Chapter 1

Getting Started on Computational Musicology and Music Information Research: An Indian Art Music Perspective

Ajay Srinivasamurthy[(a),1], Sankalp Gulati[(a),2], Rafael Caro Repetto[(b)], and Xavier Serra[(a)]

[(a)]Music Technology Group, Universitat Pompeu Fabra (UPF), Barcelona, Spain
[(b)]Institute of Ethnomusicology, Kunstuniversität Graz, Austria

Abstract

Culture-aware and culture-specific approaches to computational musicology and music information research (MIR) have been shown to be effective for analysis of a music culture. Recent efforts as a part of the CompMusic project argued that it is essential to consider the sociocultural specifics of a music tradition to effectively define research problems, collect data and propose methods for analysis. The project also demonstrated the use of such approaches, leading to a collective body of work for MIR in Indian art music. However, it is a considerable effort to define relevant research tasks, collect data and develop specific methods for analysis for each music culture, which often poses a significant entry barrier to start work in the field. One approach in such a scenario is

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

[2]Sankalp Gulati is currently with Eka.care, India and contributed to the work described in the chapter during his PhD at Music Technology Group, UPF, Barcelona.

to seek and identify parallel tasks, data and methods from the current state of the art in other music cultures and use them for a preliminary and basic analysis of culture-specific tasks, extending them with culture-specific methods to be more effective and relevant. While it is a sub-optimal compromise, such a perspective will enable preliminary analysis of a music culture using existing methods, and integrate it as a use-case with existing common frameworks and approaches in MIR and data-driven computational musicology. In this chapter, we aid such an approach by describing common concepts, frameworks, approaches, resources, data and methods for computational analysis of music from a perspective that could be useful for the analysis of Indian Art music. With this perspective, it is hypothesized that the currently established methods and tools, along with the datasets and corpora built within the CompMusic project will encourage accelerated research into different research problems relevant to Indian art music. The content of the chapter is targeted at musicians, music students, technology enthusiasts, engineers and researchers to provide them a context of the current state of the art that could help them start their work with computational analysis of Indian art music.

1.1 Introduction

Recent advances in signal processing and machine learning have provided us with computational tools and methods for complex automatic analysis. In the process however, recent machine learning methods and efforts have shifted their focus on applications and tools, rather than the underlying data. Recent approaches consider data as one of the inputs and aim to learn any complex relationship as long as data is annotated and labeled with useful metadata. Such an approach assumes that the research problems being addressed are well-defined, the data labels are accurate and objective, and the users (both producers and consumers) of such data and systems agree on the outputs of the system.

The availability of music in the form of audio enables us to take up a data-driven approach to Music Information Research and computational musicology. MIR aims to build automatic analysis methods to extract musically relevant events, metadata, labels and tags from audio. The applications include music discovery and enriched listening of music, complemented by metadata derived from the content of the music. Data-driven computational musicology aims to analyse large corpora of music by means of newly built computational tools. These tools for musicologists aim to complement manual analysis of carefully curated music pieces to larger corpora to yield valid statistical analyses that support and extend musicological hypotheses with quantitative data.

Music is a sociocultural phenomenon that poses limitations to a tool-driven approach by breaking some of the assumptions the existing modeling tools make. Any research problem posed on music needs to consider the sociocultural aspects of music to arrive at a well-defined problem that is musically relevant, grounded in theory and has a real world application for both producers and consumers of music. It is also necessary to understand the nuances of the problem and hence provide suitable approaches to solve them. Each music culture poses some unique problems that are not shared by others due to differences and nuances particular to the music culture. These differences are not deviations from a standard, but integral part of the music performance practice. A comprehensive analysis of music needs to be aware of such differences and nuances for a complete description of the music culture. Hence research problems need to be culture-specific or at least be culture-aware for a musically well-defined analysis of music. Even when the music concepts are well-defined in theory and well-agreed by experts, converting musical concepts to well-formed engineering definitions is a complex task.

While there is growing availability of music to be consumed on demand, most of that data is poorly labeled, or labeled only with editorial metadata, which is not useful for content based analysis. Manually annotating music audio with useful content metadata is an intensive process that needs music experts and involves a fair amount of subjectivity depending on the semantic level of music concept being annotated. Some basic music concepts are labels attached to a piece of music and hence clearly defined in the music theory of a music culture (at least in music traditions that have significant music theory literature), e.g. a song is in C Major or that a song is in rāg bhairav , while some concepts are complex descriptions and interpretations of music performers/listeners, which are hard to define and harder to codify into a common annotation schema, e.g., "a song is happy", "this song is the best example of rag bhairav". Automatic analysis methods hence tend to build approaches to analyze, extract and describe well-defined music concepts from audio. The lack of annotated data due to the effort involved, and a lack of agreement on certain subjective concepts in music make it especially hard to obtain well-labeled audio data for analysis. Beyond the approaches that only need some basic metadata, any meaningful analysis in MIR will need to be considered as a small-data supervised learning problem at the moment.

It is with this context that culture-aware and culture-specific approaches to building data, tools and methods have been proposed in the past to work on musically relevant research tasks with the involvement of the community that produces and consumes the specific music culture. Such approaches to computational musicology and music information research (MIR) have also been shown to be effective for analysis of a music culture. A clear case for culture-specific methods is the CompMusic (Computational Models for the Discovery of the World's Music) project [Ser11], which was a research project funded by the European Research Council from 2011 to 2017 and coordinated by Xavier Serra from the Music Technology Group of the Universitat Pompeu Fabra in Barcelona (Spain). It aimed to advance in the automatic description of music by emphasizing cultural specificity, carrying research within the field of music information processing with a domain knowledge approach. The project focused on five music traditions of the world: Hindustani (North India), Carnatic (South India), Turkish makam (Turkey), Arab-Andalusian (Maghreb, though the project focused on the Moroccan tradition), and Jingju (China).

CompMusic aimed to contribute to our multi-cultural society and focused on the advancement in the field of Music Information Research by approaching a number of current research challenges from a culture-specific perspective. CompMusic focused on the extraction of features from audio music recordings related to melody and rhythm, and on the semantic analysis of the contextual information of those recordings. The goal was to characterize culture-specific musical facets of each repertoire and to develop musically meaningful similarity measures with them.

Taking a data-driven approach to research, a major effort of CompMusic was to build a research corpus for each music tradition. The types of data gathered are mainly audio recordings and editorial metadata, which are then complemented with descriptive information about the items we have, and in some cases with music scores and/or lyrics. CompMusic focused on open research, with publications, code, and data available under open licenses. The project clearly demonstrated the use of culture-specific methods leading to a collective body of work for MIR in the music traditions considered within the project.

However, it is a considerable effort to define relevant research tasks, collect data and develop specific methods for analysis for each music culture. This often poses a significant entry barrier to start work on a music culture, especially in the absence of dedicated efforts like CompMusic. Even within the music cultures studied under CompMusic, only a handful of research problems could be solved within the duration of the project, while many identified research problems could not be addressed due to lack of resources.

A second approach in such a scenario for under-resourced music cultures is to seek and identify parallel concepts, tasks, data and methods from the current state of the art in other music cultures. Such parallels can be used for a preliminary and basic analysis of culture-specific tasks, which could then be extended to culture-specific methods to be more effective and relevant. While it seems like a sub-optimal compromise, such a perspective will reduce the entry barrier to define and take up research tasks in music cultures that have not been studied due to lack of resources. It will enable a preliminary analysis of a music culture using existing methods, identifying new challenges and limitations of current methods. It also puts the music culture into perspective within a larger body of MIR work, integrating it further as a use-case with existing common frameworks and approaches in MIR and data-driven computational musicology.

In this chapter, we take the second approach as applied to MIR and data-driven computational musicology tasks for Indian Art Music (IAM). We rely on our experience from the CompMusic project, where we build several culture-specific methods for the analysis of Indian art music. We identify parallel concepts and describe the available tools, techniques, frameworks and methods from the current state of the art for analysis of Indian art music. We also describe how in this process, Indian art music could be a use-case to test common frameworks and approaches in MIR. We aim to provide a series of perspectives on current approaches and methods that can be useful for analysis of Indian art music, while utilizing corpora and datasets that have been built for different MIR tasks in Indian art music.

Through this chapter, we wish to encourage MIR in Indian art music by providing parallel tasks and methods that could be used to bootstrap a new research task. The content of the chapter is targeted at musicians, music students, technology enthusiasts, engineers and researchers to provide them a context of the current state of the art that could help them start their work with computational analysis of Indian art music.

It is important to note that culture-aware approaches are necessary to consider the socio-cultural context for any music analysis task, as argued within the CompMusic project. The approach we take in this chapter seems contrary to that argument, but we introduce parallel tasks and generic tools and methods only for basic analyses. The parallels we draw and general approaches we present are limited in their scope, only provide limited equivalency, and don't capture the nuances necessary for in-depth analysis. These approaches are mainly intended for new researchers to start with existing tools to arrive at a basic approach that can then be built upon to solve a musically relevant problem.

We start with a brief introduction to Indian art music, followed by describing some parallels in music concepts between Indian art music and other music cultures that can help to identify and define parallel research tasks. We then describe the audio signal characteristics of Indian art music and list multiple research tasks in MIR and data-driven computational musicology that can benefit from existing methods. We then introduce the corpora and datasets available for analysis of Indian art music, followed by different tools and frameworks that can be used for analysis.

1.2 Parallels in Musical Concepts and Dimensions

While the issue of the existence of universal concepts in all of the world's music traditions is a matter of strong academic discussion and a common agreement is still far to be reached, we focus in this chapter on three musical dimensions, such as melody (including by extension harmony), rhythm and timbre. We further note that these dimensions are not independent and orthogonal, but a significant interplay exists across the dimensions that come together in music performance. Another exercise of identifying parallels is based on functional aspects of music, looking at how music is composed, learnt, performed and produced, and the tools/methods used therein. We will apply both these approaches in the chapter.

Since the focus of the chapter is to identify parallel tasks and tools useful for IAM analysis, it is necessary to find parallels with concepts that have extensive literature in the current state of the art. The current state of the art in MIR focuses mostly on Eurogenetic music, hence we look for parallels to analytical categories developed in European music theory and performance. This would span an entire gamut of musics commonly called western popular and classical music.

We reiterate that the parallel concepts identified here are not interchangeably equivalent due to the differences in nuances. The definitions across music culture extend only to a limited extent. In the process of describing parallels, we also identify some of the limitations of those parallel concepts, which will help us to identify the limitations of the methods that do an automatic analysis of these concepts. We start with a brief introduction to the main music concepts in Indian art music.

1.2.1 Introduction to Indian Art Music

A detailed introduction to Indian art music (IAM) is beyond the scope of the chapter [Bag98b; Sam98b]. However, we provide a gentle introduction to concepts that are necessary for the chapter. In the context of this chapter, 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 (also known as 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, and Indian cultures [Sar11] and is a well studied music tradition with sophisticated and grounded music theory. It has a large audience, continues to evolve in current sociocultural context and has attracted high interest from music scholarship. The presence of a large dedicated audience and research literature forms a solid basis for studying this music culture from both a musicological and computational perspective.

Both Hindustani and Carnatic music are performance oriented, heterophonic, with a main melody being sung or played by the lead artist, and mainly improvisatory in nature. Vocal music is dominant in both traditions but more in Carnatic music. Instrumental music exists in both traditions. It is important to note that in Carnatic music, the objective of instrumental musicians is to reproduce what a vocalist would produce, while in Hindustani tradition, even though it also uses the gāyaki (singing) style, instrumental music has developed a style distinct from vocal music.

A typical arrangement in a performance of IAM consists of a lead performer, a rhythm accompaniment (typically mridangam and tabla/pakhāvaj in Carnatic and Hindustani music, respectively), a constant sounding drone in the background by the tānpurā and instrumental melodic accompaniment (typically violin and harmonium or sāraṅgī in Carnatic and Hindustani music, respectively). In both traditions, rāga is the melodic framework and tāla is the rhythmic framework. Functional harmony, as defined for eurogenetic musics is not used in Indian art music.

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 [Bag98b; Sam98b], but it is emphasized that these concepts are fundamental in 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, it is not standardized and mainly serves as a mnemonic aid for melodies and lyrics, and is not directly used during a concert performance. As a result, there are recognizable stylistic differences between different interpretations of a composition, melody or a musical concept.

Further detailed descriptions of these music concepts are presented in other chapters of the monograph. We focus only on drawing parallels in the rest of the section, from the sole perspective of noting the possible extensions and limitations of these parallels to Indian art music.

1.2.2 Melody

A musical note, or a svara in IAM is a fundamental concept around melody. The idea of a svara is similar to the eurogenetic definition of a note, but differs in some aspects. The svara in IAM do not rely on absolute pitch positions, but are defined relative to a reference pitch frequency (called the ṣaḍja or ādhāra shadja, referred henceforth as tonic frequency) that is arbitrarily (or better described as conveniently) by the artists in a performance. All pitch positions and music scales are defined based on this reference pitch, making it a fundamental quantity to be estimated for any melodic analysis. The tonic is not based on a tuning reference unlike popular eurogenetic music (e.g. A4 = 440Hz), but rather based on the convenience of the lead artist in a performance.

Given a tonic, the pitch position of a note or a svara is defined based on the tuning system used. It's widely accepted that IAM uses just intonation, with all pitch positions defined in relation to the tonic frequency [KI12a]. However, a svara in IAM does not refer only to a discrete pitch position, but also includes its intonation, which broadly describes how the note is to be performed. Loosely, a svara hence could be explained as a pitch position relative to the tonic, along with the transitions coming into and leaving the pitch position, including any additional ornaments while performing the svara. The transitions and the ornaments are not a characteristic of a svara in isolation, but are typically a property of the svara within a specific rāga. To summarize, a svara is not a frequency value in IAM, but a contour of frequency values.

Ganguli and Rao [GR18] 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. Musicians and musicologists often consider that a rāga can only be learned by getting familiar with landmark compositions (hence the typical phrases) in it. A typical concert starts with the rendition of characteristic phrases of the rāga being performed. Characteristic melodic phrases are also the most prominent cues used by human listeners for identifying rāgas [KI12a]. Due to the ornamentations and continuous note transitions, a melody in IAM is better represented as a pitch contour rather than an abstract sequence of notes that it represents. Given the nature of a svara, commonly interpreted as a melodic contour and not a discrete pitch, the segmentation of melodies in IAM into its component svara is not straightforward (even for humans).

Melodic Patterns

Melodic patterns, in their simplest form, are sequences of notes along with their ornaments that frequently repeat in performance. These patterns could be characteristic of a rāga or of a composition and can also depend to an extent on the specific melodic-rhythmic context in which it occurs [GR21]. If melodies are represented with pitch contours, melodic patterns would be short segments of those contours, along with dynamics information if available.

Based on these melodic parallels, we identify that tonic recognition, predominant melody tracking, extracting and characterizing melodic patterns to be fundamental melodic analysis problems in IAM. Transcription of an audio recording into notes hence is not a meaningful task for IAM - note sequences are only skeletal and prescriptive, music itself is highly improvised, and ornaments make it difficult to summarize a melodic phrase just by a sequence of notes.

1.2.3 Rhythm and Meter

Eurogenetic music typically identifies three metrical levels for organization of rhythmic events in music: tatum, tactus/beat, measure/bar. Tatum is the smallest atomic time interval in a music piece or the fastest pulse present in a music piece. Tactus (or beat) is typically referred to as the pulse that listeners entrain to as they tap their foot or dance along with a piece of music [Han89]. A bar/measure is a group of beats to indicate grouping and a periodic accent in the music. The first beat of a bar is called a downbeat and is often associated with significant rhythmic events. A time-signature defines the structure of a bar and the number of beats it comprises. This hierarchy of different metrical levels helps to organize rhythmic patterns within a music piece.

Rhythmic organization in IAM relies on the tāla framework, which comprises fixed-length hierarchical time cycles. The hierarchical time cycles (cycles of a tāla, or tāla cycles in short) provide a structure for rendition and repetitions of melodic and rhythmic phrases and patterns. The tempo of a music piece (measured as the time difference between the start of two successive metrical cycles) is independent of tāla, and is often independently chosen by the musician based on the composition and other considerations. It is hence likely that the same composition rendered by different artists or even the same artist at different concerts can differ quite significantly in their tempo.

The hierarchy of events in a cycle is provided through sub-structures defined within a tāla cycle that help to track progression through the cycle. These sub-structures help to identify the different points of time in the cycle and hence aid in multiple tasks, such as tracking musical time intervals in a musical piece, synchronize the performances by different artists in a concert, demonstrate progression through the cycles (either with hand gestures in Carnatic music, or through canonical rhythmic patterns indicative of a position in the cycle). While the sub-structures are well-defined to provide a clear framework, the musicians are free to improvise within that framework.

Typically three different sub-structures are identified: sub-divisions, beats and a tāla cycle, often with parallels drawn to tatum, tactus and measure. However, there are clear differences in both Carnatic and Hindustani music with such parallels, which will be elaborated further. The equivalence of a tatum-tactus-bar metrical hierarchy to that in IAM is not an exact exercise, but is useful since we could build on methods to analyze the rhythmic events associated with those metrical levels and apply them to analyze similar structures in IAM. Broadly speaking, tatum, tactus and bar are often analyzed with onset detectors, beat trackers and downbeat trackers and drawing possible parallels will help bootstrap these algorithms for metrical analysis of IAM.

For drawing parallels, we use the definitions that tatum is the shortest pulse tracked in the meter, tactus is the "foot-tapping" pulse, and bar as a group of beats identified with a downbeat that is the first beat of the bar. This definition has an implicit assumption that a tactus is a pulse that listeners entrain to, and hence the span of such a pulse is often associated with working memory [HS11]. The definition of a tactus also assumes isochronicity, which is not a necessity in IAM [SHS14].

Table 1.1: Musical parallels in rhythm and meter

Eurogenetic music	**Carnatic Music**	**Hindustani Music**
tatum	akṣara	mātrā or sub-mātrā
tactus/beat	beat or aṅga	mātrā or vibhāg
measure/bar (downbeat)	āvartana (sama)	āvart (sam)

The metrical hierarchy in IAM is identified through the sub-structures of a tāla in the form of mātrā/akṣara, beats/sections and āvartana (cycle). As we see from Table 1.1, the parallels in IAM to the corresponding eurogenetic entities are not absolute or one-to-one, but depend on the tempo of the music piece and the structure of the tāla.

The parallel to a bar/measure is the tāla cycle (āvartana or āvart). The downbeat, or the first pulse of a bar is known to be significant. Similarly, the instant of beginning of each tāla cycle (also the end of the previous) is referred as sama (or sam in Hindustani music) and 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 [Cla00a, p. 81]. A tāla cycle is further divided into sections and beats.

Carnatic music defines the smallest pulse as akṣara (a syllable), which are grouped into beats, which are further grouped into sections (aṅga) that form an āvartana. The number of akṣara in a beat is based on tempo classes and the sub-division structure of the tāla. The sections of a tāla are typically unequal in length and progression through the sections are shown through hand gestures indicative of different types of sections. The definition of an isochronous "beat" as defined at the tactus (the pulse to which music listeners tap their foot to) metrical level in eurogenetic music traditions does not extend directly to Carnatic music. We hence have two possible definitions for a "beat" in Carnatic music. We could emphasize on the isochronicity property of a beat and define it as longest isochronous pulse in the music piece. We could also emphasize on the tactus property of a beat and define the Carnatic beat as the foot tapping times (not necessarily isochronous) in a particular tāla. In such a case, the Carnatic beat can align either with the isochronous pulse or a non-isochronous grouping of such pulses forming a section. While the latter definition is more musically suitable, we use the former isonchronous definition in this chapter from an MIR perspective, since it enables us to bootstrap and use isochronous pulse trackers (beat trackers, e.g.) from eurogenetic music.

Within the scope of this chapter, we define the beats of a tāla in Carnatic music to be isochronous, with sections that are typically non-isochronous. Based on the tāla structure, the tactus pulse is perceived at an ischronous beat level or at the aṅga level (typically, non-isochronous), e.g. the tactus of adi tāla is at the beat level, with 8 isochronous beats in a cycle (with 3 sections), while the tactus in mishra chapu tāla is perceived at the section level with beats grouped into 3+2+2 structure in a 7 beat cycle. We can estimate isochronous beats using MIR methods, and then, with the information about tāla structure, group such beats to estimate the foot-tapping times that may be musically relevant.

Hindustani music also defines hierarchy through mātrā, vibhāg and āvart. A mātrā is the smallest rhythmic sub-unit of a tal. Mātrā are isochronous and are grouped into possibly unequal sections (called vibhāg) that make up the whole āvart (cycle). There are also tempo classes called lay in Hindustani music which can vary between ati-vilaṁbit (very slow), vilaṁbit (slow), madhya (medium), drut (fast) to ati-drut (very fast). Depending on the lay, the mātrā may be further subdivided into shorter time-span pulsations, indicated through additional filler strokes of the tabla. However, since these pulses are not well defined in music theory, we consider mātrā to be the lowest level pulse as a parallel to the tatum. However, depending on lay, the mātrā duration can vary over a wide range of 200ms to 5 seconds. This means that the tactus pulse that listeners are entrained can align with the sub-mātrā fillers (in vilambit lay), or mātrā (in madhya lay), or with even vibhāg (in drut lay).

To summarize, the parallels to tatum-tactus-bar metrical levels using their common definitions does not directly extend to IAM though some parallels exist. Due to the structure of the tāla and tempo classes in IAM, tactus pulse could be non-isochronous and could align with different metrical levels defined in IAM. The common definition of a beat assumes isochronicity and hence does not extend directly to IAM, which can have a non-isochronous pulse to be tracked as a beat. However, isochronous pulse trackers (beat trackers, e.g.) from eurogenetic music can be bootstrapped to track the isochronous Hindustani mātrā or Carnatic beat but in a typically extended range of tempi, and then grouped based on the knowledge of tāla structure to derive relevant non-isochronous tactus pulse.

Rhythmic Patterns

Rhythmic patterns in IAM are closely aligned with the tāla cycles, and are played on melodic or percussion instruments. The rhythmic patterns serve multiple functions in addition to showcasing the rhythmic structure of the tāla and the composition. Within the scope of this chapter, we wish to draw attention to two specific kinds of rhythmic patterns, canonical rhythmic patterns played to indicate progression through a tāla cycle and different hierarchical metrical events within the cycle, and the improvisatory rhythmic patterns that showcase the variety of rhythmic patterns that could be played within the framework of a specific tāla.

The hierarchical metrical structure of the tāla is indicated through canonical rhythmic patterns that also indicate progression through the cycle. These canonical rhythmic patterns are well-defined in Hindustani music (called the theka) and played on the tabla, while they are less well-defined in Carnatic music. The canonical patterns aim to establish the fundamental metrical framework of the tāla through instances of characteristic rhythmic patterns and hence could be used by all musicians to perform together within the tāla framework. The function of these patterns is predominantly to provide the metrical structure for performance.

Improvisatory rhythmic patterns played during improvisatory sections help to showcase the variety of rhythmic patterns that could be played within the framework of a specific tāla. These patterns are typically played during percussion solos (or even the solo performance of a melodic instrument or vocals). The improvisatory patterns focus on variety and involve complex calculations that can be accommodated within the tāla, deviating from canonical structures and provide ample opportunity for demonstrating skill and expertise in rhythm. Improvisatory patterns can often last multiple tāla cycles with long and complex patterns that finally resolve at the specific point in the tāla (most often, the sama).

Percussion Patterns

A distinctive feature of IAM relates to the use of percussion patterns, which are mostly the patterns intended to be played on percussion instruments of IAM. They are distnct from rhythm patterns, which mainly refer to the temporal arrangement of different events with different accents, while percussion patterns include a temporal arrangement of different percussion timbres.

Percussion patterns in IAM are described and learnt through syllabic onomatopoeic oral mnemonics that relate to the sound of the strokes played on the mridangam/tabla. Most of the time, the percussion syllables also encode the dynamics, intonation and interaction between neighbouring strokes. Hence, as an analogy, the spoken percussion syllables (called the bōl in Hindustani music and solkaṭṭu in Carnatic music) form a language for percussion in IAM and the art of reciting these syllables is often a feature of many solo music and dance performances. Indian art dance forms such as Kathak and Bharatanātyam make extensive use of percussion syllables to define and perform movement patterns. Percussion training extensively uses these syllables to learn and perform percussion patterns on instruments. The analogy of the percussion patterns to speech enables us to utilize a multitude of speech technologies for analysis of percussion patterns in Indian art music.

While percussion instruments used in eurogenetic music could also have their strokes mapped to oral mnemonic syllables, IAM provides an existing system that is familiar to all practitioners. Further, the rich and complex set of timbres played on tabla/mridangam can be mapped to different bōl/solkaṭṭu used in IAM and used for representation, analysis and synthesis of percussion patterns. While the exact set of percussion syllables used in practice is not standardized and varies across music schools and musicians, it is still possible to map timbres to syllables and use them for automatic analysis in transcription problems. These characteristics provide a unique opportunity to study and analyze percussion as a language in IAM.

1.2.4 Timbre and Instrumentation

The timbre and instrumentation in IAM is useful to understand and contrast to help with choice of tools for signal processing. We first describe the instrumentation and timbre from a musical perspective and then describe how this manifests in audio signals.

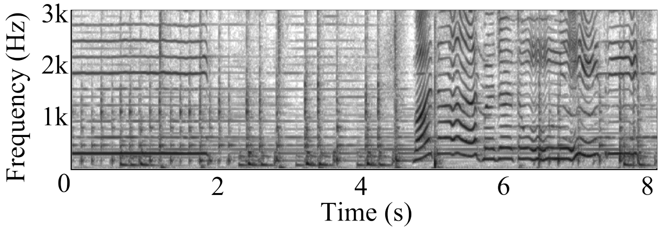
IAM performances typically have limited number of performers, with a lead musician (often vocals), a melodic accompaniment (mostly violin in Carnatic and harmonium/sāraṅgī in Hindustani music) and a (set of) percussion accompaniments (tabla in Hindustani and mridangam in Carnatic music, respectively).

This makes Indian art music predominantly melodic and heterophonic, with usually two (sometimes more) simultaneous melodic voices, leading to two or more overlapping melodic lines. Furthermore, a tānpura (drone) playing in the background reinforces the tonic frequency and often the fifth. All melodic instruments are tuned to the tonic of the lead artist. The primary percussion accompaniments are pitched, and are also tuned to the tonic of the performer. There is no polyphony in the sense of the word defined in eurogenetic music.

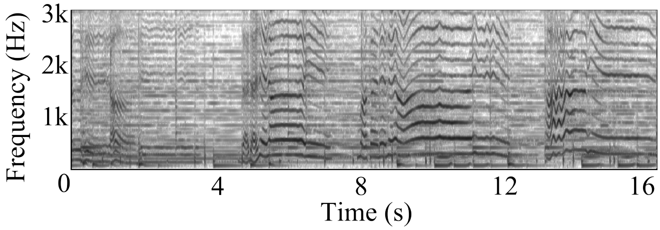
The drone in the background could be used for analysis of tonic. The heterophonic nature of music makes analysis of melody relatively simpler, but it is still complicated due to the melodic accompaniment closely following the lead melody with variations with a small delay. The pitched percussion also often overlaps in pitch with the lead instrument, making the task of percussion separation harder.

Audio and Signal Characteristics

A piece of music in an audio recording is a primary artifact for analysis of Indian art music. Hence it is useful to identify the characteristics of IAM from a signal processing perspective and describe how different IAM concepts manifest in audio. This will enable us to choose preprocessing steps and feature extraction for many MIR tasks, as well as aid in source separation tasks. Several basic MIR tasks that extract melodic and rhythmic events can benefit from generic signal processing approaches by mapping the music concepts to signal processing terminology.



(a) Carnatic music



(b) Hindustani music

Figure 1.1: Audio signal characteristics of Indian art music signals. The figure shows the spectrogram of the audio excerpt of Carnatic and Hindustani music, showing the frequencies up to 3 kHz. We can clearly observe the horizontal lines in the spectrogram that remain stationary in time from the tānpura playing in the background. The percussion strokes appear as strong vertical lines across the entire spectrum. The notes from the lead vocals and the accompanying melodic instruments can be seen as time-frequency varying harmonic series in the foreground.

Figure 1.1 shows an illustrative example of the spectrogram of audio excerpts of both Carnatic music and Hindustani music. To focus on melody and percussion, the spectrogram is plotted only up to a frequency of 3 kHz, though there is information from higher harmonics and percussion strokes at higher frequencies. The time durations of the excerpts are also different, with the Carnatic music example being shorter in duration.

The lead melody and its harmonics can be clearly seen in both the figures, along with the continuously changing fundamental frequency (F0). There is a drone in the background that provides the tonic for the performance, provided by the tānpura (Hindustani music) or tambūra (Carnatic music). The drone can also be seen as the unchanging set of spectral frequencies in both the spectrograms. The melodic accompaniment (violin) in Carnatic music closely follows the lead voice and can be seen with a lower amplitude. Harmonium is the melodic accompaniment in the Hindustani music excerpt, which can also be seen with a lower amplitude in the figure. These observations also have an important bearing on melodic analysis of IAM. Due to the ornamentations and typical intonation of svaras, pitch contours are more relevant melodic representations of melody in IAM than a sequence of notes. Even melodic phrases are better encoded as pitch contours rather than a short sequence of svaras. Segmentation of a melody into its svaras, understood as melodic contour, could be interesting to observe, for example, how the same svara is performed in relation to its melodic context, or in different rāga.

The percussion instruments tabla and mridangam have a bass drum head, and a treble drum head that is pitched. The pitched drum head is tuned to the tonic of the lead musician. The pitched strokes can be sharp or sustained. Both the figures show the percussion strokes as vertical lines, showing their broad spectrum. We can identify the strokes from the left and right drums distinctly in different frequency ranges in both cases. In addition, we can see some of the harmonics of the pitched percussion strokes, and the quick decay of the percussion strokes. The Hindustani excerpt is taken from a madhya lay piece and we can see the longer notes and sparser tabla strokes, indicating lower rhythmic density due to lower tempo.

1.2.5 Functional Parallels

The functional parallels that we discuss pertain to the different functions around Indian art music, mostly discussing the composition, performance, learning, production and consumption. These characteristics primarily help us to put the important facets of IAM into perspective and hence identify and prioritize research problems, choose artifacts, build analysis tools that are musically relevant and meaningful within the IAM community.

Music Learning

Music learning and teaching is predominantly oral, learnt through repetition and following a teacher. Music scores in sargam notation are used for learning, primarily as a mnemonic aid for learning and and not as an exhaustive representation of the melody to be performed. The oral transmission also leads to stylistic schools based on lineages of masters or geographic regions. IAM percussion has corresponding onomatopoeic oral mnemonic syllables that represent individual strokes/timbres that can be played on a mridangam/tabla. These syllables are used for teaching/learning as well as a form of art in performances. As discussed earlier, availability of syllabic representation means clear parallels to mature speech technologies that can be used for analysis of IAM.

Music Composition and Performance

IAM performances include both compositions and improvised pieces. However, even the performance of a pre-composed music piece includes significant improvisation within the framework of the rāga and tāla of the piece. The music compositions provide the lyrics (for vocal performances) and skeletal score (mostly in the form of sargam) that is mostly prescriptive and improvised upon. This primarily implies that analysis of audio performance is more relevant for IAM than analysis of music scores. While music scores do contain useful information, their use is limited in music learning and performance, and hence in detailed analysis due to extensive improvisation. This also implies that an audio recording along with its metadata is the primary artifact for analysis of IAM.

Music Consumption, Appreciation and Critique

Historically, IAM was consumed by listeners through live concerts and more recently through recordings of live concerts. The consumption platforms are changing slowly to digital platforms as is the trend with most music recently. The current practice is that most of the concerts in IAM are organized by music foundations and organizations throughout the year, with specific music festivals with a multi-fold activity in concerts. A few example festivals include the Madras music festival season (Chennai, India) and the Ramanavami music season (Bangalore, India) in Carnatic Music and ITC SRA Sangeet Sammelan, Sawai Gandharva Bhimsen Festival (Pune, India) in Hindustani music. Both these music traditions have a dedicated audience spanning different demographics and expertises. Outside of the concerts, information around IAM is consumed through listener music fora, informal discussions and media critiques/reviews.

The involvement of a large community of people in consumption and appreciation of IAM implies that a large amount of textual information (in the form of news articles, forum discussions, reviews, critiques) is available in addition to formal literature and music metadata. The music concepts are shared across the corpora for analysis, formal musicological and music theory literature, and informal discussion around music. This provides ample opportunities for research into organizing text information around IAM through existing linked data and semantic web approaches.

Music Production and Distribution

A basic unit of distribution of a recorded Carnatic music performance is a concert that consists of several pieces, each in a specific tāla and rāga. The concert has a structure with compositions and improvisations performed in specific forms of Carnatic music in a predictable order. A Carnatic music concert is usually seen to comprise a well-formed narrative and order, with the performing artists choosing appropriate forms, rāga and tāla performed in a specific order and built up as the concert progresses.

Hindustani music albums comprise one or two longer pieces, in a specific rāga followed by short pieces. A structure hence exists in both Carnatic and Hindustani music, unlike an album of popular music, which mostly has independent tracks. Segmentation of a concert into relevant individual pieces (or into shorter sections within a music piece) is hence a relevant problem for IAM. IAM corpora are hence better organized in such units (concerts) with segmented audio with associated metadata for each music piece. In addition, the corpora could have additional metadata at the release or concert level.

Since most audio recordings available in IAM are recorded from live concerts, the quality of audio can vary considerably, especially with older concert recordings. Given the limited instrumentation in IAM, it is possible to record and store multi-track recordings, which are a very useful source for multiple MIR tasks including melodic and rhythmic analysis, source separation and music transcription. An effort to collect multi-track corpus is described in Section 1.4 of this chapter.

1.2.6 Indian Art Music: Important Metadata

The previous sections drew parallels on multiple aspects of Indian Art Music. We summarize the learnings in this section. The primary object of analysis is an audio recording of a music performance and associated metadata. The rāga and tāla related metadata comprise the most crucial metadata for a piece of music.

Within the frameworks of rāga and tāla, we have analysis of certain concepts that could benefit from existing engineering formulations and hence can be explored as direct extensions to current methods. However, many concepts are socioculturally distinct and pose additional challenges to engineering formulations for their analysis. Key examples include the presence of improvised sections, the absence of a fixed tonic, non-isochronous beats, hierarchical metrical structures with long cycles, melodic representation through pitch contours, absence (or low utility) of written music scores, the presence of syllabic percussion systems, orally transmitted repertoire, and unique aspects of music production consumption and appreciation. With these distinctions, we can now discuss the research problems around MIR and computational musicology from an IAM perspective.

The remainder of the chapter will focus on tasks, data and tools, mostly from the perspective of getting started on MIR and computational musicology for IAM - tasks that could use existing tools for initial analysis and availability of data for analysis.

1.3 Research Problems in Indian Art Music

A detailed list of MIR problems relevant for IAM along with current approaches is presented in other chapters of the monograph. Hence, we won't formulate the tasks in detail and review existing methods though we discuss the broad challenges and directions that have been already explored. The research problems discussed are broadly divided into MIR problems and data-driven musicology problems. The MIR problems aim to analyze and extract musically relevant metadata and descriptors from audio, scores and text. The research problems on data-driven musicology aim to provide tools, methods and musicological insights on large corpora of music data.

1.3.1 Music Information Research (MIR) Problems

The MIR problems consider one of musical artifacts as inputs and analyze it for musically relevant metadata and automatic annotations. Since an audio recording is the primary artifact for analysis of IAM, most relevant MIR problems focus on audio, while we also discuss in brief MIR problems that focus on prescriptive music scores and text. We categorize the problems based on the artifact consumed and the musical concept the problem addresses.

Melodic Analysis

Considering the rāga as the central element, MIR problems on melodic analysis for IAM include tonic and predominant melody extraction, melodic transcription and the analysis of melodic patterns [Gul16]. Though the concept of a rāga is central to IAM, there is not an equivalent concept that has been explored within previous MIR literature. While modal analysis and melodic analysis of Turkish makams could be considered parallels, most of the melodic analysis work that considers rāga as a central concept are limited to IAM. The melodic analysis problems revolve around the concept of rāga, either contributing to its recognition through other low level melodic descriptors, or its characterization through higher musical abstractions such as melodic scales, phrases and intonation.

Predominant melody extraction is arguably the most useful melodic analysis problem given that the predominant melody is the input to many different melodic analysis tasks. Given the heterophonic nature of IAM, predominant melody estimation is less challenging than polyphonic music, though the possibility of two closely aligned melodic lines (vocal+violin for CM and vocal+harmonium/sarangi) can often pose challenges. A predominant melody in IAM is arguably best represented as a pitch contour (continuous valued) computed at each audio frame (typically a few ms), and then quoted either in absolute frequency (Hz) or normalized by tonic (and represented in cents relative to the tonic). The pitch contour can then be used downstream for all subsequent tasks. Existing methods for predominant melody extraction, e.g. melodia [SG12b] have been experimented to work with acceptable accuracy with some tuning for the task in IAM, and hence can be good initial solution for culture-specific methods if need to be developed. However, predominant melody estimation for IAM is not a completely solved problem and is still a challenging task with scope for improvement and needs further exploration.

Given the primary importance of the shadja (tonic note frequency) in IAM melody, estimation of its frequency in an audio recording of IAM is a fundamental problem in melodic analysis. Since the frequency of the reference pitch (tonic) depends on the specific needs of each leading artist who choose it, the estimation of tonic in IAM is not so much of a tonic note estimation but a tonic frequency estimation. Tonic can be reliably estimated with current methods in clean high quality recordings, with most estimation errors corresponding to estimating a related note frequency (a fifth or an octave, e.g.) as the tonic. Existing methods for tonic estimation broadly use a predominant melody extractor (e.g. melodia) and a representation that summarizes the melodic content (e.g. a pitch histrogram). From such a representation, based on svara relationships, we can extract the tonic frequency [Gul11].

Intonation analysis refers to an analysis of expression of different notes of a music piece. Intonation captures the positions (on a frequency scale) of different notes in the music piece, while also providing additional information on how those notes were performed. Given the improvisatory nature of IAM with a plethora of melodic ornamentation, intonation provides a snapshot of how specific svaras are performed within a rāga. The simplest representation of intonation in IAM is the histogram of all pitch content (predominant melody contour, e.g.) in the music piece. Multiple existing methods can be used to parametrize the histogram to arrive at useful models for intonation analysis [Kod+14a].

Analysis of melodic phrases aim to characterize sequences of svaras (encoded often as short continuous pitch contours) and infer similarity relationships between them. A music piece in IAM typically consists of composition-specific music phrases and rāga-specific music phrases, both performed within the framework of a rāga. Defining similarity metrics between melodic phrases help us to cluster melodic patterns, which can eventually help to compute similarity between the sections of a single music piece (intra-piece) or between two music pieces (inter-piece) in a large music collection. Melodic phrases are fundamental hence to organize a music collection based on melodic similarity. Analysis and clustering of melodic phrases also help us to identify the rāga of a music piece and help to provide a musically meaningful summary of the music piece. When using predominant melody representation using pitch contours, analysis of melodic phrases can be formulated as repeating sequence matching problem in a 1-D data, though time normalization using Dynamic Time Warping might be necessary[Gul16] to compute distance between different pattern instances. Once we have these distances computed, we could use multiple available clustering methods to group phrases and retrieve them.

Automatically identifying the rāga of a music piece is an important MIR task in IAM, helping us to organize large music collections on relevant metadata. Depending on the application, it could be formulated as a recognition task (no prior information or hypotheses on the rāga) or a verification task (where a rāga label is presented with the audio recording and the goal is to verify if the label is correct). rāga related information is present in multiple melodic descriptors we can extract from music recordings, and hence different methods have been employed for rāga recognition. While characteristic melodic phrases are often the best indicators of a rāga, a combination of intonation, melodic phrases and an alignment of these descriptors with rāga grammars from music theory is expected to provide the best performance for rāga recognition.

Rhythmic Analysis

Automatic rhythm analysis problems in IAM could be broadly categorized into problems of automatic rhythm annotation and analysis of rhythm patterns. Automatic rhythm annotation for eurogenetic music could be defined as estimating different elements of meter from an audio recording - with onset detection [Bel+05], tempo estimation [PP11], beat tracking [DP07] and downbeat tracking [DP06; HDF12; KBW15] as primary explored problems.

Automatic rhythm annotation, in the context of Indian art music can be defined as the estimation of the characteristic components of the tāla from audio [Sri16]. For Carnatic music, the most important rhythm related tags include estimating the median tempo of the piece (in akṣara per minute or beats per minute), the length of the cycle (in number of beats or akṣara), the tāla label (and hence implicitly the underlying metrical structure) and the subdivision structure. For Hindustani music, the most important rhythm related tags include the median tempo of the piece (in mātrā per minute), the lay class, the cycle length (in mātrās) and the taal label. Estimating the time varying tempo curve, the akṣara pulse locations, the beats, the tāla cycle section boundaries and the sama instants are the important time aligned annotation problems in Carnatic music. The most important problems in Hindustani music are the estimation of the mātrā pulsation, the vibhāg boundaries, and the sam instants.

Automatic rhythm annotation is an important rhythm analysis topic, and there are several applications in which these rhythm annotations are useful, such as music autotagging, rhythm based segmentation of audio, beat aligned processing of music, audio summarization, music transcription, and different rhythmic pattern analyses. Tracking the components of the tāla through a music piece is essential for most other rhythm description tasks such as segmentation and extraction of rhythmic patterns to define similarity. It is to be noted that many rhythm annotation problems can be jointly addressed, estimating several components together, e.g. the tāla, tempo, beats and the sama can be jointly estimated, in a task that we could be called as automatic meter analysis.

Automatic rhythm annotation could benefit from existing rhythm MIR tasks, though some modifications might be necessary. The equivalence of tactus to the mātrā in Hindustani music could be problematic since the typical durations of tactus used for beat tracking might not align with the mātrā in vilaṃbit pieces. Beat tracking methods often assume an isochronous beat - and as discussed before, beats in IAM are not isochronous. However, we could use a beat tracking algorithm to estimate an isochronous pulse in an IAM audio recording, and based on the structure of the tāla, we could pick a subset of the estimated isochronous pulse to arrive at the musically relevant beats for the tāla. Downbeat tracking in eurogenetic music has mostly been applied to short measure durations (a few seconds, typically). However, it has been shown that such methods do not extend well [SHS17] when cycle durations could be longer than a minute (e.g. in vilaṃbit pieces in Hindustani music).

While tāla provides a framework and structure, the rhythm and percussion patterns form the content through which the metrical structures and rhythms are illustrated, and hence form the other main component of rhythm analysis. Rhythm patterns mainly refer to the temporal arrangement of different events with different accents, while percussion patterns include a temporal arrangement of different percussion timbres. To contrast, percussion patterns are rhythmic patterns, but rhythmic patterns need not contain only percussion, and can be formed by melodic and/or percussion instruments.

A pattern is defined as a temporal sequence of events and hence it is necessary to estimate onsets of various instruments in music, since that creates the time-aligned sequence of note/stroke events that can be further used to obtain both rhythmic and percussion patterns. Some important sub-problems within pattern analysis are instrument-wise onset detection, pattern transcription, and pattern discovery, each of which is described further. Transcription aims to map an audio recording into a time aligned sequence of symbols (strokes, accents, e.g.). The problem of discovery is more open ended and aims to automatically retrieve interesting patterns and insights about those patterns, in a data-driven way.

The task of instrument-wise onset detection refers to detecting the onsets of specific instruments from an audio signal that is a mixture of many music instruments. For rhythm analysis in Indian art music, instrument-wise onset detection can help to obtain cues for both meter tracking and for analysis of percussion patterns. The onsets of percussion instruments mridangam and tabla provide cues to the tāla and delimit percussion patterns. A differentiation between the left (bass) and right drum onsets in both instruments is additionally insightful and useful. Instrument-specific onsets are often not estimated explicitly, but are estimated as a part of a bigger task, such as percussion transcription.

Rhythm patterns extracted from audio recordings are representative patterns of the tāla, and hence useful for both automatic tāla recognition and meter analysis. The most relevant rhythmic patterns are cycle length rhythmic patterns - patterns that are played in one full cycle of the tāla. Shorter patterns, played within a cycle mostly act as rhythmic atoms to make up the whole cycle and are played more often. However, there are long rhythmic patterns played on mridangam/tabla and accentuated through melody that can last many cycles. Automatic discovery of rhythm patterns can be used to define content based rhythmic similarity between pieces of music, which is expected to be more relevant than metadata based similarity. Automatic extraction of rhythm patterns can also be a tool for musicologists to study various rhythm patterns in larger corpora.

Percussion pattern transcription is mainly applied on audio recordings with percussion solo, and aims to transcribe the audio recording into a time-aligned sequence of drum stroke labels, and in the case of Indian art music, into percussion syllables. Percussion transcription is a subproblem of the more general music transcription. Transcription of a solo into symbolic syllables is an example of audio segmentation at a fine grain level. Transcription of solo performances are useful for percussion training. Since Indian art music is mostly improvised, the need for such a fine grained transcription system is limited, except for music education and performance analysis applications. However, a transcription can be used to automatically discover percussion patterns and develop rhythm similarity measures using such discovered patterns that could be used for other tasks, e.g. recognition of tabla gharānā [Gow+21].

The benefits of using oral syllabic systems from an MIR perspective are both the cultural specificity of the approach and the accuracy of the representation of timbre, articulation and dynamics. The characterization of these percussion traditions need to consider elements that are essential to them such as the richness of their palettes of timbres, subtleties of articulation, and the different degrees and transitions of dynamics, all of which is accurately transmitted by the oral syllables.

As discussed earlier, the onomatopoeic percussion system in Indian art music provides a language for percussion and hence is the most musically meaningful way to represent percussion patterns of tabla and mridangam. Such a link between drum patterns and natural language has been explored [MD12]. Percussion pattern transcription can be formulated as a supervised learning task, using labeled training data to build syllable stroke models, which can then be used to transcribe a test recording [Ana+14a; Gup+15; Kur+15]. Percussion pattern discovery is an unsupervised task, aiming to extract percussion patterns from audio and/or scores in an unsupervised way, though some priors can be used.

Rhythm similarity measures aim to use rhythm descriptors, metadata and segmentation information to provide an objective similarity value between two phrases, two music pieces, or two parts of the same piece. Since rhythmic similarity is not a very concrete notion, we need definitive and objective measures of similarity, especially in a multicultural setting. This would necessitate the use of knowledge based approaches for similarity modeling. The onus of developing new similarity measures clearly lie on the choice of metrics that correspond to rhythm similarity as perceived through musically relevant concepts - based on tāla and the rhythmic patterns.

Structural Analysis

Structural analyses typically deal with the problem of segmentation, referring to a broad category of problems that involve the annotation of segments of audio with a label/tag. Segmentation can be done at several levels, based on different music concepts. Segmentation problems are useful since they provide additional metadata to navigate through music collections (and within a single recording), and to further develop similarity measures.

Segmentation can be done of a full concert into the pieces that were performed in it, which is useful for archival. Segmentation at a structural level within a piece aims to segment the piece into different sections of the piece, and is useful for navigation and similarity. Different music forms in IAM are amenable to different structural segmentation, e.g. an ālāp would benefit from a segmentation based on melodic phrases, while a percussion solo could be segmented based on different sections and percussion patterns played within the solo. A music recording can also be segmented based on the instruments that are playing, a problem that can also be described as instrument tracking in audio. At a fine grained level, a music piece could also be segmented at the level of melodic and rhythmic phrases, which could be used to highlight different parts of a music piece.

A basic idea for segmentation of music pieces has been based on the assumption that change-points in different structural elements (e.g. sections) are accompanied by a change in music characteristics along different musical dimensions (melody, rhythm, timbre). This implies that change points could be detecting changes within and between structural elements (e.g. sections). A well-explored way of doing this is by extracting a novelty function [Foo00a] from self-similarity matrices [PMK10] computed from relevant features from an audio recording. The features are chosen and the time scale of similarity computation depend on the segmentation task, e.g. we could use beat similarity matrix for tracking metrical structure of a music piece [Sri+12].

Analysis of Timbre

The analysis of timbre mainly refers to modeling different instruments present in a recording. A few timbre analysis tasks would include percussion transcription (described above), diarization of instruments played (instrument-based segmentation of an audio music recording) and source separation. Source separation for music is a well-explored topic with both classical methods and deep learning methods applied to music [ABM13; LPS19; RS15; Sma04; Sri+14a] and hence not discussed in depth in this chapter. However, it is to be noted that these methods haven't been extensively applied to IAM, which might pose unknown challenges due to the complex resonances in instruments used in IAM such as the tānpurā, sitār and the vīṇa (due to the bridge) or in mridangam/tabla (due to the loaded membrane).

While we presented problems on each dimension separately, the musical concepts do not exist in isolation and neither do the research problems and approaches. There is an overlap between the melodic, rhythmic, structural and timbral analysis tasks and hence we could devise approaches that aim to analyze and extract different subsets of descriptors across these dimensions. In addition, methods for analysis of one music concept could benefit from available prior annotations (manual or automatically extracted) from other dimensions, e.g. a segmented concert into music pieces with a single tāla will help to track tāla better, and in turn the availability of tāla cycle markers on the music recording will help structural segmentation of a music piece into different sections (since most sections are aligned with tāla cycle boundaries).

Analysis of Scores

Symbolic music scores in Indian art music are not comprehensive and are only indicative, often only skeletal and prescriptive. They are seldom used in performance, but used to a limited extent in music training and archival through sargam notation for melody and syllabic notation for percussion. There are no standard notation systems for melody or percussion, in both Hindustani and Carnatic music, which are widely accepted and used.

Due to the large deviation of performed music from the indicative scores, score analysis can at best be good starting points towards further analysis. There are no standardized tools (such as humdrum [Hur02], music21[3]) for the analysis of IAM scores. Further, there is no comprehensive collection of machine readable music scores in Indian art music, though there are small collections built for analysis [SC12a].

Automatic score analysis research in the context of melody, rhythm, and percussion for Indian art music is scarce. Symbolic scores have been used for different melodic analyses in Carnatic music [Kod+14a; Ran+19a; RS13a] for Carnatic music and for Hindustani music vocal compositions [Cho07; SC12b], creating a machine readable Hindustani melodic music score dataset. Tabla bōl sequences [Cho06] have been used for predictive modeling of solo tabla performances using the multiple viewpoint modeling framework [Cho+10; CSA10].

[3] https://web.mit.edu/music21/

However, symbolic representations offer multiple advantages. Music scores can be used for symbolic analysis of melodic/percussion patterns, and used to discover patterns from score corpora, a task that much less complex than extracting them from audio. We can then use these patterns and search for them in longer transcribed recordings. Such an approach with pattern discovery from scores followed by pattern search in audio has been explored in the past within a speech recognition + keyword search framework. Since the music scores form the language to disseminate music, we can also utilize the recent advances in language modeling and natural language processing for analysis of music scores.

Analysis of Text

Textual information for IAM mostly comprises metadata available with audio recordings, discussions about the music on different fora (for audience, musicians, music labels), commentaries, critiques, artist biographies and other information. The text information could be free form text or tagged text (linked text) with marked entities. Both forms of data could be analyzed through semantic web technologies that aim to integrate human knowledge into computer systems to solve complex problems that require human expertise. An ontology specifies concepts, attributes, relations, constraints, and instances in a domain [BHL01; BL04; GFC04]. Since music is a complex and varied phenomenon with many perspectives, a cultural domain specific ontology is needed to define the relationships that pertain to a specific type of music from which we could then derive different relationships between music concepts and community.

Built using the knowledge of music theory and practice, the ontologies would be useful for querying complex musical relationships between the pieces. The ontologies complement the features derived from the data with music knowledge based relationships that can be used for defining similarity, e.g. using a tāla ontology and the knowledge of cycle length, it might be easier to identify the tāla from audio. Further, the ontologies will also be useful to create specific models with priors obtained from the ontology. In summary, ontologies can be built both for a direct use in navigation and inference, and for building domain specific machine learning algorithms.

1.3.2 Data-driven Musicology Research Problems

Recent advances in digital humanities have brought forward aspects of human behavior using large corpora. The focus of current studies in digital humanities lies largely on language and social data corpora while music corpora have received less exploration. Performance analysis of music corpora can provide us with several insights into different aspects of music and show us the contrasts and similarities between music theory and practice. Such analyses on larger corpora can yield us additional insights that are often difficult to obtain with traditional manual analysis.

Corpus studies are in general driven by the common motivation of contributing empirical results that improve the understanding of a specific property of data in the corpus. In music, typically, these properties are melody, harmony, and rhythm. Manual analyses of such properties in music corpora have been performed as long as the related disciplines, such as ethnomusicology or music theory, have existed. However, in the last decades, the availability of computational methods enables the evaluation of larger amounts of data more easily. Data-driven analysis of large corpora is especially amenable to computational methods and can provide additional tools for statistical analysis. Such analyses can provide broad corpus level inferences for a musicologist, complementing a manual detailed analysis of small set of representative pieces.

Data-driven musicology utilizes sizeable corpora (annotated or otherwise) of music - audio, scores, text for statistical analysis to infer musicological insights. For IAM, performance analysis of large set of audio recordings can provide us useful insights to verify common knowledge and in addition gather additional insights into music performances of IAM. The large corpora used for analysis help to provide statistical validity, capturing both generalities through averaging over the whole corpus, but also measure nuances as deviations from the average. There are a multitude of opportunities for data-driven musicology with existing corpora of IAM, but we illustrate specific studies as examples of the scope and utility of the approach.

Srinivasamurthy et al. [Sri+17b] provides insights into aspects of tempo and rhythmic elaboration in Hindustani music, based on a study of a large corpus of recorded performances. Focusing on aspects of tempo and rhythm, the paper demonstrated the value of a computational methodology for the analysis of large music corpora by revealing the range of tempi used in performances, intra-cycle tempo dynamics and percussion accents at different positions of the taal cycle. In the article, typical tempo developments and stress patterns within a metrical cycle were computed, referred to as tempo and rhythm patterns, respectively. Rhythm patterns were obtained by aggregating spectral features over metrical cycles. They reflect percussion patterns that are frequent in the corpus and enable a discussion of the relation between such patterns and the underlying metrical framework, the taal. Tempo patterns, on the other hand, are computed using reference beat annotations. They document the dynamic development of tempo throughout a metrical cycle and revealed insights into the flexibility of time in Hindustani music for the first time using quantitative methods on a large set of performances. This study was further extended to Carnatic music using the same principles and methodology [SIS19].

Similar statistical analysis of melodic concepts is also possible, as evidenced from the work by Ganguli et al. [Gan+16], which aimed to discover implicit patterns or regularities present in the temporal evolution of vocal melodies of Hindustani music. The paper applied existing MIR tools and techniques to a collection of ālāp performance. Using svara-based and svara duration-based melodic features, the paper quantified the manifestation of melodic concepts such as vādi, samvādi, nyās and graha svara in vocal performances.

[Kod+12a] aimed to characterize intonation in Carnatic music by parametrizing pitch histograms. The article obtained a pitch histogram for a recording, and parametrized the prominent peaks. The parametrization followed by a qualitative assessment on a larger collection of rāga showed discriminative power of the entire representation that is also useful in musically relevant tasks such as performer and instrument characterization.

These examples are only illustrative and there is high potential for data-driven musicological studies that can complement current literature and extend the methods to large corpora of music.

1.4 Corpora, Datasets and Annotations

A significant part of data-driven research in MIR and computational musicology needs good quality data. Data corpora that are representative of the music culture under study are critical for building and testing analysis methods. To develop such MIR approaches and advance knowledge, there is a need for research corpora that can be considered authentic and representative of the real world. The data sources comprise of audio, metadata accompanying audio, music scores, lyrics, manual and automatic annotations, and linked (semantic) data.

Building such data corpora scientifically for MIR itself is a research problem [PF12; Ser14]. Setting up criteria for selection and curation of music, and designing datasets for research are to be done with objective parameters that can then be used to measure the goodness of a corpus for a particular research task. Collection of good quality data and easy access to both audio and metadata is essential for reproducibility of research.

A research corpus is an evolving collection of data that is representative of the domain under study and can be used for relevant research problems. A good data corpus includes data from multiple sources and can even be community driven. In the context of MIR, since it is practically infeasible to work with the whole universe of music, a research corpus acts as a representative subset for research. Hence, algorithms and approaches developed and technologies demonstrated on the research corpus can be assumed to generalize to real world scenarios.

A test corpus or a test dataset is often a subset of the research corpus, possibly with additional metadata for use in a specific research task. In experiments, test datasets are used to develop tools, and to evaluate and improve their performance. Computational approaches are developed using these datasets and then extended to the research corpus. Hence test datasets can even consist of synthetic data that can be used for testing. Unlike a research corpus, a test corpus is fixed for use in a specific experiment. A test corpus can evolve, but each version of the dataset used in a specific experiment is retained for better reproducibility of research results.

Serra [Ser14] 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. [Wil+16] recently formulated a set of guidelines to improve the findability, accessibility, interoperability, and reuse (FAIR) of digital assets. While Serra 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. The CompMusic project considered the tenets and guiding principles by Serra to build research corpora with good quality and coverage while emphasizing FAIR principles to store, navigate and reuse data. In addition, data corpora are most useful when there is open access to audio, metadata, and annotations, while allowing for the datasets to grow both in size and quality through community-driven efforts.

1.4.1 Research Corpora and Test Datasets for Indian Art Music

All the music cultures under study can be described in terms of musical concepts, music content and the music community. The elements of the corpora can be associated with one or more of these categories and hence useful for computational tasks in these three aspects. Central to each corpus is an audio music recording with its metadata. For both Carnatic and Hindustani music, the CompMusic project aimed to build an annotated audio sub-collection that is representative of the real world performance practices[4] [Sri+14b].

The CompMusic Carnatic music research corpus mainly comprises audio recordings, associated editorial metadata, lyrics, scores, contextual information on music concepts, and community (social) information from online music forums and other sources. Audio recordings, editorial metadata, scores, and lyrics are the content used by signal processing and machine learning approaches. Contextual information and the forum discussions form the music concepts and community information used for semantic analysis. The primary metadata associated with each concert is the name of the release, the lead and the accompanying artists, and the musical instruments in the concert. For each audio recording contained in the release, the relevant metadata are the artists performing on the track, the name of the composition/s and the composer, rāga/s, tāla/s, musical form/s.

The CompMusic Hindustani music research corpus comprises commercially available music releases from several music labels. It mainly consists of khayāl and dhrupad vocal music releases, though a significant number of instrumental music releases are present. The metadata associated with each release is the name of the release, the lead and the accompanying artists, and the musical instruments in the concert. For each audio recording in the release, the relevant metadata are the artists performed on the track, the name of the composition/s (bandiṡ) and the composer/s (if composed), rāg/s, tāl/s, lay/s (tempo class), form/s, and section/s.

The editorial metadata associated with each release in both HM and CM corpora are stored in MusicBrainz[5]. MusicBrainz assigns a unique MBID for each entity in MusicBrainz, such as the artist, composer, instrument, recording, work, and a release. This helps to organize the metadata in an effective way. All the editorial metadata was entered using Latin alphabet and a uniform and consistent Latin transliteration [ISO01] was used when the language of the release was not English.

[4] https://compmusic.upf.edu/corpora
[5] https://musicbrainz.org/

The test corpora (or test datasets) are designed for specific tasks and contain additional information such as annotations and derived data. They are useful for various melody and rhythm analysis tasks. There are several test datasets for different music cultures built within CompMusic[6].

1.4.2 Saraga - Open Data Collections for Indian Art Music

The CompMusic Carnatic and Hindustani research corpora are representative and vast, but they comprise copyrighted audio. While the audio can be used for personal use, we cannot easily disseminate it further. To avoid this bottleneck, CompMusic project also built Saraga - two large open data collections [Sri+21] of Indian Art Music. The Saraga Hindustani and Carnatic collections comprise 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 multi-track recordings, making it a valuable collection for research on melody extraction, source separation, automatic mixing, and performance analysis. An easy access to the audio, metadata, and the annotations in the collections is provided 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, a mechanism is provided for both the research and music community to contribute additional data and annotations to the collections. The paper also presents applications with the collections for music education, understanding, exploration, and discovery.

The data resources described in the section are collectively one of the largest curated and annotated sources of data for MIR and computational musicology research on Indian Art Music. They are carefully built to support multiple research problems and hence are valuable to the community to further research on IAM. They are disseminated through multiple publications [Gul16; Sri+14b; SS14a] that provide details about the content and potential applications of the datasets. The data is also accessible through different tools such mirdata, Dunya and PyCompMusic via easy to use APIs, each of which are described further in the next section.

1.5 Tools and Frameworks

The previous sections introduced relevant research problems in IAM and data corpora available to work on those problems. Though the scope of this chapter did not permit extensive description of the problems or the data available, we hope that the pointers provided and references will help to gather additional information on specific problems readers are interested in. Further, the

[6] http://compmusic.upf.edu/datasets

introductory scope of this chapter also did not let us describe different methods taken up by the community so far for the research problems of interest. Other chapters of the monograph are expected to provide a survey on different problems and approaches explored so far.

In this section, we provide complementary information by introducing tools and frameworks we can use to annotate, store, archive, organize, retrieve, manually analyze and consume the datasets in different MIR experiments on the CompMusic and Saraga datasets. These tools and frameworks are more of enablers for research rather than methods of analysis, but nonetheless important to the access of data and some basic analysis. We do not claim to be comprehensive with the tools and frameworks presented here, but we describe the tools that have been used for the mentioned tasks in the past and have been adequate according to our experience with them.

1.5.1 Tools for Annotations and Analysis

Tools for annotation help us to annotate different tags and time-aligned events on an audio recording. In the process, they also enable us to do manual analysis on the recordings, which could be helpful to build hypotheses, propose analysis methods and design large scale experiments. Sonic Visualiser[7] [CLS10] is a free, open-source application useful for visualization and annotation of audio. It is designed for musicologists, archivists, signal-processing researchers, and anyone who wishes to carefully study an audio music recording. Sonic Visualiser is a program for highly configurable detailed visualisation, analysis, and annotation of audio recordings.

Sonic Visualizer supports the Vamp audio analysis plugin system[8], an audio processing plugin system for plugins that extract descriptive information from audio data - typically referred to as audio analysis plugins or audio feature extraction plugins. Some of the most common feature extraction methods (e.g. melodia for predominant melody extraction) are provided as Vamp plugins that could be used within Sonic Visualizer. For batch audio feature extraction, we can use Sonic Annotator, which is a non-interactive command-line program application that can batch apply the the same feature extraction plugins as Sonic Visualiser.

1.5.2 Tools for Archival, Retrieval, Data Access and Evaluation

The frameworks for archival and retrieval are large linked data collections that can store different music related metadata and features and provide APIs for accessing the collections. The CompMusic data collections extensively use MusicBrainz, AcousticBrainz and Dunya for the purpose, supported by APIs.

MusicBrainz is an open music encyclopedia that collects music metadata and makes it available to the public. It is slated to the community curated and community maintained source

[7] https://www.sonicvisualiser.org/
[8] https://www.vamp-plugins.org/

of music information releasing the data under open licenses. It provides a reliable and unambiguous form of music identification through unique IDs called MusicBrainz ID. The AcousticBrainz[9] project aims to crowd source acoustic information for all music in the world and to make it available to the public. This acoustic information describes the acoustic characteristics of music and includes low-level spectral information and information for genres, moods, keys, scales and much more. The goal of AcousticBrainz is to provide music technology researchers with a massive database of information about music. AcousticBrainz is a joint effort between Music Technology Group at Universitat Pompeu Fabra in Barcelona and the MusicBrainz project. At the heart of this project lies the Essentia[10] [Bog+13] toolkit from the MTG - this open source toolkit enables the automatic analysis of music. The output from Essentia is collected by the AcousticBrainz project and made available to the public. AcousticBrainz organizes the data on a recording basis, indexed by the MusicBrainz ID for recordings.

Dunya[11][PSS13a] comprises the music corpora and related software tools that have been developed as part of the CompMusic project. These corpora have been created with the aim of studying particular music traditions and they include audio recordings plus complementary information that describes the recordings. Each corpus has specific characteristics and the developed software tools allow to process the available information in order to study and explore the characteristics of each musical repertoire. The entire Carnatic and Hindustani research corpora are available through Dunya, which internally uses MusicBrainz IDs for identification and linking entities.

The Saraga[12] data collections distributed under Creative Commons licenses can be accessed in two primary ways. The data in the collections can be accessed through the PyCompMusic API built to access Dunya collections. In addition, the Saraga repository[13] provides a dump of all editorial metadata and some annotations. More recently, the Saraga collections are also available through mirdata[14], which is an open-source Python library that provides tools for working with common MIR datasets, including tools for downloading datasets to a common location and format, validating that the files for a dataset are all present, loading annotation files to a common format, consistent with mir_eval [Raf+14] and parsing track level metadata for detailed evaluations.

1.6 Summary

The chapter, targeted at musicians, music students, technology enthusiasts, engineers and researchers, aimed to provide a context of the current state of the art that could help them start

[9] https://acousticbrainz.org/
[10] https://essentia.upf.edu/
[11] https://dunya.compmusic.upf.edu/
[12] https://mtg.github.io/saraga/
[13] https://github.com/MTG/saraga
[14] https://mirdata.readthedocs.io/

their work with computational analysis of Indian art music. While emphasizing the need for laborious culture-aware and culture-specific approaches to computational musicology and music information research (MIR), we argued that we could alternatively also seek and identify parallel tasks, data and methods from the current state of the art in other music cultures and use them for a preliminary and basic analysis of culture-specific tasks. We noted that such a perspective will enable preliminary analysis of a music culture using existing methods, and integrate it as a use-case with existing common frameworks and approaches in MIR and data-driven computational musicology.

Within this context, the chapter provided a gentle introduction describing common concepts, frameworks, approaches, resources, data and methods for computational analysis of music from a perspective that could be useful for the analysis of Indian Art music. We hope that with the pointers provided in the chapter on the parallels between music concepts, relevant research problems, data corpora available along with tools available to access them would encourage further community-driven efforts and research into various open research problems in Indian Art Music.

1.7 References

[ABM13] Anantapadmanabhan, A., Bellur, A., and Murthy, H. A. "Modal analysis and transcription of strokes of the mridangam using non-negative matrix factorization". In: *Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*. Vancouver, Canada, May 2013, pp. 181–185.

[Ana+14a] Anantapadmanabhan, A. et al. "Tonic-Independent Stroke Transcription of the Mridangam". In: *Proceedings of the 53rd AES International Conference on Semantic Audio*. London, UK, Jan. 2014.

[Bag98b] Bagchee, S. *Nād: Understanding Raga Music*. Business Publications Inc., 1998.

[Bel+05] Bello, J. P. et al. "A Tutorial on Onset Detection in Music Signals". In: *IEEE Transactions on Speech and Audio Processing* vol. 13, no. 5 (Sept. 2005), 1035–1047.

[BHL01] Berners-Lee, T., Hendler, J., and Lassila, O. "The semantic web". In: *Scientific American* vol. 284, no. 5 (2001), pp. 34–43.

[BL04] Brachman, R. and Levesque, H. *Knowledge Representation and Reasoning (The Morgan Kaufmann Series in Artificial Intelligence)*. Morgan Kaufmann, May 2004.

[Bog+13] Bogdanov, D. et al. “ESSENTIA: an Audio Analysis Library for Music Information Retrieval”. In: *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR 2013)*. Curitiba, Brazil, Nov. 2013, pp. 493–498.

[Cho+10] Chordia, P. et al. “Multiple viewpoints modeling of tabla sequences”. In: *Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR 2010)*. Utrecht, Netherlands, Aug. 2010, pp. 381–386.

[Cho06] Chordia, P. “Automatic transcription and representation of solo tabla music”. In: *Computing in Musicology* vol. 14 (Aug. 2006), pp. 123–138.

[Cho07] Chordia, P. “A System for the Analysis and Representation of Bandishes and Gats Using Humdrum Syntax”. In: *Proceedings of Frontiers of Research in Speech and Music Conference*. 2007.

[Cla00a] Clayton, M. *Time in Indian Music : Rhythm, Metre and Form in North Indian Rag Performance*. Oxford University Press, 2000.

[CLS10] Cannam, C., Landone, C., and Sandler, M. “Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files”. In: *Proceedings of the ACM Multimedia 2010 International Conference*. Florence, Italy, Oct. 2010, pp. 1467–1468.

[CSA10] Chordia, P., Sastry, A., and Albin, A. “Evaluating Multiple Viewpoints Models of Tabla Sequences”. In: *Proceedings of ACM Multimedia Workshop on Music and Machine Learning*. Florence, Italy, Oct. 2010, pp. 21–24.

[DP06] Davies, M. E. P. and Plumbley, M. D. “A spectral difference approach to downbeat extraction in musical audio”. In: *Proceedings of the 14th European Signal Processing Conference (EUSIPCO 2006)*. Florence, Italy, Sept. 2006.

[DP07] Davies, M. E. P. and Plumbley, M. D. “Context-Dependent Beat Tracking of Musical Audio”. In: *IEEE Transactions on Audio, Speech, and Language Processing* vol. 15, no. 3 (Mar. 2007), pp. 1009–1020.

[Foo00a] Foote, J. “Automatic audio segmentation using a measure of audio novelty”. In: *Proceedings of the IEEE International Conference on Multimedia and Expo 2000*. Vol. 1. New York, USA, Aug. 2000, pp. 452–455.

[Gan+16] Ganguli, K. K. et al. “Data-driven Exploration of Melodic Structures in Hindustani Music”. In: *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR 2016)*. New York (USA), Aug. 2016, pp. 605–611.

[GFC04] Gómez-Pérez, A., Fernández-López, M., and Corcho, O. *Ontological Engineering with Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web. Series: Advanced Information and Knowledge Processing*. 1st. Springer, 2004.

[Gow+21] Gowriprasad, R. et al. "Tabla Gharana Recognition from Audio music recordings of Tabla Solo performances". In: *Proceedings of the 22nd International Society for Music Information Retrieval Conference* (Online). Online, Nov. 2021, pp. 547–554.

[GR18] Ganguli, K. K. and Rao, P. "On the Distributional Representation of Ragas: Experiments with Allied Raga Pairs". In: *Transactions of the International Society for Music Information Retrieval* vol. 1, no. 1 (2018), pp. 79–95.

[GR21] Ganguli, K. and Rao, P. "A study of variability in raga motifs in performance contexts". In: *Journal of New Music Research* vol. 50, no. 1 (2021), pp. 102–116.

[Gul11] Gulati, S. "A Tonic Identification Approach for Indian Art Music". Master's Thesis. Barcelona, Spain: Universitat Pompeu Fabra, 2011.

[Gul16] Gulati, S. "Computational Approaches for Melodic Description in Indian Art Music Corpora". Doctoral dissertation. Barcelona, Spain: Universitat Pompeu Fabra, 2016.

[Gup+15] Gupta, S. et al. "Discovery of Syllabic Percussion Patterns in Tabla Solo Recordings". In: *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)*. Malaga, Spain, Oct. 2015, pp. 385–391.

[Han89] Handel, S. *Listening: An Introduction to the Perception of Auditory Events*. MIT Press, 1989.

[HDF12] Hockman, J. A., Davies, M. E. P., and Fujinaga, I. "One in the Jungle: Downbeat Detection in Hardcore, Jungle, and Drum and Bass". In: *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*. Porto, Portugal, Oct. 2012, pp. 169–174.

[HS11] Holzapfel, A. and Stylianou, Y. "Scale transform in rhythmic similarity of music". In: *IEEE Transactions on Speech and Audio Processing* vol. 19, no. 1 (2011), pp. 176–185.

[Hur02] Huron, D. "Music information processing using the humdrum toolkit: Concepts, examples, and lessons". In: *Computer Music Journal* vol. 26, no. 2 (2002), pp. 11–26.

[ISO01] ISO/TC. *ISO/TC 15919:2001 Information and documentation – Transliteration of Devanagari and related Indic scripts into Latin characters*. Geneva, Switzerland: International Organization for Standardization, 2001.

[KBW15] Krebs, F., Böck, S., and Widmer, G. “An Efficient State-Space Model for Joint Tempo and Meter Tracking”. In: *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)*. Malaga, Spain, Oct. 2015, pp. 72–78.

[KI12a] Krishna, T. M. and Ishwar, V. “Karnātik music: Svara, Gamaka, Phraseology and Raga Identity”. In: *Proceedings of the 2nd CompMusic Workshop*. Istanbul, Turkey, July 2012, pp. 12–18.

[Kod+12a] Koduri, G. K. et al. “Rāga recognition based on pitch distribution methods”. In: *Journal of New Music Research* vol. 41, no. 4 (2012), pp. 337–350.

[Kod+14a] Koduri, G. K. et al. “Intonation analysis of ragas in Carnatic music”. In: *Journal of New Music Research* vol. 43, no. 1 (2014), pp. 73–94.

[Kur+15] Kuriakose, J. et al. “Akshara Transcription of Mrudangam Strokes in Carnatic Music”. In: *Proceedings of the 21st National Conference on Communication (NCC)*. Mumbai, India, Feb. 2015.

[LPS19] Lluís, F., Pons, J., and Serra, X. *End-to-end music source separation: is it possible in the waveform domain?* 2019. arXiv: 1810.12187 [cs.SD].

[MD12] Mauch, M. and Dixon, S. “A Corpus-based Study of Rhythm Patterns”. In: *Proceedings of the 13th International Conference on Music Information Retrieval (ISMIR 2012)*. Porto, Portugal, Oct. 2012, pp. 163–168.

[PF12] Peeters, G. and Fort, K. “Towards a (Better) Definition of the Description of Annotated MIR Corpora”. In: *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*. Porto, Portugal, Oct. 2012, pp. 25–30.

[PMK10] Paulus, J., Müller, M., and Klapuri, A. “Audio-based Music Structure Analysis”. In: *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*. Utrecht, The Netherlands, 2010, pp. 625–636.

[PP11] Peeters, G. and Papadopoulos, H. “Simultaneous Beat and Downbeat-Tracking Using a Probabilistic Framework: Theory and Large-Scale Evaluation”. In: *IEEE Transactions on Audio, Speech and Language Processing* vol. 19, no. 6 (2011), pp. 1754–1769.

[PSS13a] Porter, A., Sordo, M., and Serra, X. “Dunya: A System for Browsing Audio Music Collections Exploiting Cultural Context”. In: *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR 2013)*. Curitiba, Brazil, Nov. 2013, pp. 101–106.

[Raf+14] Raffel, C. et al. “mir_eval: A Transparent Implementation of Common MIR Metrics”. In: *Proceedings of the 15th International Society for Music Information Retrieval (ISMIR) Conference*. Taipei, Taiwan, Oct. 2014.

[Ran+19a] Ranjani, H. G. et al. "A compact pitch and time representation for melodic contours in Indian art music". In: *The Journal of the Acoustical Society of America* vol. 145, no. 1 (2019), pp. 597–603.

[RS13a] Ranjani, H. and Sreenivas, T. "Grouping carnatic music notes using a multi-gram language model". In: *Proceedings of Acoustics 2013*. New Delhi, India, Nov. 2013.

[RS15] Ranjani, H. G. and Sreenivas, T. V. "Multi-instrument detection in polyphonic music using Gaussian Mixture based factorial HMM". In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2015, pp. 191–195.

[Sam98b] Sambamoorthy, P. *South Indian Music Vol. I-VI*. The Indian Music Publishing House, 1998.

[Sar11] Saraf, R. *Development of Hindustani Classical Music (19th & 20th centuries)*. Vidyanidhi Prakashan, 2011.

[SC12a] Srinivasamurthy, A. and Chordia, P. "A Unified System for Analysis and Representation of Indian Classical Music using Humdrum Syntax". In: *Proceedings of the 2nd CompMusic Workshop*. Istanbul, Turkey, July 2012, pp. 38–42.

[SC12b] Srinivasamurthy, A. and Chordia, P. "Multiple Viewpoint Modeling of North Indian Classical Vocal Compositions". In: *Proceedings of the 9th International Symposium on Computer Music Modeling and Retrieval (CMMR)*. London, UK, June 2012, pp. 344–356.

[Ser11] Serra, X. "A multicultural approach in Music Information Research". In: *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*. Miami, USA, Oct. 2011, pp. 151–156.

[Ser14] Serra, X. "Creating Research Corpora for the Computational Study of Music: the case of the CompMusic Project". In: *Proceedings of the 53rd AES International Conference on Semantic Audio*. London, Jan. 2014.

[SG12b] Salamon, J. and Gómez, E. "Melody Extraction From Polyphonic Music Signals Using Pitch Contour Characteristics". In: *IEEE Transactions on Audio, Speech, and Language Processing* vol. 20, no. 6 (Aug. 2012), pp. 1759–1770.

[SHS14] Srinivasamurthy, A., Holzapfel, A., and Serra, X. "In Search of Automatic Rhythm Analysis Methods for Turkish and Indian Art Music". In: *Journal of New Music Research* vol. 43, no. 1 (2014), pp. 97–117.

[SHS17] Srinivasamurthy, A., Holzapfel, A., and Serra, X. "Informed Automatic Meter Analysis of Music Recordings". In: *Proceedings of the 18th International Society for Music Information Retrieval (ISMIR) Conference*. Suzhou, China, Oct. 2017, pp. 679–685.

[SIS19] Srinivasamurthy, A., Ishwar, V., and Serra, X. “Statistical Analysis of Tempo and Rhythmic Elaboration in Carnatic Music”. In: *Proceedings of the 24th International Symposium: Frontiers of Research in Speech and Music (FRSM - 2019)*. Kanpur, India, July 2019.

[Sma04] Smaragdis, P. “Non-negative Matrix Factor Deconvolution; Extraction of Multiple Sound Sources from Monophonic Inputs”. In: *Independent Component Analysis and Blind Signal Separation*. Ed. by Puntonet, C. G. and Prieto, A. Vol. 3195. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, pp. 494–499.

[Sri+12] Srinivasamurthy, A. et al. “A Beat Tracking Approach to Complete Description of Rhythm in Indian Classical Music”. In: *Proceedings of the 2nd CompMusic Workshop*. Istanbul, Turkey, July 2012, pp. 72–78.

[Sri+14a] Srinivasamurthy, A. et al. “Transcription and Recognition of Syllable based Percussion Patterns: The Case of Beijing Opera”. In: *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014)*. Taipei, Taiwan, Oct. 2014, pp. 431–436.

[Sri+14b] Srinivasamurthy, A. et al. “Corpora for Music Information Research in Indian Art Music”. In: *Proceedings of Joint 42nd International Computer Music Conference/13th Sound and Music Computing Conference*. Athens, Greece, Sept. 2014, pp. 1029–1036.

[Sri+15] Srinivasamurthy, A. et al. “Particle Filters for Efficient Meter Tracking with Dynamic Bayesian Networks”. In: *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR 2015)*. Malaga, Spain, Oct. 2015, pp. 197–203.

[Sri+16] Srinivasamurthy, A. et al. “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 (ICASSP 2016)*. Shanghai, China, Mar. 2016, pp. 76–80.

[Sri+17b] Srinivasamurthy, A. et al. “Aspects of Tempo and Rhythmic Elaboration in Hindustani Music: A Corpus Study”. In: *Frontiers in Digital Humanities* vol. 4 (2017).

[Sri+21] Srinivasamurthy, A. et al. “Saraga: Open Datasets for Research on Indian Art Music”. In: *Empirical Musicology Review* vol. 16, no. 1 (2021), pp. 85–98.

[Sri16] Srinivasamurthy, A. “A Data-driven Bayesian Approach to Automatic Rhythm Analysis of Indian Art Music”. Doctoral dissertation. Barcelona, Spain: Universitat Pompeu Fabra, 2016.

[SS14a] Srinivasamurthy, A. and Serra, X. “A Supervised Approach to Hierarchical Metrical Cycle Tracking from Audio Music Recordings”. In: *Proceedings of the 39th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*. Florence, Italy, May 2014, pp. 5237–5241.

[Wil+16] Wilkinson, M. D. et al. “The FAIR Guiding Principles for scientific data management and stewardship”. In: *Scientific Data* vol. 3, no. 1 (2016).