previous sections that could potentially use the Saraga collections, we present other applications using the collections. We present two examples that use Saraga collections for different applications in music education, understanding, exploration, and discovery.

### **Musical Bridges**

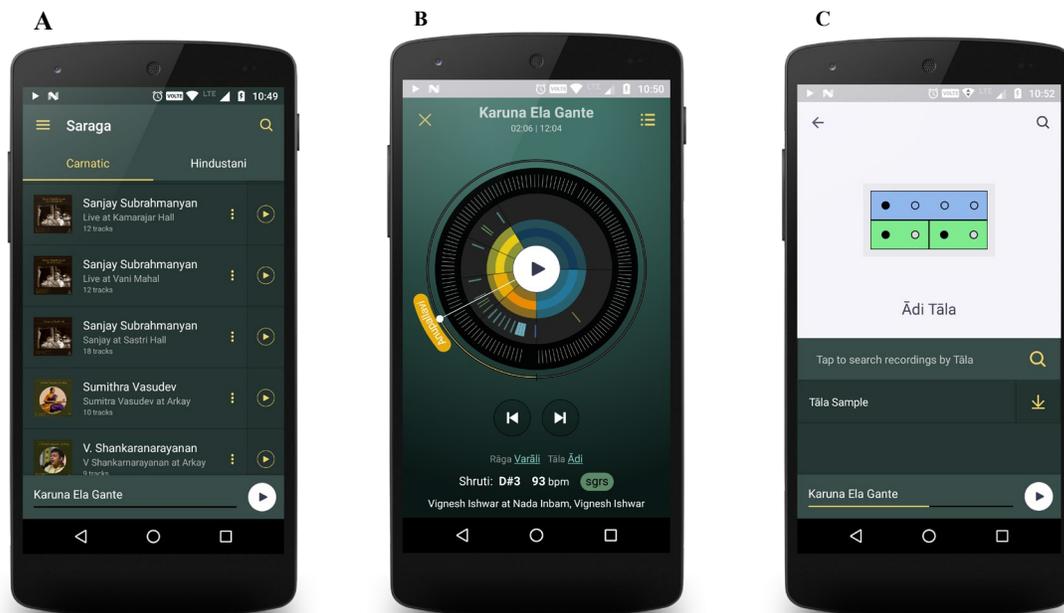
The Saraga collections offer great potential for music education. The Musical Bridges project [24] aims to bring together different cultures through music understanding. To that aim, it develops computational tools to be used in all sorts of educational settings to facilitate the comprehension by the student of some of the key elements of selected music traditions, including concepts such as *tāla*, *rāga*, *ṣaḍja* (tonic) or characteristic melodic phrases from Hindustani and Carnatic art music. The tools offer interactive visualizations synchronized to recordings from the Saraga collections and draw on their corresponding annotations, with the aim of focusing the listening attention of the student to those key elements.

### **The Saraga App**

While the Saraga collections are useful for corpus studies in musicology or for MIR tasks, the datasets are large and comprehensive enough to be used in the context of music exploration too. For demonstrational purposes, we have built an android mobile app that is a music exploration tool and provides an enhanced listening experience with a subset of Saraga Hindustani and Carnatic collections.

The app demonstrates a culturally relevant approach to browse and navigate through an audio collection of IAM using metadata such as *rāga*, *tāla*, and form labels (which are often missing in album covers but are available in our dataset), and the annotations. The app visualizes time-aligned expert manual annotations synchronized with the streamed audio. In addition, information on different IAM concepts, such as *rāgas* and *tālas*, along with audio examples, is provided on the app, making the app a useful tool to learn and explore IAM. The user can easily access different realizations of a particular musical concept (such as typical phrases of a *rāga*) simply by clicking a visual display representing the concept.

In Figure 1, we provide some screenshots of the application. The app is free, and the readers are invited to download and try the app through the link on the companion website.



**Fig. 1.** The Saraga application (screen views). A: recording list view for the two collections, B: representation for a recording (displays structural, rhythmic, and melodic annotations), C: theoretical representation of a tāla.

## SUMMARY

Recently, there have been many contributions in developing open datasets for music analysis and exploration. However, most resources created are restricted in one way or another. Datasets containing annotations often refer to inaccessible audio, or when audio is available, they are often short snippets that fail to represent a real-world scenario. The scarcity of large and well-annotated open music collections continues to hinder reproducible research.

With this paper, we described the largest open-access annotated data collections for Indian Art Music. We also provide a companion website and example scripts to facilitate access and use of the data by other researchers. The audio in the datasets are representative of current performance practices and are released under Creative Commons, enabling reproducible research with the audio data. Further, a part of the Carnatic music collection is available in multitrack and provides a rare opportunity for researchers to analyze interactions between the tracks while also enabling more accurate analyses and extraction of melodic and rhythmic descriptors.

The annotations provided in the Saraga datasets have the potential to serve many research tasks in MIR and computational musicology. For example, the melodic and rhythmic annotations in the datasets are particularly useful for studies focusing on meter tracking, structural segmentation, pattern similarity, pattern detection, and pattern discovery tasks. These tasks are amongst the most popular and relevant in the MIR domain.

Metadata stored in an open encyclopedia such as MusicBrainz in linked form is of great use to study relationships such as stylistic similarities, time evolution of specific characteristics (using date information), and accessing other linked resources (such as Wikipedia) to build knowledge bases (Koduri, 2016). Open metadata is also preferred for enabling the extension of its size through collaborative efforts of the MusicBrainz community. Apart from artist and album information, the metadata we have curated involves tags for rāga, tāla, and form. This categorical information is a basic need in most computational studies but yet, hardly accessible in most other corpora.

The audio recordings in the collections are organized into source and derived files, with version control and easy access through the PyCompMusic API and a git repository. The git repository allows for a complete, up-to-date view of the data collection and provides version control for audio, metadata, manual

and automatic annotations. Envisioned as a living and growing collection, we have provided mechanisms to contribute open audio and annotations to the collections. Both the research and music community are encouraged to use the datasets and contribute to it with more audio, better manual annotations, and algorithms for extraction of automatic annotations.

Other than research applications, we demonstrated two use cases that utilize the datasets for music understanding and exploration. Musical Bridges provides interactive tools for understanding and appreciating Hindustani music. The Saraga app is a mobile application for music exploration that streams audio and visualizes the annotations to help users view specific instances of particular aspects for IAM (e.g., typical melodic phrases of a rāga). The Saraga app can be used by music listeners for an enhanced experience when interacting with the Saraga music collections.

Open datasets are important for conducting collaborative work between different disciplines as they serve as resources collected collaboratively, adding different layers of annotations and information from different disciplines to enrich the data. Such open datasets are one of the needs of interdisciplinary fields of music analysis where annotations/labels by musicologists, musicians, and engineers are required to build large databases indispensable for data-driven knowledge extraction. We sincerely hope researchers from many different disciplines will use the Saraga collections to trigger collaborative interdisciplinary studies.

### ACKNOWLEDGMENTS

This work has been funded by the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ERC grant agreement 267583, The CompMusic Project (<http://compmusic.upf.edu>). The authors thank Baris Bozkurt for his inputs on an earlier version of the manuscript; music experts Vignesh Ishwar and Kaustuv Kanti Ganguli for their help in annotating Saraga collections; Alastair Porter and Gopala Krishna Koduri for their support in setting up the infrastructure for storing and accessing Saraga collections; Arkay Convention Center (Chennai, India), Shrutinandan (Kolkata, India), ITC-SRA (Kolkata, India) and all the musicians who have contributed audio recordings to the Saraga collections. This article was copyedited by Annaliese Micallef Grimaud and layout edited by Diana Kayser.

### NOTES

[1] Correspondence can be addressed to: Ajay Srinivasamurthy ([ajays.murthy@upf.edu](mailto:ajays.murthy@upf.edu)) or Xavier Serra ([xavier.serra@upf.edu](mailto:xavier.serra@upf.edu)), Music Technology Group, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain.

[2] Ajay Srinivasamurthy is currently with Amazon Alexa, India and contributed to the work described in the paper during his PhD at Music Technology Group, UPF, Barcelona before joining Amazon.

[3] Sankalp Gulati ([sankalp.gulati@upf.edu](mailto:sankalp.gulati@upf.edu)) is currently with Synaptic, India and contributed to the work described in the paper during his PhD at Music Technology Group, UPF, Barcelona.

[4] Rafael Caro Repetto ([rafael.caro@upf.edu](mailto:rafael.caro@upf.edu)) is currently with the Institute of Ethnomusicology, Kunstuniversität Graz, Austria.

[5] FAIR Principles: <https://www.go-fair.org/fair-principles/>

[6] CompMusic project website: <https://compmusic.upf.edu>

[7] Different research corpora built within the CompMusic project: <https://compmusic.upf.edu/corpora>

[8] Annotated CompMusic datasets: <http://compmusic.upf.edu/datasets>

[9] Dunya: <https://dunya.compmusic.upf.edu/>

- [10] PyCompMusic tools to organize, access and analyze Dunya and Saraga collections: <https://github.com/MTG/pycompmusic>
- [11] Creative Commons Licenses: <https://creativecommons.org/licenses/>
- [12] Companion website of Saraga collections: <https://mtg.github.io/saraga>
- [13] Arkay Convention Center: <http://www.arkayconventioncenter.com/>
- [14] MusicBrainz, a community maintained open encyclopedia of music metadata: <https://musicbrainz.org/>
- [15] Vignesh Ishwar: <https://www.facebook.com/vigneshishwarofficial/>
- [16] Kaustuv Kanti Ganguli: <https://kaustuvkanti.webs.com/>
- [17] Sonic Visualizer, a tool for visualisation, analysis, and annotation of music audio recordings: <https://www.sonicvisualiser.org/>
- [18] Essentia is an open-source library and tools for audio and music analysis: <https://essentia.upf.edu/>
- [19] CompMusic collections on MusicBrainz: <https://musicbrainz.org/user/compmusic/collections>
- [20] Saraga Carnatic collection on MusicBrainz: <https://musicbrainz.org/collection/a163c8f2-b75f-4655-86be-1504ea2944c2>
- [21] Saraga Hindustani collection on MusicBrainz: <https://musicbrainz.org/collection/6adc54c6-6605-4e57-8230-b85f1de5be2b>
- [22] JSON format: <https://www.json.org>
- [23] Saraga Git repository: <https://github.com/MTG/saraga>
- [24] Musical Bridges project page: <https://www.upf.edu/web/musicalbridges>. The interactive tools for the aid of music understanding and appreciation, including those related to Indian Art Music, can be found in the Resources section: <https://www.upf.edu/web/musicalbridges/recursos>.

## REFERENCES

- Bagchee, S. (1998). *Nād: understanding raga music*. Business Publications Inc.
- Bertin-Mahieux, T., Ellis, D., Whitman, B., & Lamere, P. (2011). The million song dataset. In *Proceedings of the 12th International Society for Music Information Retrieval Conference* (pp. 591–596). Miami, USA: ISMIR. <https://doi.org/10.7916/D8NZ8J07>
- Bhatkhande, V. N. (1990). *Hindustani Sangeet Paddhati: Kramik Pustak Maalika*, Vol. I-VI, Sangeet Karyalaya.
- Bhatkhande, V. N. (1984). *Music Systems in India*, Ed. S. N. Ratanjankar, S. Lal & Co Delhi.
- Cannam, C., Landone, C., & Sandler, M. (2010). Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files. In *Proceedings of the 18th ACM International Conference on Multimedia* (pp. 1467–1468). Florence, Italy. <https://doi.org/10.1145/1873951.1874248>

- Clayton, M. (2007). Observing entrainment in music performance: Video-based observational analysis of Indian musicians' tanpura playing and beat marking. *Musicae Scientiae*, 11(1), 27. <https://doi.org/10.1177/102986490701100102>
- Clayton, M. (2008). *Time in Indian Music: Rhythm, Metre and Form in North Indian Rag Performance*. Oxford, UK: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195339680.001.0001>
- Danielou, A. (2010). *The ragas of Northern Indian music*, Munshiram Manoharlal Publishers, New Delhi.
- Deshpande, V. H. (1989). *Indian Musical Traditions: An Aesthetic Study of the Gharanas in Hindustani Music* (2nd ed.), Popular Prakashan.
- Ganguli, K.K. & Rao, P., (2018). On the Distributional Representation of Ragas: Experiments with Allied Raga Pairs. *Transactions of the International Society for Music Information Retrieval*, 1(1), pp.79–95. <https://doi.org/10.5334/tismir.11>
- Gulati, S., Bellur, A., Salamon, J., Ranjani, H. G., Ishwar, V., Murthy, H. A., Serra, X. (2014). Automatic tonic identification in Indian art music: approaches and evaluation. *Journal of New Music Research*, 43(1), 53–71. <https://doi.org/10.1080/09298215.2013.875042>
- Gulati, S. (2016). *Computational Approaches for Melodic Description in Indian Art Music Corpora*. PhD dissertation. Music Technology Group, Universitat Pompeu Fabra. (<https://www.tdx.cat/handle/10803/398984>)
- Janakiraman, S. R. (2008). *Essentials of musicology in south Indian music*, Indian Music Publishing House, Chennai.
- Jha, R. (2001). *Abhinav Geetanjali Vol. I-V*, Sangeet Sadan.
- Koduri, G. K. (2016). *Towards a multimodal knowledge base for Indian art music: A case study with melodic intonation*. PhD dissertation. Music Technology Group, Universitat Pompeu Fabra. <https://www.tdx.cat/handle/10803/402439>
- Krishna, T. M. (2013). *A Southern Music: The Karnatik Story*, Harper Collins.
- Krishna, T. M. & Ishwar, V. (2012). Carnatic music: Svara, Gamaka, Motif and Raga identity. In *Proceedings of the 2nd CompMusic workshop* (pp. 12–18) Istanbul, Turkey. <https://repositori.upf.edu/handle/10230/20494>
- Meer, W. V. D. (1980). *Hindustani music in the twentieth century*, Springer.
- Porter, A., Sordo, M., & Serra, X. (2013). Dunya: A System for Browsing Audio Music Collections Exploiting Cultural Context. In *Proceedings of the 14th International Society for Music Information Retrieval Conference* (pp. 101–106). Curitiba, Brazil: ISMIR. <https://doi.org/10.5281/zenodo.1417355>
- Porter, A., Bogdanov, D., Kaye, R., Tsukanov, R., Serra, X. (2015). Acousticbrainz: a community platform for gathering music information obtained from audio. In *Proceedings of the 16th International Society for Music Information Retrieval Conference* (pp. 786–792). Malaga, Spain: ISMIR. <https://doi.org/10.5281/zenodo.1414938>
- Raja, D. S. (2012). *Hindustani Music Today* (1st ed.), D.K. Printworld (P) Ltd.
- Ramanathan, N. (1999). *Musical forms in the Sangita Ratnakara*, Sampradaya, Chennai.

- Salamon, J., & Gomez, E. (2012). Melody Extraction from Polyphonic Music Signals Using Pitch Contour Characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6), 1759–1770. <https://doi.org/10.1109/TASL.2012.2188515>
- Sambamoorthy, P. (1998). South Indian Music Vol. I-VI. The Indian Music Publishing House.
- Saraf, R. (2011). Development of Hindustani Classical Music (19th & 20th centuries) (1st ed.). Vidyanidhi Prakashan.
- Serra, X. (2011). A multicultural approach to music information research. In *Proceedings of the 12th International Society for Music Information Retrieval Conference* (pp. 151–156). Miami, USA: ISMIR. <https://doi.org/10.5281/zenodo.1416592>
- Serra, X. (2014). Creating Research Corpora for the Computational Study of Music: the case of the CompMusic Project. In *AES 53rd International Conference on Semantic Audio*. London, UK: AES.
- Serra, X. (2017). The computational study of a musical culture through its digital traces. *Acta Musicologica*, 89(1), 24–44.
- Srinivasamurthy, A., Koduri, G. K., Gulati, S., Ishwar, V., & Serra, X. (2014). Corpora for Music Information Research in Indian Art Music. In *Proceedings of the Joint International Computer Music Conference/Sound and Music Computing Conference*. Athens, Greece. <https://doi.org/10.5281/zenodo.850759>
- Srinivasamurthy, A., Holzapfel, A. T. Cemgil, and X. Serra. (2016). A generalized Bayesian model for tracking long metrical cycles in acoustic music signals. In *Proceedings of the 41st IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 76-80). Shanghai, China. <https://doi.org/10.1109/ICASSP.2016.7471640>
- Srinivasamurthy, A. (2016). A Data-driven Bayesian Approach to Automatic Rhythm Analysis of Indian Art Music. PhD Dissertation. Music Technology Group, Universitat Pompeu Fabra. <https://www.tdx.cat/handle/10803/398986>
- Srinivasamurthy, A., Holzapfel, A., Ganguli, K. K., & Serra, X. (2017). Aspects of Tempo and Rhythmic Elaboration in Hindustani Music: A Corpus Study. *Frontiers in Digital Humanities*, 4, 1–16. <https://doi.org/10.3389/fdigh.2017.00020>
- Subramaniam, L. (1999). The reinvention of a tradition: Nationalism, Carnatic music and the Madras Music Academy, 1900–1947. *Indian Economic & Social History Review*. 36 (2): 131–163. <https://doi.org/10.1177/001946469903600201>
- Trivedi, R. (Ed.) (2008), Bharatiya Shastriya Sangit: Shastra, Shikshan Va Prayog, Sahitya Sangam, Allahabad, India.
- Tzanetakis, G. (2014). Computational ethnomusicology: a music information retrieval perspective. In *Proceedings of the Joint International Computer Music Conference/Sound and Music Computing Conference* (pp. 112-117). Athens, Greece.
- Viswanathan, T. & Allen, M. H. (2004), Music in South India, Oxford University Press.
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>