Towards a multimodal knowledge base for Indian art music: A case study with melodic intonation

Gopala Krishna Koduri

TESI DOCTORAL UPF / 2016

Director de la tesi

Dr. Xavier Serra i Casals
Music Technology Group
Department of Information and Communication Technologies

The doctoral defense was held on …… at the Universitat Pompeu Fabra and scored as ……

Dr. Xavier Serra Casals
Thesis Supervisor
Universitat Pompeu Fabra, Barcelona

Dr. Anja Volk
Thesis Committee Member
Utrecht University, Netherlands

Dr. Baris Bozkurt
Thesis Committee Member
Koç University, Turkey

Dr. George Fazekas
Thesis Committee Member
Queen Mary University of London, UK

To Amma.

This thesis has been carried out between Oct. 2011 and Oct. 2016 at the Music Technology Group (MTG) of Universitat Pompeu Fabra (UPF) in Barcelona (Spain), supervised by Dr. Xavier Serra Casals. The work in ch. 3 has been conducted in collaboration with Dr. Preeti Rao (IIT-B, Mumbai, India). The work in ch. 5 and part of ch. 6 has been conducted in collaboration with Dr. Joan Serrà (Telefónica R&D, Barcelona, Spain). The work in ch. 7 has been carried out with help from the CompMusic team, mainly Vignesh Ishwar, Ajay Srinivasamurthy and Hema murthy. The work in ch. 8 has been conducted in collaboration with Sertan Senturk in the CompMusic team. This work has been supported by the Dept. of Information and Communication Technologies (DTIC) PhD fellowship (2011-16), Universitat Pompeu Fabra and the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

Acknowledgements

IIIT-H (International Institute of Information Technology - Hyderabad), my alma mater, had a lasting impact on my thought processes and life goals. More importantly, my experiences during the time there helped me realize the relevance of inclusive growth and development to Indian society, which is etched with diversity in almost every aspect of life. My further academic pursuit is greatly influenced by this. I'm deeply thankful to everyone who has been part of that journey, especially Prof. Bipin Indurkhya who encouraged us to pursue our interests in a stress-free environment and provided us ample opportunities to excel, and Pranav Kumar Vasishta, who with his constructive discourses has shed much light on engaging topics ranging from individual freedom to societal structures.

It was during my internship with Preeti Rao in IIT-B (Indian Institute of Technology - Bombay) during 2010, that I really found camaraderie in my academic work. I'm very thankful to her for providing that opportunity. It was also there that I first met Xavier Serrra who was there to present the work done at MTG and an introduction to CompMusic project. It is also there that I first met Sankalp Gulati, whose guidance in melodic-analysis had been immensely helpful when I was just getting started. After the internship that Xavier offered me at MTG, it was crystal clear that the objectives of CompMusic project completely aligned with whatever I planned after my stint at IIIT-H.

My stay at MTG has been a thoroughly rewarding experience to say the least. To have worked with/alongside the most brilliant minds in the MIR domain is humbling. Xavier's exceptional foresight and prowess in planning are something that I will always reflect on as a constant source of learning. Without his guidance and support, this work would not have been the same. I'm deeply grateful to him for this opportunity which was provided with plentiful freedom and en-

couragement. I thank Joan Serrà who guided my work on intonation description during the initial years. His attention to detail and academic rigor are inspiring. Thanks to Emilia Gómez for letting me audit her course on basics of signal processing which later helped me greatly in my work. Thanks to Hendrik Purwins who always had been very approachable and the goto guide for anything related to machine learning.

The CompMusic project has brought together an amazing group of people from around the world, and it is fortunate for having been part of it. They have not only been my colleagues at work, but also the social circle outside work who have lent every help in calling Barcelona my home. Special thanks to (in no particular order) Ajay Srinivasamurthy, Sankalp Gulati, Sertan Senturk, Vignesh Ishwar, Kaustuv Kanti Ganguli, Marius Miron, Swapnil Gupta, Mohamed Sordo, Alastair Porter, Sergio Oramas, Rafael Caro Rapetto, Rong Gong, Frederic Font, Georgi Dzhambazov, Andrés Ferraro, Vinutha Prasad, Shuo Zhang, Shrey Dutta, Padi Sarala, Hasan Sercan Atlı, Burak Uyar, Joe Cheri Ross, Sridharan Sankaran, Akshay Anantapadmanabhan, Jom Kuriakose, Jilt Sebastian and Amruta Vidwans. I had an opportunity to interact with almost everyone at MTG which helped me in one way or another. Thanks to Srikanth Cherla, Dmitry Bogdanov, Juan José Bosch, Oscar Mayor, Panos Papiotis, Agustín Martorell, Sergio Giraldo, Nadine Kroher, Martí Umbert, Sebastián Mealla, Zacharias Vamvakousis, Julián Urbano, Jose Zapata, Justin Salamon and Álvaro Sarasúa. I thank Cristina Garrido, Alba Rosado and Sonia Espí for making my day-to-day life at MTG a lot easier. Also thanks to Lydia García, Jana Safrankova, Vanessa Jimenez and all other university staff for helping me wade through the administrative formalities with ease. Special thanks to my multilingual colleagues, Frederic Font and Rafael Caro Rapetto, for helping me with Catalan and Spanish translations of the abstract.

I thank Hema murthy, Preeti rao, T. M. Krishna, Suvarnalata Rao, N. Ramanathan and M. Subramanian for their feedback and help in several important parts of the thesis. I thank all my co-authors and collaborators for their contributions. I thank the (anonymous or otherwise) reviewers of several publications for their detailed feedback which helped me constantly improve this work. I also thank the members of the thesis committee for offering their invaluable time and feedback.

Barcelona is definitely not the same without the friends I made in this vibrant city. For all the treks, tours, dinners, cultural exchanges, volleyball and mafia games, many thanks to Ratheesh, Manoj, Shefali, Jordi, Indira, Laia, Alex, Srinibas, Praveen, Geordie, Princy, Kalpani, Windhya, Ferdinand, Pallabi, Neus and Waqas. Warmest gratitude to Eva, the most empathetic person I have ever known. To a part-time introvert like me, it is just magic how she makes friends in a jiffy

and make them feel comfortable as if they have known each other for their whole life!

Finally, none of this would have been possible but for the iron will of my mother who religiously sacrificed anything and everything for my sister's and my education. For somebody who is barely exposed to the world outside our home, I'm always puzzled by her mental stamina. Nothing can possibly express my gratitude to her, and my father who firmly stood by her. Thanks to the more-courageous-more-patient version of me - my sister, and the wonderful being she hooked me up with - my wife, for bearing and handling my sudden disappearances in the hours of the need during these years.

Gopala Krishna Koduri

2^{nd} November 2016

Abstract

This thesis is a result of our research efforts in building a multi-modal knowledge-base for the specific case of Carnatic music. Besides making use of metadata and symbolic notations, we process natural language text and audio data to extract culturally relevant and musically meaningful information and structuring it with formal knowledge representations. This process broadly consists of two parts. In the first part, we analyze the audio recordings for intonation description of pitches used in the performances. We conduct a thorough survey and evaluation of the previously proposed pitch distribution based approaches on a common dataset, outlining their merits and limitations. We propose a new data model to describe pitches to overcome the shortcomings identified. This expands the perspective of the note model in-vogue to cater to the conceptualization of melodic space in Carnatic music. We put forward three different approaches to retrieve compact description of pitches used in a given recording employing our data model. We qualitatively evaluate our approaches comparing the representations of pitched obtained from our approach with those from a manually labeled dataset, showing that our data model and approaches have resulted in representations that are very similar to the latter. Further, in a raaga classification task on the largest Carnatic music dataset so far, two of our approaches are shown to outperform the state-of-the-art by a statistically significant margin.

In the second part, we develop knowledge representations for various concepts in Carnatic music, with a particular emphasis on the melodic framework. We discuss the limitations of the current semantic web technologies in expressing the order in sequential data that curtails the application of logical inference. We present our use of rule languages to overcome this limitation to a certain extent. We then use open information extraction systems to retrieve concepts, entities and their relationships from natural language text concerning Carnatic music. We evaluate these systems using the concepts and relations from knowledge rep-

resentations we have developed, and groundtruth curated using Wikipedia data. Thematic domains like Carnatic music have limited volume of data available online. Considering that these systems are built for web-scale data where repetitions are taken advantage of, we compare their performances qualitatively and quantitatively, emphasizing characteristics desired for cases such as this. The retrieved concepts and entities are mapped to those in the metadata. In the final step, using the knowledge representations developed, we publish and integrate the information obtained from different modalities to a knowledge-base. On this resource, we demonstrate how linking information from different modalities allows us to deduce conclusions which otherwise would not have been possible.

Resum

Aquesta tesi és el resultat de la nostra investigació per a construir una base de coneixement multimodal per a la música Carnàtica. A part d'utilitzar metadades i representacions simbòliques musicals, també processem text en llenguatge natural i l'àudio mateix per tal d'extreure informació que sigui rellevant tant des d'un punt de vista cultural com musical i que puguem estructurar amb representacions formals de coneixement. El procés que seguim està compost principalment de dues parts. En la primera part analitzem les gravacions d'àudio per descriure'n l'entonació de les altures tonals utilitzades. Comparem i avaluem aproximacions existents basades en histogrames d'altures tonals utilitzant una base de dades comuna de referència i en subratllem els avantatges i les limitacions. Proposem un nou model de dades per descriure l'altura tonal de les notes i superar les limitacions prèviament identificades. Aquest model va més enllà dels ja establerts i permet acomodar la conceptualització de l'espai melòdic en la música Carnàtica. Utilitzant el nostre model de dades proposem tres mètodes diferents per extreure descripcions compactes de les altures tonals de les notes d'una gravació. Fem una avaluació qualitativa a través de la comparació de descripcions generades amb els mètodes proposats i descripcions generades manualment, i comprovem que els nostres mètodes generen descripcions molt semblants a les generades manualment. També comprovem com els nostres mètodes són útils per a la classificació de raga avaluant amb la base de dades més gran de música Carnàtica que s'ha creat fins al dia d'avui. Dos dels nostres mètodes obtenen puntuacions més altes que els millors mètodes existents, amb marges de millora estadísticament significatius.

En la segona part de la nostra investigació desenvolupem representacions de coneixement sobre diversos conceptes de la música Carnàtica, posant un èmfasi especial en aspectes melòdics. Parlem sobre les limitacions de les tecnologies de la web semàntica pel que fa a la representació del concepte d'ordre en dades se-

qüencials, fet que limita les possibilitats d'inferències lògiques. Proposem l'ús de llenguatges de normes per, fins a cert punt, superar aquestes limitacions. Després utilitzem sistemes d'extracció d'informació per recuperar conceptes, entitats i les seves relacions a partir de l'anàlisi de text natural sobre música Carnàtica. Avaluem aquests sistemes utilitzant conceptes i relacions extretes de representacions de coneixement que nosaltres mateixos hem desenvolupat i també utilitzant dades curades provinents de la Wikipedia. Per temàtiques com la música Carnàtica hi ha un volum de dades limitat accessible en línia. Tenint en compte que aquests sistemes estan pensats per funcionar amb grans volums de dades on les repeticions són importants, en fem una comparació qualitativa i quantitativa emfatitzant aquelles característiques més rellevants per casos amb volums de dades limitats. Els conceptes i entitats recuperades són emparellats amb conceptes i entitats presents a les nostres metadades. Finalment, utilitzant les representacions de coneixement desenvolupades, integrem les informacions obtingues de les diferents modalitats i les publiquem en una base de coneixement. Utilitzant aquesta base de coneixement demostrem com el fet de combinar informacions provinents de diferents modalitats ens permet arribar a conclusions que d'una altra manera no haurien estat possibles.

(Translated from English by Frederic Font)

Resumen

Esta tesis es resultado de nuestro trabajo de investigación para construir una base de conocimiento multimodal para el caso específico de la música carnática. Además de hacer uso de metadatos y notación simbólica, procesamos texto de lenguaje natural y datos de audio para extraer información culturalmente relevante y musicalmente significativa, y estructurarla con representaciones formales de conocimiento. En líneas generales, este proceso consiste en dos partes. En la primera parte, analizamos grabaciones de audio para describir la entonación de las alturas usadas en las interpretaciones. Llevamos a cabo un exhaustivo análisis y evaluación de los métodos basados en distribución de altura propuestos anteriormente, señalando sus ventajas y limitaciones. Proponemos un nuevo modelo de datos para la descripción de alturas con el fin de superar las limitaciones identificadas. Esto amplía la perspectiva del modelo actual de nota para contribuir a la conceptualización del espacio melódico en música carnática. Ofrecemos tres propuestas diferentes para la extracción de una descripción compacta de las alturas usadas en una grabación dada utilizando nuestro modelo de datos. Evaluamos cualitativamente nuestras propuestas comparando las representaciones de alturas obtenidas según nuestro método con aquellas procedentes de un conjunto de datos anotado manualmente, con lo que mostramos que nuestro modelo de datos y nuestras propuestas resultan en representaciones muy similares a estas últimas. Además, en una tarea de clasificación de raagas en el mayor conjunto de datos de música carnática hasta la fecha, dos de nuestras propuestas muestran mejor rendimiento que el estado del arte con un margen estadístico significativo.

En la segunda parte, desarrollamos representaciones de conocimiento para varios conceptos en música carnática, con un particular énfasis en el marco melódico. Discutimos las limitaciones de las tecnologías de web semántica actuales para expresar el orden de datos secuenciales, lo que restringe la aplicación de inferencia lógica. Presentamos nuestro uso de lenguajes de reglas para superar hasta cierto

punto esta limitación. A continuación utilizamos sistemas abiertos de extracción de información para extraer conceptos, entidades y sus relaciones a partir de texto de lenguaje natural relacionado con música carnática. Evaluamos estos sistemas usando los conceptos y las relaciones de las representaciones de conocimiento que hemos desarrollado, así como información de referencia contrastada con datos de Wikipedia. Dominios temáticos como el de música carnática tienen un volumen limitado de datos disponibles en internet. Considerando que estos sistemas están construidos para datos a escala de la web, en la que es posible beneficiarse de las repeticiones, comparamos sus rendimientos cualitativa y cuantitativamente, enfatizando las características deseadas para casos como este. Los conceptos y entidades extraídas son mapeadas a aquellos existentes en los metadatos. En el paso final, usando las representaciones de conocimiento desarrolladas, publicamos e integramos la información obtenida por diferentes modalidades en una base de conocimiento. Con este recurso demostramos como la conexión de información de diferentes modalidades nos permite deducir conclusiones que de otra manera no habrían sido posibles.

(Translated from English by Rafael Caro Repetto)

Contents

List of Figures

List of Tables

Part I

Setting the stage

Chapter 1

Introduction

"The main challenge is to gather musically relevant data of sufficient quantity and quality to enable music information research that respects the broad multi-modality of music."

— Serra et al. (2013)

Music information research (MIR) is an interdisciplinary domain requiring coordinated efforts between various fields of research including but not limited to information technology, music theory, musicology, audio engineering, digital signal processing and cognitive science (Futrelle and Downie (2002)). So far, majority of the work in this domain has been focused on analyses of audio-data, placing a special emphasis on digital signal processing, machine learning and information retrieval that facilitate feature extraction and creation of data models, aimed at addressing specific needs such as query by singing, structural analysis and so on. However, music data is not limited to audio recordings and is manifest across diverse modalities. The other equally important sources of music data are: i) scores/symbolic notations, ii) music-context that includes metadata, lyrics, artist biographical information and musicological resources, and iii) user-context including data that helps in modeling user behavior and in personalization of various systems such as a recommendation engine (see ch. 2 in Serra et al., 2013). We believe that music data available in these other modalities beyond audio music recordings has not been exploited to a desirable extent.

There does exist a sizeable body of work that addresses different problems in MIR domain putting forward hybrid approaches that feed on audio data, domain knowledge, user behavior and folksonomies (see Celma, 2010, and the references therein). However, the data sources under consideration remain islands of information in a vast majority of such approaches. We first explore the plausible reasons for this situation, and then discuss what fundamental problems arise as

a consequence. Data schemas are the most common means to structure data, ranging from data stored in XML documents to relational databases. Often, such schemas lack coherence even when comparing schemas of seemingly similar data sources by virtue of their content. To some degree, this is expected as such schemas are often driven by adhoc needs within an organization or a specific application. They are clearly not intended to be coherent in their scope or meaning. For instance, one schema may define a class of artists as people who play an instrument or sing. Another schema might also include people who compose music. Such differences are even more deeper in the case of music concepts. Another important reason for the status quo of the data sources is the way in which they are deployed in hybrid approaches. Often the data models are trained using the data sources separately from each other, and are only combined by weighing, switching and cascading their predictions/results (Song et al. (2012)).

This leads to a fundamental issue. For instance, the conceptual description or interpretation of a music concept for a specific need by itself is not usually the end goal of information researchers. It forms a basis to build their data models, but does not become a part of those models due to the aforementioned reasons. Consequently it is lost, making it difficult to be reproducible/accessible for reuse by other researchers and developers. This is mainly a result of the lack of semantics in their data models. It limits their comprehensibility, and poses difficult challenges to their comparison or integration. Further, the cultural differences only add more complexity to these already striking problems.

Advances in the domain of Semantic Web, and as a consequence in Knowledge Representation subdomain of Artificial Intelligence have resulted in software infrastructure that can be used in addressing these aforementioned issues. Researchers from Natural Language Processing and Bio-informatics communities have taken advantage of these developments with reasonably good success (Stevens et al., 2000; Fellbaum, 1998, are some examples). This thesis is a case-study of our end-to-end efforts in building a multi-modal knowledge-base (KB) for the specific case of Carnatic music. Through this, we hope to discover the problems involved in its pragmatic implementation in the music domain and address a few during the course.

We first discuss why this effort is important in sec. 1.1. Then in sec. 1.2, we present a concrete set of problems we chose to address in order to make our goal a viable one within the time constraints of this work. We then picture an overview of our process in sec. 1.3 and the various datasets and repertoires that were used during the course of this work in sec. 1.4. Finally, in sec. 1.5, we summarize our contributions and in sec. 1.6, discuss how we organized rest of the thesis.

1.1 Motivation

Our musical experiences are a combination of a variety of phenomena in our day-to-day lives. In today's connected world, such experiences have only multiplied. Each such experience results in a data trace that leads us to understand more about cultural and musical aspects of a community. These can include anything from a gossip on a public forum about a music entity, to a scholarly treatment of a music concept. Linking and integrating those data traces unleashes the latent potential of their semantics.

Further, the world wide web in its current form has blurred the boundaries between data producers and consumers. A typical example is an e-commerce portal. Vendors and customers in such portals both simultaneously play the roles of data producers and consumers. Vendors put up a catalog which consumers use to browse and make purchase decisions. On the other hand, the browsing and buying patterns generated by consumers are used by vendors in optimizing their offerings. This is a mutually rewarding data cycle. Other such examples include social networks, media portals and so on. This fundamental behavior where the erstwhile consumers actively engage with the services to produce more data, has resulted in data of mammoth proportions, often openly accessible on the web. Exploitation of such data in its various modalities becomes even more important when seen in this context.

On the other hand, the cultural aspect of the music assumes great importance in higher-level tasks such as judging the similarity between music entities and recommending new music (Serra (2011)). Analyses of audio music recordings can only deliver us insights into the musical properties. The rest has to be gathered from the aforementioned data traces that people generate in their musical experiences. The information extracted from such different sources, including audio music recordings, are often complementary in nature and mutually enrich their value.

Therefore, there is an impending need for multi-modal KBs that have an immense potential in many MIR-related higher level tasks. The objective of this work is to present a case-study that discusses and documents our efforts in building a KB. We have chosen Carnatic music for this case-study for two important reasons: i) It is culturally very different from the music cultures that MIR community frequents to. Therefore, from our understanding of this tradition, we believe it poses a set of new challenges that will need solutions different from the state-of-the-art, ii) It is one of the most thoroughly explored music tradition from a musicological point of view. There exist compendiums and works that span the last two millennia, and

it is an active contemporary form with scholarly conventions and music festivals held throughout the year.

1.2 Problem statement

Keeping the viability of our goal in mind, we have constrained it to address the following specific problems. Understanding some of these may require the reader to go through the introduction to Carnatic music concepts in ch. 2 and/or the introduction to semantic web in ch. 4.

- *Extract a musically meaningful representation of svara intonation in Carnatic music, analyzing audio music recordings.*

  Svaras in Carnatic music are analogous to, but quite different from notes in western popular music. In this part of our work, we develop approaches to model the svaras and obtain a representation out of which we can extract musically meaningful information. It is well known that a svara is not a fixed frequency position, unlike a note. The continuous melodic movements are integral to the identity of melodic atoms used in Indian art music. Therefore, in order to better understand the nature of pitches used in a recording, our data models need to go beyond studying intervals or steady notes.

- *Extract music concepts and the corresponding individual entities along with relevant relations from natural language text.*

  We make use of the state-of-the-art Natural Language Programming (NLP) techniques to identify musical concepts and their corresponding individual entities that find a mention in natural language text. Examples include *Composer* concept, in which *Tyagaraja* and *Annamacharya* are individual entities. We also extract the relations between them that are of relevance. For instance, one such extraction can be Tyagaraja *composed* Endaro Mahanubhavulu. Here, *composed* is a relation between a composer and a composition. We exploit multiple semi-structured and unstructured text resources available online, such as Wikipedia, Rasikas.org forum and raaga listings on several music portals. The main challenge in this task is the recall rate which is severely limited by the quantity of text available. Therefore, our contribution is in adapting state-of-art concept and relation extraction engines to a niche domain such as a specific music tradition.

- *Build an ontology to structure information extracted from audio music recordings and natural language text.*

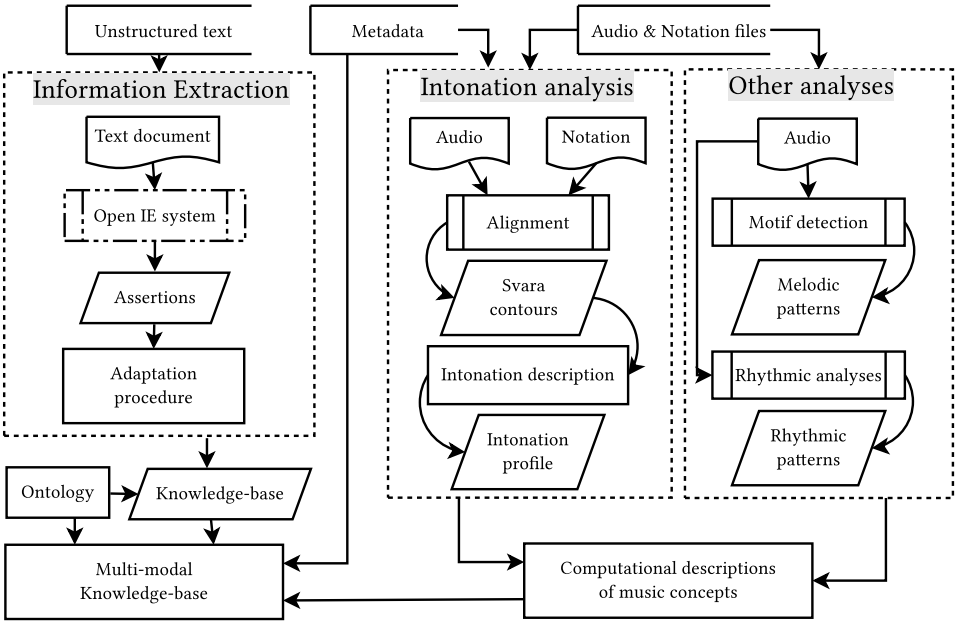


Figure 1.1: Overview of our work. Notice that the following modules - Alignment, Motif detection and Rhythmic analyses are not part of the work reported in this thesis. They are part of CompMusic project. Also, the Open Information Extractions systems that we use are neither part of the thesis, nor the project.

Ontology is a formal specification of a data schema that makes the semantics of the schema explicit. To our knowledge, there are no ontologies developed for Indian art music so far. The formalization of music concepts in this tradition, which is contrastingly different compared to western popular/classical music, pose challenges in designing suitable patterns. Therefore, while reusing several existing ontologies, we also define new vocabularies for modeling the semantics of various Carnatic music concepts. This is later used in building the KB, thus publishing the information extracted so far, while simultaneously enriching it.

1.3 An overview of the proposed approach

Fig. 1.1 shows a complete overview of the work reported in this thesis. We explain the flow presented in the figure in a bottom-up manner. Remember that the end goal of the thesis is creation of a multi-modal KB. For this, we build ontologies necessary for structuring information that results from different processes.

These processes include the three blocks marked as i) Information extraction, ii) Intonation analysis and iii) Other analyses. The modules in the last block, viz., Motif detection and Rhythmic analyses are not part of our work, but are carried out as part of the CompMusic project (Serra (2011))[1]. The information extracted using those processes is to be published to the KB, and hence are an important consideration in building ontologies.

In the first block, we take in a set of text documents and convert them to structured set of facts published using the ontologies. This is done using an Open Information Extraction system to identify concepts and relations in the text, which are further filtered using an adaptation procedure put in place specifically for this domain. In the second block, we analyze the audio music recordings in conjunction with corresponding (approximate) notation to obtain intonation description of constituent svaras. In the final step, this information is published using the ontologies and is merged with the one obtained using the processes in the first block.

1.4 Datasets

CompMusic corpora

Throughout the duration of the CompMusic project, one of the major efforts has been to improve the quantity and quality of the data collection curated for research. This corpora covers five distinct music traditions: Carnatic and Hindustani from India, Beijing opera from China, Makam music of Turkey and Arab-andalusian music from Maghreb region. It includes the metadata, audio recordings and various features and data extracted from the audio such as melody, tonic, onsets and so on. All the metadata has been uploaded to MusicBrainz coordinating the efforts to make sure that their schema supports these music traditions.

Further, through regular systematic checks, the data is verified for wrong labels and other errors. Except the audio, all the data is available online through a navigation portal, Dunya (Porter et al. (2013))[2] and also through an API attached to the portal[3]. Srinivasamurthy et al. (2014b) gives a thorough account of the efforts that have gone into the creation, maintenance and dissemination of this corpora. The reference also explains how this corpora compares with data available on commercial data portals in terms of size and completeness. All the datasets used

[1] http://compmusic.upf.edu
[2] http://dunya.compmusic.upf.edu
[3] https://github.com/MTG/pycompmusic

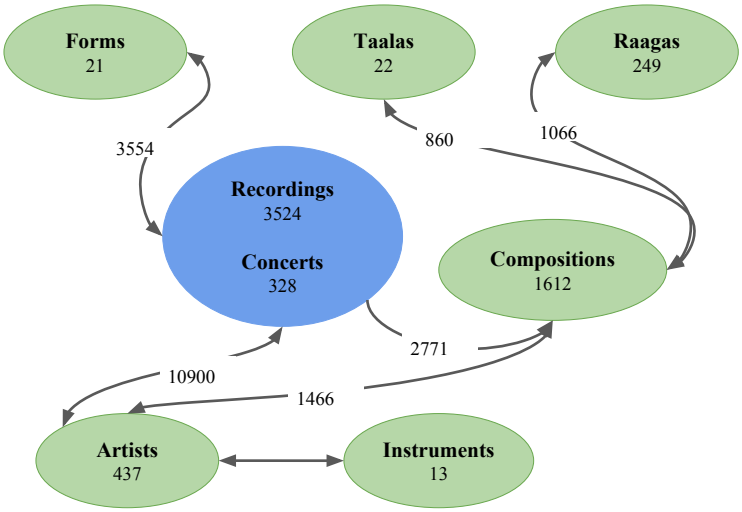


Figure 1.2: Statistics of the Carnatic corpus, which is a part of CompMusic corpora. The numbers on connecting arrows indicate the relations between corresponding music concepts as annotated on MusicBrainz.

in this work for svara intonation description are drawn from the Carnatic corpus. So far, this is the largest collection available for research on Carnatic music. Fig. 1.2 gives an idea of its size.

Raaga dataset 1

During the course of this thesis, we have drawn several datasets from the Carnatic corpus each meeting a certain set criteria. This criteria depends on both the task at hand and the size of the corpus at that point in time. Raaga dataset 1 is one of the first datasets drawn from the Carnatic corpus. We chose only those rāgas for which there are at least 5 recordings. Table 1.1 gives the complete details of the dataset. We use this dataset in the evaluation of approaches developed early in the thesis (ch. 5).

Rāga	Recordings	Duration (minutes)	Lead artists	Concerts
Ābhōgi	5	104	4	5
Ānandabhairavi	10	85	10	10
Asāvēri	14	134	13	12
Aṭāṇa	6	42	6	6
Bēgaḍa	10	74	8	8
Behāg	8	47	7	8
Bhairavi	18	411	15	17
Bilahari	7	107	6	6
Dēvagāndhāri	6	42	6	6
Dēvamanōhari	5	53	4	4
Dhanāśri	7	8	2	2
Dhanyāsi	8	141	6	8
Dhīraśankarābharanaṁ	16	367	10	15
Haṁsadhvani	11	95	10	11
Hari kāṁbhōji	8	115	8	8
Hindōlaṁ	8	113	6	6
Jaunpuri	5	17	5	5
Kāpi	7	31	6	6
Kalyāṇi	14	303	11	13
Kamās	19	230	13	13
Kāmbhōji	11	265	10	11
Karaharapriya	9	195	8	8
Kēdāragauḷa	5	58	5	5
Madhyamāvati	10	100	10	9
Mānji	5	49	5	5
Mōhanaṁ	8	127	8	8
Mukhāri	8	81	8	8
Nagasvarāḷi	5	28	5	5
Nāṭakuranji	7	88	6	7
Pantuvarāḷi	17	257	16	15
Pūrvīkalyāṇi	9	177	7	9
Ranjani	5	74	4	4
Rītigaula	5	70	5	5
Sahāna	7	103	6	6
Saurāṣṭraṁ	40	44	9	9
Senchuruṭṭi	7	28	6	6
Ṣanmukhapriya	5	96	5	4
Śrīranjani	5	47	5	5
Śudhdha sāvēri	6	96	6	6
Sindhu bhairavi	6	28	5	5
Suraṭi	9	78	8	9
Tōḍi	27	841	19	21
Vāchaspati	5	46	1	1
Vasanta	6	42	5	5
Yadukula kāṁbhōji	5	54	5	5
45 rāgas	**424**	**5617**	**38**	**62**

Table 1.1: Detailed statistics of the raaga dataset 1 (RD1).

Raaga dataset 2

This dataset is drawn post consolidation of the Carnatic corpus. To our knowledge, it is by far the largest Carnatic music dataset used for research. It is first used in an evaluation by Gulati et al. (2016b). We use this dataset for most of the evaluations performed towards the end of the thesis (ch. 6). It consists of 40 raagas, each with 12 recordings. These recordings together account for 65 vocal artists and 311 compositions in different forms such as varnam, kriti, keertana and so on. Table. 1.2 gives the list of all raagas in the dataset, total duration of songs, number of unique lead artists and concerts per raaga. More details of this dataset are available online[4].

Varnam dataset

Varṇaṁ[5] is a compositional form in Carnatic music. They are composed in different rāgas (melodic framework) and tālas (rhythmic framework). Though they are lyrical in nature, the fundamental emphasis lies in a thorough exploration of the melodic nuances of the rāga in which it is composed. Hence, varṇaṁs are indispensable in an artist's repertoire of compositions. They are an invariable part of the Carnatic music curriculum, and help students to perceive the nuances of a rāga in its entirety. The coverage of the properties of svaras and gamakas covered in a varṇaṁ within a given rāga is quite exhaustive. This makes the varṇaṁs in a particular rāga a good source for many of the characteristic phrases of the rāga.

We recorded 28 varṇaṁs in 7 rāgas sung by 5 young professional singers each of whom received training for more than 15 years. To make sure we have clean pitch contours for the analysis, all the varṇaṁs are recorded without accompanying instruments, except the drone. The structure of varṇaṁ allows to attribute each part shown in Figure 7.1 to a select few taala cycles depending on the speed. We take advantage of this information to semi-automate the synchronization of the notation and the pitch-contour of a given varṇaṁ. For this, we annotated all the recordings with tāla cycles. Also, in order to further minimize the manual intervention in using the annotations, all the varṇaṁs are chosen from the same tāla (adi tāla, the most popular one (Viswanathan and Allen, 2004)). Table. 1.3 gives the details of the varṇaṁ collection recorded for this analysis. This dataset is made publicly accessible for download[6].

[4] http://compmusic.upf.edu/node/328

[5] This Sanskrit word literally means color, and varṇaṁs in Carnatic music are said to portray the colors of a rāga

[6] http://compmusic.upf.edu/carnatic-varnam-dataset

Rāga	Duration (hours)	Lead artists	Concerts
Ṣanmukhapriya	3.05	12	12
Kāpi	1.19	9	12
Bhairavi	5.51	9	12
Madhyamāvati	3.45	12	12
Bilahari	3.37	11	12
Mōhanaṁ	4.71	8	12
Sencuruṭṭi	0.92	10	12
Śrīranjani	1.93	12	12
Rītigauḷa	3.23	11	12
Hussēnī	1.25	10	12
Dhanyāsi	3.37	8	12
Aṭāna	1.67	11	12
Behāg	1.26	10	12
Suraṭi	2.3	11	12
Kāmavardani	3.51	11	12
Mukhāri	3.3	12	12
Sindhubhairavi	1.01	10	12
Sahānā	2.63	11	12
Kānaḍa	2.82	9	12
Māyāmāḷavagauḷa	2.58	11	12
Nāṭa	1.74	11	12
Śankarābharaṇaṁ	4.76	8	12
Sāvēri	2.97	10	12
Kamās	2.39	8	12
Tōḍi	7.23	9	12
Bēgaḍa	2.98	9	12
Harikāmbhōji	3.8	9	12
Śrī	1.51	10	12
Kalyāṇi	5.42	9	12
Sāma	1.16	10	12
Nāṭakurinji	1.8	10	12
Pūrvīkaḷyāṇi	5.87	9	12
Yadukula kāṁbōji	2.13	11	12
Dēvagāndhāri	2.27	11	12
Kēdāragauḷa	4.08	11	11
Ānandabhairavi	1.84	9	12
Gauḷa	2.05	6	11
Varāḷi	3.92	10	12
Kāṁbhōji	6.2	9	12
Karaharapriya	7.28	11	12

Table 1.2: Details of the Raaga dataset 2 (RD2) with 40 raagas, each with 12 recordings.

Rāga	#Recs	Duration (min)	#Taala annotations
Ābhōgi	5	29	158
Bēgaḍa	3	27	147
Kalyāṇi	4	27	143
Mōhanaṁ	4	24	158
Sahāna	4	28	156
Sāvēri	5	36	254
Śrī	3	26	138
Total	**28**	**197**	**1154**

Table 1.3: Details of the varṇaṁ dataset.

Kriti dataset

Varnam dataset allowed us to gain valuable insights into svara intonation. However, it comes with a limitation that all the performances of a given raaga are of the same composition. Therefore, the representation computed for a svara can be specific to either the raaga or the composition. In order to eliminate this ambiguity and cross-verify the learnings from varnam dataset, we have put a similar effort in building a dataset with the most performed musical form in Carnatic music - kriti.

A kriti differs from varnam in many ways, starting from their preferred time slot during the course of concert and their intent. Unlike a varnam exposition where patterns sung with svara syllables are pre-composed, kriti often accommodates improvised patterns with svara syllables. Kriti also makes room for a certain kind of improvisation on lyrical phrases called Neraval where the performer brings out subtle variations in a selected line from a composition (see ch. 2). Such differences call for another dataset that can potentially offer more insights on svara intonation.

Table. 1.4 gives the details of the kriti dataset. This dataset is a collection of commercial audio recordings. Hence, unlike varnam dataset, the recordings can not be put in the public domain. Hence, we made the predominant melody available. The notations of these Kritis are taken from several books including Rao (1997b,a, 1995), and are manually transformed to machine-readable format. Like the varnam dataset, this dataset is publicly accessible as well[7]

[7] `http://compmusic.upf.edu/node/314`.

Raaga	#Comp.	#Singer	#Rec.
Anandabhairavi	3	5	7
Atana	4	5	5
Bhairavi	5	7	8
Devagandhari	5	5	5
Kalyani	4	4	5
Todi	9	15	15
Total	**30**	**24**	**45**

Table 1.4: A more diverse dataset compared to the varnam dataset. This consists of 45 recordings in 6 raagas performed by 24 unique singers encompassing 30 compositions.

1.5 Summary of contributions

Following are the main contributions of this thesis.

- Novel approaches to svara intonation description.
- Ontologies for various Carnatic music concepts, importantly svara, raaga and melodic phrases.
- A multi-modal KB, and identifying the limitations and challenges in the process of building it.
- A thorough comparison of raaga classification approaches based on pitch histograms.
- Machine readable varnam and kriti datasets.

And following are the other contributions stemming from our work, which we believe need further research.

- A novel semantic distance measure to be used with linked open data.
- A set of guidelines to improve the recall of Open IE systems in domains with scarce data.
- An approach to bring out the salient aspects of different music cultures analyzing natural language text.

1.6 Thesis organization

We divide the thesis into three parts: I. Setting the stage, II. Audio music analysis for intonation description, and III. A multi-modal KB of Carnatic music. In the first part (the current one), we introduce the basics of Carnatic music that form a necessary background to understand the contents herein. Then we survey the computational work related to melodic framework in Indian art music. As part of this, we present a thorough evaluation of histogram-based approaches for raaga classification, on a common dataset derived from an earlier snapshot CompMusic Carnatic corpus. We then introduce the semantic web concepts and present a survey of work concerning ontologies and linked data in the music domain.

In the second part, we present four approaches, each building on observations from the previous one, for intonation description of svaras in Carnatic music. In the third part, we first discuss the ontologies developed for Carnatic music concepts. We then present and document the details of the KB. We also discuss an approach to extract salient features of a music tradition using natural language text to distinguish it from others, and propose a novel semantic distance measure that can take advantage of this knowledge in providing better recommendations.

Chapter 2

Indian art music

"Though one devote himself to many teachers, he must extract the essence. As the bee from flowers."

— Kapila

This chapter is a brief primer on the music theory necessary to understand this work and its context. We begin with a succinct introduction to the social and cultural context of the Indian art music to help the reader in sensing the ecosystem this music thrives in. It is followed by a discussion on the fundamental music concepts and the relevant terminology. We then go over the concert format, which helps us understand the relevance of those musical concepts in the actual performance context. This chapter is not intended to be an exhaustive source of information about the music traditions we discuss, rather we limit ourselves to the scope of our work. We do point the reader to several musicological resources to gain further insights. Also, we defer some of the relevant discussion on a few aspects to chapter 9, where we present the knowledge representations we developed for this music.

2.1 Geographical, social and cultural context

There are two prominent art-music traditions in the Indian subcontinent: Carnatic in the Indian peninsular and Hindustānī in north India, Pakistan and Bangladesh. Both are actively practiced and have a very relevant place in the current societal and cultural context. They are taught orally in a "guru-shishya parampara", which is a lineage of teachers and students. Typically a student trains from as early as 2-3 years for at least about two decades before they are considered ready to perform. The role of notation in these music traditions is limited to a memory aid in

recalling the melody and lyrics. It is not meant to represent the music to be reproduced as is. This, combined with oral transmission of the heritage, has resulted in several "banis" or "gharanas"[1] which have recognizable stylistic differences in how they interpret a composition, a melody or a musical concept.

Both the music traditions share many musical concepts at a higher level as they have evolved together throughout much of the past. Historically, the northern part of the subcontinent has witnessed several foreign invasions which had a significant impact on their culture - music in particular. Of those, Persian culture of the Moghuls have had arguably the most visible and lasting influences - ranging from instruments and compositions to the music itself (Kaul (2007)). On the contrary, the Indian peninsular has been relatively untouched but for the international trade. As a consequence, today there are considerable differences in these two music traditions (Narmada (2001) compares how and in what manner the melodic framework differs in Carnatic and Hindustani music).

Indian art music and dance have been under the patronage of kings, temples, zamindars[2] and politicians until the early years of democracy (Sriram (2007)). Except the temples, all of them are now replaced by corporate funds, crowd-funded organizations and paid concerts. Both the art forms have been greatly influenced and inspired by the Indian philosophy, mythology and religious thought besides the life of the people. They are intertwined with people's lives in their many activities such as weddings, religious rituals and various celebrations (see fig. 2.1). Their telling influence can also be felt in the contemporary music forms such as Film music and dance (e.g: Bollywood).

Numerous annual events and festivals through out the year are testimonials to a flourishing ecosystem of Indian art music. These also include scholarly conventions that engage musicians and scholars in the community on a common platform[3]. We have took part in some of them attempting to further our understanding of this music.

2.2 Music concepts and terminology

While there are several treatises in the last two thousand years that are considered land mark references for musical aspects in Indian art music, Dikshitar et al. (1904) is the one that is closest to discussing the concepts as they are prac-

[1] Loose translation is schools of music

[2] Indian aristocrats, who held huge tracts of land which were lent to other people.

[3] One example for such an event is Margazhi festival also known as Madras Music Season - http://en.wikipedia.org/wiki/Madras_Music_Season

Figure 2.1: This is a typical ensemble in south Indian Hindu weddings. The two artists in the center play Nadaswaram, a reed instrument while the other two players play Thavil, a rhythm instrument.

ticed in the contemporary Carnatic music. Though most of our definitions and discussion of music concepts are largely in line with the aforementioned compendium, our understanding of the music is more influenced by the contemporary musicologists and musicians. In the context of this thesis, Jairazbhoy (1971); Shankar (1983); Bhagyalekshmy (1990); Sambamoorthy (1998); Viswanathan and Allen (2004); Janakiraman (2008) are few of the important musicological references. The concepts and terminology used throughout this work refer to Carnatic music unless mentioned otherwise. Wherever necessary, the differences with Hindustani music would be explicitly stated. We use the term Indian art music to refer to both.

Melody

Raga

Rāga is the melodic framework of the Indian art music. There are hundreds of different raagas in use in the contemporary Indian art music. Though literature in the past drew parallels between raaga and scale to help the reader understand the former, it must be clarified that scale is just one structural component of raaga. Moreover from a musical point of view, it is not the primary differentiating aspect between raagas. Each raaga is characterized by a set of properties which are largely a result of the functional roles different svaras play, melodic phrases that carry the identity of the raaga and the organic growth of compositions/repertoire that enrich such vocabulary of melodic phrases.

Mātanga, in his epic treatise Bṛhaddēśī, defines rāga as "that which colors the mind of good through a specific svara (defined below) and varṇa (literally color/shade) or through a type of dhvani (sound)" (Sharma and Vatsayan (1992)). Each rāga therefore, can be thought of as a musical entity that leaves an impression on the minds of listeners which is shaped by the properties of its substructures. Taking cues from the past computational works concerning raagas (mainly Chordia and Rae (2007) and Krishnaswamy (2004)), we define raaga as such - "Rāga is a collection of melodic atoms and a technique for developing them. These melodic atoms are sequences of svaras that are inflected with various micro-pitch alterations and articulated with expressive sense of timing. Longer musical phrases are built by knitting these melodic atoms together ". Notice that one must comprehend that rāga is more than a sequence of discrete svaras for understanding it, especially so in developing a computational representation for analytical purposes.

There are several classification systems of raagas based on varied criteria ranging from objective properties of its substructures (eg: the number of svaras) to subjective or cultural notions (eg: part of the day the raaga is intended for). In ch. 9, we go over many such classifications as we attempt to represent such knowledge in an ontology. Here however, we go over two types of classifications in the interest of comprehensibility of this thesis. Remember that we referred to the organic growth of a raaga's repertoire as one of most important aspects in defining a raaga's identity. There are raagas which have been around for a long time, and there are newer raagas which have been created/revived by recent composers. Musicians and scholars differentiate them as phraseology-based raagas and scale-based raagas respectively (Krishna and Ishwar (2012)). The latter do not often posses a strong identity but for their scale structure. Therefore, most of the sung music is based on the former class of raagas. We have taken measures to ensure that the audio music collections used in our work reflect this. Another classification of raagas calls a set of raagas sharing the same set of svaras as allied raagas. Note that the musical characteristics of raagas in a given allied group differ owing to their phraseology and properties of svaras.

Svara

We defined raaga as a collection of melodic atoms. Svaras and gamakas (discussed in a short while) are inseparable constituents that together make up these melodic atoms. It is common to find analogies between svara in Indian art music and note in western classical music. However, much like the analogy between raaga and scale, this obscures the actual purpose of svara. A note can be understood as a pitch-class, which is a set of points, each an octave distant from its immediate

successor and predecessor on the frequency spectrum. A svara on the other hand, manifests itself as a region on the frequency spectrum than a point, the extent of which depends mainly on its immediate neighbors in a given raaga. Their identity as a region rather than a point arises due to the fact that svaras are sung inseparably with gamakas.

There are seven symbols used in Indian art music for svaras - Sa, Ri, Ga, Ma, Pa, Da, Ni. A svara can have two or three variants called svarasthanas, which literally mean locations of the svara. For example, Ma has two variants Ma_1, and Ma_2. In a given raaga, each svara in it assumes one such location[4]. Table. 2.1 gives the list of svarastānas with their Karṇāṭaka and Hindustānī names and the ratios they share with tonic (Shankar (1983)). Although there are 16 svarastānas in all, 4 of them share ratios with others (In the Table. 2.1, the svarastānas sharing the ratios are indicated with same *Position* value). Tonic frequency is chosen according to the singer's comfort, and all the accompanying instruments are tuned accordingly. Note that the transposition of a set of svaras, i.e., shifting all of them linearly by a given interval, do not change the rāga. But making another svara Sa can result in a different rāga.

Gamaka

Given a svara, a rapid oscillatory movement about it is one of the several forms of movements, which are together called as gamakas. Another form of gamaka involves making a sliding movement from one svara to another. There are a number of such movements discussed in musicological texts (Dikshitar et al. (1904)). Gamakas bear a tremendous influence on how a tune is perceived, and eventually on the identity of the raaga itself. They are often considered the soul of these art-music traditions (Krishna and Ishwar (2012)). The melodic shape and the extent of gamakas sung with svaras determine the identity of melodic atoms that constitute a raaga.

Though gamakas are used in both Carnatic and Hindustānī, the pattern of usage is very distinct. Besides gamakas, there are alankāras (literally ornamentations) which are patterns of svara sequences which beautify and enhance the listening experience. On this note, we would like to emphasize that gamakas are not just decorative patterns or embellishments (whereas alankāras are), they are very essential to the definition of rāga. Krishna and Ishwar (2012) discuss various manifestations of the most important gamaka in Carnatic music, called Kaṁpita.

[4]Hence, svara and svarastāna are normally used interchangeably in this article, as elsewhere, except when the distinction is necessary.

Symbol	Position	Ratio	Karṇāṭaka/Hindustānī name
Sa	1	1	Ṣaḍjama
R1	2	16/15	Śuddha/Kōmal Riṣabha
R2	3	9/8	Chatuśr̥ti/Tīvra Riṣabha
G1	3	9/8	Śuddha Gāṇdhāra
G2	4	6/5	Sādāraṇa/Kōmal Gāṇdhāra
R3	4	6/5	Ṣaṭśr̥ti Riṣabha
G3	5	5/4	Aṇtara/Tīvra Gāṇdhāra
M1	6	4/3	Śuddha/Kōmal Madhyama
M2	7	64/45	Prati/Tīvra Madhyama
Pa	8	3/2	Pañchama
D1	9	8/5	Śuddha/Kōmal Daivata
D2	10	5/3	Chatuśr̥ti/Tīvra Daivata
N1	10	5/3	Śuddha Niṣāda
N2	11	16/9	Kaisiki/Kōmal Niṣāda
D3	11	16/9	Ṣaṭśr̥ti Daivata
N3	12	15/8	Kākali/Tīvra Niṣāda

Table 2.1: The list of svarastānas used in Karṇāṭaka and Hindustānī music, along with the ratios shared with tonic. Note that the positions 3, 4, 10 and 11 are shared by two svarastānas each.

Unlike several other music traditions where music notation is a crucial source of information during learning as well as performing, notation is very sparingly used. Indeed, it is considered just a memory aid. One of the multitude of possible reasons can be the difficulty in notating gamakas, owing to the complexity of movements.

Melodic phrases

It is often noted by musicians and musicologists that a rāga can only be learned by getting familiar with several compositions in it. Any phraseology-based rāga is endowed with a generous repertoire of characteristic phrases, each of which encapsulates its properties. These phrases are known by the names prayogas and svara sancharas in Carnatic and pakads in Hindustani. Typically in a concert, the artist starts with singing these phrases. They are also the main clues for listeners to identify rāga. This pool of phrases for a rāga keeps evolving over time, often taken from landmark compositions in that rāga.

Rhythm

Taala in Carnatic music, is one of the highly developed rhythmic frameworks around the world. The most typical representation of a taala in musicological texts shows its angas (literally parts). There are three angas[5] laghu (I), drutam (O), anudrutam (U). The time-measures of the latter two are 2 and 1 akshara (syllable) respectively. The time-measure corresponding to a laghu depends on the jaati (literally class) of corresponding taala, which can be one of the five that makes it 3, 4, 5, 7 or 9 aksharas long. Suppose we are given a taala with an anga structure of I_3UO, where the subscript 3 on I indicates the jaati of the taala, and consequently of the laghu. The total number of aksharas in this taala can be counted as 3+1+2. In theory, there are seven basic taala classes knows together as suladi sapta taalas, corresponding to the severn basic templates of anga structures (Eg: IOI, OI, IUO etc). These, in conjunction with the five jaatis of laghu give 35 taalas in that category. On top of this, there is a layer that determines the gati/nade (literally gait) of the taala, which takes one of the five levels that divide each akshara into 3, 4, 5, 7 and 9 minor units. Therefore, the 35 taalas in conjunction with different gatis/nades result in 175 taalas. There are other taalas as well, which have origins from folk music and fall outside this structure. Chapu taalas are examples of these. Srinivasamurthy et al. (2014a) discusses Carnatic taala structure in a detailed manner with an emphasis on exploring them using computational approaches. Clayton (2000) gives a thorough account of Hindustani taalas.

Forms

Broadly, forms in Indian art music are classified into two categories: compositional and improvisational forms. However, they can also be categorized based on the characteristics of lyrical content and/or musical properties. Though there exist several forms and corresponding classification schemes historically, it suffices to present a subset of forms that are currently practiced in Carnatic music in order to convey the overall picture.

There are various compositional forms with differing musical content and the intended purpose within the social/musical context: geetham, varnam, padam, javali, swarajati, kirtana, krti, and thillana etc. There are several classification schemes of which we discuss a few. For a more complete list of forms and their classifications, the reader may refer to Janakiraman (2008); Shankar (1983). One such classification scheme distinguishes vocal, instrumental and dance forms, with a possibility that several forms can be simultaneously classified to more

[5] There are six historically, but the other three are very seldom used in contemporary art music.

than one category. Another classification scheme distinguishes pure and applied music. The forms which emphasize the melodic aspects to portray the rāga are classified under pure music, whereas compositions used for a specific purpose (eg: thillana which is often used in classical dance forms) are classified under applied music. Another classification scheme divides them into abhyasa gana and sabha gana. The forms intended for learning technicalities of the music (such as geetams, varnams, various svara exercises) are classified under abhyasa gana. The forms which are intended for performance are classified under sabha gana.

There are four kinds of improvisational forms in Carnatic music: alapana, taanam, neraval and kalpana-svara. Alapana is an unmetered, free form of improvisation to elaborate on properties of the rāga. It is sung with nonsensical syllables such as vowels and does not contain lyrics. Taanam is a metered form of improvisation sung with syllables such as tom, nam, tam, anantam and namtam, but it is usually unaccompanied by percussion instruments. For neraval, the artist picks up a line from the composition, and sings it with various melodic variations allowed within the rāga, subject to rhythmic constraints. The kalpana-svara is a form where the artist sings solfege in groups of varying lengths of rhythmic cycles, which usually end on a line chosen from a composition. Rhythmic accompaniment is allowed in both neraval and kalpana-svara.

2.3 Kutcheri: the concert format

Indian art music is heterophonic in nature and is characterized by elaborate improvisations that are often interspersed with composed bits of a song. A typical Carnatic music concert has a lead artist, a violinist (melodic accompanist), a mridangam player (rhythmic accompanist), a ghatam player and a kanjira player (supporting rhythmic accompanists), and a tanpura player (drone). See Fig. 2.2.

The violinist often closely imitates the lead artist with a lag of few milliseconds, improvising on it as s/he finds fit. The role of rhythmic accompaniment is to enhance and accentuate the melodic presentation. One major difference between Carnatic and Hindustani concert format is that, the rhythmic accompanists in Carnatic music are free from the duty of keeping the time for the lead artist, while in Hindustani it is their primary role. The tanpura player strums the strings of the instrument producing frequencies of tonic and fourth/fifth and their many harmonics. This is used as the melodic reference for the performance.

The concert can last anywhere between 2-4 hours. There is no strict standard as such for a lead artist to conduct a concert. However, one that was popularized by Ariyakudi Ramajuna Iyengar during the 1970s has become a relatively dominant

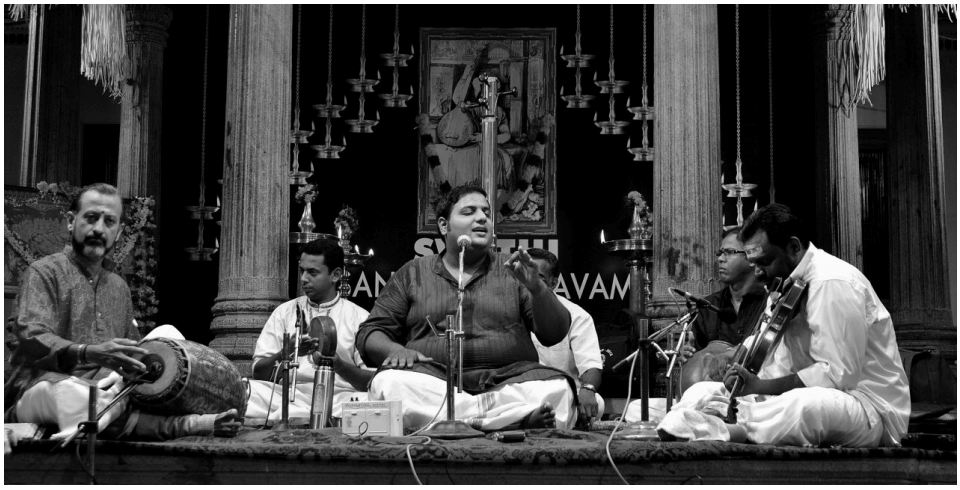


Figure 2.2: A typical ensemble for a Carnatic music concert. From left to right are mridangam player (the main rhythmic accompaniment), kanjira player (co-rhythmic accompaniment), lead vocalist, tanpura player (provides drone), ghatam player (co-rhythmic accompaniment) and violin player (melodic accompaniment).

choice in the current years (Krishna (2013)). In this format, the concert starts with a varnam, which is usually sung as it was composed. It is followed by a few kritis and keertanas, performed with short bursts of improvisation such as alapana and kalpana-svara, within each song. The main piece of the concert is what is known as a Raagam-Taanam-Pallavi (RTP). In this, the artist sings an elaborate alapana and taanam in a raaga of her/his choice, and performs a pallavi (resembling neraval) using a lyrical line from an existing composition in that raaga. The violinist also gets her share of time in this piece alongside the lead artist. The RTP is followed by Tani-avaratanam which is when the rhythmic accompanists showcase their improvisatory skills.

The lead artist typically communicates her/his choice of compositions and the concert structure beforehand with the accompanists. It is common that the artists do not practice together before the concert. Usual venues for the concerts include renowned music sabhas, hindu temples, corporate, social and cultural events, and small gatherings of neighborhoods. Of these, sabhas are community-funded not for profit organizations and more importantly, contribute to major chunk of the activity.

2.4 Summary

The computational study of Carnatic music offers a number of problems that require new research approaches. Its instruments emphasize sonic characteristics that are quite distinct and not well understood. The concepts of Raaga and Taala are completely different from the western concepts used to describe melody and rhythm. Their music scores serve a different purpose than the ones of western music. The tight musical and sonic coupling between the singing voice, the other melodic instruments and the percussion accompaniment within a piece, requires going beyond the modular approaches commonly used in music information research (MIR). The tight communication established in concerts between performers and audience offer great opportunities to study issues of social cognition. The study of the lyrics of the songs is also essential to understand the rhythmic, melodic and timbre aspects of the Carnatic music.

Chapter 3

A review of past research concerning raagas in Indian art music

There is an abundance of musicological resources that thoroughly discuss the melodic and rhythmic concepts in Indian art music (see Dikshitar et al., 1904; Sambamoorthy, 1998; Bagchee, 1998; Clayton, 2000; Narmada, 2001; Viswanathan and Allen, 2004; Ramanathan, 2004, and the references therein). However, only a few which approach from a computational perspective are available. In this chapter, we present a survey of scientific literature that concern melodic framework in Indian art music traditions, with a specific emphasis on work related to Carnatic music.

In sec. 3.1, we present computational and analytical approaches proposed to understand different characteristics of the raaga framework, categorizing them based on what aspects of the melodic framework they pursue. Of these, one that garnered more attention is classification of raagas, arguably the most relevant task seen from an application perspective. Therefore, we have put additional effort in surveying work related to this. In sec. 3.2, we discuss the relevance of this task in the practical context, and discuss how people with varying levels of exposure to this music identify raagas. This helps us understand the relevance of different characteristics of raagas in their classification. We then consolidate approaches that are based on various features extracted from pitch-distributions in sec. 3.3. In the next section, we present our evaluation of these approaches on a common dataset to understand better their strengths and drawbacks. Finally, we summarize the chapter by presenting our understanding of the state-of-the-art, and

outlining few potential ways forward.

3.1 A categorical overview of the past work

Different aspects of raagas which interested the music information research community include validation of microtonal intervals, tuning and svara positions, identifying raaga-specific patterns, perception of emotion, and the automatic classification of raagas. In this section, we summarize past work concerning these with an emphasis on work related to understanding raagas through their svaras.

Microtonal intervals

The topic of positions of svaras and their relation to shrutis has been a matter of intense debate among musicologists and musicians in the last millennium. The term shruti is overloaded with several interpretations of musicologists trying to understand and fit it into their contemporary systems of music. Over the last century however, there is a growing consensus that the purpose of shrutis varied much since the time they are first defined ((Read Ramanathan, 1981; Rao, 2004; Meer and Rao, 2009, for different interpretations of sruti historically for the past few hundred years, and their relevance to contemporary art music)).

However, the long-standing debate has not been settled completely yet. Some of the important questions that were addressed in this are: i) How does one define shruti? ii) Do shrutis form a basis for svara positions in contemporary music? iv) Are shrutis still relevant? Since the advances in signal processing, researchers have been exploring ways to get answers to these through an objective and qualitative analysis of the audio music content. We summarize some of the representative portions of the work here and the stand point they assume by the virtue of their findings and/or arguments.

Ramanathan (1981) has carried out an extensive study of the musicological literature concerning sruti spanning the last two millennia. In his work, the author quotes musicologists from various centuries to clarify each of their interpretations of the term shruti. He clarifies that there are at least two distinct interpretations prevalent for shruti at various periods of time. One of them, tracing back to Natyashastra (Rangacharya, 2010, is an English translation of the magnus opus on theater arts originally written in Sanskrit), treats shruti as a unit that is used in defining intervals of varying sizes between svaras. The other interpretation which he quotes from Naradiya Siksha, another magnus opus on music which predates Natyashastra, treats shruti as a musical concept that refers to the flexible yet distinct intonation of svaras in different modes of music. He exemplifies a few

common pitfalls scholars and researchers often are unaware of in understanding shruti (such as the legacy names of svaras, for example chatushurti rishaba, that can be misleading). He also goes on to explain why the ancient interpretations are not relevant to today's practice.

Rao (2004); Meer and Rao (2009); Rao and Wim van der Meer (2010) extend this work and defend the latter interpretation to be the most relevant to the contemporary music. They substantiate this by juxtaposing the analyses of audio samples showing how maestros sang when asked to sing the same svara in different raagas in their respective shrutis. On the other hand, researchers who advocate the former interpretation of shruti have worked towards conducting quantitative analyses to show how shruti forms the basis for the svara positions in the contemporary music. Datta et al. (2006) have curated a collection of over 116 recordings in 4 raagas by 23 singers to verify this. They have considered 8 different tuning systems (with differing ratios between svara positions) as possible references of svara positions. They identify stead segments from the pitch contour extracted from these audio music recordings and extract the scale used for the recording. The intervals between the svara positions are then represented in units of shrutis. The authors claim that the results support the hypothesis that shrutis form the basis for the svara positions, and further calculate the sizes of intervals between the svara positions in terms of shrutis. However, this work assumes an intervallic size of a shruti, which was said to be never defined in Natyashastra, or later works that carry the same interpretation (Ramanathan (1981)).

Tuning

Besides the work on shruti, another aspect that was also well-explored is the tuning system and svara positions. In Indian art music, tonic can correspond to any frequency value chosen by the lead performer. Hence, there is no given standard. Therefore, automatically identifying the tonic (Sa) of a recorded performance becomes fundamental to any melodic analyses. Chordia et al. (2013) propose an approach that estimates the tonic and the raaga of the audio music sample. In this work, shifted versions of the pitch distribution of the given audio is compared against samples in the training set. The nearest neighbor gives both the tonic and the raaga label for the test sample. Their best result reports an accuracy of 93.2% in tonic estimation. Gulati et al. (2014) compare and thoroughly evaluate different approaches to automatic tonic identification in both Carnatic and Hindustani music traditions, proposed in Bellur et al. (2012); Gulati (2012); Ranjani et al. (2011). Seven different approaches are evaluated over six diverse datasets in varying contexts such as presence/absence of additional metadata, gender of the

singer, quality and duration of the audio samples and so on. This comparative study concludes that methods which combine multi-pitch analysis with machine learning techniques perform better compared to those which are mainly based on expert knowledge.

Another line of work concerning tuning is about verifying the hypothesis that svara positions can be explained by and correlated to one or the other known tuning systems such as just-intonation or equal-temperament. Even for other melodic analyses that go beyond the tuning aspects, this hypothesis has been a starting point. Levy (1982) conducted a study with Hindustani music performances to understand the aspects of tuning and intonation. This work concluded that the svaras used in the analyzed performances did not strictly adhere to either just-intonation or equal-tempered tuning systems. Further, pitch consistency of a svara was shown to be highly dependent on the nature of gamaka usage. The svaras sung with gamakas were often found to have a greater variance within and across performances and different artists. Furthermore, the less dissonant svaras were also found to have greater variance. However, it was noted that across the performances of the same rāga by a given artist, this variance in intonation was minor. More recently, Swathi (2009) conducted a similar experiment with Carnatic music performances and draws similar conclusions about the variance in intonation.

Krishnaswamy (2003) discusses various tuning studies in the context of Carnatic music, suggesting that they use a hybrid tuning scheme based on simple frequency ratios plus various tuning systems, especially equal temperament. His work also points out the lack of empirical evidence for the same thus far. Recently, Serrà et al. (2011) have shown existence of quantitative differences between the tuning systems in the current Carnatic and Hindustani music traditions. In particular, their results seem to indicate that Carnatic music follows a tuning system which is very close to just-intonation, whereas Hindustani music follows a tuning system which tends to be more equal-tempered. Like in the case of work done by Datta et al. (2006), their work only considers the steady regions in the pitch contour in coming to the aforementioned conclusions.

Throughout these efforts, we come across a recurring observation, that understanding the identity of a svara has much more to it than their exact positions (which roughly translate to steady regions in the pitch contour). For this very reason, Rao (2004) criticizes the limited view through which the analysis was conducted in the past. Rao (2004); Krishnaswamy (2004) reemphasize that a svara is a multidimensional phenomenon, and an exact position does not have much relevance in these art music traditions. This potentially is the reason why Levy

(1982); Meer and Rao (2009); Serrà et al. (2011) observe that there is a lot of variability in the usage of pitches for the same svara in different raagas. In the light of these works, a statistic that was observed by Subramanian (2007) assumes high relevance. In an analysis of a sample rendition in Carnatic music, the author finds that only 28% of the rendition constitutes a steady melody, in which the tonic (15%) and the fifth (9%) are the dominant svaras. This alone reflects on the fact that there is much more musically meaningful information in the melody that has not been yet paid attention to. The work done by Belle et al. (2009) is indicative of the potential this information holds as they classify a limited set of five Hindustani rāgas imagining svara as a region than a point.

Melodic motifs

Motifs are repeating patterns in the melody. In IAM, they often emerge in a composition or an improvisation as a reflection of, and/or a reinforcement to characteristics of the respective raaga. Ross et al. (2012) propose an approach to detect motifs in Hindustani music taking advantage of the fact that a specific set of raaga-characteristic phrases called mukhdas are always aligned to the beginning of a taala cycle. This fact is used in restricting the search space for finding the candidate motifs. SAX and dynamic time warping (DTW) for measuring similarity between such candidates of non-uniform length. Rao et al. (2014) employ DTW based template matching and HMM based statistical modeling on labeled datasets in both Hindustani and Carnatic for matching and classifying phrases.

Ishwar et al. (2013) use a two-pass dynamic programming approach to extract matching patterns in a melody given a query. In the first pass, they use rough longest common subsequence (RLCS) in matching the saddle points of the query against those in the complete melody. In order to get rid of the false alarms amongst the resulting matches, they employ RLCS to match the continuous pitch contours of the results with the query. This approach is used in locating raaga specific motifs in alapanas. They obtain a mean f-measure of 0.6 on fours queries over 27 alapanas in Kambhoji raaga, and 0.69 over 20 alapanas in Bhairavi raaga. Dutta and Murthy (2014) addresses some of the issues reported by Ishwar et al. (2013) in their approach, the main being the number of false alarms which are still evident even after the second pass. Three different measures are proposed to weed out them: density of the match, normalized weighted length and linear trend of the saddle points. Hypothesizing that the beginning lines in different sections of a song are more potent in containing raaga-specific patterns, they employ this approach in finding the common patterns between those lines and further filtering them based on their occurrence across compositions in a given

raaga, to extract raaga-specific motifs from among them.

Gulati et al. (2015, 2016a,b) first identify melodic motifs in a large collection of 1764 audio recordings combining four variants of DTW with techniques to limit the computational complexity of the task. Then they pick a smaller subset of 160 recordings which are distributed in 10 ragas, to automatically characterize raaga specific motifs. For this they use network analysis techniques like the network's topological properties and community detection therein, arriving at non-overlapping clusters of phrases which are then characterized as belonging to a raaga based on its properties. Expert review of the results show that 85% of the motifs are found to be specific to the raaga cluster they are assigned to. Further, considering each audio recording as a document with its motifs as terms, this network is used in a raaga classification task which now becomes analogous to topic modeling in text-based information retrieval. The results show 70% accuracy on a 480 recording collection having 40 raagas, and 92% accuracy on a 10 raaga subset.

Perception of emotion

Wieczorkowska et al. (2010); Chordia and Rae (2009); Devadoss and Asservatham (2013); Balkwill and Thompson (1999), Koduri and Indurkhya (2010) Indian art forms use a sophisticated classification of mental states based on the *rasa* theory, which has been evolving since it was discussed in Natyashastra. In a nutshell, the theory defines a set of mental states, and each state has associated emotions with it. For instance, Raudram (literally Fury) is one such category. There can be different emotions associated to this mental state such as love, hatred, anger and so on.

Past treatises on music have discussed at length the association between raagas and rasas. It is in this context that researchers have tried to validate and/or establish if raagas do elicit specific emotions[1]. Balkwill and Thompson (1999) study the emotional responses of enculturated and non-enculturated listeners to Hindustani raagas.They have used the alapana recordings with a mean duration of 3 minutes, and chose four emotions for the study - joy, sadness, anger and peace. Their results seem to indicate that despite having no cultural and musical exposure to Hindustani music, the responses of non-enculturated listeners were by and large in agreement with the responses of enculturated listeners. Further, they indicate that the listeners responded with the intended emotion in most cases.

[1]Notice that we have used the term emotion and not rasa here.

Wieczorkowska et al. (2010) also conducted a similar experiment with results somewhat different from those reported by Balkwill and Thompson (1999). They find that short melodic sequences of 3 seconds do elicit emotions in both enculturated and non-enculturated listeners, and the similarity in their respective responses is significant. However the results show that the emotional responses recorded by the listeners are largely not in conformity with those prescribed for the raagas in the literature. Results from the study carried out by Chordia and Rae (2009) also indicate that the responses from enculturated and non-enculturated listeners largely agree and also that the responses within a raaga are consistent with each other. Koduri and Indurkhya (2010) conducted a survey on Carnatic raagas and gathered responses from enculturated listeners. They also find that the raagas elicited specific emotional responses, and they are largely consistent across listeners. The later two studies also attempt to correlate these emotional responses with various musical attributes automatically extracted from the audio music recordings.

There are a few assumptions that run across the research on this subject which often go unstated. It is important to take note of these before making an assessment of their results. The rasa theory almost always has been discussed within the context of theatrical arts, of which music is only a part. In trying to 'validate' the raaga-rasa associations from such treatises, one can be said to automatically assume the relevance of such association between raagas and rasas outside the theater arts. Further, the rasa theory clearly distinguishes between a mental state (rasa) and an emotion. However, this distinction is often ignored in the related work which leads to another assumption as follows: When a raaga is said to elicit a rasa, it is assumed to elicit the emotions associated with that rasa.

To put the first assumption into perspective - rasa theory as discussed in the treatises has more practical relevance to Indian art dance forms today than the music forms. This is however not to say that emotional associations to music are not of relevance. For instance, in Bharatanatyam or Kuchipudi[2], the students are trained to elicit the various rasa and emotion combinations in the audience using their facial expressions and body movements. Such explicit associations with rasa are not part of teaching raagas. The second assumption we discussed is also an important one, especially when the distinction between rasa and emotion is clearly spelled out.

The final takeaway from the past research on this topic seems to be that raagas do elicit specific emotions, which are consistent across enculturated and non-enculturated listeners. However, due the aforementioned implicit assumptions

[2]Two of the widely known classical dance forms of India

that are made, other aspects of the work including the association between raagas and rasas as discussed in the past musicological treatises, need further investigation.

3.2 Raaga classification

Seasoned listeners develop a keen ear over the years that help them develop an ability to identify a raaga when it is performed. In this section, we first go over how this is done in practical context by both listeners and musicians. This will help us in understanding how different characteristics of raaga allow people to recognize it. Indeed, most computational approaches developed to understand raaga are often evaluated over a classification test that involves labeling a test sample with a raaga. We briefly review some of these in chronological order in the latter part of the section.

Practical relevance of the task

Given the intricacies of the melodic framework, the task of identifying a rāga can seem overwhelming. But the seasoned rasikas[3] identify rāga within a few seconds of listening to a performance. During a performance, there is an interplay between the performer and the audience in communicating the identity of the raaga being performed. Rasikas enjoy the unveiling of the raaga's identity, and also the performer's ability to reinforce it throughout. Though there are no rules of thumb in identifying rāga, expert musicians believe that broadly there are two procedures by which people identify it from a performance. This normally depends on whether the person is a trained musician or a rasika. People who do not have much knowledge of rāgas cannot identify them unless they memorize the compositions and their rāgas.

The procedure followed by rasikas typically involves correlating two tunes based on how similar they sound. Years of listening to tunes composed in various rāgas gives listener enough exposure. A new tune is compared with the known ones and is classified depending on how similar it sounds to a previous tune. This similarity can arise from a number of factors: characteristic phrases, rules in transition between svaras, usage-pattern of few svaras and gamakas.

This process depends heavily on the cognitive abilities of a person. Without enough previous exposure, it is not feasible for a person to attempt identifying

[3] A term often used for a seasoned Carnatic music listener, which literally means *the one who enjoys art.*

rāga. There is a note-worthy observation in this approach. Though many people cannot express in a concrete manner what the properties of rāga are, they are still able to identify it. This very fact hints at a possible supervised classifier, that can take advantage of properties of rāga.

On the other hand, a trained musician tries to identify the characteristic phrases of rāga. These are called svara sañchāras in Carnatic and pakaḍs in Hindustānī. If the musician finds these phrase(s) in the tune being played, rāga is immediately identified. In some cases, musicians play the tune on an instrument (imaginary or otherwise) and identify the svaras being used. They observe the gamakas used on these svaras, locations of various svaras within the melodic phrases and the transitions between svaras.

This process seems to use almost all the characteristics of rāga. It looks more systematic in its structure and implementation. The procedures used by trained musicians and non-trained listeners provide useful insights to implement a rāga recognition system. As we will see, the existing approaches try to mimic them as much as possible. They can broadly be classified as example-based or knowledge-based or both. The example-based approaches correspond to the intuitive approach used by rasikas to identify a rāga, such as matching similar phrases. The knowledge-based approaches reflect the analytic approach which is employed by the trained musicians, such as identifying the svaras, their roles and gamakas.

Given several attributes of a rāga that serve to distinguish it from all the other rāgas, and the fact that rāgas are learnt by listening and imitation rather than by an analytical application of rules, there appear to be no clear-cut guidelines for the machine recognition of rāgas. The lay listener's intuitive approach suggests a loosely constrained machine learning strategy from a large rāga-labeled audio database of compositions.

Computational approaches to raaga classfication

Chakravorty et al. (1989) proposed recognition of rāgas from notation. They use a scale-matching method to group the raagas into scale-groups[4] in the first-level. This involves matching set of permitted svaras and forbidden svaras in a test sample against those of the training samples. Next within each scale-group, they match a lexicon of phrases of each rāga against those in the training set. The input notation is segmented into approximate ārōhaṇa-avarōhaṇa sections for each candidate rāga considered. Then lexical matching is carried out, first with exact sequences, then with coarse search allowing partial matching. The system

[4]Rāgas in each scale-class or scale-group have identical svaras but different phrases.

is evaluated on 75 rāgas distributed over 45 scale groups in all. They report perfect accuracy for grouping raagas into their scale-groups and 73% accuracy for raaga classification within each scale-group.

Inspired by Sahasrabuddhe and Upadhye (1992) on the use of a finite automaton to generate svara sequences characteristic of a raaga, Pandey et al. (2003) used a generative statistical model in the form of a hidden markov model (HMM) for each rāga. Sequence of svaras was automatically extracted from solo vocal recording applying heuristics driven note-segmentation technique. The individual svaras form the states. The HMM (actually, just MM since nothing is "hidden") that best explained the observed svara sequence was the detected rāga. Thus sequential information is exploited subject to the limitations of a first-order Markov model. The same work also proposed phrase matching expressed as an approximate substring search for the pakaḍ (catch phrase) of the rāga. In another method, the rāga was identified by counting the occurrences of n-grams of svaras in the pakaḍ. The evaluation was restricted to discriminating 31 samples in 2 rāgas. An average accuracy of 77% and 87% is reported in raaga classification using only HMMs and HMMs along with pakad matching. The central idea in this work, which is to model a rāga as HMM, was also used by Sinith and Rajeev (2006). The same idea was used in an attempt to automatically generate Hindustānī music (Das and Choudhury (2005)), albeit with less success.

Chordia and Rae (2007) used pitch-class profiles to represent the distributions and hence the relative prominences of the different svaras. They also used svara bigram distributions to capture some sequential information. Using just the pitch-class profiles to classify 17 rāgas (142 audio segments of 60s each), the system achieves an accuracy of 78%. Using only the bi-grams of pitches, the accuracy is 97.1%. Using both was reported to give an almost perfect result. Gedik and Bozkurt (2010); Bozkurt (2011) present a similar approach for makam classification in Turkish-makam music, by matching pitch histograms with higher bin resolution.

Sridhar and Geetha (2009) have followed an approach where the set of svaras used in an audio recording is estimated, and compared with the templates in the database. The rāga corresponding to the best matched template is selected as the class label. Their test data consisted of 30 tunes in 3 rāgas sung by 4 artists, out of which 20 tunes are correctly labeled by the system. The tonic is manually input, and the other svaras are identified based on the respective ratio with the tonic.A similar approach based on detecting the svaras used in ārōhaṇa and avarōhaṇa to find the rāga is presented by Shetty and Achary (2009).

Dighe et al. (2013) use the energy information from chromagram to extract the

swara sequence from a given audio recording. *Vadi* of the raaga of that piece is used in identifying the label of the most occurring svara. This consequently allows them to label the rest of the swaras found in the sequence. A svara histogram is computed to be used to be used with random-forest classifier for raaga classification. An average accuracy of 94.28% is reported over 10-fold cross validation experiment on a dataset comprising of 8 raagas and 127 recordings. Kumar et al. (2014) propose kernels for similarity using pitch-class profiles and n-grams, and employ them with an SVM classifier to test how they perform in a raaga classification test. On a dataset of 170 samples in 10 raagas[5], they have reported highest average accuracies of 70.51%, 63.41% and 83.39% when using kernel for pitch-class profiles, n-grams and both in a linear combination respectively.

3.3 Consolidating pitch-distribution based approaches

The approaches surveyed in sec. 3.2 can be broadly categorized as the following, based on the source of the features computed: pitch-distributions, phrase detection and svara transitions. Of these, pitch-distributions are the most explored. In this section, we discuss few experiments which fall into this category, testing the impact of various sources of information and also the effect of changing different parameters specific to each approach. In 3.4, we report the results of these experiments conducted on a much comprehensive dataset compared to the ones on which they were tested in the past literature. This will help us to better understand the merits and limitations of each approach.

The datasets used in the surveyed rāga recognition approaches are not representative enough for several reasons. Pandey et al. (2003) and Sridhar and Geetha (2009) used datasets which had as few as 2 or 3 rāgas. The datasets were also constrained to some extent by the requirement of monophonic audio for reliable pitch detection. The dataset used by Chordia and Rae (2007) is also quite limited. The data available on a popular commercial online music portal such as raaga.com ($>$ 500 performers, $>$ 300 rāgas)[6], shows that there is a scope to improve the quality and size of the data used for the task. Therefore the conclusions drawn from the existing experiments can not be easily generalized.

First, we discuss different approaches to obtain rāga-specific measurements from a pitch distribution computed from the continuous pitch contour. Then we describe few common computational steps needed for an evaluation of these approaches. Finally, the results are reported and discussed in sec. 3.4.

[5]This is the dataset we used in 3.3, and later shared it with authors of this paper.
[6]Observations made on 25/05/2012.

Template matching

In this approach, the set of svaras identified in a given recording are matched against the rāga templates and then the rāga corresponding to the template that scores the highest is output as the class label. We use two different methods to determine the svaras present in a given recording and to obtain the rāga templates.

In both the methods, from a given recording, pitch is extracted and a high-resolution octave-folded histogram (1200 bins) aligned to the tonic of the recording is obtained. In the first method (A_{th}), we consider the set of svaras of a rāga as defined in theory as template for the respective rāga. From the histogram of a given recording, we obtain the values of bins at locations corresponding to the 12 just intonation intervals. The top 7 are taken as the svaras used in the recording. In the second method (A_{de}), the rāga template is obtained by averaging tonic-aligned histograms corresponding to individual recordings in the rāga, and picking the most salient peaks, a maximum of 7. The svaras used in an individual recordings are also inferred in the same manner.

Distributions constrained to "steady regions"

The pitch contour obtained from the recording may be used as such to obtain a pitch-class distribution. On the other hand, given the heavy ornamentation in Indian art-music (see fig. 6.4), marking or estimating the boundaries of svaras is impractical. Therefore in the past work, pitch-class distributions are often computed using only the stable pitch-regions in the melody.

In order to determine a stable region in pitch-contour, the local slope of the pitch-contour is used to differentiate stable svara-regions from connecting glides and ornamentation (Pandey et al. (2003)). At each time instant, the pitch value is compared with its two neighbors to find the local slope in each direction. If the magnitude of either of the local slopes lies below a threshold value T_{slope}, the current instant is considered a stable svara region:

$$\begin{aligned} |F(i-1) - F(i)| < T_{slope} \;\; Or \\ |F(i+1) - F(i)| < T_{slope} \end{aligned} \tag{3.1}$$

where F is the pitch contour converted to cent scale. All the instances where the slope is beyond T_{slope} are discarded as they don't belong to the stable regions. Finally, the pitch values in the segmented stable svara regions are quantized to the nearest available svara value in the just intonation scale using the tonic. This step helps to smooth-out the minor fluctuations within intended steady svaras.

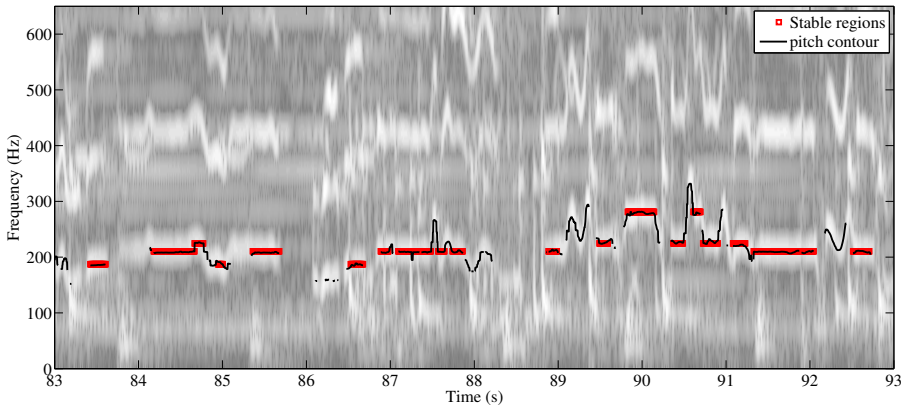


Figure 3.1: The pitch contour is shown superimposed on the spectrogram of a short segment from a Carnatic vocal recording along with the identified stable pitch-regions.

Fig. 6.4 shows a continuous pitch contour with the corresponding segmented and labeled svara sequence superimposed.

A pitch-class profile is computed using only the stable svaras thus obtained, and hence is a 12-bin histogram corresponding to the octave-folded values quantized to just intonation intervals. There are two choices of weighting for histogram computation. We call the pitch-class profiles corresponding to those two choices as $P_{instances}$ and $P_{duration}$, where former refers to weighting a svara bin by the number of instances of the svara, and the latter refers to weighting by total duration over all instances of the svara in the recording. In each case, results are reported for different values of T_{slope}. Further, we also experimented setting a minimum time threshold (T_{time}) to pick the stable regions.

Distributions obtained from full pitch contour

In this approach, we consider the whole pitch-contour without discarding any pitch values. We call this $P_{continuous}$. In this case, we consider different bin resolutions for quantization in constructing the histogram to observe its impact. This step is motivated by the widely discussed microtonal character of Indian art music, which in particular identifies svara as a region rather than a point (Krishnaswamy (2004)).

For all the classification experiments of sections. 3.3 and 3.3, we need a distance measure and a classifier to perform rāga recognition. A good distance measure

for comparing distributions should reflect the extent of similarity between their shape. Further, we would also like to observe the impact of adding tonic information. Therefore, we conduct experiments twice: with tonic and without it. For this to be possible, the distance measure should also facilitate comparing pitch-class profiles in the absence of tonic information. Hence, we choose Kullback-Leibler (KL) divergence measure as a suitable measure for comparing distributions. Symmetry is incorporated into this measure by summing the two values as given below (Belle et al. (2009)).

$$D_{KL}(P,Q) = d_{KL}(P|Q) + d_{KL}(Q|P) \tag{3.2}$$

$$d_{KL}(P|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \tag{3.3}$$

where i refers to the bin index in the pitch-distribution, and P and Q refer to pitch-distributions of two tunes. In the cases where tonic information is not available, we consider all possible alignments between P and Q, and choose the one that scores best in terms of minimizing the distance measure.

k-NN classifier is used in conjunction with the selected distance measure. Results are reported over several values of *k*. In a leave-one-out cross-validation experiment, each individual tune is considered a test tune while all the remaining constituted the training data. The class label of the test tune is estimated by a simple voting method to determine the most recurrent rāga in the *k* nearest neighbors. The selection of the class label C is summarized in the following equation:

$$C = arg\ max_c \sum_i \delta(c, f_i(x)) \tag{3.4}$$

where c is the class label (rāga identity in our case), $f_i(x)$ is the class label for the i^{th} neighbor of x and $\delta(c, f_i(x))$ is the identity function that is 1 if $f_i(x) = c$, or 0 otherwise.

Common computational steps

In all our experiments, we use pitch-contour, tonic information and histogram analysis. Here we briefly explain the computational steps to obtain them.

Pitch extraction To accurately mark the F0 of the stable pitch-regions, the estimation errors that are generated by all F0 detection methods need to be minimized. In many portions of the vocal recording used, accompanying violinist fills

the short pauses of vocalist, and also very closely mimics vocalist with a small time lag. This is one of the main problems we encountered when using pitch tracking algorithms like YIN (de Cheveigné et al. (2002)): violin was also being tracked in a number of portions. As it is usually tuned an octave higher, this resulted in spurious pitch values. To overcome this, we use predominant melody extraction (Salamon and Gomez (2012)) based on multi-pitch analysis. But this has an inherent quantization step which does not allow high bin resolutions in histogram analysis. So we use a combination of both. In each frame, we transform the estimated pitch value from both methods into one octave and compare them. In those frames where they agree within a threshold, we retain the corresponding YIN pitch transforming it to the octave of the pitch-value from multi-pitch analysis. We discard the data from frames where they disagree with each other. On an average, data from 53% of the frames is retained. Though it is a computationally intensive step, this helps in obtaining clean pitch tracks, which have less f0 estimation errors. The frequencies are then converted to cents. We use tonic information as base frequency when it is available, otherwise we use 220Hz. The octave information is retained.

Figure 6.4 shows the output pitch track superimposed on the signal spectrogram for a short segment of Carnatic vocal music where the instrumental accompaniment comprised violin and mr̥daṅgaṁ (percussion instrument with tonal characteristics). We observe that the detected pitch track faithfully captures the vocal melody unperturbed by interference from the accompanying instruments.

Tonic identification Tonic is the base pitch chosen by a performer that allows to fully explore the vocal (or instrumental) pitch range in a given rāga exposition. This pitch serves as the foundation for melodic tonal relationships throughout the performance and corresponds to Sa svara of rāga. All the accompanying instruments are tuned in relation to tonic of the lead performer. The artist needs to hear tonic throughout the concert, which is provided by the constantly sounding drone that plays in background and reinforces tonic. The drone sound is typically provided by Tāmpura (both acoustic and electric), sr̥tī box or by sympathetic strings of instrument such as Sitār or Vīṇa.

For computational analysis of Indian art music, tonic identification becomes a fundamental task, a first step towards many melodic/tonal analyses including rāga recognition, intonation analysis and motivic analysis. There is not much research done in the past on tonic identification. However recent studies have reported encouraging results. Ranjani et al. (2011) explores culture-specific melodic characteristics of Carnatic music that serve as cues for tonic (Sa svara). A more general approach applicable to both Carnatic and Hindustānī music is proposed

by Gulati (2012), which takes advantage of the presence of drone sound to identify tonic. However each of these approaches have there own limitations and requirements. We followed the approach proposed by Gulati (2012), which is based on multi-pitch analysis of the audio data, and automatically learned set of rules (decision tree) to identify the tonic pitch. We evaluated our implementation on the same database that the author had used and achieved nearly the same results (93% accuracy for 364 vocal excerpts of Hindustānī and Carnatic music).

Histogram Computation The pitch contour in cents is folded to one octave. Given the number of bins and the choice of weighting, the histogram is computed:

$$H_k = \sum_{n=1}^{N} m_k, \tag{3.5}$$

where H_k is the k-th bin count, $m_k = 1$ if $c_k \leq F(n) \leq c_{k+1}$ and $m_k = 0$ otherwise, F is the array of pitch values and (c_k, c_{k+1}) are the bounds on k-th bin. If the histogram is weighted by duration, N is the number of pitch values. If it is weighted by the number of instances, the pitch contour is first segmented and N corresponds to the number of segments.

3.4 Evaluation over a common dataset

A well annotated and comprehensive database is of fundamental value for this field of research. The Carnatic and Hindustānī datasets, taken from the growing collections of CompMusic project (Serra (2011)), provide a convenient mechanism of organization and retrieval of audio and metadata (Serra (2012)). We use full-length recordings, which range in length from 2 minutes to 50 minutes. The dataset encompasses all the possible improvisational and compositional forms of Carnatic music, and Dṛpad and Khayāl genres of Hindustānī music[7].

Table. 3.1 shows the configurations, the tasks in which they are used and their short handles which are used henceforth. The discussion on relevance of using these configurations is deferred to subsequent sections.

The first experiment is based on matching just the scale templates. It is followed by an experiment which matches pitch-class profiles with and without availing the tonic information. Later, we test the same approach choosing finer bin resolutions in the distribution. In the last experiment, we briefly discuss how the

[7]The other genres in Hindustānī include Ghajal, Ṭhumrī, Kavvāli etc., which are classified as semi-classical in nature.

Task	Collection	Recordings/rāga	Rāgas
Template matching	Carnatic	7	14
	Hindustānī	4	17
Pitch-distribution based approaches	Carnatic	5	43
		5	12
		10	12
	Hindustānī	5	16
		5	8
		8	8

Table 3.1: Different datasets derived from CompMusic collections.

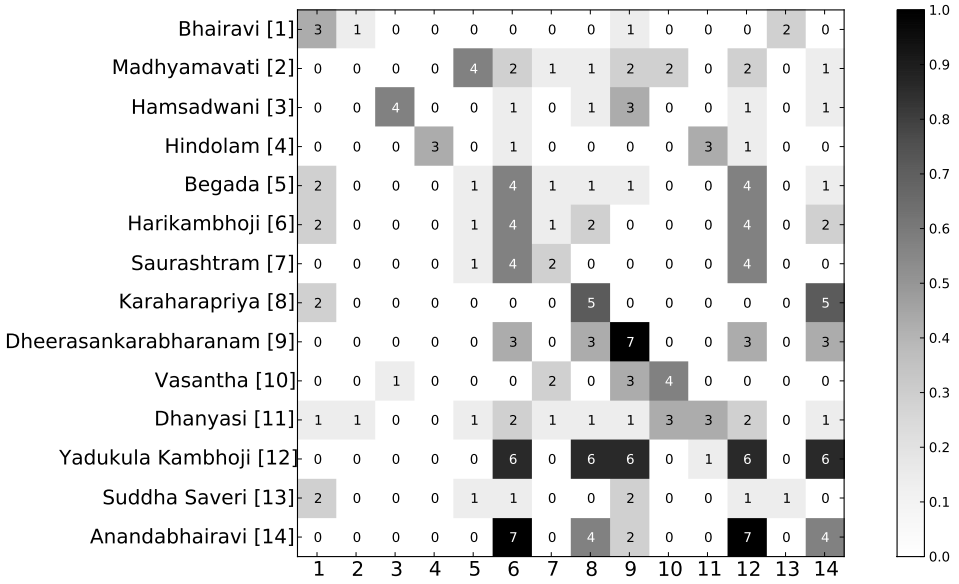
intonation information helps to disambiguate rāgas which are confused when using the pitch-class profiles alone. Before reporting the results of the evaluation, we present the details of the datasets used, and the common computational steps involved.

Template matching

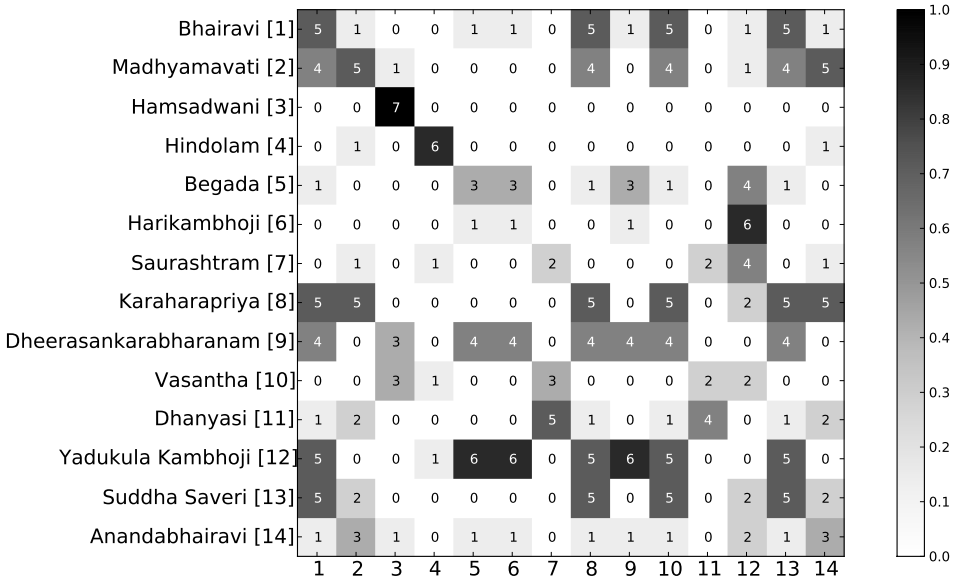
The template matching approaches rely on just the svara positions. To evaluate such approaches, it is necessary to have a dataset which is representative of 72 mēḷakarta rāgas for Carnatic music and ten tāṭ families in Hindustānī music. In mēḷakarta scheme of rāgas, each one differs by a svara from its neighbors. Hence, having several pairs of rāgas in the dataset such that they are also neighbors in the mēḷakarta scheme contributes to the dataset's completeness. However, it is difficult to find recordings for most of the mēḷakarta rāgas. The Carnatic and Hindustānī datasets we chose for this purpose consist of 14 and 17 rāgas respectively (see. Table. 3.1). The rāgas are chosen such that they differ in constituent svaras.

Figs. 3.2 and 3.3 show the confusion matrices for both the methods (A_{th}, A_{de}, ref. to sec. 3.3) over Carnatic and Hindustānī datasets respectively. F-measures for (A_{th}, A_{de}) over Carnatic and Hindustānī datasets are (0.26, 0.23) and (0.35, 0.39) respectively. In both the methods, there are cases where the actual rāga is correctly matched, and also the cases where there is another rāga which scores equal to the actual rāga.

It is interesting to note that A_{th} suits Carnatic and A_{de} suits Hindustānī in comparative terms though the difference in their performances is marginal. This can be a consequence of the fact that svaras in Carnatic music are rarely sung without

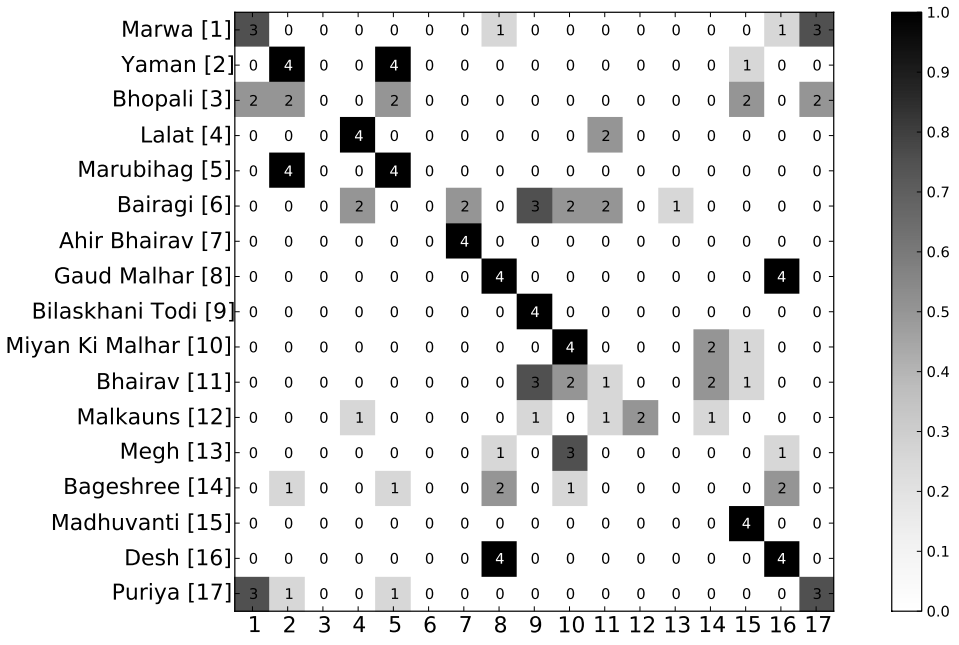


(a) A_{th} on Carnatic dataset

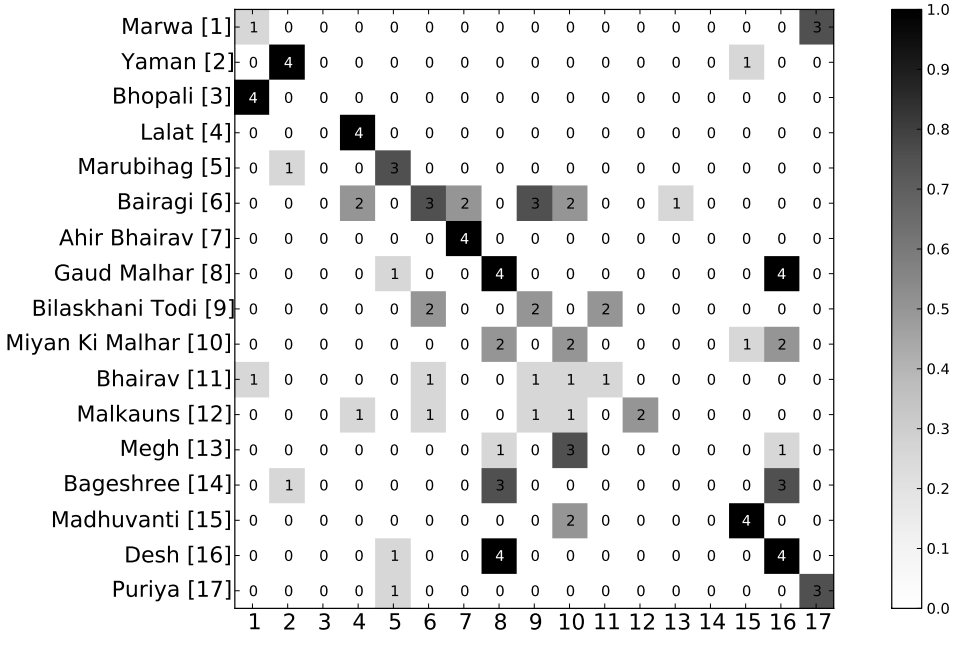


(b) A_{de} on Carnatic dataset

Figure 3.2: Confusion matrices for the two template matching methods (A_{th} and A_{de}) on Carnatic datasets. The grayness index of (x,y) cell is proportional to the fraction of recordings in class y labeled as class x.



(a) A_{th} on Hindustānī dataset



(b) A_{de} on Hindustānī dataset

Figure 3.3: Confusion matrices for the two template matching methods (A_{th} and A_{de}) on Hindustani datasets. The grayness index of (x,y) cell is proportional to the fraction of recordings in class y labeled as class x.

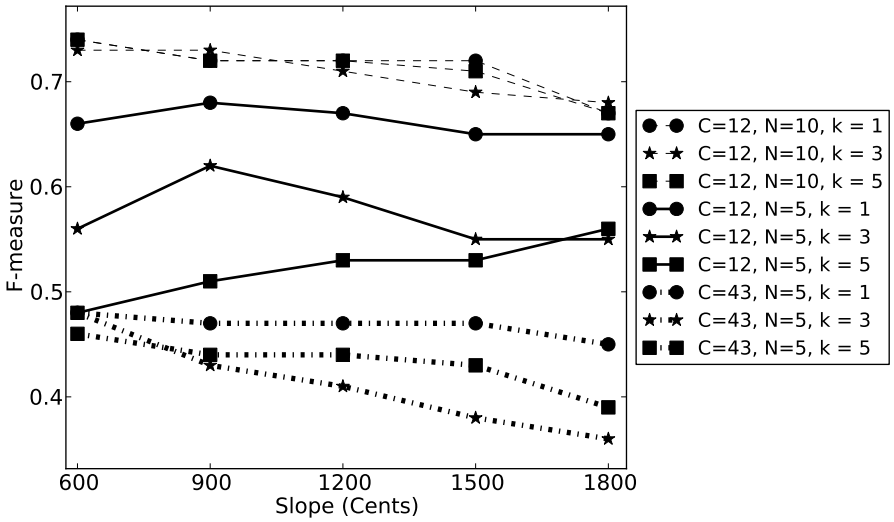
gamakas, which influence the peak characteristics of svaras in histogram. As as result, the peaks corresponding to such svaras often appear as slides with a little bump, which can not be identified using a conventional algorithm to find local maxima and have a good probability to be accounted for only in A_{th}. Where as in Hindustānī music where the svaras are held relatively steady, A_{de} performs marginally better than A_{th}. On the other hand, it is to be noted that the methodologies which we have adopted to obtain the svaras used in a given recording are not the best and can be improved. A further step ahead would be to obtain histograms from stable pitch-regions. In order to quickly test if this helps, from the stable pitch-regions obtained, we picked the most recurrent intervals from each recording and matched those against templates obtained in A_{th} and A_{de}. The accuracies obtained are not very different from those reported in Figures. 3.2 and 3.3. This observation reinforces our belief that rāga classification using template matching approach alone cannot be scaled to classify, say, even just the mēḷakarta rāgas.

Distributions constrained to "steady regions"

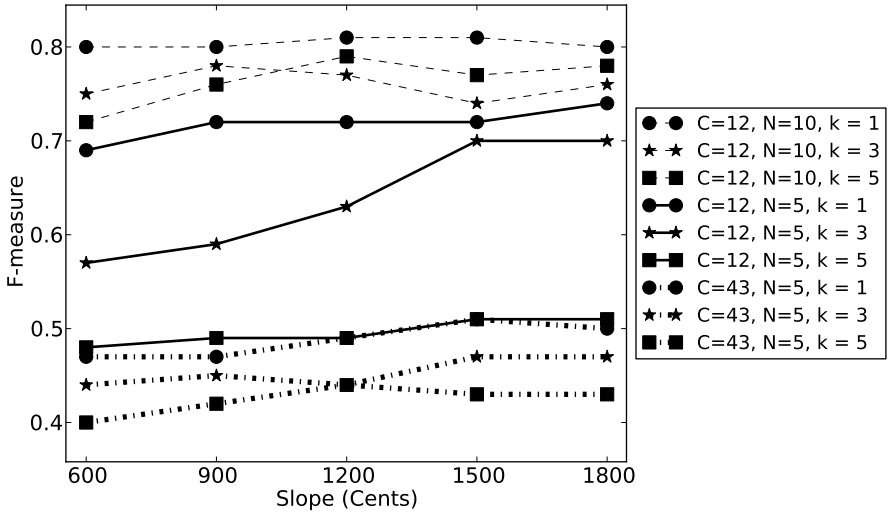
When the number of rāga classes also include janya rāgas of a few of the mēḷakarta rāgas in the dataset, we will require additional information beyond svara positions. Pitch-class profiles contain information about relative usage of svaras besides their positions. Though several of such mēḷakarta-janya rāga groups can possibly be distinguished using the template-matching approaches, for most cases we expect the additional information from pitch-class profiles to contribute for a better classification.

There are two crucial factors in determining stable svara regions: slope threshold and time-duration threshold (T_{slope} and T_{time}, 3.3). We conducted three experiments to check their impact on the performances of $P_{instance}$ and $P_{duration}$. The datasets chosen for this task are listed in Table. 3.1. The datasets are chosen to facilitate observation of the impact of number of rāgas, and number of recordings per rāga in all the experiments.

In the first experiment, T_{time} is set to 0 and T_{slope} is varied from 600 to 1800 cents in steps of 300. Figs. 3.4 and 3.5 show the results. The performance of $P_{duration}$ stays the same while that of $P_{instance}$ slightly degrades with increasing T_{slope}. With lower T_{slope}, we observed that a svara sung with even a slight inflection is divided into multiple segments, which primarily effects $P_{instance}$. Better performance of $P_{duration}$ over $P_{instance}$ in general, also explains the slight increase in the performance of $P_{instance}$ at lower T_{slope}.

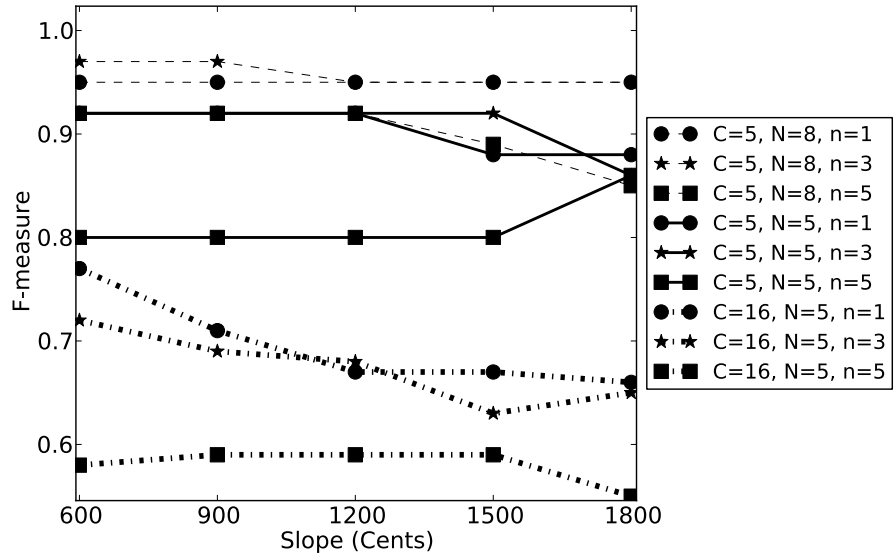


(a) Carnatic datasets with $P_{instance}$

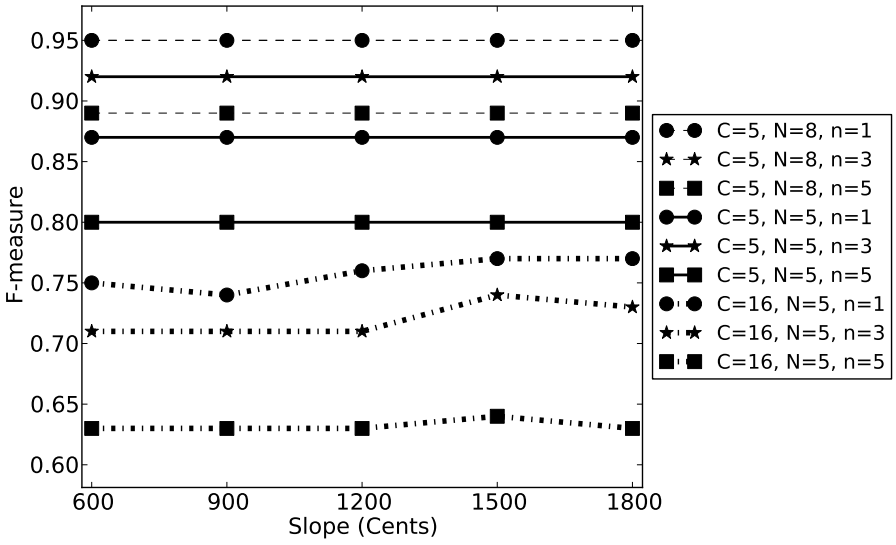


(b) Carnatic datasets with $P_{duration}$

Figure 3.4: F-measures for performances of $P_{instances}$ and $P_{duration}$ on Carnatic and Hindustānī datasets, with T_{time} set to 0 and T_{slope} varied between 600 to 1800. C & N denote number of rāgas, number of recordings per rāga in the dataset. k denotes number of neighbors in k-NN classification.



(a) Hindustānī datasets with $P_{instance}$



(b) Hindustānī datasets with $P_{duration}$

Figure 3.5: F-measures for performances of $P_{instances}$ and $P_{duration}$ on Carnatic and Hindustānī datasets, with T_{time} set to 0 and T_{slope} varied between 600 to 1800. C & N denote number of rāgas, number of recordings per rāga in the dataset. k denotes number of neighbors in k-NN classification.

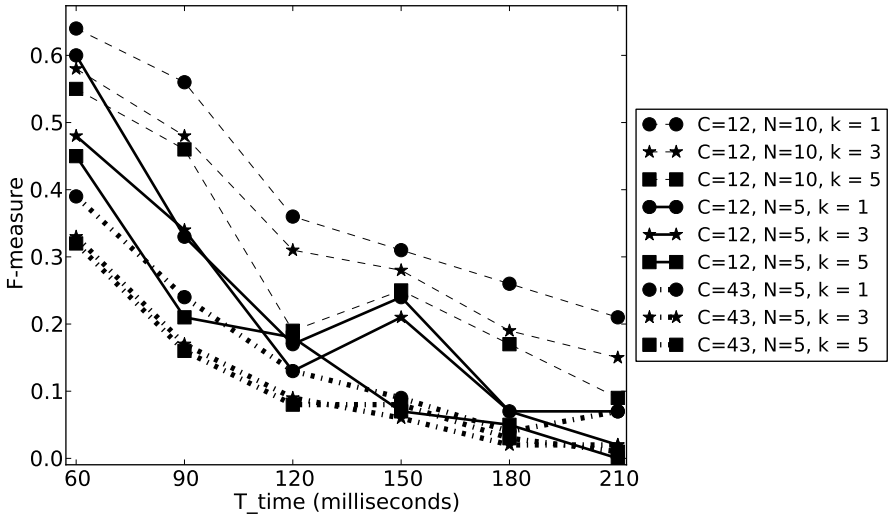
In the second experiment, T_{slope} is set to 1500 and T_{time} is varied from 60 to 210 milliseconds, in steps of 30. Figs. 3.6 and 3.7 show the results. With increasing T_{time}, the amount of pitch-data shrinks drastically for Carnatic recordings. This taxes the classification performance heavily (see. fig. 3.6a, 3.6b). Further, the effect is even more pronounced on the performance of $P_{instance}$. On the other hand, these observations are not as strong in the results over Hindustānī datasets (see. fig. 3.7a, 3.7b). The can be explained by presence of long steady svaras in Hindustānī music, which aligns with our observations in sec. 3.3.

In the third experiment, we set T_{slope} to 1500 and T_{time} to 0, and classified Carnatic and Hindustānī datasets using $P_{instances}$ and $P_{duration}$ and $P_{continuous}$ $(24\,bins)$ to compare their performances. Fig. 3.8 shows the results. $P_{duration}$ outperforms the other two, which is more evident in the classification of Carnatic rāgas. This implies that svara durations play an important role in determining their relative prominence for a particular rāga realization. This is consistent with the fact that long sustained svaras like dīrgha svaras play a major role in characterizing a rāga than other functional svaras which occur briefly in the beginning, the end or in the transitions. The benefit of identifying stable svara-regions is seen in the superior performance of $P_{duration}$ over $P_{continuous}$ $(24\,bins)$.

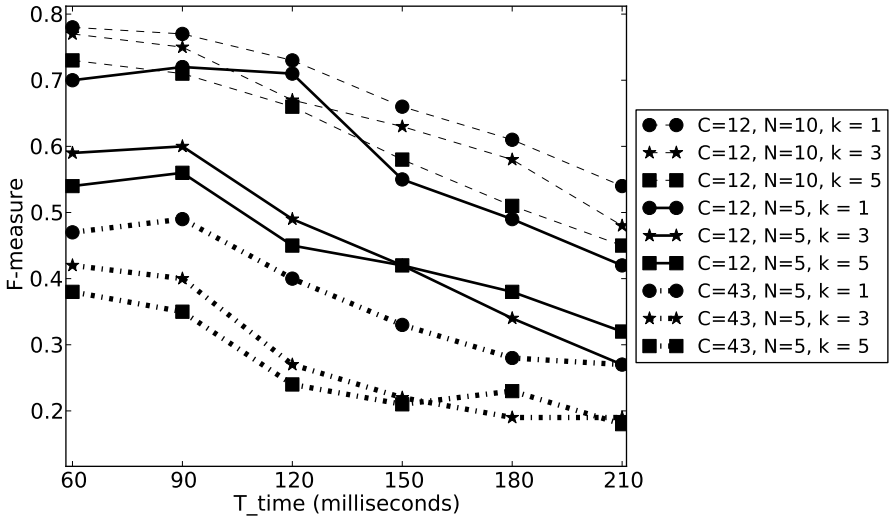
From fig. 3.8b, it can be seen that as the number of classes increase, the performance of $P_{duration}$ is less affected compared to others. It is possible that a fall in number of samples per class causes it, but this argument can be ruled out as there is no noticeable difference between the results based on the three pitch-distributions when we keep number of classes at 5 for Hindustānī and 12 for Carnatic, and vary the number of samples per class (see fig. 3.8). Further, a considerable rise in f-measures with an increase in number of samples per class, for all the three distributions, indicate that there is a large diversity in the performances of a rāga. This also falls in line with the general notion that one has to get familiar with a number of compositions in a rāga in order to learn it.

Distributions obtained from full pitch contour

In this experiment, we would like to see the impact of bin resolution and tonic information on classification performance. We vary bins from 12 to 1200, and evaluate the performance of pitch-distributions thus obtained, with and without tonic information. Figs. 3.9 and 3.10 show the results. For both Carnatic and Hindustānī datasets. when the tonic information is available, pitch-distributions with higher number of bins performed better. But it is also clear that beyond 24 bins, the accuracies for each combination of k and dataset, more or less remain saturated. In the case where tonic information is not given, however, there is a

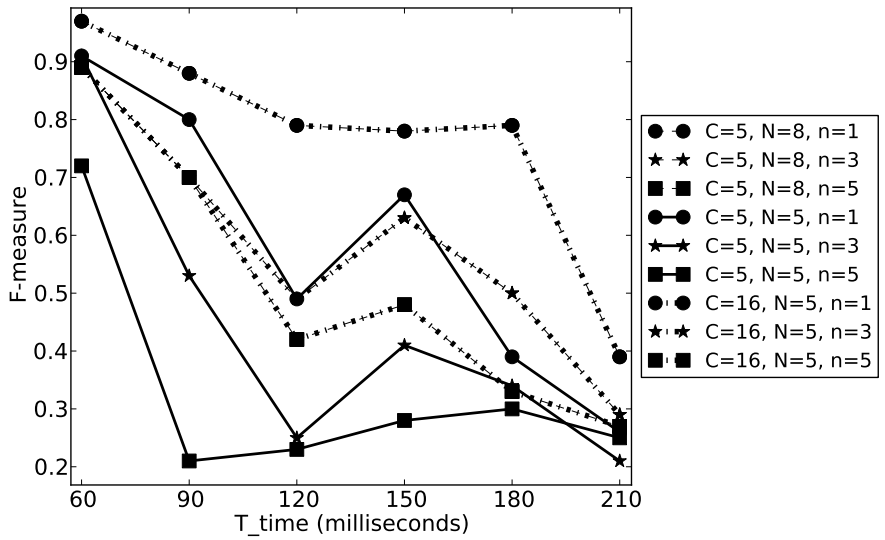


(a) Carnatic datasets with $P_{instance}$

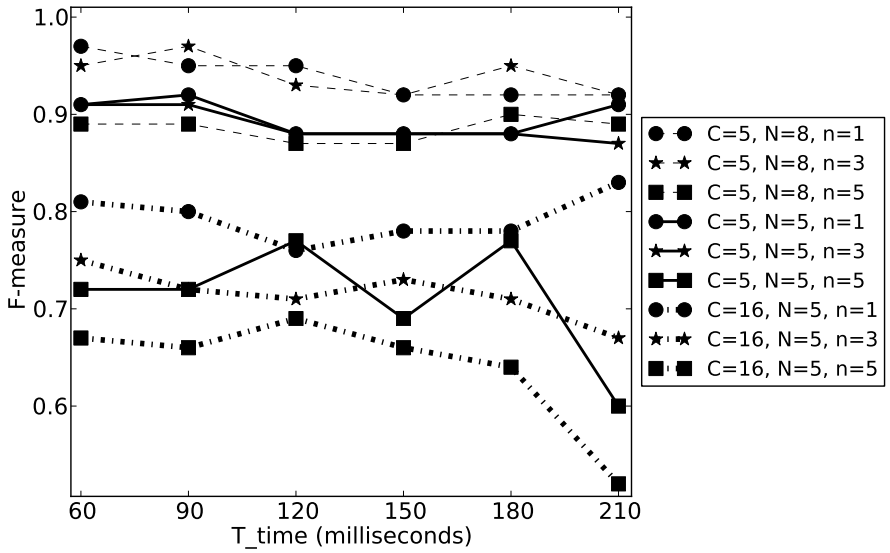


(b) Carnatic datasets with $P_{duration}$

Figure 3.6: F-measures for performances of $P_{instances}$ and $P_{duration}$ on Carnatic datasets, with T_{slope} set to 1500 and T_{time} varied between 60 to 210. C & N denote number of rāgas, number of recordings per rāga in the dataset. k denotes number of neighbors in k-NN classification.

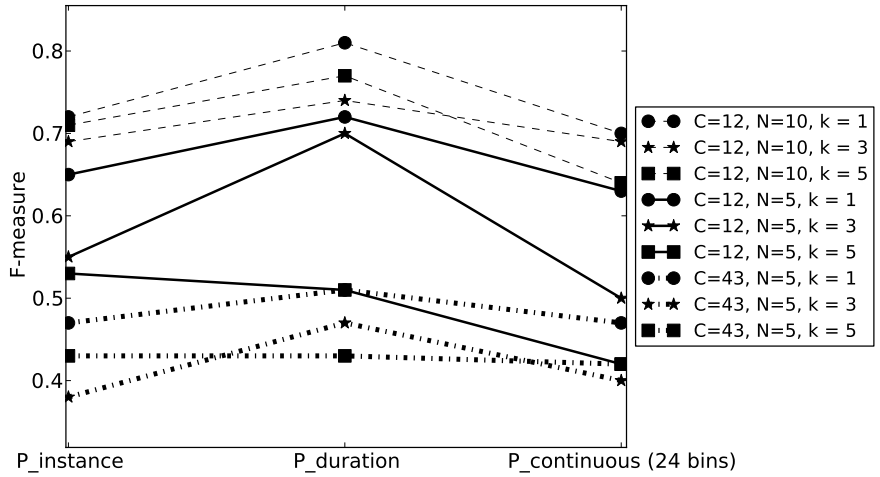


(a) Hindustānī datasets with $P_{instance}$

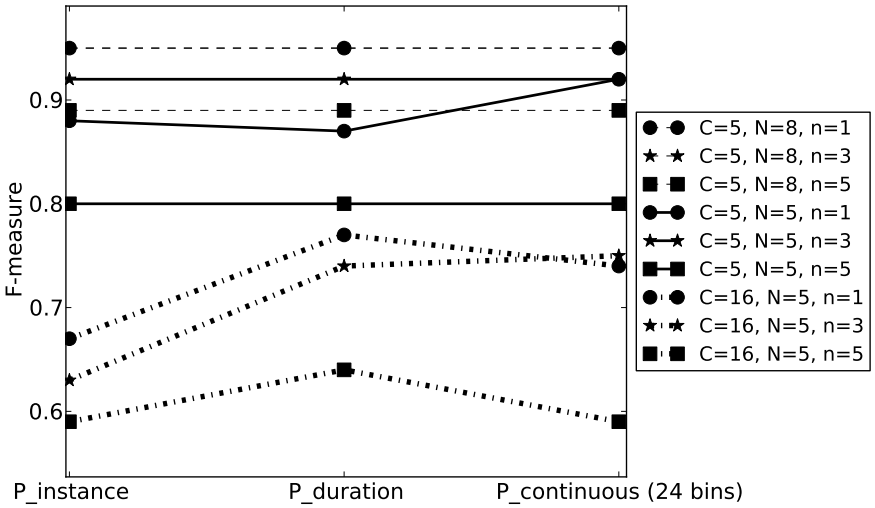


(b) Hindustānī datasets with $P_{duration}$

Figure 3.7: F-measures for performances of $P_{instances}$ and $P_{duration}$ on Hindustānī datasets, with T_{slope} set to 1500 and T_{time} varied between 60 to 210. C & N denote number of rāgas, number of recordings per rāga in the dataset. k denotes number of neighbors in k-NN classification.



(a) Carnatic



(b) Hindustānī

Figure 3.8: Comparison of the performances of different pitch class profiles ($P_{instances}$, $P_{duration}$, $P_{continuous}$ $(24\,bins)$ on Carnatic and Hindustānī datasets. C & N denote number of rāgas, number of recordings per rāga in the dataset. k denotes number of neighbors in k-NN classification.

slight but comparatively steady increase in the accuracies with increasing number of bins.

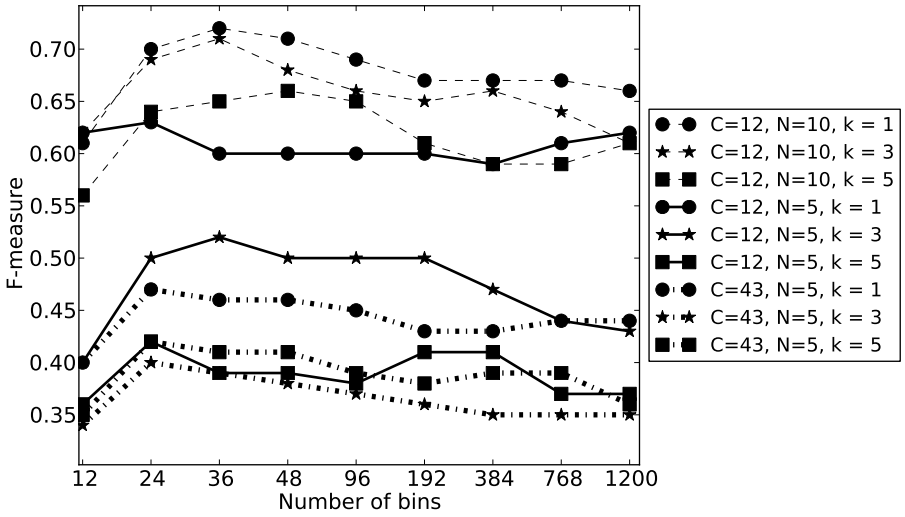
However, the tonic identification method we used has an error rate of 7%. The results reported in figs. 3.9a & 3.10a, therefore carry this error too. In order to realize the impact of tonic information on the performance of the system, we have analyzed the cases where the systems with, and without tonic information failed. In the cases where a correct class label is output, let T_s and N_s be the set of cases where the tonic information is used and not used respectively. Then, $\frac{|Ts-Ns|}{N}$ where N is total number of recordings in the dataset, is the proportion of cases where the availability of tonic information has helped in improving the accuracy. This comes out to be 5% for most configurations run with $P_{continuous}$ (12 bins). As there is a 7% inherent error in the tonic identification method, we expect this proportion to go up further.

3.5 Summary and conclusions

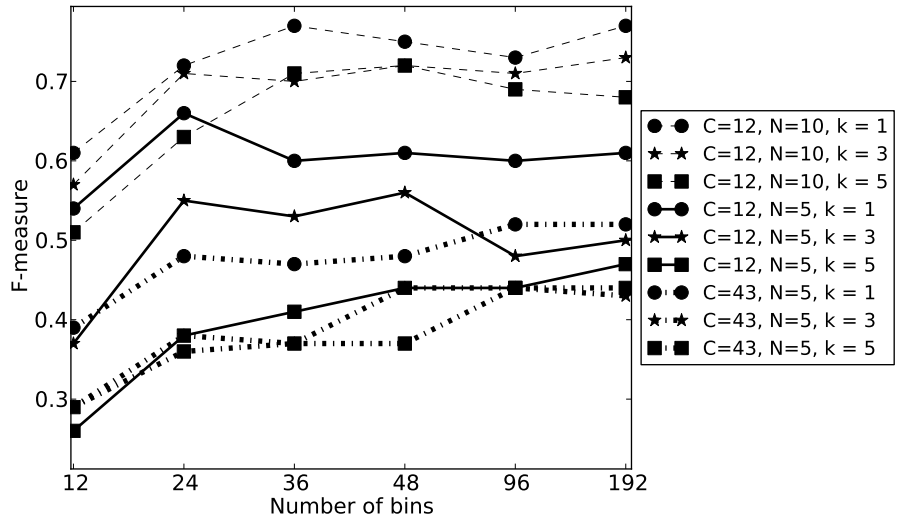
In this chapter, we have presented a brief categorical review of computational and analytical approaches to understanding the melodic framework of Indian art music. Of the different aspects that are studies, viz., validation of microtonal intervals, tuning and svara positions, identifying raaga-specific patterns, perception of emotion, and the automatic classification of raagas, the last one is the most explored one. Within the raaga classification approaches, we have thoroughly evaluated the ones based on pitch-distributions on a larger and comprehensive database. In template matching, we have explored two methods to determine rāga templates from pitch-contour. In the approaches based on distributions constrained to stable-regions, we have reported our experiments varying the parameters involved in determining stable-regions. Further, on unconstrained pitch-distributions, the impact of different bin resolutions and the effect of adding tonic information are reported.

In all the experiments reported thus far, the overall best accuracy among each of the datasets is way higher than chance, indicating the effectiveness of pitch-class profile as a feature vector for rāga identification. It is encouraging to find that a simple pitch distribution based approaches are able to exploit considerable information about the underlying rāga. However, we soon observed the limitations of these different approaches on introducing more classes/raagas.

Including the non-steady regions in the pitch-class distribution did not help in the approaches surveyed. However as mentioned before, the gamakas play very important role in characterizing the rāga as evidenced by performance as well as

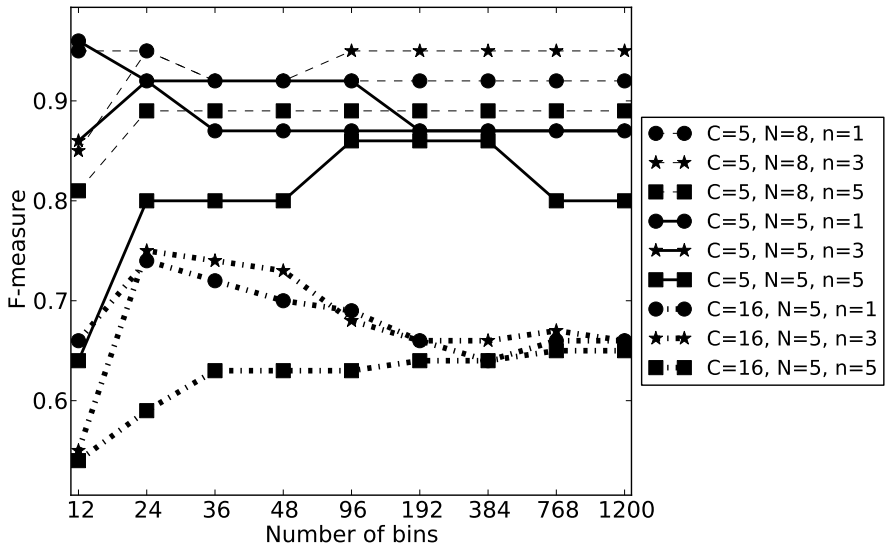


(a) Carnatic datasets with tonic information

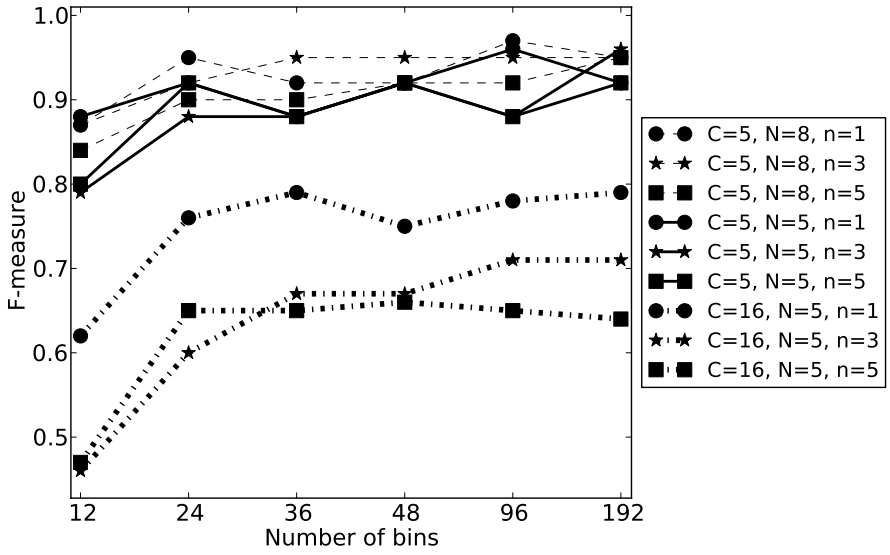


(b) Carnatic datasets without tonic information

Figure 3.9: Comparison of the performances of $P_{continuous}$ with different bin-resolutions on Carnatic datasets. C & N denote number of rāgas, number of recordings per rāga in the dataset. k denotes number of neighbors in k-NN classification.



(a) Hindustānī datasets with tonic information



(b) Hindustānī datasets without tonic information

Figure 3.10: Comparison of the performances of $P_{continuous}$ with different bin-resolutions on Hindustānī datasets. C & N denote number of rāgas, number of recordings per rāga in the dataset. k denotes number of neighbors in k-NN classification.

listening practices followed. Therefore, for gamakas to be effectively exploited in automatic identification, it is necessary to tune the approaches in ways that allow capturing this information in a musically meaningful manner. An aggregate pitch distribution which discards all time sequence information seems inadequate for the task. One way is to directly capture their temporal characteristics such as the actual pitch variation with time. Another viable way is to abstract the movements in the pitch contour such that the abstraction itself makes musical sense.

Approaches aimed at understanding the pitches used in Indian art music (such as Levy, 1982; Serra, 2011; Krishnaswamy, 2003), or for that matter those which use information about pitches to further understand raaga have usually followed a 'stable region' approach, which inherently assumes that svaras are points on frequency spectrum. However, from our discussion in 3.1, it is clear that such assumption does not help in understanding the identify of svaras and therefore of raagas. So far, tuning analysis has been employed to explain the interval positions of Carnatic music with one of the known tuning methods like just-intonation or equal-temperament Krishnaswamy (2003); Serra (2011), and this has been criticized owing to the fundamental difference between popular western music traditions in conceptualizing the way pitches are employed in the performance. Therefore, future research on aspects concerning raaga must be grounded in its conceptualization as it is practiced today. A study that seeks to find the exact locations of the svaras does not seem to address a musically sound problem, whereas studying the nature of usage of pitches about each svara, and within the greater context of a given raaga seem to be of value.

Chapter 4

Music knowledge representation

Knowledge representation is a subdomain of research within Artificial Intelligence. The primary objective of this field is to enable machines to understand and perform tasks which would otherwise need human analytical capabilities. This further requires a computational representation of knowledge concerning the domain of interest. Grounded in formal logic, such representations would allow reasoning over a knowledge-base (KB) allowing machines to perform complex tasks.

With an exponential rise of the data available on the World Wide Web (WWW), the domain of knowledge representation has seen a revived attention in the last decade. In this direction, semantic web is an extension to the current web that facilitates machine consumption of the data alongside humans, and their interoperability (Berners-Lee et al. (2001)). The linked open data movement is another step towards this which leverages the semantic web technologies to promote pragmatic adaptability of semantic web principles in the web community (Bizer et al. (2009a)). In this chapter, we first introduce the semantic web technologies and discuss the relevance of knowledge representation for music in the context of semantic web. We review the current state of the art within the intersection of MIR and semantic web, specifically in two categories: ontologies (sec. 4.2) and linked data (sec. 4.3).

4.1 Introduction to Semantic web

Web to Semantic web

WWW has evolved through many milestones since its infancy of passively sharing information (Aghaei (2012)). Broadly speaking, various stages in this evolution are marked by the shift in its emphasis over the information shared, the users and the machine itself. In its formative years, the web had been a place for sharing information, a one-way communication platform where the users of the web were passive consumers of the information hosted. During this time, the emphasis has solely been on the information shared. This period is loosely termed as Web 1.0 where various sources of the information are inter-connected by hyperlinks allowing users to navigate between them. With an advent of web logging services which allowed users to write and publish on the web, the status-quo was dramatically transformed to place its emphasis now on users and the content they generate. The web services have also access to better technology that allowed them to constantly improve and adapt to the needs. The social networking and media-sharing services have further added to this cause, leading up to Web 2.0 (O'Reilly (2007)).

With users being engaged in content creation, the size of the web has multiplied rather quickly. Most of the content is either unstructured, or structured often in relational databases with an ad-hoc schema. This has restricted the consumption of this data mostly to human users and also resulted in data islands. There is a discerning need for this data to be machine-readable for further-reaching applications. As a result, the primary emphasis of Web 3.0, also known as Semantic web, is on structuring the data and interlinking different sources of data, with an objective of making the web machine readable (Berners-Lee et al. (2001)).

Semantic web is based on a set of standards and technologies shown in fig. 4.1. We discuss those which are W3C recommendations[1]. The first layer has URI (Uniform Resource Identifier)[2] and Unicode as the constituent blocks. The latter allows any human language to be used on the semantic web under a common standard. This is an important fundamental consideration. Because in the early days of the web, languages from around the world have their own coding schemes often clashing with others. This had resulted in a practical impossibility of a multi-lingual web. URI, which manifests as a URL (Uniform Resource Locator)

[1] The technologies or standards that acquire W3C recommendation status are usually considered to be extensively reviewed, tested and are stable.

[2] IRI (Internationalized Resource Identifier) is an international variant of URI that is not limited to the set of ASCII characters. We use IRI and URI interchangeably.

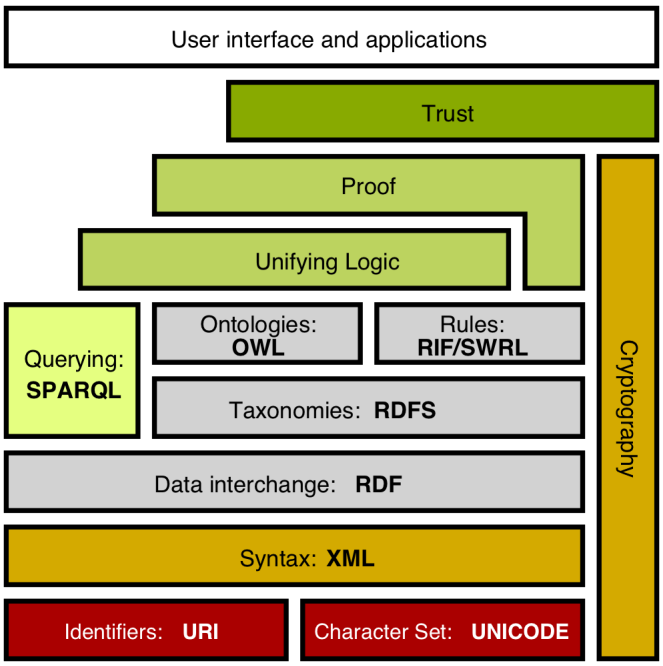


Figure 4.1: A stack of standards and technologies that make up the Semantic web.

in the context of web or URN (Uniform Resource Name) elsewhere, is a string of Unicode characters that helps in identifying any resource. An example URL is http://dbpedia.org/resource/Purandara_Dasa (a Carnatic music composer), and an example URN is urn:isbn:978-2-9540351-1-6 (a publication).

The second layer consists of XML (Extensible Markup Language) which provides a common format to structure data into documents. A data model is defined to structure the data from a given source using XML Schema[3], whereas XML Namespace[4] allows referring to information in multiple sources as is often the necessity on the semantic web.

The RDF (Resource Description Framework)[5] constitutes the third layer in the stack. Knowledge in a given domain is represented as a set of statements. RDF defines a simple framework which allows creation of such statements in the form of triplets consisting a subject, an object, and a predicate that establishes relation between them. These statements result in an interconnected graph of such

[3]https://www.w3.org/XML/Schema
[4]https://www.w3.org/TR/1999/REC-xml-names-19990114/
[5]http://www.w3.org/RDF/

statements. We consider the class of *Composers* to illustrate the use of RDF and other semantic web technologies. Listing. 4.1 shows these sentences using RDF: i) Tyagaraja is a Composer, ii) He composed Endaro Mahanubhavulu, iii) It is a composition and iv) It is composed in Sri raga.

Listing 4.1: Sample RDF statements.

```
<?xml version="1.0"?>

<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:so="http://abc.org/demo#">

<rdf:Description
rdf:about="http://abc.org/demo/Tyagaraja">
  <rdf:type rdf:resource="http://abc.org/demo/Composer" />
  <so:has_composition rdf:resource="http://abc.org/demo/
      Endaro_Mahanubhavulu" />
</rdf:Description>

<rdf:Description
rdf:about="http://abc.org/demo/Endaro_Mahanubhavulu">
  <rdf:type rdf:resource="http://abc.org/demo/Composition" />
  <so:has_raaga rdf:resource="http://abc.org/demo/Sri_raga" />
</rdf:Description>

</rdf:RDF>
```

RDF Schema[6] in the next layer provides constructs that allow us to define a set of classes and properties, more generally a vocabulary appropriate to an application/purpose using RDF. As an example, we can define a class called Artist and make statements about it (like, all Artists are Person, Composer is an Artist where Person, Composer are other classes). This allows to create class and property hierarchies. The code snippet from listing. 4.1 shows the usage of constructs from RDF Schema to describe the class of *Composer*. We omit the header declaring namespaces (i.e., rdf and so from listing. 4.1) from snippets hereafter.

[6] http://www.w3.org/TR/rdf-schema/

Listing 4.2: Class hierarchy defined using RDF Schema constructs.

```
<rdfs:Class rdf:ID="Artist" />
<rdfs:Class rdf:ID="Composer" />
<rdfs:Class rdf:ID="Person" />

<rdf:Description
rdf:about="#Composer">
<rdf:subClassOf rdf:resource="#Artist" />
</rdf:Description>

<rdf:Description
rdf:about="#Artist">
<rdf:subClassOf rdf:resource="#Person" />
</rdf:Description>

</rdf:RDF>
```

Next in the stack are OWL (Web Ontology Language), RIF (Rule Interchange Format)/SWRL (Semantic Web Rule Language) and SPARQL (SPARQL Protocol and RDF Query Language). OWL[7] is a class of semantic markup languages, which is grounded in description logics of varying expressiveness. It builds on top of RDF and RDF Schema, and includes more expressive constructs such as universal and existential quantifiers, and cardinality constraints. For instance, as the example from listing. 4.1 shows, OWL constructs facilitate expressing that a class called Person can only have one birth place. In this listing, in continuation to previous listings, we define a Place class, then a property called has_birthplace, confining its domain and range to Person and Place classes respectively. Then we use owl:Restriction construct to enforce the constraint that a person can have only one birth place.

Listing 4.3: Usage of cardinality constraints in OWL.

```
<rdfs:Class rdf:ID="Place" />

<owl:ObjectProperty rdf:about="has_birthplace">
  <rdfs:domain rdf:resource="#Person"/>
  <rdfs:range rdf:resource="#Place"/>
</owl:ObjectProperty>

<owl:Restriction>
  <owl:onProperty rdf:resource="#has_birthplace" />
  <owl:maxCardinality rdf:datatype="&xsd;nonNegativeInteger">1</
      owl:maxCardinality>
</owl:Restriction>
</rdf:RDF>
```

[7] http://www.w3.org/TR/owl-ref/

There are certain semantics that are conditional. These need to be expressed as rules over the hierarchies and statements created with the constructs provided in OWL and RDFS. RIF[8] and SWRL[9] are two such standards/rule languages defined for this purpose. Like SWRL, there exist several rule languages. RIF is a standard that facilitates integration of synthesis of rulesets across such languages. These rules allow inferring facts that can be true given a set of RDF statements, and can also trigger actions that result in modifications to KB. For instance, let us say that we would like to classify an artist who is both a singer and a composer of a song as a singer-songwriter of that song. This can be achieved using RIF rule as specified in listing. 4.1.

Listing 4.4: Example RIF rule.

```
Document(
Prefix(rdf <http://www.w3.org/1999/02/22-rdf-syntax-ns#>)
Prefix(rdfs <http://www.w3.org/2000/01/rdf-schema#>)
Prefix(so <http://abc.org/demo#>)

Group(
    Forall ?Artist ?Song (
    If   And(so:singer(?Artist ?Song) so:composer(?Artist ?Song))
    Then so:singer_songwriter(?Artist ?Song)
    )
  )
)
```

SPARQL[10] allows to efficiently query over the resultant knowledge-graph. This knowledge graph can be a set of simple RDF statements, or ontologies and KBs built using RDFS and OWL. Listing. 4.1 shows an example SPARQL query that retrieves the names of composers who are born in India. These queries also help in adding new statements to the KB based on a set of observations. For instance, for any *Person* who *composed* at least one *Composition*, we add a statement that is the *Person* is a *Composer*.

[8] https://www.w3.org/TR/rif-overview/
[9] https://www.w3.org/Submission/SWRL/ - not a w3c recommendation yet.
[10] https://www.w3.org/TR/sparql11-overview/

Listing 4.5: Example RIF rule.

```
PREFIX rdf: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX so:  <http://abc.org/demo#>

SELECT ?Composer
WHERE {
  ?Composer rdf:type so:Composer .
  ?Composer so:has_birthplace so:India .
}
```

These technologies and standards allow sharing complex knowledge about a domain with great flexibility in their form, yet facilitating interoperability between different sources of data. Ontologies further facilitate evolution of data schemas without impeding the interaction between applications. MusicBrainz[11], an online public music metadata cataloging service, is one such example. It publishes the editorial metadata about various entities of an audio recording together with relationships between those entities, as XML serialized RDF. Client applications such as MusicBrainz Tagger and MusicBrainz Picard use this medium to communicate with the data server. RDF facilitates easy merger of data from different sources, for example, from MusicBrainz and DBpedia[12], as it dereferences each entity using an URI on the web.

Ontologies

Relational databases have been arguably the most popular choice for structuring data served and created on the web as of date. The major difference between a relational database schema and an ontology lies in exposing the semantics of things they are defined to describe. Take the specific case of music metadata. The intended definition of *Artist* changes from one relational database to another across different services. For instance, a database might list only people who perform music as artists, while another might also include people who compose. This practice in itself is harmless, and in fact a necessity for diverse needs of different cultures, or even different applications. However, as the semantics of *Artist* are formally undefined in a database schema, there is no logical way to verify if the intended semantics are mutually compatible, and consequently if the two given sources of data can be interlinked.

The first requirement in structuring the data while interconnecting multiple sources, is to define a common vocabulary or a schema, that all data sources can comply

[11] http://musicbrainz.org/
[12] http://dbpedia.org/

with. Such a vocabulary/schema is referred to as ontology. More formally, it is defined as "the manifestation of a shared understanding of a domain that is agreed between a number of agents and such agreement facilitates accurate and effective communications of meaning, which in turn leads to other benefits such as interoperability, reuse and sharing" (Uschold and Gruninger (1996)). Note that it is not necessary to use identical ontologies for two data sources to be integrable, it is only necessary that the conceptualization of different things in the ontologies used are not logically conflictive. Which in case they are, the data models now spell it out explicitly unlike in the case of a relational database schema.

There have been several ontologies built to describe most common objects/things on the web such as people, time, biological concepts and so on[13]. Schema.org[14] is an early effort in designing light-weight ontologies for things often found on commercial websites such as electronic goods, places, people etc. The content producers on the web are encouraged to structure their content using these ontologies for better visibility and aggregation, especially on search engines such as google, yahoo or bing.

Friend of a friend (FOAF)[15] is one of the most commonly used ontologies on the web. It is used in describing people and their relationships which result in a social network. This ontology has several use cases. Let's say we are defining the class *Artist* in an ontology, where every *Artist* is a person. The first step would be to define semantics of the *Person* class. But FOAF already has a *Person*[16] class. A good practice in the spirit of Semantic web would be to reuse this class and build on it as necessary. Several ontologies and KBs indeed reuse the existing ontologies in describing things and structuring data concerning them. Other commonly used ontologies include Dublin Core (DC)[17] for metadata, Semantically Interlinked Online Communities (SIOC)[18], Simple Knowledge Organization System (SKOS)[19] and geonames[20] for geospatial information.

[13] Ontology search engines are a common way to find if ontologies exist for a concept one is looking for. Some of the popular ontology search engines are listed here - https://www.w3.org/wiki/Search_engines

[14] http://schema.org

[15] http://www.foaf-project.org

[16] http://xmlns.com/foaf/spec/#term_Person

[17] http://dublincore.org/documents/dcmes-xml

[18] http://sioc-project.org

[19] https://www.w3.org/2004/02/skos/

[20] http://www.geonames.org/ontology/documentation.html

Data to Linked data

As a result of such efforts in developing ontologies for the commonly described things, an increasing number of sources are publishing their data structured using these ontologies. This has also resulted in reuse of ontologies between related data sources. For instance, one can borrow the concept of *Artist* from an existing ontology and build on top it as necessary, so long as there is no logical conflict in conceptualization of *Artist* in both the ontologies. Such proliferation of structured data on the web meant that the concerned sources of data are now linked with each other. This results in a transformation of the current web into a giant knowledge graph that subsumes all the different sources of data.

Fig. 4.2 shows the different sources of data in the linked open data (LOD) cloud, and the links between them which signify cross-referencing among the sources. This was the state of linked open data on the web as of 2014[21]. Notice that central to this network of sources is DBpedia[22]. It is a collection of all the structured data available on Wikipedia in its different language editions. Majority of this comes from *infoboxes*[23] and categories of Wikipedia. The media related datasets amount to around 2.2% of this data, while the majority (around 18%) consist of datasets coming from different governments thanks to their open data initiatives across the world. Elaborate statistics of the LOD cloud can be accessed at this link[24].

Linked data further fostered applications that can take advantage of richer semantics of the data and interoperability between various sources (Bizer et al. (2009a)). Some of the first applications were linked data browsers akin to the traditional web browsers that helped us navigate hyperlinked HTML pages. Examples include Tabulator[25] (Berners-lee et al. (2006)), Disco Hyperdata Browser[26], FOAF-naut[27] and DBpedia Mobile[28]. The other class of early applications were linked data search engines. A few like Falcons take a search query from the user and return results that match the query. This pattern mimics what the users are accustomed to traditional search engines like Google, Yahoo! and Bing. They further allow narrowing the search to specific type of object (viz., top-level classes in ontologies), or a concept (corresponding to a class in ontology), or a document. Typically these search engines are used in retrieving documents containing in-

[21] http://lod-cloud.net
[22] http://wiki.dbpedia.org
[23] https://en.wikipedia.org/wiki/Infobox#Wikipedia
[24] http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state
[25] https://www.w3.org/2005/ajar/tab
[26] http://wifo5-03.informatik.uni-mannheim.de/bizer/ng4j/disco
[27] http://jibbering.com/foaf
[28] http://wiki.dbpedia.org/projects/dbpedia-mobile

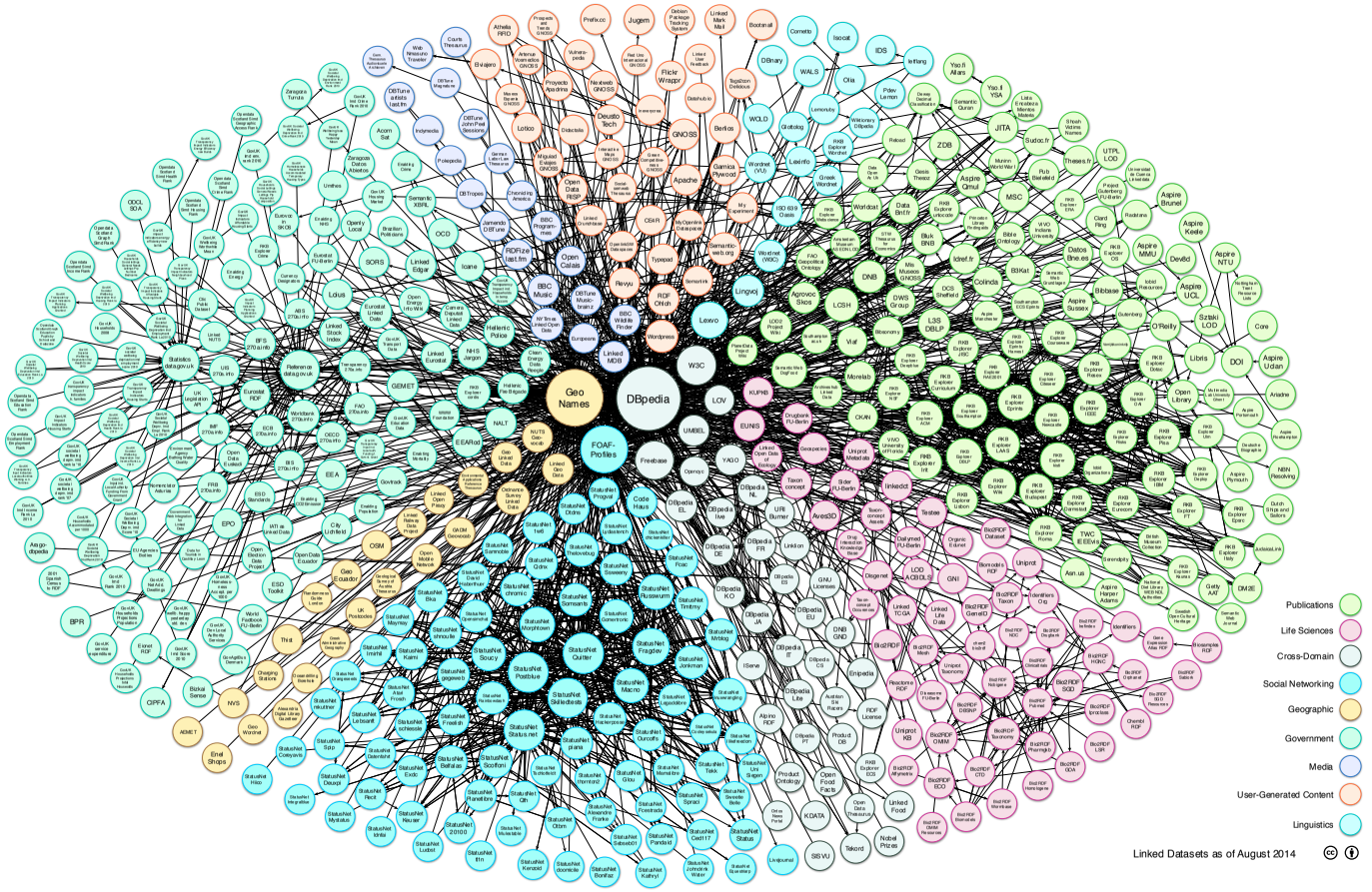


Figure 4.2: Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak.

stance data matching a given query, or the ontologies which have the concepts of interest. Few other search engines like Swoogle or Sindice offer access to their underlying indexes through an Application Programming Interface (API). These are intended for other applications built on top of linked data for easy access to RDF data, eliminating the need to built indexes for each application.

Commercial applications have also tapped into linked data to offer sophisticated functionalities. For instance, tasks like comparing products offerings to make a purchase decision can now be automated as product details across ecommerce portals can be mapped, their reviews analyzed and prices compared (Eg: Google shopping[29]). Applications can also combine information from multiple sources, for instance restaurant history and geographical location, to generate personalized recommendations of higher relevance to the user (Eg: Zomato[30]). Linked data has also been used as a means to data integration in information portals (Eg: British Broadcasting Corporation[31])

4.2 Role of ontologies in music information research

Why do we need ontologies?

In the MIR community, there is a growing interest towards developing culture-specific approaches Serra (2011); Serra et al. (2013). In order to better understand the interpretation of music concepts involved in computational modeling, researchers collaborate with musicologists and musicians. The conceptual description of music arising out of such collaborations by itself is not usually the end goal of information researchers. It forms a basis to build their models, but does not become a part of them. Consequently it is lost, making it difficult to be reproducible/accessible for reuse by other researchers. This is mainly a result of the lack of semantics in their data models. It limits the comprehensibility of those models, and poses difficult challenges to compare or integrate different models of a given musical concept.

The knowledge representation technologies discussed in the previous section can be utilized to bridge the gap between music information research and music theory. This will greatly enhance the impact they have on each other. Further, it will facilitate a knowledge-guided comparison and integration of different models of music concepts.

[29] https://www.google.com/shopping
[30] https://www.zomato.com
[31] https://www.w3.org/2001/sw/sweo/public/UseCases/BBC

Ontologies have been successfully used under similar context in other domains, like bioinformatics Stevens et al. (2000). Besides addressing the problems stated, ontologies, in the context of semantic web Berners-Lee et al. (2001), serve a host of purposes relevant to information retrieval. Advances in the development of efficient inference procedures for highly expressive description logics, have enabled them to provide a suitable logical formalism to ontology languages Baader (2007). This enables ontologies to be KBs, which combined with reasoning engines can be integrated into information systems. Further, an ontology specification allows the agents on semantic web to communicate with each other using common terminology and semantics. This in turn opens avenues to a multitude of machine-intelligent applications Berners-Lee et al. (2001).

A summary of past work

Consequently, there has been a noticeable interest in ontologies in the information research community (see ch.1 in Staab and Studer, 2009). Within the domain of MIR, ontologies that concern different aspects of music, like production, cataloging, consumption and analysis have been developed: some of them are specific to applications in which the developer intends to use the ontology (Eg: Playback Ontology[32]), and some others are a bit more general in their scope (Eg: Timeline[33] and Event[34] ontologies).

Garcia and Celma (2005); Celma (2006) developed a music ontology that is later used in recommending new songs to users. This ontology converts the XML schema of metadata covered in the MPEG-7 standard to an OWL ontology. Later this ontology is used in integrating three different schemas including MusicBrainz, Simac music ontology[35] and Kanzaki music ontology[36]. This ontology is populated with information extracted from the analysis of RSS feeds covering the listening habits of users (from last.fm), new music releases and events, podcasts, blogs and reviews. The resulting KB in conjunction with users' FOAF profiles is used in generating music recommendations. In summary, this work combines the users' personal information like demographics, socio-economics, relationships etc with their explicit music preferences to generate personalized recommendations.

[32] http://smiy.sourceforge.net/pbo/spec/playbackontology.html
[33] http://purl.org/NET/c4dm/timeline.owl
[34] http://purl.org/NET/c4dm/event.owl
[35] Both MPEG-7 and Simac music ontologies cannot be accessed anymore.
[36] http://www.kanzaki.com/ns/music#

Ferrara et al. (2006) builds on the so called MX formalism for music representation to develop and populate a music ontology. The MX format, an XML-based one, is a multilayer structure for music resource description with the following layers: structural, music logic, notational, performance and audio. This is shown to allow for an accurate representation of scores. Then they propose MX-Onto ontology which has two layers - Context[37] and Genre classification[38]. To capture the semantics of a music resource, information related to ensemble, rhythm, harmony and melody are abstracted by analyzing scores represented using MX formalism. The abstraction process involves simplifying data into pieces of information albeit with some loss of accuracy in description. For instance, say a given song is performed by an ensemble of two violins, a viola and a cello. Then this is abstracted as a string quartet rather than representing the ensemble using exact numbers and instruments. The context layer, once it is populated using the aforementioned abstraction processes over the scores, is used in populating the genre classification layer using SWRL rules.

Raimond (2008); Raimond et al. (2007) proposed Music Ontology (MO) which builds on top of several other ontologies. It subsumes Timeline ontology for representing the varied range of temporal information, including but limited to release dates, durations, relative positions and so on. It depends on Event ontology for representing events such as performances and compositions. For the editorial metadata, it builds on Functional Requirements and Bibliographic Records (FRBR)[39] (specifically for concepts like *Work*, *Manifestation*, *Item* and *Expression*) and Friend Of A Friend (FOAF)[40] (for its *Person* and *Group* concepts) ontologies, defining music specific vocabulary using the concepts and properties therein. As a result, MO allows representing music creation workflows, temporal information, events and editorial metadata. MO is also extended by other ontologies including Key Ontology[41], Chord Ontology[42], Tonality Ontology[43], Symbolic Music Ontology[44], Temperament Ontology[45], Instrument Ontology[46], and Audio Features Ontology[47] (see Fazekas et al., 2010, and the references there in for a de-

[37] http://islab.dico.unimi.it/ontologies/mxonto-context.owl
[38] http://islab.dico.unimi.it/ontologies/mxonto-genre.owl
[39] http://vocab.org/frbr/core
[40] http://xmlns.com/foaf/spec/
[41] http://purl.org/net/c4dm/ keys.owl
[42] http://purl.org/ontology/chord/
[43] http://purl.org/ontology/tonality
[44] http://purl.org/ontology/symbolic-music
[45] http://purl.org/ontology/temperament
[46] http://purl.org/ontology/mo/mit, this is a SKOS transformation of MusicBrainz instrument taxonomy
[47] http://motools.sourceforge.net/doc/audio_features.html

scription of these ontologies). It has been used as a means to integrate and publish data from MusicBrainz[48], BBC (Kobilarov et al. (2009)), DBTune[49] and so on.

Gracy et al. (2013) conducted an extensive case study in mapping several linked music data projects and further linking these to library data that are structured using more broader vocabularies which are not specific to music. In this regard, they pointed out the limitations of MO in representing the content creation workflow for music. In their work, they identify three sphere of activity in the life cycle of a musical work - composition, production and use. The first two are well-representable using MO. The user generated data like their listening history/habits and folksonomies which are part of the third sphere of activity, are not made a part of the creation workflow as envisaged in MO. However, this can at least be partially overcome by using ontologies like SIOC in tandem with MO. They also review how users search for information related to different facets fo music including artists, works, performance and so on, placing context as a valuable source to expand the scope of content-based music information retrieval.

Jacobson and Raimond (2009); Jacobson (2011) proposed Similarity Ontology (MuSim) which builds on top of MO to describe the connections between various music entities. Such connections are modeled as directed or undirected associations between two given entities, which include similarity. Further, these associations can be detailed with a description including the method used for creating them. This helps in not only expressing whether two given entities are connected, but also the method used for deciding the association. This meant that the MuSim provided provenance information alongside the description of connections.

Fazekas and Sandler (2011) proposed Studio Ontology[50] for representing the aspects concerning music production in detail. It builds on several existing ontologies such as Event and MO. It defines Device and Connectivity ontologies for describing devices used in music production and connections between them. It is designed in a modular way so as to be extended by ontologies specific to an application such as audio mixing, recording, effects and editing. Wilmering et al. (2013) extend this with Audio Effects ontology which helps in structuring the data about implementation and usage of audio effects.

The Ordered List Ontology[51], Counter Ontology[52] and Playback Ontology [53] are

[48] http://linkedbrainz.c4dmpresents.org/
[49] http://dbtune.org
[50] http://isophonics.net/content/studio-ontology
[51] http://purl.org/ontology/olo/core#
[52] http://purl.org/ontology/co/core#
[53] http://purl.org/ontology/pbo/core#

defined for representing playback related aspects of music such as playlists which are an ordered sequence of music recordings, and play count or skip. Ordered List and Counter ontologies can also be applicable in other domains where sequential information and counting objects are necessary for knowledge representation.

Tables. 4.1 and 4.2 summarize the ontologies discussed so far[54]. We have included the total number of axioms, classes and properties of each ontology which we have been able to resolve and/or find on the web. The 'Expr' column indicates expressivity, which tells us what variety of Description Logic was used in creating the ontology. We also summarize in a few words what the ontology is intended for.

The work discussed so far follows the top-down approach in annotating web resources with their semantics, where the ontologies are built manually before being used in describing data on the web. While this has the obvious benefits, it is also difficult to scale, especially to the web context ((Bizer et al., 2007, discusses the same and sketches the beginnings of DBpedia as a step to address this)). There is another line of work which follows a bottom-up approach that complements the former. These approaches, like the ones discussed by Wu et al. (2006); Specia et al. (2007), infer semantic models or partial ontologies from folksonomies - user given tags and categories. Wang et al. (2010); Sordo et al. (2012); Oramas et al. (2015) use such data about music resources to extract semantic information which is further used in creating an ontology and/or generating music recommendations.

Wang et al. (2010) proposes an approach that combines a music taxonomy, such as the one taken from AllMusic[55], and WordNet[56] to predict mood of a song based on its social tags. First the terms in the taxonomy are transformed to an ontology, and each class is mapped to a concept level in the WordNet. Then the terms in the ontology are expanded using hyponyms, synonyms and sibling words obtained from the wordnet. Following this the tags are matched with the ontology with expanded terms. A seed set of songs are tagged with mood tags based on a content-based mood tagging system. A set of DL-safe rules further allow in inferring mood of target songs from the set of labeled seed songs. The results are shown to be more accurate than a few alternate SVM-based methods.

Sordo et al. (2012) discuss a method to automatically extract meaningful information from the analysis of natural language text in the online music discussion

[54] Some of them cannot be traced as of writing this thesis and have been left out in this summary

[55] http://www.allmusic.com

[56] WordNet is a lexical database of English, where words are grouped in cognitive synonyms, each of which is a distinct concept. It can be accessed at https://wordnet.princeton.edu

Ontology	Reference	Axioms	Classes	Properties	Expr.	Description
MPEG-7 Ontology	Garcia and Celma (2005)	-	-	-	-	OWL port of MPEG-7 metadata schema
MXOnto-Context	Ferrara et al. (2006)	608	51	29	ALEON(D)	Contextualizes a music resource abstracting information related to ensemble, melody, rhythm and harmony using symbolic score analysis
MXOnto-Genre	Ferrara et al. (2006)	197	72	4	ALN(D)	Defines a class hierarchy used for genre classification, in conjunction with MXOnto-Context and SWRL rules
Timeline ontology	Raimond (2008)	743	52	135	SHOIN(D)	Extends owl-time ontology[a] to support multiple timelines and allow interlinking them
Event ontology	Raimond (2008)	285	19	63	SHOIN(D)	Allows representation of musical events ranging from chorus in an audio file to performance in a city
Music Ontology	Raimond (2008)	3577	127	337	SHOIN(D)	Subsumes Timeline, Event and other ontologies to define vocabulary for describing a variety of music resources such as artists, albums and tracks

Table 4.1: A summary of ontologies proposed so far in MIR domain.

Ontology	Reference	Axioms	Classes	Properties	Expr.	Description
Music Similarity ontology	Jacobson (2011)	139	7	16	ALHN(D)	Allows expressing associations between music resources, similarity being one such association
Chord Ontology	Harte et al. (2005)	490	9	10	ALU(D)	Context-free representation of chords
Tonality Ontology	-	2239	23	31	ALOIF(D)	Describes tonal content information of musical works
Tempera-ment Ontology	Raimond (2008)	258	19	11	ALOF(D)	Describes tuning of an instrument
Ordered List Ontology	-	204	3	6	ALIN(D)	Provides vocabulary for representing sequential information as a semantic graph
Counter Ontology		477	21	69	SHOIN(D)	Provides vocabulary for describing a general counter concept
Playback Ontology	-	-	-	-	-	Subsumes Ordered List and Counter ontologies to define vocabulary for representing aspects of playback domain such as playlist and play or skip count.

Table 4.2: Continuation to table. 4.1

forums such as rasikas.org. First a dictionary of common music terms in the respective music tradition is defined. This dictionary is used in creating a complex network by matching the terms therein against the forum posts. The resultant network is studied by observing various characteristics like node relevance, co-occurrences and term relations via few semantically connecting words. Examples of potent relations exposed using this approach include the lineage information of an artist, musical influence and similarity.

Oramas et al. (2015) proposes an approach similar to the one proposed by Wang et al. (2010) in leveraging knowledge graphs in a hybrid recommendation engine. In this approach, the tags and textual descriptions are analyzed for concept identification which are then linked to semantically rich sources like WordNet and DBpedia[57]. A knowledge graph built using this data, in conjunction with data about users' listening habits is used in generating recommendations, which are shown to be significantly better than those obtained using state-of-the-art collaborative filtering approaches.

4.3 Linked open data in the domain of music

DBpedia since its beginnings in 2007 (Bizer et al. (2007)) have remained an important effort that bootstrapped activity in the semantic web by helping realizing the potential of interlinked information. It is a community effort in which the goal is to publish the structured data content in Wikipedia as an RDF dataset. This has become the nucleus to which nearly every open data project linked to resulting in a giant web of data (Bizer et al. (2009b)). As noted earlier, this can be observed in the snapshot of the web of data presented in fig. 4.2. DBpedia 2014 boasts 3 billion RDF triples in which 580 million are from English edition of Wikipedia while the 2.46 billion come from 124 other language editions. In all, the dataset has 38.3 million unique things, a majority of which also have multilingual presence. In this release, there were about 123,000 music albums. There are also other parallel efforts to DBpedia, such as Freebase[58] which serves 1.9 billion triples, YAGO which serves 120 million triples (Mahdisoltani et al. (2014)) and WikiData (Erxleben et al. (2014)).

In the domain of music, MO has been instrumental in integrating schemas and linking data across a large number of sources including Jamendo, Magnatune, BBC music, AudioScrobbler data, MySpace, MusicBrainz and Echonest (Kobi-

[57] http://wiki.dbpedia.org

[58] https://developers.google.com/freebase/, now discontinued and data migrated to Wikidata - https://www.wikidata.org/wiki/Wikidata:WikiProject_Freebase

Dataset	Triples in millions	Interlinked with
Magnatune	0.322	DBpedia
Jamendo	1.1	Geonames, MusicBrainz
BBC peel	0.277	DBpedia
Last.fm	600	MusicBrainz
MySpace	12000	-
MusicBrainz	60	DBpedia, MySpace, Lingvoj
Isophone	0.457	MusicBrainz

Table 4.3: Datasets published and interlinked as part of DBtune project as summarized in Fazekas et al. (2010)

larov et al. (2009); Raimond et al. (2009)). DBtune project[59] is part of Linking Open Data community project[60], which undertook the aforementioned integration. Table. 4.3 gives an overview of the status of this project and the related statistics. Cumulatively over all the sources of data, DBtune now provides access to around 14 billion triples. Another set of tools such as Sonic Visualizer, Sonic Annotator and the VAMP audio analysis plugin API use MO in publishing the feature information as RDF (Cannam et al. (2010b)).

These efforts in linking data have lead to a valuable KB of huge proportions, even if it is still at a nascent stage especially considering that a majority of the web data still needs to be structured and linked using semantic web standards. This KB has been exploited by research and development community in building new or better information retrieval algorithms and applications. Freebase had been central to the knowledge graph that Google uses to improve and personalize the search results. DBpedia, WordNet and other rich linked data sources have been used for improving the performance of Natural Language Processing and information retrieval algorithms by facilitating entity linking and relation extraction (Mendes and Jakob (2011)). There are also more direct usages of linked data such as those which Passant (2010a,b) suggest. In this work, they proposed a set of distance measures over linked data capitalizing on i) direct links between two given resources and ii) shared links between such resources via other resources. These set of distance measures are used in recommending music and literature resources and the results are found to be comparable to those of collaborative filtering approaches used on commercial platforms such as last.fm.

[59] http://dbtune.org

[60] https://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

However, there are certain limitations to publishing structured data, interlinking and using it. To begin with, ontologies in music domain are not complete enough and need to be updated frequently to include more vocabulary or modify the existing classes and properties. This may result in a conflict with data sources which already have published their data using the previous version of the ontology. At the web scale, this problem is not easily tracked to be able to address it. Though there is a limited support to versioning ontologies in the form of annotations, this does not strictly result in a compliance that is needed to alleviate this problem altogether. Another important limitation is the lack of infrastructure, i.e., proliferation of standards and software tools, that allows a reliable, seamless access and provides querying of linked data across multiple sources. The federated query capabilities of SPARQL 1.1 specification[61] is a step in a direction to address this limitation. With increasing reliability of linked data endpoints, we believe this issue can be overcome in near future.

Cannam et al. (2010b) summarizes other issues that were faced in publishing audio features using RDF. As RDF is text-heavy, it results in wastage of space, processing time and network resources which impedes the consumption of this data. Also, there is no efficient means of representing numerical data such as feature vectors or matrices. Further, inconsistencies in encoding of even simple numerical data pose a problem to querying the data, especially more so when using a federated setup. This problem gets worse with more sophisticated data types. And lastly, as RDF data does not enforce order of statements, query optimization for tasks such as distance computation is not straight-forward.

4.4 Summary and conclusions

In this chapter, we have discussed how the web evolved through various stages leading up to the semantic web. We have briefly introduced the technologies and standards that are part of it, emphasizing ontologies and linked data sources that played an important role in bootstrapping its take off. We have also stressed on the need for knowledge representation in the domain of music, discussing the related past work in this domain. The work so far has been instrumental in developing ontologies for varied aspects of music relevant for both structuring/integration of data sources, and computational tasks in music information research such as recommendation and similarity. There is also considerable interest in the research community in leveraging folksonomies and natural language data in or-

[61] https://www.w3.org/TR/sparql11-federated-query/

der to build bottom-up approaches to automatically extract semantic information which can further lead to partial ontologies.

It is clear that in the domain of MIR, we are just beginning to understand the potential of knowledge representation and linking data sources, in the context of semantic web. While the existing ontologies are a right step in getting started, to our knowledge, work on developing ontologies for deeper musicological information is yet to be realized. Further, we need to take into account that the current ontologies which model the music concepts such as scales or chords, do so in the context of western popular music. This does not mean that an ontology needs to be all engulfing to include musical aspects of different cultures (Read Veltman, 2004, for knowing the challenges in modeling cultural differences). But, we need music related knowledge in those cultures to be represented using semantic web technologies, especially for the reasons we have argued in sec. 4.2. This becomes even more important considering the recent surge of interest in the MIR community for developing culture-aware approaches (Serra (2011); Serra et al. (2013)).

Part II

Audio music analysis for intonation description

Chapter 5

Parametrizing pitch histograms

Intonation is a fundamental music concept that has a special relevance in Indian art music. It is characteristic of a rāga and key to the musical expression of the artist. Computational description of intonation is of importance to several music information retrieval tasks such as developing similarity measures based on rāgas and artists. For computational purposes, we define intonation as characteristics of pitches and pitch modulations used by an artist in a given musical piece. From this definition, our approach will consider a performance of a piece as our unit of study. In Carnatic music practice, it is known that the intonation of a given svara varies significantly depending on the style of singing and the rāga (Swathi, 2009; Levy, 1982). Our work is only concerned with intonation as a property of raaga, and therefore differences that arise from various styles of singing are beyond the scope of this work.

The study of svara intonation differs from that of tuning in its fundamental emphasis. The later refers to the discrete frequencies with which an instrument is tuned, thus it is more of a theoretical concept than intonation, in which we focus on the pitches used during a performance. The two concepts are basically the same when we study instruments that can only produce a fixed set of discrete frequencies, like the piano. On the other hand, given that in Indian art music even when the instruments are tuned using a certain set of intervals, the performances feature a richer gamut of frequencies than just the steady notes (the harmonium in Hindustani music is an important exception). Therefore, an understanding of tuning alone cannot explain the real performances. Therefore, the study of intonation as we just defined it assumes a greater importance. We maintain the distinction between these terms, tuning or intonation, as necessary through out the rest of the thesis.

In this part of the thesis, we document our efforts chronologically in understanding this task and developing automated approaches that help us get closer to a meaningful description from audio music signals. Pitch histograms have been used with fair degree of success in classifying raagas in Indian art music (Chordia et al. (2013); Chordia and Rae (2007)) and makams in Makam music of Turkey (Bozkurt (2011)). The conclusions drawn in our evaluation of pitch-histogram based approaches (3) point us to the fact that despite the importance of gamaka-filled svara movements, they do not seem to add more information that can help discriminate raagas.

In our first approach which we discuss in this chapter, we further verify this conclusion. Our hypothesis in doing so is as follows: If each peak in the histogram accounts to a svara sung in the performance, then the movements involving the svara influence the properties of pitch distribution around it. For instance, if a given svara is always sung taking off from the previous semitonal position, then we assume that the pitch distribution of the svara ought to have certain measurable predisposition towards the take off point. In this chapter, we propose parametrization of svara peaks in the pitch histograms to partly capture this information which will help us verify the aforementioned conclusion. This can also potentially help us understand more about intonation of svaras.

5.1 Overview of the approach

From the observations made by Krishnaswamy (2003) and Subramanian (2007), it is apparent that steady svaras only tell us part of the story that goes with a given Carnatic music performance. However, the gamaka-embellished svaras pose a difficult challenge for automatic svara transcription. Therefore, alternative means of deriving meaningful information about the intonation of svaras becomes important. Gedik and Bozkurt (2010) present a survey of histogram analysis in music information retrieval tasks, and also emphasize the usefulness of histogram analysis for tuning assessment and makam recognition in Makam music of Turkey. As we already noted earlier, the gamakas and the role of a svara are prone to influence the aggregate distribution of a svara in the pitch histogram of the given recording. We believe that this information can be derived by parametrizing the distribution around each svara (cf. Belle et al., 2009).

Our intonation description approach based on histogram peak parametrization involves five steps. In the first step, prominent vocal segments of each performance are extracted (sec. 5.2). In the second step, the pitch corresponding to the voice is extracted using multipitch analysis (sec. 5.3). In the third step, a pitch

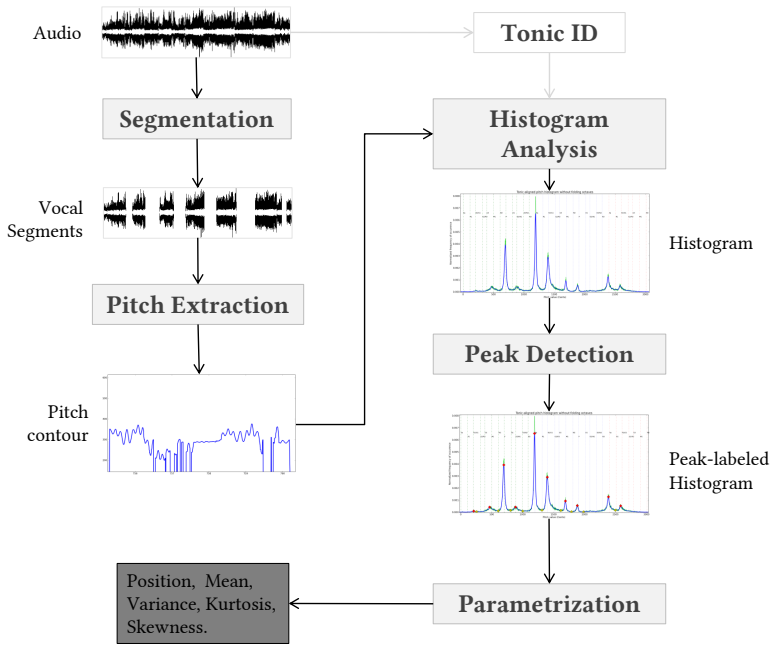


Figure 5.1: Block diagram showing the steps involved in Histogram peak parametrization method for intonation analysis.

histogram for every performance is computed (sec. 5.4) and in the final step, its prominent peaks are detected and each peak's distribution is characterized by using the valley points and an empirical threshold. Then the parameters that characterize each of the distributions are extracted (sec. 5.5). Figure 5.1 shows the steps in a block diagram.

5.2 Segmentation of the audio music recording

In a typical Carnatic ensemble, there is a lead vocalist who is accompanied by a violin, drone instrument(s), and percussion instruments with tonal characteristics (Raman, 1934). Based on the instruments being played, a given performance is usually a mix of one or more of these: vocal, violin and percussion. The drone instrument(s) is heard throughout the performance. The order and interspersion of these combinations depend on the musical forms and their organization in the performance. For different melodic and rhythmic analysis tasks, it is required to distinguish between these different types of segments. Therefore, it is necessary to have a segmentation procedure which can automatically do this.

In this study, we do not address the intonation variations due to artists. However, as we consider each recording as a unit for describing intonation, there is a need to assert the artist and the rāga which characterize the intonation of the recording. For this reason, we have considered those recordings in which only one rāga is sung, which is the case for most of the recordings. Furthermore, we also distinguish between the segments where the lead artist exerts a dominant influence and the segments in which the accompanying violin is dominant. We choose the pitch values only of the former segments. In order to do this, we consider three broad classes to which the aforementioned segments belong to: vocal (all those where the vocalist is heard, irrespective of the audibility of other instruments), violin (only those where the vocalist is not heard and the violinist is heard) and percussion solo.

To train our segmentation algorithm to classify an audio excerpt into the three classes, we manually cropped 100 minutes of audio data for each class from commercially available recordings[1], taking care as to ensure diversity: different artists, male and female lead vocalists, clean, clipped and noisy data, and different recording environments (live/studio). The audio data is split into one-second fragments. There are few fragments which do not strictly fall into one of the three classes: fragments with just the drone sound, silence, etc. However, as they do not affect the intonation analysis as such, we did not consciously avoid them.

After manual segmentation we extract music descriptors. Mel-frequency cepstral coefficients (MFCCs) have long been used with a fair amount of success as timbral features in music classification tasks such as genre or instrument classification (Tzanetakis and Cook, 2002). Jiang et al. (2002) proposed octave based spectral contrast feature (OBSC) for music classification which is demonstrated to perform better than MFCCs in a few experiments with western popular music. Shape based spectral contrast descriptor (SBSC) proposed by Akkermans et al. (2009) is a modification of OBSC to improve accuracy and robustness by employing a different sub-band division scheme and an improved notion of contrast. We use both MFCC and SBSC descriptors, along with a few other spectral features that reflect timbral characteristics of an audio excerpt: harmonic spectral centroid, harmonic spectral deviation, harmonic spectral spread, pitch confidence, tristimulus, spectral rolloff, spectral strongpeak, spectral flux and spectral flatness (Tzanetakis and Cook, 2002).

A given audio excerpt is first split into fragments of length 1 second each. The sampling rate of all the audio recordings is 44100 Hz. Features are extracted

[1]These recordings are also derived from CompMusic collection, some of which also are part of the sub-collection we chose for evaluation.

for each fragment using a framesize of 2048 and a hopsize of 1024 (double sided Hann window is used). The mean, covariance, kurtosis and skewness are computed over each 1-second fragment and stored as features. MFCC coefficients, 13 in number, are computed with a filterbank of 40 mel-spaced bands from 40 to 11000Hz (Slaney, 1998). The DC component is discarded, yielding a total of 12 coefficients. SBSC coefficients and magnitudes, 12 each in number, are computed with 6 sub-bands from 40 to 11000Hz. The boundaries of sub-bands used are 20 Hz, 324 Hz, 671 Hz, 1128 Hz, 1855 Hz, 3253 Hz and 11 kHz (see Akkermans et al., 2009). Harmonic spectral centroid (HSC), harmonic spectral spread (HSS) and harmonic spectral deviation (HSD) of the i^{th} frame are computed as described by Kim et al. (2006):

$$HSC_i = \frac{\sum_{h=1}^{N_H}(f_{h,i}A_{h,i})}{\sum_{h=1}^{N_H}A_{h,i}} \tag{5.1}$$

$$HSS_i = \frac{1}{HSC_i}\sqrt{\frac{\sum_{h=1}^{N_H}[(f_{h,i}-HSC_i)^2A_{h,i}^2]}{\sum_{h=1}^{N_H}A_{h,i}^2}} \tag{5.2}$$

$$HSD_i = \frac{\sum_{h=1}^{N_H}\left|\log_{10}A_{h,i}-\log_{10}SE_{h,i}\right|}{\sum_{h=1}^{N_H}\log_{10}A_{h,i}} \tag{5.3}$$

where $f_{h,i}$ and $A_{h,i}$ are the frequency and amplitude of h^{th} harmonic peak in the FFT of the i^{th} frame, and N_H is the number of harmonics taken into account, ordering them by frequency. For our purpose, the maximum number of harmonic peaks chosen was 50. $SE_{h,i}$ is the spectral envelope given by:

$$SE_{h,i} = \begin{cases} \frac{1}{2}(A_{h,i}+A_{h+1,i}) & \text{if } h=1 \\ \frac{1}{3}(A_{h+1,i}+A_{h,i}+A_{h-1,i}) & \text{if } 2 \le h \le N_H-1 \\ \frac{1}{2}(A_{h-1,i}+A_{h,i}) & \text{if } h=N_H \end{cases}$$

All the features thus obtained are normalized to the 0-1 interval. In order to observe how well each of these different descriptors perform in distinguishing the aforementioned three classes of audio segments, classification experiments are conducted with each of the four groups of features: MFCCs, SBSCs, harmonic spectral features and *other* spectral features. Furthermore, different classifiers are employed: naive Bayes, k-nearest neighbors, support vector machines, multilayer perceptron, logistic regression and random forest (Hall et al., 2009). As the

smallest group has 12 features, the number of features in other groups is also limited to 12 using information gain feature selection algorithm (Hall et al., 2009). The classifiers are evaluated in a 3-fold cross validation setting in 10 runs. All three classes are balanced. Table 5.1 shows the average accuracies obtained.

MFCCs performed better than the other features, with the best result obtained using a k-NN classifier with 5 neighbors. The *other spectral features* and SBSCs also performed considerably well. Using paired t-test with a p-value of 0.05, none of the results obtained using harmonic spectral features were found to be statistically significant with respect to the baseline at 33% using zeroR classifier.

From among all features, we have selected 40 features through a combination of hand-picking and information-gain feature selection algorithm. These features come from 9 descriptors: MFCCs, SBSCs, harmonic spectral centroid, harmonic spectral spread, pitch confidence, spectral flatness, spectral rms, spectral strong-peak and tristimulus. The majority of these features are means and covariances of the nine descriptors. Table 5.1 shows results of classification experiments using all the features. In turn, k-NN classifier with 5 neighbors, performed significantly better than all the other classifiers.

5.3 Predominant melody extraction

With the segmentation module in place, we minimize to a large extent the interference from accompanying instruments. However, there is a significant number of the obtained voice segments in which the violinist fills short pauses or in which the violin is present in the background, mimicking the vocalist very closely with a small time lag. This is one of the main problems we encountered when using pitch tracking algorithms like YIN (de Cheveigné et al., 2002), since the violin was also being tracked in quite a number of portions. To address this, we obtain the predominant melody using a multi-pitch analysis approach proposed by Salamon et al. (2012). In this, multiple pitch contours are obtained from the audio, which are further grouped based on auditory cues like pitch continuity and harmonicity. The contours which belong to the main melody are selected using heuristics obtained by studying features of melodic and non-melodic contours.

The frequencies are converted to cents and normalized with the tonic frequency obtained using the approach proposed by Gulati (2012). In Carnatic music, the lead artist chooses the tonic to be a frequency value which allows her/him to explore close to three octaves. The range of values chosen for tonic by the artist usually is confined to a narrow range and does not vary a lot. Hence, we take advantage of this fact to minimize the error in tonic estimation to a large extent,

	k-NN	Naive Bayes	Multilayer perceptron	Random Forest	SVM	Logistic regression
MFCCs	91.51	72.22	81.44	90.44	83.56	73.59
SBSCs	88.41	72.64	79.64	87.93	79.71	73.58
Harmonic spectral features	66.75	60.93	69.56	74.19	69.68	67.30
Other spectral features	87.45	70.22	84.56	89.15	84.0	80.79
All combined (40 features picked using feature-selection)	93.88	74.44	91.85	92.44	91.58	85.26
All combined (40 hand-picked features)	96.94	83.30	95.42	96.08	95.90	89. 42

Table 5.1: Accuracies obtained in classification experiments conducted with features obtained from four groups of descriptors using different classifiers.

using a simple voting procedure. A histogram of the tonic values is obtained for each artist and the value which is nearest to the peak is obtained. This is considered to be the *correct* tonic value for the artist. The tonic values which are farther than 350 cents from this value are then set to the *correct* tonic value thus obtained. After these preprocessing steps, we go ahead to obtain the intonation description.

5.4 Histogram computation

As Bozkurt et al. (2009) point out, there is a trade-off in choosing the bin resolution of a pitch histogram. A high bin resolution keeps the precision high, but significantly affects the peak detection accuracy. However, unlike Turkish makam music, where the octave is divided into 53 Holdrian commas, Carnatic music uses roughly 12 svarastānas (Shankar, 1983). Hence, in this context, choosing a finer bin width is not as much a problem as it is in Turkish makam music. In addition, we employ a Gaussian kernel with a large standard deviation to smooth the histogram before peak detection. However, in order to retain the preciseness in estimating the parameters for each peak, we consider the values from the distribution of the peak before smoothing, which has the bin resolution as one cent. We compute the histogram H by placing the pitch values into their corresponding bins:

$$H_k = \sum_{n=1}^{N} q_k, \tag{5.4}$$

where H_k is the k-th bin count, N is the number of pitch values, $q_k = 1$ if $c_k \leq P(n) \leq c_{k+1}$ and $q_k = 0$ otherwise, P is the array of pitch values and (c_k, c_{k+1}) are the bounds on k-th bin.

5.5 Svara peak parametrization

Peak detection

Traditional peak detection algorithms can broadly be said to follow one of the three following approaches (Palshikar, 2009): (a) those which try to fit a known function to the data points, (b) those which match a known peak shape to the data points, and (c) those which find all local maximas and filter them. We choose to use the third approach owing to its simplicity.

The important step in such an approach is filtering the local maximas to retain the peaks we are interested in. Usually, they are processed using an amplitude

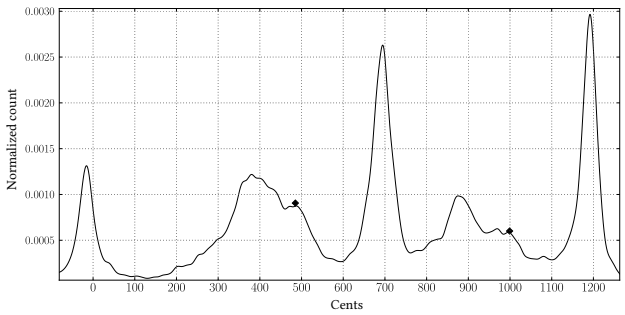


Figure 5.2: A sample histogram showing the peaks which are difficult to be identified using traditional peak detection algorithms. X-axis represents cent scale.

threshold (Palshikar, 2009). However, following this approach, the peaks such as the ones marked in Figure 5.2 are not likely to be identified, unless we let the algorithm pick up a few spurious peaks. The cost of both spurious and undetected peaks in tasks such as intonation analysis is very high as it directly corresponds to the presence/absence of svaras.

To alleviate this issue, we propose two approaches to peak detection in pitch histograms which make use of few constraints to minimize this cost: peak amplitude threshold (A_T), valley[2] depth threshold (D_T) and intervallic constraint (I_C). Every peak should have a minimal amplitude of A_T, with a valley deeper than D_T on at least one side of it. Furthermore, only one peak is labelled per musical interval given by a predetermined window (I_C).

The first one of the peak detection approaches is based on the slope of the smoothed histogram. A given histogram is convolved with a Gaussian kernel to smooth out jitter. The length and standard deviation of the Gaussian kernel are set to 44 and 11 bins respectively. The length of the histogram is 3600 (corresponding to 3 octaves with 1 cent resolution). The local maximas and minimas are identified using slope information. The peaks are then found using D_T, and with an empirically set intervallic constraint, I_C. A local maxima is labelled as a peak only if it has valleys deeper than D_T on both sides, and it is also the maxima at least in the interval as defined by I_C.

The second one is an interval based approach, where the maximum value for every musical interval (I_C) is marked as a peak. The interval refers to one of the just-intonation or the equal temperament intervals. In the case of a just-intonation interval, the window size is determined as the range between the mean

[2]Valley is to be understood as the deepest point between two peaks.

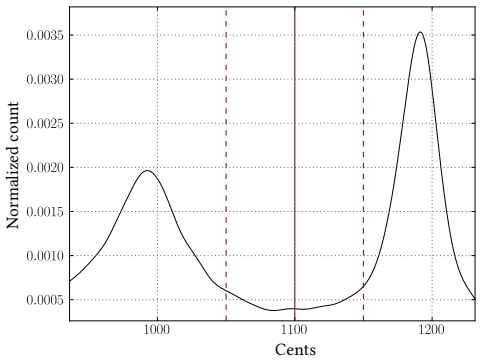


Figure 5.3: A semi-tone corresponding to 1100 cents is shown, which in reality does not have a peak. Yet the algorithm takes the point on either of the tails of the neighbouring peaks (at 1000 and 1200 cents) as the maxima, giving a false peak.

values obtained with the preceding and succeeding intervals. In the case of an equi-tempered interval, it is constant for all the intervals, which is input as a parameter. The window is positioned with the current interval as its center. The peaks thus obtained are then subject to A_T and D_T constraints. In this approach, it is sufficient that a valley on either side of the peak is deeper than D_T.

Among all the points labelled as peaks, only a few correspond to the desired ones. Figure 5.3 shows three equi-tempered semi-tones at 1000, 1100 and 1200 cents. There are peaks only at 1000 and 1200 cents. However, as the algorithm picks the maximum value in a given window surrounding a semi-tone (window size is 100 cents in this case), it ends up picking a point on one of the tails of the neighbouring peaks. Therefore, we need a post-processing step to check if each peak is a genuine local maxima. This is done as follows: the window is split at the labelled peak position, and the number of points in the window that lie to both sides of it are noted. If the ratio between them is smaller than 0.15, there is a high chance that the peak lies on the tail of the window corresponding to a neighbouring interval[3]. Such peaks are discarded.

In order to evaluate the performance of each of these approaches, we have manually annotated 432 peaks in 32 histograms with pitch range limited from -1200 cents to 2400 cents. These histograms correspond to the audio recordings sampled from the dataset reported in table. 1.4. As there are only a few parameters, we performed a limited grid search to locate the best combination of parameters

[3]This value is empirically chosen.

for each approach using the given ground-truth. This is done using four different methods: one method from slope based approach (M_S), two methods from interval based approach corresponding to just-intonation (M_{JI}) and equi-tempered intervals (M_{ET}), and a hybrid approach (M_H) where the results of M_S and M_{JI} are combined. The intention of including M_H is to assess whether the two different approaches complement each other. The reason for selecting M_{JI} in the hybrid approach is explained later in this section.

Table A.1 shows the ranges over which each parameter is varied when performing the grid search. For the search to be computationally feasible, the range of values for each parameter are limited based on the domain knowledge of the intervals and their locations, and empirical observations Shankar (1983); Serra (2011). A maximum F-measure value of 0.96 is obtained using M_H with A_T, D_T and I_C set to $5.0 \cdot 10^{-5}$, $3.0 \cdot 10^{-5}$ and 100 respectively. In order to further understand the effect of each parameter on peak detection, we vary one parameter at a time keeping the values for the other parameters as obtained in the optimum case. Figure A.1 in Appendix. A.1 shows the impact of varying different parameters on different methods.

The kernel size for Gaussian filter was also evaluated, giving optimal results when set to 11. Higher and lower values are observed to have poor impact on the results. In the case of the window size, the larger it is, the better has been the performance of M_H and M_S. We suppose it is because the large window sizes handle deviations from the theoretical intervals with more success. Unlike equi-tempered intervals, just-intonation intervals are heterogeneous. Hence, a constant window has not been used. In M_{ET}, there does not seem to be a meaningful pattern in the impact produced by varying the window size. From Figure A.1, we observe that D_T and A_T produce an optimum result when they are set to $5.0 \cdot 10^{-5}$, $3.0 \cdot 10^{-5}$ respectively. Further increasing their values results in the exclusion of many valid peaks.

As Serra (2011) have shown, Carnatic music intervals align more with just-intonation intervals than the equi-tempered ones. Therefore, it is expected that the system achieves higher accuracies when intervals and I_C are decided using just-intonation tuning. This is evident from the results in Figure A.1. This is also the reason why we chose M_{JI} over M_{ET} to be part of M_H. Serra (2011) also show that there are certain intervals which are far from the corresponding just-intonation intervals. As slope-based approach does not assume any tuning method to locate the peaks, in the cases where the peak deviates from theoretical intervals (just intonation or equi-tempered), it performs better than interval-based approach. In the interval based approach, the peak positions are presumed to

be around predetermined intervals. As a result, if a peak is off the given interval, it will be split between two windows with the maximums located at extreme position in each of them, and hence are discarded in the post-processing step described earlier. This is unlike the slope based approach, where the local maximas are first located using slope information, and I_C is applied later. The results from Figure A.1 emphasize the advantage of a slope-based approach over an interval-based approach.

On the other hand, the interval based approach performs better when the peak has a deep valley only on one side of the peak. As a result, methods from the two approaches complement each other. Hence, M_H performs better than any other method. Therefore we choose this approach to locate peaks from pitch histograms. Most peaks are detected by both M_{JI}/M_{ET} and M_S. For such peaks, we preferred to keep the peak locations obtained using M_S. The source code corresponding to the three peak detection approaches is made openly available online[4].

Parametrization

In order to parametrize a given peak in the performance, it needs to be a bounded distribution. We observe that usually two adjacent peaks are at least 80 cents apart. The valley point between the peaks becomes a reasonable bound if the next peak is close by. But in cases where they are not, we have used a 50 cent bound to limit the distribution. The peak is then characterized by six parameters: peak location, amplitude, mean, variance, skewness and kurtosis. We extract parameters for peaks in three octaves. Each peak corresponds to a svarastāna. For those svarastānas which do not have a corresponding peak in the pitch histogram of the recording, we set the parameters to zero. Since for each octave there are 12 svarastānas, the total number of features of a given recording is 216 (3 octaves $\times$ 12 svarastānas $\times$ 6 parameters).

5.6 Evaluation & results

Intonation is a fundamental characteristic of rāga. Therefore, automatic rāga classification is a plausible way to evaluate computational descriptions of intonation. The two parameters from histogram analysis that have been used for rāga classification task in the literature are position and amplitude of the peaks (3). We devise an evaluation strategy that tests whether the new parameters we propose

[4] https://github.com/gopalkoduri/pypeaks

are useful, and also if they are complementary and/or preferred to the ones used in the literature.

The evaluation strategy consists of two tasks: feature selection and classification. The feature selection task verifies if the new parameters are preferred to the features from position and amplitude parameters. In this task, we pool in the features from all the parameters and let the information gain measure and support vector machine feature selection algorithms pick the top n features among them (Hall et al., 2009; Witten and Frank, 2005). We then analyze how often features from each pool of the parameters get picked.

The rāga classification task allows us to check if the features from the new parameters bring in complementary information compared to the features from position and amplitude. For this, we divide this task into two subtasks: classification with features obtained from the position and amplitude parameters, and classification with features obtained from all the parameters (position, amplitude and new parameters: mean, variance, skewness and kurtosis). We compare the results of the two subtasks to check if the features from the new parameters we propose carry complementary information to distinguish rāgas.

To ensure that the comparison of results in the two subtasks is fair, we use top n features in each subtask picked by information gain algorithm in feature selection task. Furthermore, six different classifiers were used: naive Bayes, k-nearest neighbours, support vector machines, logistic regression, multilayer perceptron and random forest (Hall et al., 2009; Witten and Frank, 2005)[5], and the accuracies obtained for each of them are checked if they stabilize after a few runs of the experiment.

The number of raagas/classes is large in our dataset (table. 1.4), and each svara in a given raaga can be equally important in its identity. Feature selection over all the classes together will not result in meaningful behavior due to competitive tradeoff that would ensue. To address this issue, we perform numerous classification experiments each of which has 3 classes. As ${}^{45}C_3$ is a huge number, for the sake of computational feasibility, we listed all the possible combinations and picked 800 of them in a random manner. Each such combination is further subsampled thrice so that all the classes represented in that combination have equal number of instances, which is 5 as it is the minimum number of instances in a class in our music collection. As the total number of instances in each case is 15, we limit the number of features picked by the feature selection algorithms to 5.

[5]The implementations provided in Weka were used with default parameters.

	Position		Amplitude		Mean		Variance		Skewness		Kurtosis	
	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.
Information gain	0.9	0.7	1.4	0.8	1.2	0.8	0.4	0.4	0.8	0.6	0.4	0.4
SVM	1.7	0.9	0.9	0.6	1.1	0.7	0.5	0.4	0.4	0.3	0.5	0.4

Table 5.2: Results of feature selection on three-class combinations of all the rāgas in our music collection, using information gain and support vector machines. Ratio of total number of occurrences (abbreviated as Occ.) and ratio of number of recordings in which features from a given parameter are chosen at least once (abbreviated as Rec.), to the total number of runs are shown for each parameter. Note that there can be as many features from a parameter as there are number of svaras for a given recording. Hence, the maximum value of Occ. ratio is 5 (corresponding to 5 features selected per recording), while that of Rec. ratio is 1.

Features/Classifier	Naive Bayes	3-Nearest Neighbours	SVM	Random forest	Logistic regression	Multilayer Perceptron
Position and Amplitude	79.13	78.52	68.91	81.26	78.65	78.75
All features	78.26	78.46	71.79	81.16	78.61	78.78

Table 5.3: Averages of accuracies obtained using different classifiers in the two rāga classification experiments, using all the rāgas. The baseline calculated using zeroR classifier lies at 0.33 in both experiments.

Table. 5.2 shows the statistics of outcomes of the two feature selection algorithms. For each parameter, two ratios are shown. The first one, abbreviated as Occ., is the ratio of total number of occurrences of the parameter to the total number of runs. The second one, abbreviated as Rec., is the ratio of number of recordings in which the parameter is chosen at least once, to the total number of runs. The former lets us know the overall relevance of the parameter, while the latter allows to know the percentage of recordings to which the relevance scales to. Clearly, the position and amplitude of a peak are the best discriminators of rāgas given the high values for both ratios. It is also an expected result given the success of histograms in rāga classification (Koduri et al., 2012). The mean of the peak is also equally preferred to the position and amplitude, by both the feature selection algorithms.

Mean, variance, skewness and kurtosis are chosen in nearly 40-50% of the runs. Recall that each recording has 216 features, with 36 features from each of the parameters. Therefore, in 40-50% of the runs, features from the new parameters (mean, variance skewness and kurtosis) are preferred despite the availability of features from position and amplitude. This shows that the new parameters carry important information for distinguishing rāgas, than the positions and amplitudes for few svaras.

The results from the rāga classification task help us to assess the complementariness of the features from new parameters. Table 5.3 shows the averages of all the results obtained using each classifier over all the sub-sampled combinations for the two subtasks (classification of rāgas using features from all parameters, and those of position and amplitude). There is only a marginal difference in the results of the two subtasks, with a noticeable exception in the case of results obtained using SVM which seems to indicate that the features from new parameters made a difference.

There is a class of rāgas which share exactly the same set of svaras, but have different characteristics, called allied rāgas. These rāgas are of special interest as there is a chance for more ambiguity in the positions of svaras. This prompted us to report separately the results of the feature selection and rāga classification tasks described earlier, on 11 sets of allied rāgas which together have 332 recordings in 32 rāgas. For those allied rāga sets which have more than two rāgas per set (say n), we do the experiments for all ${}^{n}C_{2}$ combinations of the set.

Table 5.4 shows the statistics over the outcomes of feature selection algorithms. One noteworthy observation is that the relevance of variance and kurtosis parameters is more pronounced in the classification of the allied rāgas, compared to the classification of all the rāgas (ref. table 5.2). This is in line with our hypothesis

	Position		Amplitude		Mean		Variance		Skewness		Kurtosis	
	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.
Information gain	0.9	0.7	1.3	0.8	0.8	0.7	0.6	0.5	0.7	0.6	0.7	0.5
SVM	1.2	0.8	1.0	0.7	1.0	0.8	0.7	0.5	0.4	0.3	0.7	0.6

Table 5.4: Results of feature selection on sub-sampled sets of recordings in ${}^{n}C_{2}$ combinations of allied rāgas using information gain and support vector machines. Ratio of total number of occurrences (abbreviated as Occ.) and ratio of number of recordings in which the parameter is chosen at least once (abbreviated as Rec.), to the total number of runs are shown for each parameter.

Features/Classifier	Naive Bayes	3-Nearest Neighbours	SVM	Random forest	Logistic regression	Multilayer Perceptron
Position and Amplitude	86.94	88.84	86.87	85.84	82.70	86.37
All features	87.66	89.28	87.67	85.93	83.69	87.75

Table 5.5: Accuracies obtained using different classifiers in the two rāga classification experiments, using just the allied rāga groups. The baseline calculated using zeroR classifier lies at 0.50 in both experiments.

owing to the special property of allied rāgas. Table 5.5 shows the classification results. However, the results from table 5.3 show only a marginal increase in the accuracies of classification using features from all the parameters, compared to the case of using features from just position and amplitude parameters. This indicates that though the new parameters are preferred to position and amplitude parameters, they do not bring in much complementary information.

5.7 Summary & conclusions

We started out with an objective to verify the conclusion from our comprehensive evaluation of histogram-based raaga classification approaches. It states that despite the importance of gamaka-filled svara movements in Carnatic music, they do not seem to add more information that can help discriminate raagas. We have proposed a histogram peak parametrization approach in order to verify this and evaluated it qualitatively using two tasks. The new parameters describing the pitch distribution around each svara peak were shown to be useful in discriminating rāgas. However, as observed in the general rāga classification task, the information contained in the new parameters obtained through this approach do not seem to out do or add to the information given by position and amplitude parameters. Therefore, we conclude that part of the earlier conclusion holds true, which is the distribution of pitches around a svara peak in the histogram do not add substantial information to peak position and amplitude parameters. However, the preference for the new parameters in feature selection algorithm does not completely rule out the potential of the information they hold.

The code resulting from these experiments is consolidated and is made available as two python modules - intonation and pypeaks. Instructions to install and use the former can be found at this link[6]. The source code for this module can be accessed at this link[7]. The pypeaks peak detection module can be installed from this link[8], and the source code can be accessed from a github repository[9].

[6] https://pypi.python.org/pypi/intonation
[7] https://github.com/gopalkoduri/intonation
[8] https://pypi.python.org/pypi/pypeaks
[9] https://github.com/gopalkoduri/pypeaks

Chapter 6

Context-based pitch distributions of svaras

There are certain limitations to the histogram peak parametrization (HPP) approach discussed in the last chapter. Few svaras, by the nature of the role they play, will not be manifested as peaks at all. Rather, they will appear as a gradual slide latched on to a neighboring peak that cannot be identified by a peak detection algorithm. Even more common than those are the melodic movements which often span a few semitones. They result in a cross distribution of pitch values among svara peaks. The HPP is an aggregate approach which does not account for the contextual information of pitches: the melodic & temporal contexts. The former refers to the larger melodic movement of which a given pitch instance is part of. The later refers to the properties of a given modulation: whether it is a fast intra-svara movement, a slower inter-svara movement, a striding glide that stretches from one svara to another, etc. In HPP, pitch value gets the same treatment irrespective of where it occurs in pitch contour.

We believe addressing this limitation can substantially improve the information held by the new parameters describing the distribution of pitch values around a svara, which we proposed as part of HPP. Consider the following two scenarios: (i) a given svara being sung steadily, and (ii) the same svara appearing as a transitory contour or even as a brief take off point. Using HPP, it is not possible to handle them differently. But in reality, the first occurrence should be part of the given svara's distribution, and the second occurrence should belong to a svara that gets the local emphasis. The objective of the approach we propose in this chapter is to handle such cases by incorporating the local melodic and temporal context of the given pitch value.

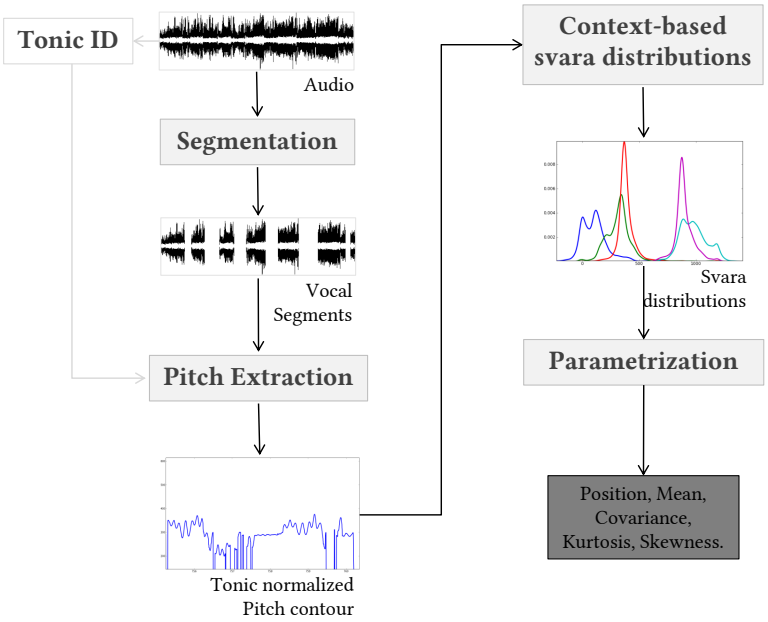


Figure 6.1: Block diagram showing the steps involved to derive context-based svara distributions.

6.1 Overview of the approach

Fig. 6.1 shows an overview of the steps involved in this approach in a block diagram. Notice that this method alleviates the need for peak detection and finding the distribution bounds as we obtain each svara distribution independently (compared with steps required for HPP shown in fig. 5.1). As we have already noted earlier, these two steps which are part of HPP have their own limitations. The peak detection algorithm is prone to pick erroneous peaks and/or leave out few relevant ones, especially so in an aggregate pitch histogram. On the other hand, in order to estimate the parameters it is necessary to determine the bandwidth of peaks from the histogram. In the cases where the valley points of a peak are not so evident and the peak distribution overlapped with that of a neighboring svara, we chose a hard bound of 50 cents on either side of the peak. This affects the parameters computed for the distribution. Therefore, it is indeed desirable to to avoid both those steps in this approach to avoid such issues altogether.

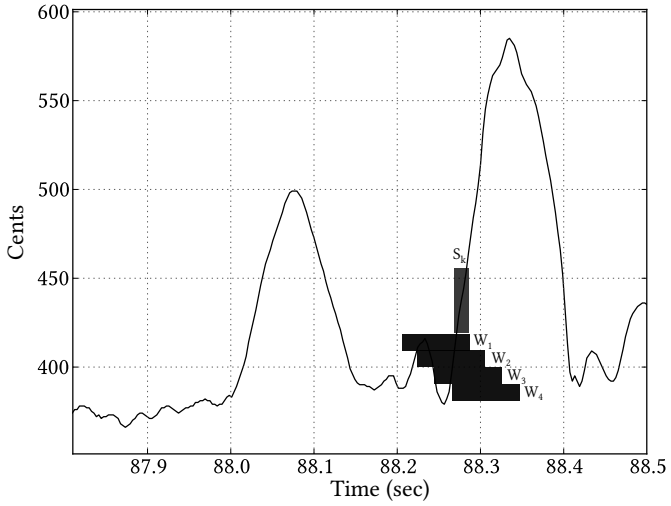


Figure 6.2: The positions of windows shown for a given segment S_k, which spans t_h milliseconds. In this case, width of the window (t_w) is four times as long as width of the segment (t_h), which is also hop size of the window. X-axis represents time and y-axis represents cent scale.

6.2 Isolating svara pitch distributions

Estimating melodic context using moving windows

In this approach, the pitches are distributed among the 12 svarastānas based on the context estimated from the pitch contour. The entire melody extracted from the song is viewed as a collection of small segments, each of a few milliseconds duration. For each segment, we consider the mean values of a few windows containing the segment. Those windows are positioned in time such that, in each subsequent hop, the segment moves from the end of the first window to the beginning of the last window. Each window is further reduced to a statistic, such as mean, median or mode. A series of such values provide us with useful contextual information. Figure 6.2 shows the positions of windows for a given segment S_k.

In the context of analysis and synthesis of gamakas in Carnatic music, Subramanian (2002, 2007) points out that the mean of a melodic movement often corresponds to the svara perceived by listeners. Therefore, we use the mean statistic as a way to compactly represent the pitch content of a given window. Further, given these series of mean values of windows containing the given segment, the next step is to make an intelligent guess of the svara being sung, to which the segment belongs to.

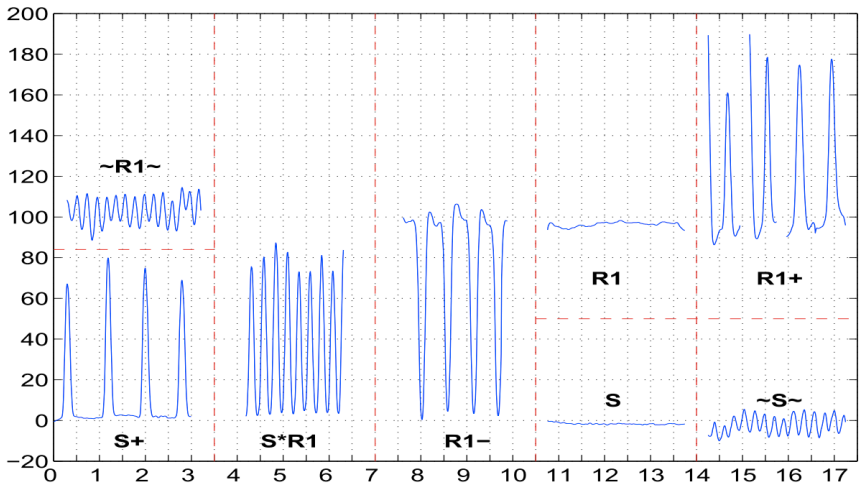


Figure 6.3: Different classes of melodic movements, reproduced as categorized by Krishnaswamy (2004).

There are two aspects that influence the process to decide the svara being sung. Remember that we noted earlier that it is not uncommon to find melodic movements which span atleast two semitones. Indeed, among the movements documented and categorized by Krishnaswamy (2004) and Subramanian (2007), they make up the majority. Therefore this is the first aspect we should account for, in deciding the svara. Fig. 6.3[1] shows a few classes of these movements in their synthetic forms. The real contours however would be a mix and continuum of these patterns.

The other important aspect is that inflexion points in the contour typically are the places which account for relatively higher duration of time, compared to transitory points. Given the possibilities of different types of contours around a svara, we assume that in general more time is spent closer to the svara being sung. However, one glance at the categories and examples shown in fig. 6.3 says that there are cases where this is not true. As this assumption allows to keep our methodology simple, we would like to verify if it can at least partially benefit our analysis.

The median statistic over the mean values of the windows running through the given segment is likely to address both these aspects if we quantize it to the nearest svara location. Choosing the median value downplays the influence of ex-

[1] This is reproduced from Krishnaswamy (2004) for convenience of the reader, as we often refer to these movements in the remainder of this chapter.

treme points in the contour especially in those where an inflexion point is far from the actual svara being sung (such kinds as the inflections at S end in R1+, R1- movements listed in fig. 6.3). This addresses the first aspect we discussed. Further, quantizing this value to the nearest svara location will help us in addressing the second aspect partially. Given that a movement can be a near symmetric oscillation around a svara location, or based off it where the oscillation extends to either the svara above or below it. A median value is a trade off between a mean value that would only address the former, and a mode that accounts for the later.

Formalizing our discussion, we define a shifting window with its size set to t_w milliseconds and hop size set to t_h milliseconds. For a k^{th} hop on pitch contour P, k=0,1,...$\frac{N}{t_h}$, where N is the total number of samples of the pitch contour, we define segment (S_k) as:

$$S_k = P(t_w + (k-1)t_h : t_w + kt_h) \tag{6.1}$$

where S_k is a subset of pitch values of P as given by Eq. 6.1. Notice that the width of the segment is t_h milliseconds. The mean of each window that contains the segment is computed as:

$$\mu_k = \frac{1}{t_w} \sum_{i=kt_h}^{t_w+kt_h} P(i) \tag{6.2}$$

The width of each window is t_w milliseconds. We now define ϵ, the total number of windows a given segment S_k can be part of, and $\bar{m}_k$, the median of the mean values of those ϵ windows as:

$$\epsilon = \frac{t_w}{t_h} \tag{6.3}$$

$$\bar{m}_k = \text{median}(\mu_k, \mu_{k+1}, \mu_{k+2} \ \ldots \ \mu_{k+\epsilon-1}) \tag{6.4}$$

Given Eqs. 6.1-6.4, a pitch-distribution $\mathbb{D}_I$ of a svara I is obtained as:

$$\mathbb{D}_I = \{S_k \ | \ \text{argmin}_i \ |\Gamma_i - \bar{m}_k| = I\} \tag{6.5}$$

where Γ is a predefined array of just-intonation intervals corresponding to four octaves. Therefore, $\mathbb{D}_I$ corresponds to the set of all those vocal pitch segments for which the median of mean values of windows containing that segment is closest to the predetermined just-tuned pitch (Γ_I) corresponding to svarastāna I. A histogram is computed for each $\mathbb{D}_I$, and the parameters are extracted as described in ch. 5. As evident, the key difference between the two approaches lies in the way parameters for each svara are obtained. In the earlier approach,

we identify peaks corresponding to each svara from the aggregate histogram of the recording. In this approach, we isolate the pitch values of each svara from the pitch contour and compute a histogram for each svara.

The crucial parameters in this approach are t_w and t_h. A Carnatic music performance usually is sung in three speeds: lower, medium and higher ((Viswanathan and Allen, 2004)). A large part of it is in the middle speed. Also, singing in higher speed is more common than in the lower speed. From our analysis of varṇaṁs in Carnatic music, we observed the average duration each svara is sung in the middle speed to be around 200-250ms, while in the higher speed it is observed to be around 90-130ms.

Therefore, based on the choice of the window size (t_w), two different contexts arise. In the cases where the window size is less than 100ms (thus a context of 200ms for each segment), the span of the context more or less will be confined to one svara. Whereas in the other cases, the context spans more than one svara. In the first set of experiments reported below, we explore the first case.

Hop size (t_h) decides the number of windows (ϵ) which a given segment in the pitch contour is part of. A higher value for ϵ is preferred as it provides more fine-grained contextual information about the segment S_k (See Eqs. 6.1 and 6.3). This helps to take a better decision in determining the svara distribution to which it belongs to. However, if ϵ is too high, it might be that either t_w is too high, or t_h is too low, both of which are not desired: a very high value for t_w will span multiple svaras which is not we explore in this set of experiments, and a very low value for t_h is not preferred as it implies more computations. Keeping this in mind, we empirically set t_w and t_h to 100ms and 20ms respectively. Figure 6.4 shows the results for $t_w = 150\,\text{ms}$, $t_h = 30\,\text{ms}$, $t_w = 100\,\text{ms}$, $t_h = 20\,\text{ms}$ and $t_w = 90\,\text{ms}$, $t_h = 10\,\text{ms}$. In the figure, the intra-svara movements tend to be associated with the corresponding svara whereas the inter-svara movements are segmented and distributed appropriately.

Using this approach, we intend that the pitch segments be attributed to the appropriate svarastānas. However, as we noted earlier in our discussion at the beginning of the section, our assumptions and the complexity of the task mean that we have only limited success in doing so. Hence we do not claim that the final distributions are representative of the actual intonation of svaras as intended by the artists. Yet, as we obtain the context for segments in every recording using the same principle, we believe there will be more intra-raaga correspondences than the inter-raaga ones.

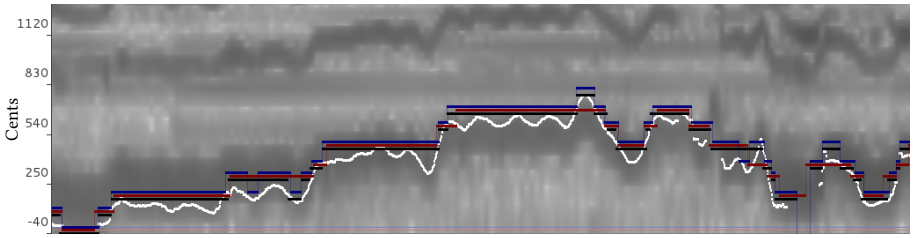


Figure 6.4: The pitch contour (white) is shown on top of the spectrogram of a short segment from a Carnatic vocal recording. The red (t_w = 150ms, t_h = 30ms), black (t_w = 100ms, t_h = 20ms) and blue (t_w = 90ms, t_h = 10ms) contours show the svara to which the corresponding pitches are binned to. The red and blue contours are shifted few cents up the y-axis for legibility.

Evaluation & results

We run the same set of tasks as we did for HPP, but with the parameters obtained using context-based svara distributions. We will regard the results from HPP as the baseline and compare with them. Tables 6.1 & 6.2 show the statistics over the outcome of feature selection on all rāgas and allied rāga groups respectively.

Unlike the statistics from tables 5.2 and 5.4, the position parameter assumes a relatively lesser role in rāga discrimination, while amplitude still is the most discriminating parameter. With an exception of kurtosis, all the newly introduced parameters (mean, variance and skewness) also are chosen by the feature selection algorithms more frequently than before. This marks the relevance of melodic and temporal context of svaras for their intonation description. Also it clearly indicates that the approach has been, at least partially, successful in leveraging such context.

In order to assess if the parameterization of context-based svara distributions bring in complementary information, here too, we conducted the same set of rāga classification experiments as we have done for HPP. Tables 6.3 and 6.4 show the averages over all the results for classification experiments conducted over all the rāgas in our music collection, and the allied rāga groups respectively. There is a notable and consistent improvement in the accuracies when all raagas are considered in the classification test, while the differences are marginal when classifying just the allied raagas. This too clearly indicates that our approach has been successful, though not to a desirable extent in which case the results over allied raagas would have clearly established it.

	Position		Amplitude		Mean		Variance		Skewness		Kurtosis	
	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.
Information gain	0.8	0.6	1.5	0.8	1.0	0.7	0.8	0.6	0.6	0.5	0.3	0.3
SVM	0.7	0.6	1.7	0.9	0.9	0.6	0.7	0.5	0.5	0.4	0.4	0.4

Table 6.1: Results of feature selection on sub-sampled sets of recordings in ${}^nC_3$ combinations of all rāgas using information gain and support vector machines. Ratio of total number of occurrences (abbreviated as Occ.) and ratio of number of recordings in which the parameter is chosen at least once (abbreviated as Rec.), to the total number of runs are shown for each parameter.

	Position		Amplitude		Mean		Variance		Skewness		Kurtosis	
	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.	Occ.	Rec.
Information gain	0.7	0.6	1.3	0.8	0.7	0.6	0.9	0.6	0.8	0.6	0.6	0.5
SVM	0.9	0.6	1.4	0.8	0.9	0.6	0.6	0.5	0.7	0.5	0.5	0.4

Table 6.2: Results of feature selection on sub-sampled sets of recordings in ${}^nC_2$ combinations of just the allied rāgas using information gain and support vector machines. Ratio of total number of occurrences (abbreviated as Occ.) and ratio of number of recordings in which the parameter is chosen at least once (abbreviated as Rec.), to the total number of runs are shown for each parameter.

Method/Classifier	Naive Bayes	3-Nearest Neighbours	SVM	Random forest	Logistic regression	Multilayer Perceptron
Histogram peak parametrization	78.26	78.46	71.79	81.16	78.61	78.78
Context-based pitch distributions	82.63	82.83	79.90	82.69	81.11	82.17

Table 6.3: Accuracies obtained using different classifiers in the rāga classification experiment with all the rāgas using histogram peak parametrization, and context-based pitch distributions. The baseline calculated using zeroR classifier lies at 0.33 in both experiments.

Method/Classifier	Naive Bayes	3-Nearest Neighbours	SVM	Random forest	Logistic regression	Multilayer Perceptron
Histogram peak parametrization	87.66	89.28	87.67	85.93	83.69	87.75
Context-based pitch distributions	88.88	89.38	85.35	87.94	83.55	86.06

Table 6.4: Accuracies obtained using different classifiers in the rāga classification experiment with the allied rāga groups using histogram peak parametrization and context-based pitch distributions. The baseline calculated using zeroR classifier lies at 0.50 in both experiments.

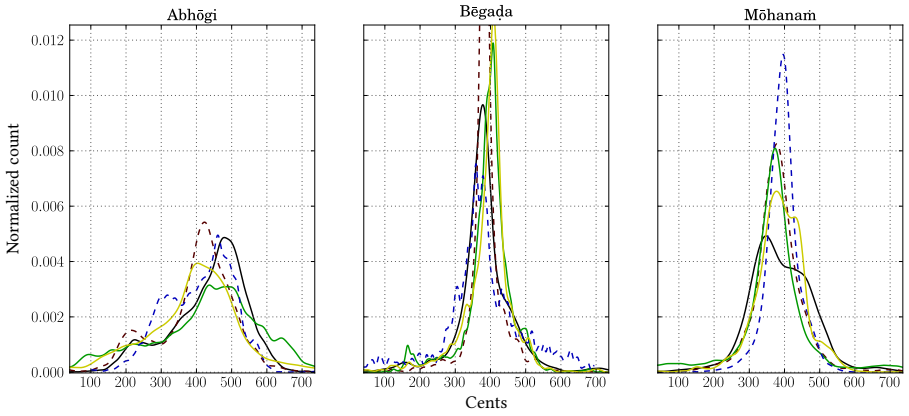
Retrospection of the approach

Music transcription in Carnatic music is a very challenging task; musicologists believe that it is harder even for expert musicians[2]. Therefore, we came up with an approach that rather than aiming at an accurate transcription, partially captures the melodic and temporal context so that the svara distributions can reflect the types of movements and the role of other svaras in a given svara's delineation in the performance. The results do indicate that the approach is partially successful in doing so. However, as the results from allied raaga classification indicate, there is still ample room for improvement (Table. 6.4).

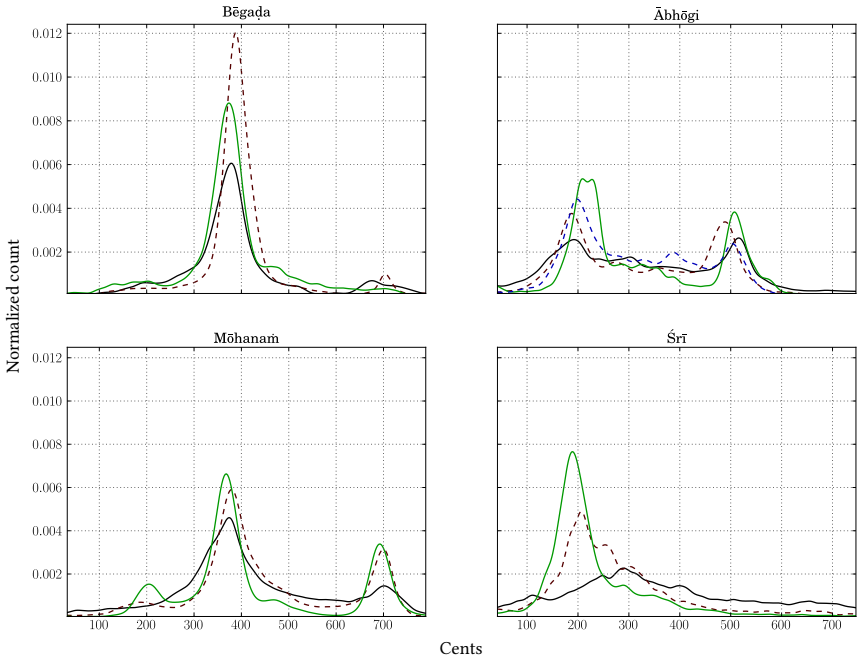
Fig. 6.5a shows the pitch histograms of Ga svara in Ābhōgi, Bēgaḍa and Mōhanaṁ rāgas, obtained using our approach. Fig. 6.5b (reproduced from ch. 7 for convenience) shows the same, but are generated using annotations from Varnam dataset as we will describe in the next chapter. For now, consider the later to be the closest we can get to the groundtruth. We can observe that in the case of Ābhōgi rāga, our method is partially successful in showing two peaks (i.e., at 200 and 500 cents), vaguely resembling the peaks in the corresponding plot in fig. 6.5b. However, this is not the case with Mōhanaṁ rāga where the pitch histograms obtained from our method failed to show peaks at 200 and 700 cents, though we see a slight bump around 450 cents. For Bēgaḍa rāga, we observe a single peak in the histograms shown in both figures. This is due to our deliberate choice, which is to constrain the window size so that it does not stretch beyond the typical duration of a svara which is 150 milliseconds. The plots clearly reflect this showing subdued presence of other svaras in a given svara's histogram.

As we already mentioned, the categories of melodic movements listed in fig. 6.3 are only representative. In the real pitch contours, such movements form a continuum that is even more challenging for analysis. Especially, if we allow the window size to stretch over 150 milliseconds, handling the transitory cases between the categories assumes higher relevance. In the next section, we try to address these issues and improve our approach further.

[2]In our interactions with musicians who had more than 15 years of training, they were confident in labeling only parts of the pitch contour when asked to do so. When it comes to the svara boundaries, they clearly convey that the transitions are fuzzy and cannot be marked objectively.



(a) Pitch histograms of Ga svara in three rāgas: Ābhōgi, Bēgaḍa and Mōhanaṁ, obtained using context-based svara distributions. Different lines in each plot correspond to different singers.



(b) Pitch histograms of Ga svara in four rāgas: Bēgaḍa, Mōhanaṁ, Ābhōgi and Śrī. X-axis represents cent scale. Different lines in each plot correspond to different singers.

Figure 6.5: Comparison of svara histogram plots obtained using our approach with those obtained using annotations in Varnam dataset.

6.3 Refining the svara description: consolidating the learnings

Changes to our approach

We first discuss ways to improve the usage of simple statistics in estimating melodic and temporal context of a given pitch segment. As a shorter window is ruled out as a suitable choice in our earlier discussion, we now work with a window size that spans more than a svara. This changes the very premise for arguments which lead us to mean and median statistics. Earlier, we used mean value over a window as its summary as it concerned movements over single svaras (see fig. 6.3). However, the rationale behind doing so cannot be extended to the present case when every window we look at almost always spans more than one svara.

As a result, the mean can no longer be a valid summary of a window that extends beyond a svara. Therefore, we suggest the following alternative procedure. We pick the two most common pitch values from the window. If their counts are comparable, we pick the mean value of the window to be its summary. Otherwise, we choose the most common pitch value (i.e., mode). To even out subtle variations in the pitch data, we quantize the values to 30 cent bins. Taking a count of the resulting values makes it more robust to subtle variations in the pitch contour. The rationale in this method is to determine the dominant svara when there are more than one in a window. If the two most commonly occurring pitch values have similar counts, we assume it is the case when the window spans an oscillatory movement over two semitones. In this case, we pick the mean value over the window, in line with our choices in the experiment before. Notice that there can also be cases where non-oscillatory movements result in being summarized by mean, but we believe there are fewer such cases, especially in Carnatic music where the oscillatory movement is the most common type (Krishna and Ishwar (2012)).

We found another shortcoming of our approach observing the svara plots such as the ones shown in fig. 6.5. As already mentioned, our aim is not the transcription of melodic contours. Therefore, not using the information about svaras that are allowed in a given raaga, has resulted in pitch values being distributed to those which are not part of the raaga. This is another potential cause which may have resulted in subdued presence of other svaras in a given svara plot. This can be easily addressed by distributing the pitch values only among the svaras allowed in the raaga.

Evaluation & results

We evaluate our approach with these modifications discussed, in a raaga classification test on a dataset drawn from the CompMusic collection (sec. 1.4). Note that, we used batches of 3 raagas in an earlier raaga classification task (see sec. 5.6) as it is carried out in conjunction with a feature selection task. Unlike that, in the current evaluation, we use all 40 classes in one classification task. We report two different sets of results. The first set listed in table. 6.5 are the results obtained using the parameters computed from the pitch distributions.

In our dataset, the number of parameters for an instance far exceeds the number of instances per class. Therefore, most classifiers are prone to overfitting problem. Given this caveat, table. 6.5 reports results using different kinds of classifiers. Our discussion concerns those which are easier to interpret and are relatively not effected by the nature of our dataset. These include the k-Nearest Neighbours and the tree-category of classifiers (Decision tree and Random Forest).

In particular, we are interested in observing the resulting set of rules in a decision tree. We already know from our previous results that the position and amplitude are the most relevant parameters (see tables. 6.1 and 6.2). We also saw that the new parameters - mean, variance, kurtosis and skew - do carry useful information for distinguishing raagas. Inspecting the decision tree model built using these parameters would give us more insights about the role each of them play in raaga classification, and therefore their relevance.

Results from table. 6.5 clearly show that our current approach outperforms histogram peak parametrization in its efficiency in raaga classification. Further, in a paired t-test with a significance level of 5%, improvement in results from the current approach across all the classifiers is shown to be statistically significant. Evidently, this proves that parameters from isolated svara histograms hold much more relevant information than those from an aggregate histogram. Further, the decision tree (reproduced in sec. A.2) shows that the new parameters introduced are more effective towards the lead nodes of the tree, which is expected as position and amplitude play a role in the beginning to segregate raagas differing by presence/absence of svaras. For those that share the svaras, the new parameters do seem to make a significant different in classification.

Having established that, sans parametrization, we need to validate whether the isolated svara histograms themselves are more informative compared to an aggregate histogram. From a musicological perspective, svara histograms do have a certain advantage wherein one can measure the presence of other svaras in its movements more precisely. Whereas an aggregate histogram does not facilitate

Method/Classifier	Naive Bayes	3-Nearest Neighbours	SVM	Multilayer Perceptron	Random Forest	Decision Tree
Histogram peak parametrization	27.71	27.92	32.08	28.54	48.75	26.46
Context-based pitch distributions	66.67	58.33	53.13	55.42	87.29	73.13

Table 6.5: Accuracies obtained using different classifiers in the rāga classification experiment with all the rāgas using histogram peak parametrization, and the improved context-based svara distributions. The baseline accuracy calculated using zeroR classifier lies at 2.5%.

Method	Accuracy
Aggregate histograms	77.3%
Isolated svara distributions	84.6%

Table 6.6: Accuracies obtained by matching pitch distributions obtained using different approaches.

this at all. However, observing how they fare in a raaga classification test may help us understand how deep the difference between them is.

The second set of results listed in table. 6.6 are accuracies obtained by directly matching pitch distributions. In this case, we use Kullback-Leibler divergence measure as given in eq. 3.2 to match two given normalized pitch distributions. Thereafter, we find k-Nearest Neighbours (k=1, 3, 5) in a leave-one-out cross-validation experiment to label a given test sample with a raaga. Though the difference in their performances is not as striking as it is in table. 6.5, isolated svara histograms still outperform aggregate histograms by a fair margin. The difference between accuracies obtained by matching aggregate histograms and using their parameters indicates that peak parametrization in histograms does result in a loss of valuable information. The isolated svara histograms obtained using the current approach however do not seem to be affected by parametrization, which seems to indicate their robustness.

6.4 Summary & conclusions

We have presented an approach to parametrize context-based svara distributions to obtain intonation description from Carnatic music recordings. In this approach, we addressed some of the major drawbacks of the histogram peak parametrization method (ch. 5): lack of melodic and temporal context, and finding the bandwidth of the peaks. Unlike the former approach where pitches for each svara are derived from a histogram, in this method, the pitches corresponding to each svara are obtained directly from the pitch contour by considering it's melodic and temporal context. For each svara, a histogram is then computed from which the parameters are obtained. Thus, it alleviates the need for estimating the location and bandwidth of peaks. As a result, this approach requires lesser number of parameters to be tuned. The results from the final evaluation task (sec. 6.3) show that this approach outperforms the earlier one, proving that this approach provides a better intonation description. The source code for both the methods

is made openly available online[3].

We believe these results are sufficient to justify the usefulness and robustness of our svara representation and parametrization compared to the earlier approaches. In the next couple of chapters, our emphasis will be on qualitative analysis that can further improve this representation so that they are suitable for extracting musically meaningful parameters.

The code resulting from this chapter will be consolidated and published to the intonation python module[4]. The source code can be accessed from the corresponding github repository[5]. The usage instructions are contained therein.

[3]URL: https://github.com/gopalkoduri/intonation
[4]https://pypi.python.org/pypi/intonation
[5]https://github.com/gopalkoduri/intonation

Chapter 7

Taking a step back: Qualitative analysis of varnams

All through our work on intonation description of svaras, there has been a need for a reference that can be used to compare the representations of svaras obtained using our approaches. For this, ideally we need an annotated dataset of Carnatic songs where each svara is clearly marked. But as we already mentioned, labeling svara boundaries turns out to be a challenging task even for an expert musician. Therefore, we look for alternate ways to get as close to the ideal reference as possible.

Varṇaṁs are a particular form in Carnatic music that succinctly capture the essence of a raaga. Further, they are often sung as composed, allowing us to take advantage of their notation. We have put together a dataset which we introduced in sec. 1.4 that has monophonic audio, notations and annotations of taala cycles. We make use of this dataset in getting to the aforementioned reference for svara representations.

First, we introduce the structure of varnams and explain why they are suitable for our analysis. Then we discuss how we semi-automatically synchronize the taala cycles in svara notation of a varnam to those in its annotations. We use the svara timestamps to compute svara histograms.

7.1 Relevance and structure of varnams

The macro structure of varṇaṁ has two parts: pūrvāṅga and uttarāṅga. The pūrvāṅga consists of the pallavi, anupallavi and muktāyi svara. The uttarāṅga

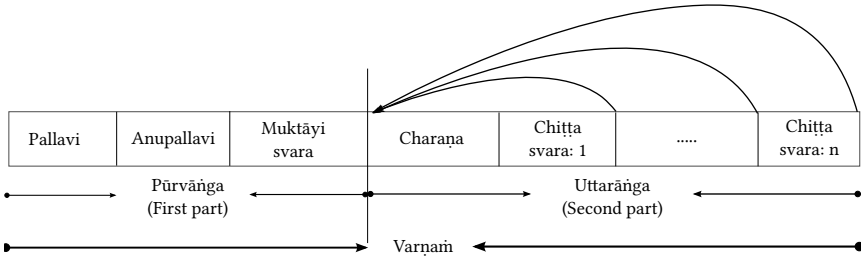


Figure 7.1: Structure of the varṇaṁ shown with different sections labeled. It progresses from left to right through each verse (shown in boxes). At the end of each chiṭṭa svara, charaṇa is repeated as shown by the arrows. Further, each of these verses is sung in two speeds.

consists of the charaṇa and the chiṭṭa svaras. Figure 7.1 shows the structure of the varṇaṁ with two parts and different sections labeled. A typical varṇaṁ performance begins with the singing of pūrvāṅga in two different speeds, followed by uttarāṅga, where in after each chiṭṭa svara, the singer comes back to charaṇa. Different variations to this macro structure give rise to various types of varṇaṁs: pada varṇaṁs, tāna varṇaṁs and dhāru varṇaṁs ((Rao, 2006)). Varṇaṁs are composed in a way such that the structure includes variations of all the improvisational aspects of Carnatic music ((for an in-depth understanding of the relevance of varṇaṁs in Carnatic music, see Vedavalli, 2013b,a)). For example, chiṭṭa svaras[1] are composed of svaras that capture all their possible combinations and structures in a given rāga. This helps singers in an improvisational form called *kalpana svaras*, where they permute and combine svaras as allowed by the rāga framework to create musically aesthetic phrases.

Due to the varṇaṁ structure and it's purpose, the rendition of varṇaṁs across musicians is fairly less variant than the variations seen in the renditions of other compositional forms. This is because most performances of the varṇaṁs deviate less from the given notation. Though the artists never use the notations in their actual performances, they have been maintained in the tradition as an aid to memory. Our work exploits this rigidity in structure of the varṇaṁ to align the notation with the melody and extract the pitch corresponding to the various svaras. Rāgas were chosen such that all the 12 svarastānas in Carnatic music are covered ((Serra, 2011; Krishna and Ishwar, 2012)). This would allow us to observe the impact of different melodic contexts (i.e., in different rāgas) on each of the svaras.

[1]Chiṭṭa svaras in Sanskrit literally mean the *svaras in the end.*

7.2 Svara synchronization

Recall that our goal is to have a reference representation for svaras in their different melodic context, viz., raagas. For this, we obtain all the pitch values corresponding to each svara in a given raaga, and analyze their distributions. The method consists of five steps: (1) The pitch contour of the recording is obtained (in the same way as described in sec. 5.3). (2). Tāla cycles are manually annotated by two musicians. (3) These tāla cycles are semi-automatically synchronized with the notation. (4). Pitch values corresponding to each svara are obtained from the pitch-contour. (5). A normalized histogram from the pitch values of each svara is computed and interpreted (as in sec. 5.4).

In order to be able to semi-automate the synchronization process, we confine our analysis to a predetermined structure of the varṇaṁ in its sung form: pūrvāṇga in two speeds, followed by a verse-refrain pattern of charaṇa and chiṭṭa svaras, each in two speeds. Using Sonic Visualizer ((Cannam et al., 2010a)), we marked the time instances which correspond to the start and end of tāla cycles which fall into this structure. A sequence of tāla cycles is generated from the notation such that they correspond to those obtained from the annotations. Hence, we now have the start and end time values for each tāla cycle (from annotations) and the svaras which are sung in that cycle (from notation).

Recall that we chose to analyze the varṇaṁs sung only in adi tāla (sec. 1.4). Each cycle in ādi tāla corresponds to 8 or 16 svaras depending on whether the cycle is sung in fast or medium speed. Each cycle obtained from annotations is split into appropriate number of equal segments to mark the time-stamps of individual svaras. The pitches for each svara are then obtained from the time locations in the pitch contour as given by these time-stamps. A normalized histogram is then computed for each svara combining all its pitch-values (see eq. 5.4).

7.3 Analysis of svara histograms

Fig. 7.2 shows pitch histograms for performances in two rāgas: Kalyāṇi and Śankarābharaṇaṁ. Even though they theoretically have all but one svara in common, the pitch histograms show that the peak locations and their characteristics are different. This is a clear indication that the rāgas cannot be differentiated by using just their svarastānas.

There are many such rāgas which have common svaras. However, their intonation is very different depending on the respective rāga's characteristics. To elaborate this, we take the example of svara Ga which is common between the

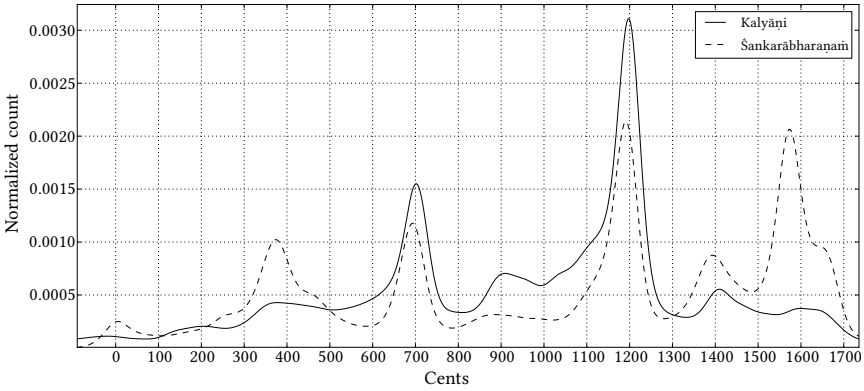


Figure 7.2: Histograms of pitch values obtained from recordings in two rāgas: Kalyāṇi and Śankarābharaṇaṁ. X-axis represents cent scale, normalized to tonic (Sa).

rāgas Mōhanaṁ and Bēgaḍa. Due to the context in which the Ga is sung in each of the rāgas, the intonation and the gamakas expressed on the svara change. Figure 7.3 shows that the svara Ga in Bēgaḍa corresponds to one sharp dominating peak at 400 cents. This concurs with the fact that the Ga in Bēgaḍa is always sung at its position with minimum gamakas. It is a steady note in the context of the rāga Bēgaḍa. On the other hand, the same figure shows that Ga in Mōhanaṁ corresponds to two peaks at 400 and 700 cents with a continuum from one peak to the other. The dominant peak is located at 400 cents (i.e., Ga's position). This is in line with the fact that Ga in Mōhanaṁ is rendered with an oscillation around its pitch position. The oscillation may vary depending on the context in which it is sung within the rāga. Ga in Mōhanaṁ, generally, starts at a svara higher (Ma or Pa) even though it may not be theoretically present in the rāga, and ends at its given position after oscillation between its own pitch and a higher pitch at which the movement started.

Another example of such svara is Ga in Ābhōgi and Śrī[2]. Fig. 7.3 shows that Ga in Ābhōgi is spread from 200 cents to 500 cents, with peaks at 200 cents and 500 cents respectively. These peak positions correspond to the svaras Ri and Ma, respectively. The inference one can make from this is that the Ga in Ābhōgi is sung as an oscillation between Ri and Ma of the rāga Ābhōgi, which is true in practice. The pitch histogram for Ga of Śrī in fig. 7.3 shows that the peak for Ga in Śrī is smeared with a peak at 200 cents which is the Ri in Śrī. This is consistent

[2]Ga in Bēgaḍa and Mōhanaṁ correspond to a svarastāna which is different from the one that Ga in Ābhōgi and Śrī correspond to.

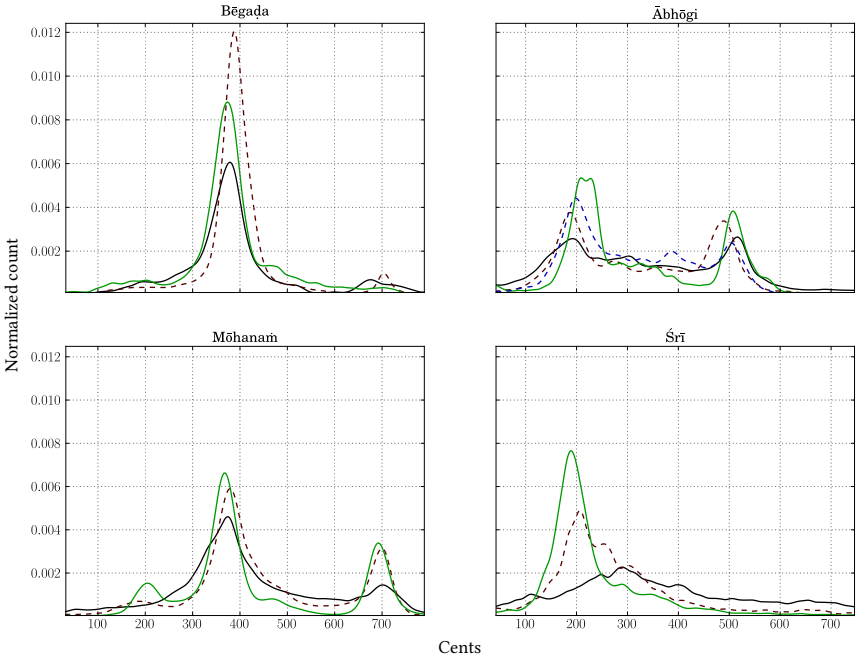


Figure 7.3: Pitch histograms of Ga svara in four rāgas: Bēgaḍa, Mōhanaṁ, Ābhōgi and Śrī. X-axis represents cent scale. Different lines in each plot correspond to different singers.

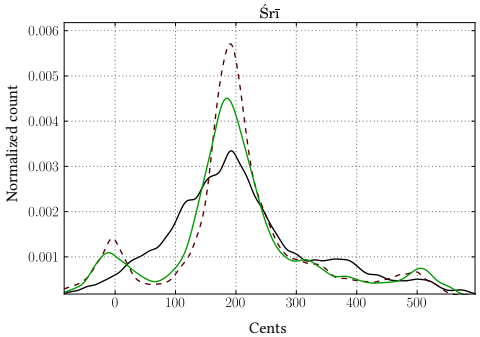


Figure 7.4: Pitch histogram for Ri svara in Śrī rāga. X-axis represents cent scale. Different lines in each plot correspond to different singers.

Svara (Rāga)	Sa	Ri	Ga	Ma	Pa	Da	Ni
Ga (Bēgaḍa)	0/14	74/56	-	80/64	0/18	2/0	0/4
Ga (Mōhanaṁ)	4/2	72/71	-	-	68/96	28/4	-
Ga (Ābhōgi)	24/0	44/68	-	55/58	-	2/0	-
Ga (Śrī)	0/2	88/88	-	0/0	0/0	0/0	2/0
Ri (Śrī)	106/132	-	88/88	52/46	6/6	0/0	26/6

Table 7.1: Transition statistics for svaras discussed in the section. Each cell gives the ratio of number of transitions made from the svara (corresponding to the row) to the number of transitions made to the svara.

with the fact that Ga in Śrī is rendered very close to Ri. A comparison of the pitch histograms of the Ri in Śrī (Figure 7.4) and the Ga in Śrī shows that the peaks of Ga and Ri almost coincide and the distribution of the pitch is also very similar. This is because the movement of Ga in Śrī always starts at Ri, touches Ga and lands at Ri again. Ga in Śrī is always a part of any phrase that ends with RGR sequence of svaras, and in this context Ga is rendered as mentioned above.

Insights such as the ones we discussed in this section require musical knowledge about the svaras and their presentation in the context of a rāga. To complement this, we have derived the transition matrices of svaras in each varṇaṁ from notations. The transition statistics of a given svara are observed to usually correspond to the pattern of peaks we see in its pitch histogram. Table. 7.1 lists the transitions involving Ga in Bēgaḍa, Mōhanaṁ, Ābhōgi and Ga, Ri in Śrī.

With the exception of Ga in Bēgaḍa, we notice that the other svaras to/from which the transitions occur are the ones which are manifest in the pitch histogram of the given svara. Combining this information with peaks in pitch histogram yields interesting observations. For instance, a svara such as Ga in Bēgaḍa rāga records a number of transitions with Ri and Ma svaras, but the pitch histogram shows a single peak. This clearly indicates that it is a svara sung steadily without many gamakas. On the other hand, in the case of svaras like Ga in Mōhanaṁ, we see that there are a number of transitions with Ri and Pa svaras, while there are also several peaks in the histogram. This is an indication that the svara is almost always sung with gamakas, and is anchored on other svara or sung as a modulation between two svaras. The transitions are also indicative of the usage of svaras in ascending/descending phrases. For instance, the transitions for Ga svara in Śrī rāga mark the limited context in which it is sung.

To keep the discussion concise and to the point, we have taken a few examples to

explain how the same svara in different raagas is manifest differently. Plots for all the svaras are available at [url] for further reference.

7.4 Summary & conclusions

We analyzed varṇaṁs to get a reference for svara representation that can act as a groundtruth for comparing the representations obtained using different approaches. We have presented arguments as to how these representations are musicologically valid. The observations juxtaposing resulting svara histograms with the corresponding svara transition statistics indicate that the former are reliable models capturing adequate details concerning svara intonation and the melodic movements involving it.

In the next chapter, we propose an alternative approach to context-based pitch distributions of svaras. In this, we make use of an automatic melodic phrase alignment approach to get timestamps of svaras in the absence of taala annotations. We compare the svara histograms obtained using these two approaches with the reference histograms obtained using varnam dataset in this chapter. Doing so will allow us to qualitatively understand the merits and limitations of context-based pitch histograms of svaras, and those obtained by automatic melodic phrase alignment.

Chapter 8

Melodic phrase alignment for svara description

In this chapter, we consolidate the svara representation we have been working with so far, and place it in context alongside other representations used in different contexts for musical notes. We then use a methodology that partially aligns the audio recording with its music notation in a hierarchical manner first at metrical cycle-level and then at note-level, to describe the pitch content using our svara representation. The intonation description of svaras using this representation is evaluated extrinsically in a classification test using the varnam and the kriti datasets (see sec. 1.4).

8.1 Consolidating svara representation

A musical note can be defined as a sound with a definite pitch and a given duration. An interval is a difference between any two given pitches. Most melodic music traditions can be characterized with a set of notes it uses and the corresponding intervals. They constitute the core subject matter of research concerning the tonality and melodies of a music system. For any quantitative analyses therein, it is required to have a working definition and a consequent computational model of notes which dictate how and what we understand of the pitch content in a music recording.

In much of the research in music analysis and information retrieval, the most commonly encountered model is one that considers notes as a sequence of points separated by certain intervals on frequency spectrum. There are different representations of the pitch content from a given recording based on this notion,

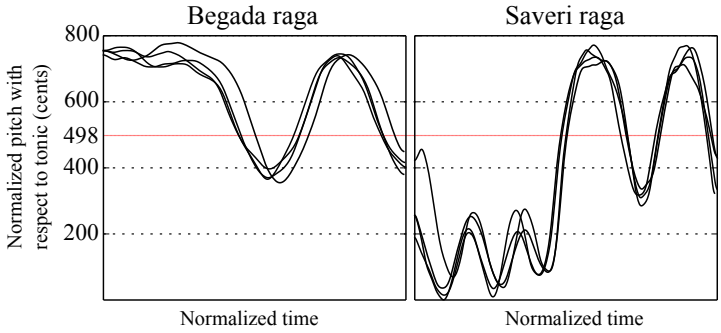


Figure 8.1: Pitch contours of M_1 svara in different raagas.

the choice among which is influenced to a great degree by the intended application. Examples include pitch class profiles (Fujishima (1999)), harmonic pitch class profiles (Gómez (2006)) and pitch histograms (Gedik and Bozkurt (2010)) besides others.

Albeit a useful model of notes used alongside several information retrieval tasks, we believe it is limited in its purview. To illustrate this, we took the case of Carnatic music, where the counterpart to note is svara, which as we know has a very different musicological formulation (sec. 2). A svara is defined to be a definite pitch value with a range of variability around it owing to the characteristic movements arising from its melodic context. It is emphasized that the identity of a svara lies in this variability (Krishna and Ishwar (2012)), which makes it evident that the former model of notes has a very limited use in this case.

Fig. 8.1 shows melodic contours extracted from the individual recordings of M_1 svara (498 cents) in different raagas. It shows that a svara is a continuum of varying pitches of different durations, and the same svara is sung differently in two given raagas. Note that a svara can vary even within a raaga in its different contexts (Subramanian (2007); Krishnaswamy (2003)). Taking this into consideration, we proposed an alternate representation of pitches constituent in a svara. In this, we define a note as *a probabilistic phenomenon on a frequency spectrum.* This notion can be explored in two complementary approaches: i) *temporal*, which helps to understand the evolution of a particular instance of a svara over time and ii) *aggregative*, which allows for studying the whole pitch space of a given svara in its various forms, often discarding the time information.

The representation we proposed takes the latter approach. From the annotations in the varnam dataset, we aggregate the pitch contours over the svara reported in fig. 8.1 for the same set of raagas. Fig. 8.2 shows its representations, which are

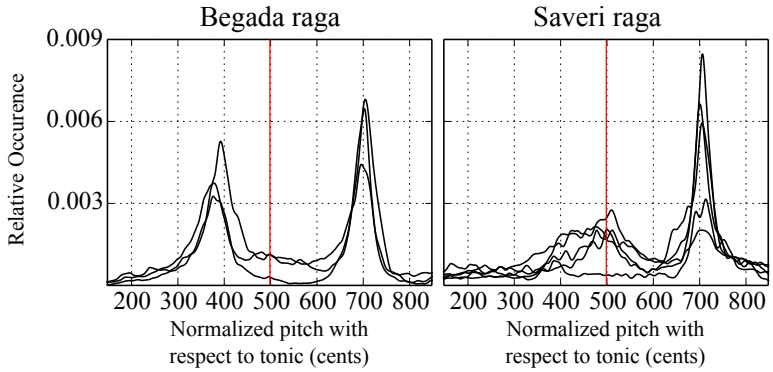


Figure 8.2: Description of M_1 svara using annotated data.

normalized distributions of values in respective pitch contours. The correspondences between the two figures are quite evident. For instance, M_1 in Begada is sung as an oscillation between G_3 (386 cents) and M_1. The representation reflects this with peaks at the corresponding places. Further, the shape of the distributions reflect the nature of pitch activity therein.

8.2 Audio-score alignment

In this section, we make use of an audio-score alignment system to build the svara representation we discussed. This alleviates the need for taala cycle annotations which we found to be necessary for our analysis in the previous chapter. Further, this approach also employs the notation information which we believe will result in a better svara representation compared to our approach discussed in ch. 6.

Audio-score alignment can be defined as *the process of finding the segments in the audio recording that correspond to the performance of each musical element in the music score.* For this task, several approaches have been proposed using techniques such as Hidden Markov models (Cont (2010); Maezawa et al. (2011)), conditional random fields (Joder et al. (2010)) and dynamic time warping (Dixon and Widmer (2005); Fremerey et al. (2010); Niedermayer (2012)).

The structural mismatch between the music score and the audio recording is a typically encountered challenge in audio-score alignment. This is also common phenomenon in the performances of varnams and kritis, where the singers tend to repeat, omit or insert cycles in the score. To overcome this problem there exists methodologies, which allow jumps between structural elements (Fremerey et al. (2010); Holzapfel et al. (2015)). However these methodologies are not designed

to skip musical events in the performance, which are not indicated in the score, such as impromptu improvisations commonly sung in kritis. Moreover, we may not need a complete alignment between the score and audio recording in order to accumulate a sufficient number of samples for each svara.

Senturk et al. (2014) introduced an audio-score alignment methodology for aligning audio recordings of Ottoman-Turkish makam music with structural differences and events unrelated to the music score. Şentürk et al. (2014) later extended it to note-level alignment. The methodology proposed by the former divides the score into meaningful structural elements using the editorial section annotations in the score. It extracts a predominant melody from the audio recording and computes a synthetic pitch of each structural element in the score. Then it computes a binarized similarity matrix for each structural element in the score from the predominant melody extracted from the audio recording and the synthetic pitch. The similarity matrix has blobs resembling lines positioned diagonally, indicating candidate alignment paths between the audio and the structural element in the score. Hough transform, a simple and robust line detection method (Ballard (1122)), is used to locate these blobs and candidate time-intervals for where the structural element is performed is estimated. To eliminate erroneous estimations, a variable-length Markov model based scheme is used, which is trained on structure sequences labeled in annotated recordings. Finally, Şentürk et al. (2014) applies Subsequence Dynamic Time Warping (SDTW) to the remaining structural alignments to obtain the note-level alignment.

The alignment methodology used in our work is based on the work of Senturk et al. (2014); Şentürk et al. (2014). Since the original methodology is proposed for Ottoman-Turkish makam music, several parameters are optimized according to the characteristics of our data. Also, several steps are modified in the original methodology for the sake of generalization and simplicity. These changes are detailed in **?**. Here, we summarize the procedure as follows:

1. Extract the predominant melody from the audio recording (Salamon et al. (2012)), normalize it using tonic.
2. Get synthetic pitch contour from the notation assuming just-intonation temperament and 70 bpm (corresponding to average tempo in varnam dataset).
3. Estimate possible partial alignments between the predominant melody and the synthetic pitch contour at cycle-level.
4. Discard erroneous estimations using k-means clustering to purge clusters with low scores.

5. Extract svara samples from within each aligned cycle assuming equal duration for each svara.

8.3 Computing svara representations

For a given recording, for each svara, σ, in the corresponding raaga, we obtain a pool of normalized pitch values, $x^\sigma = \{x_1^\sigma, x_2^\sigma, \dots\}$, aggregated over all the aligned instances from its melodic contour. Our representation must capture the probabilities of the pitch values in a given svara. We compute a normalized histogram over the pool of pitch values for each svara. For brevity sake, we consider pitch values over the middle octave (i.e., starting from the tonic) at a bin-resolution of one cent:

$$h_m^\sigma = \frac{\sum_i \lambda_m(x_i^\sigma)}{|x^\sigma|},$$

where h_m^σ is the probability estimate of the m-th bin, $|x^\sigma|$ is the number of pitch values in x^σ and λ function is defined as:

$$\lambda_m(a) = \begin{cases} 1, & c_m \leq a \leq c_{m+1} \\ 0, & \text{otherwise} \end{cases}$$

where a is a normalized pitch sample and (c_m, c_{m+1}) are the bounds of the m-th bin.

Figure 8.3 shows the representations obtained in this manner for M_1 svara (our running example from Figure 8.1) in different raagas. Notice that the representations obtained for M_1 are similar to the corresponding representations shown in Figure 8.2. This representation allows to deduce important characteristics of a svara besides its definite location (i.e., 498 cents) in the frequency spectrum. For instance, from Figure 8.3, one can infer that M_1 in Begada and Saveri are sung with an oscillation that ranges from G_3 (386 cents) to P (701 cents) in the former and M_1 to P in the latter.

8.4 Evaluation, comparison and discussion

As we already mentioned, our primary goal is to have a qualitative comparison between the svara representations obtained from our different approaches. A quick albeit opaque way to verify how they compare against each other is to employ the parameters from those representations in a raaga classification task. Post

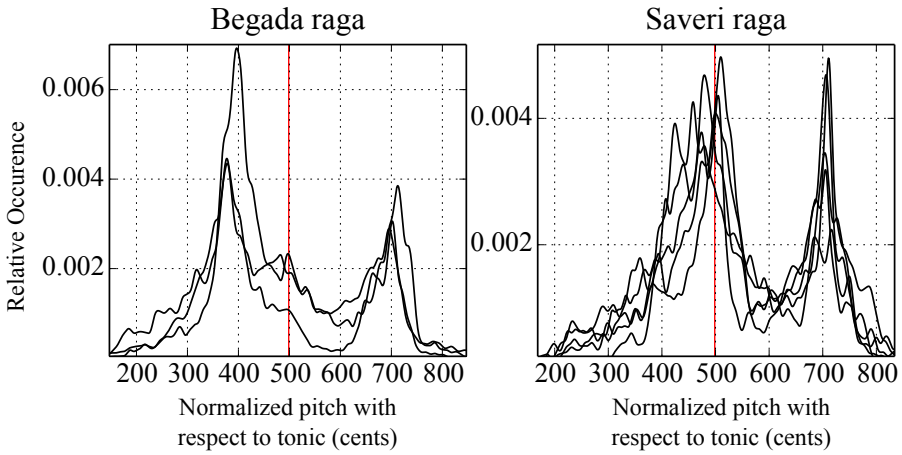


Figure 8.3: Description of M_1 svara (498 cents in just intonation) using our approach.

this, we directly compare how well the svara representations from the current approach and the context-based pitch distributions of svaras compare against those computed using annotations in the varnam dataset.

Therefore, the svara-level alignment and the computed representation are first evaluated extrinsically using a raaga classification task on the varnam and the kriti datasets introduced in ch. 1. In the varnam dataset, we aligned 606 cycles and 15795 svaras in total. Out of these cycles 490 are true positives. By inspecting the false positives we observed two interesting cases: occasionally an estimated cycle is marked as false positive when one of the boundary distances is slightly more than 3 seconds. The second case is when the melody of the aligned cycle and performance is similar to each other. In both situations considerable number of the note-level alignments would still be useful for the svara model. Within our kriti dataset, 1938 cycles and 59209 svaras are aligned in total. We parametrize the representation of each svara using the same set of features we used in our earlier approaches:

i. The highest probability value in the normalized histogram of the svara

ii. Mode, which is the pitch value corresponding to the above

iii. Mean of the distribution, which is the probability-weighted mean of pitch values

iv. Pearson's second skewness coefficient

v. Fisher's kurtosis

vi. Variance

As we know, there are 12 svaras in Carnatic music, where each raaga has a subset of them. For the svaras absent in a given raaga, we set the features to a nonsensical value. Each recording therefore has 72 features in total. The smallest raaga-class has three recordings in the varnam dataset, with few classes having more, so we subsampled the dataset thrice each time with a different seed. The number of instances per class is three. We have also subsampled kriti dataset in a similar manner, with number of instance per class set to five, and the dataset is further subsampled five times.

We performed the classification experiment over the subsampled sets of the two datasets using the leave-one-out cross-validation technique. The mean F_1-scores using the representations obtained from the annotations in the varnam dataset, the current approach and context-based pitch distributions of svaras across the subsampled datasets for the two datasets are reported in table. 8.1. Results corresponding to all the methods remain more or less the same. Owing to the limitation of the varnam dataset wherein all the recordings in a given raaga are of the same composition, the results turn out to be near perfect. The results over the kriti dataset are reported in table. 8.2. We can observe that here the context-based pitch distributions have performed marginally better than the current approach. However, this difference is statistically insignificant.

The series of figures. 8.4- 8.8 show representations of svaras in Sahana except Sa and Pa, which aren't shown here as they are mostly sung steady. We can observe that in most cases, there are clear correspondences between svara plots from different approaches. These include the overall shape of the distribution, the peak positions and their relative importance indicated by the peak amplitude. It is encouraging to see that the context-based pitch distributions approach which uses just the svara information performs very similar to our current approach which uses scores. For brevity sake, we have shown the plots of svaras in one raaga. Those corresponding to other raagas are available at [url].

8.5 Conclusions

We have consolidated our representational model for svara that expands the scope of the current note model in use by addressing the notion of variability in svaras. We have discussed an approach that exploits scores to describe pitch content in the audio music recordings, using our model. However, we seek attention to the

Method/Classifier	Naive Bayes	1-Nearest Neighbours	SVM	Multilayer Perceptron	Random Forest	Decision Tree
Context-based	100.00	100.00	100.00	100.00	95.24	74.60
Phrase-aligned	98.41	100.00	100.00	100.00	90.48	80.95
Annotations	95.24	95.24	95.24	95.24	95.24	74.60

Table 8.1: Accuracies obtained using different classifiers in the rāga classification task on the varnam dataset. The baseline accuracy calculated using zeroR classifier lies at 14.29%.

Method/Classifier	Naive Bayes	1-Nearest Neighbours	SVM	Multilayer Perceptron	Random Forest	Decision Tree
Context-based	95.33	96.00	93.33	94.67	96.00	96.67
Phrase-aligned	96.67	86.67	86.67	90.00	89.33	83.33

Table 8.2: Accuracies obtained using different classifiers in the rāga classification task on the kriti dataset. Note that there are no annotations available in this dataset, hence none is reported. The baseline accuracy calculated using zeroR classifier lies at 16.67%.

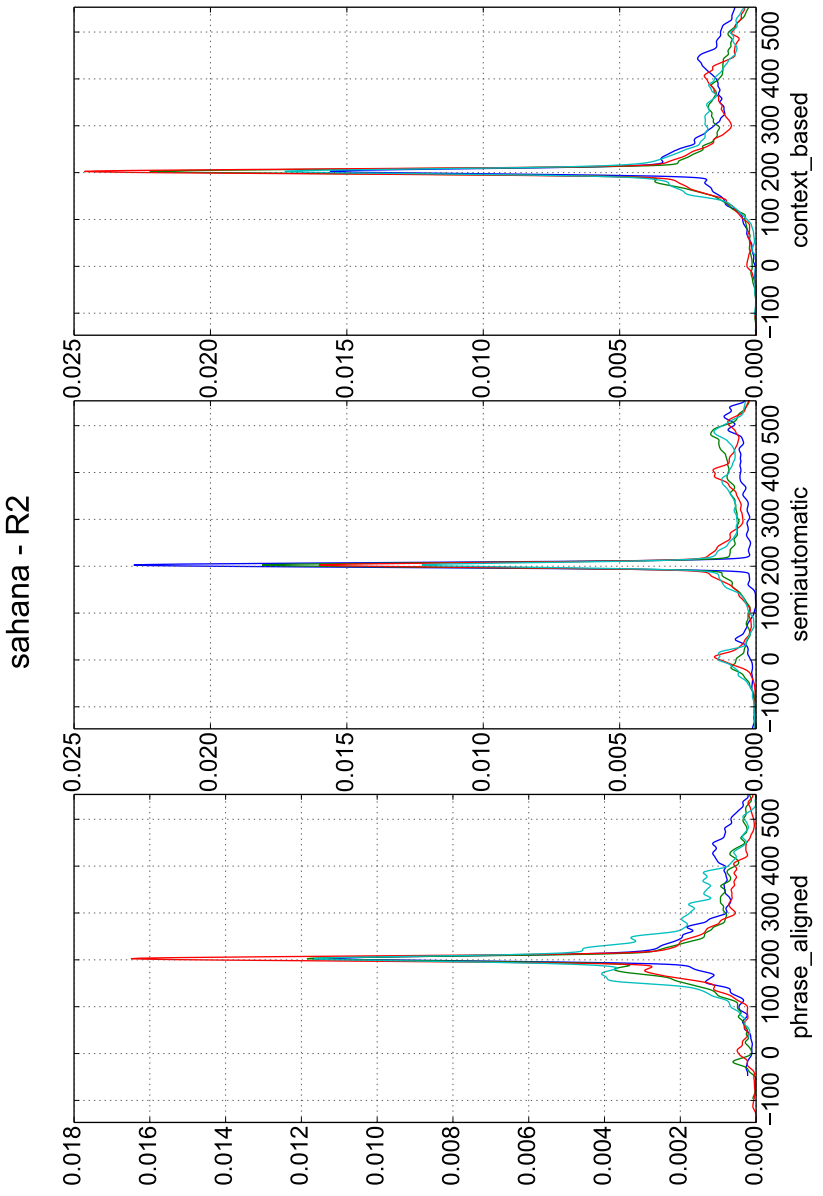


Figure 8.4: Representation for R2 svara in Sahana raaga computed using the three approaches.

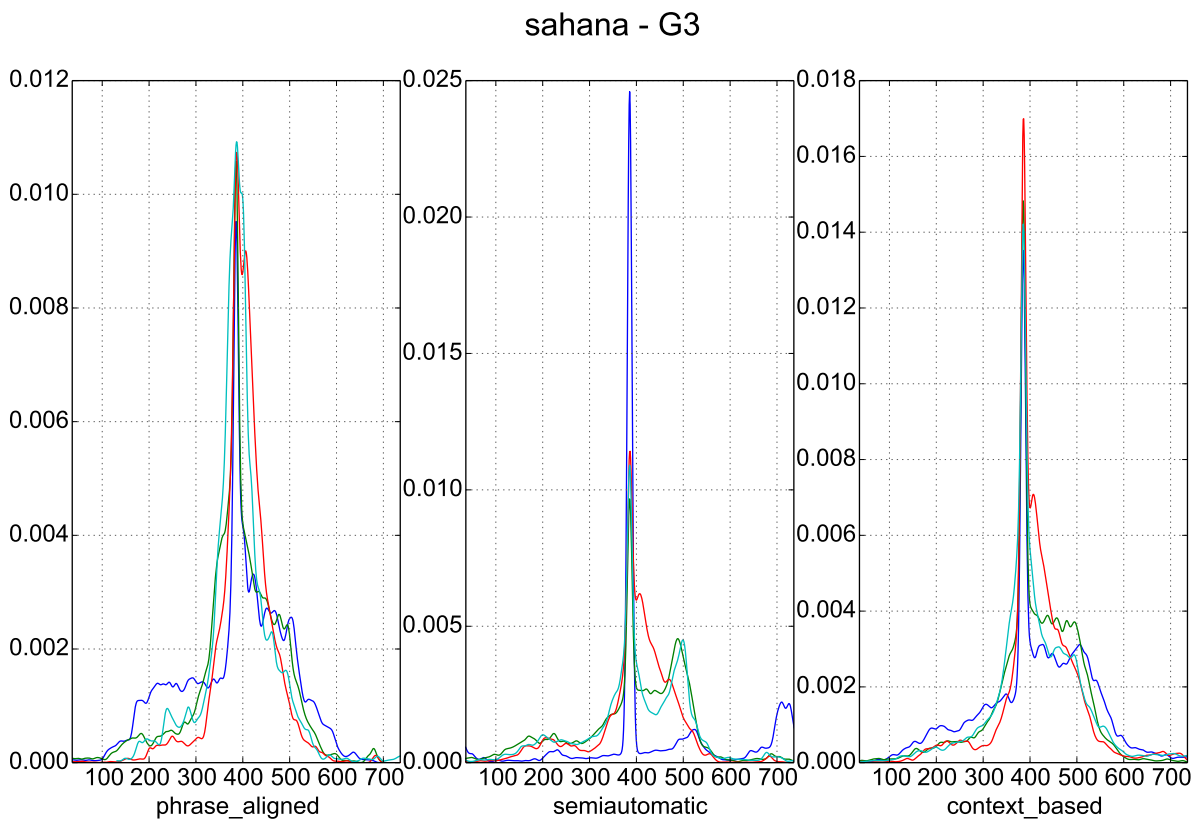


Figure 8.5: Representation for G3 svara in Sahana raaga computed using the three approaches.

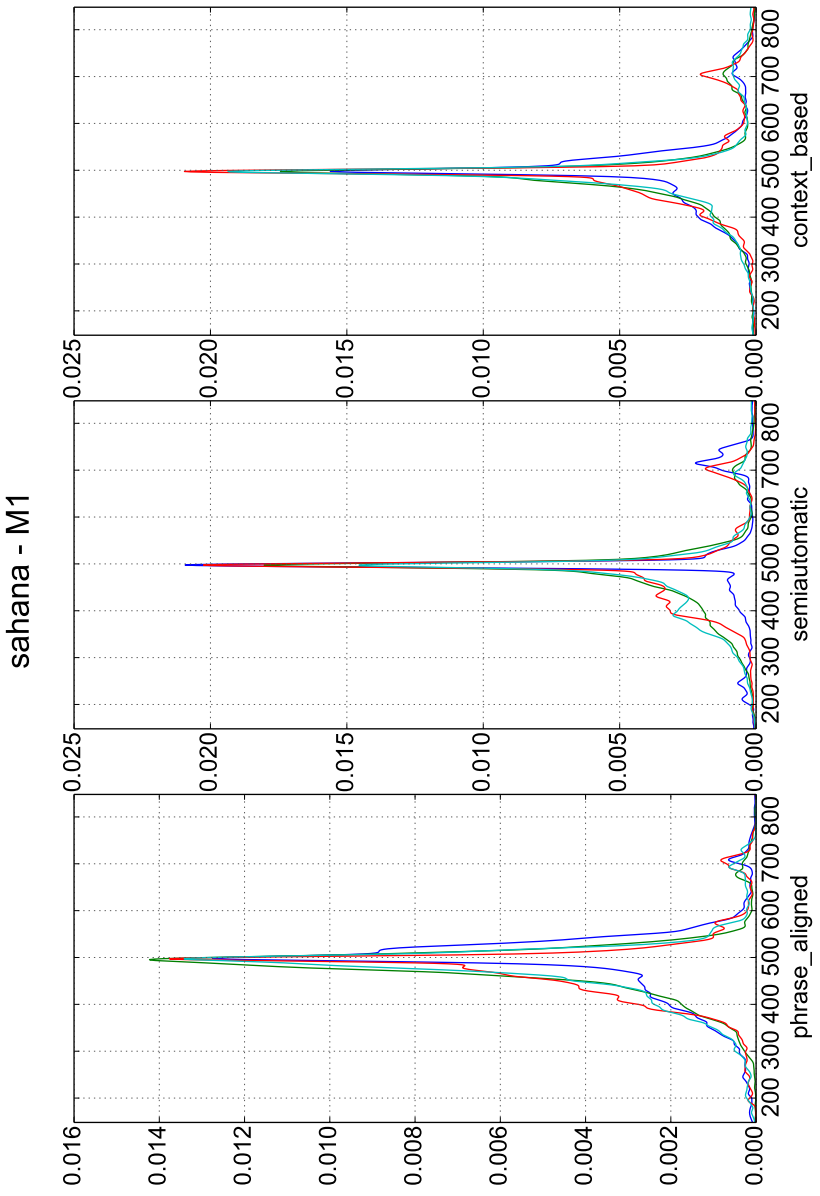


Figure 8.6: Representation for M1 svara in Sahana raaga computed using the three approaches.

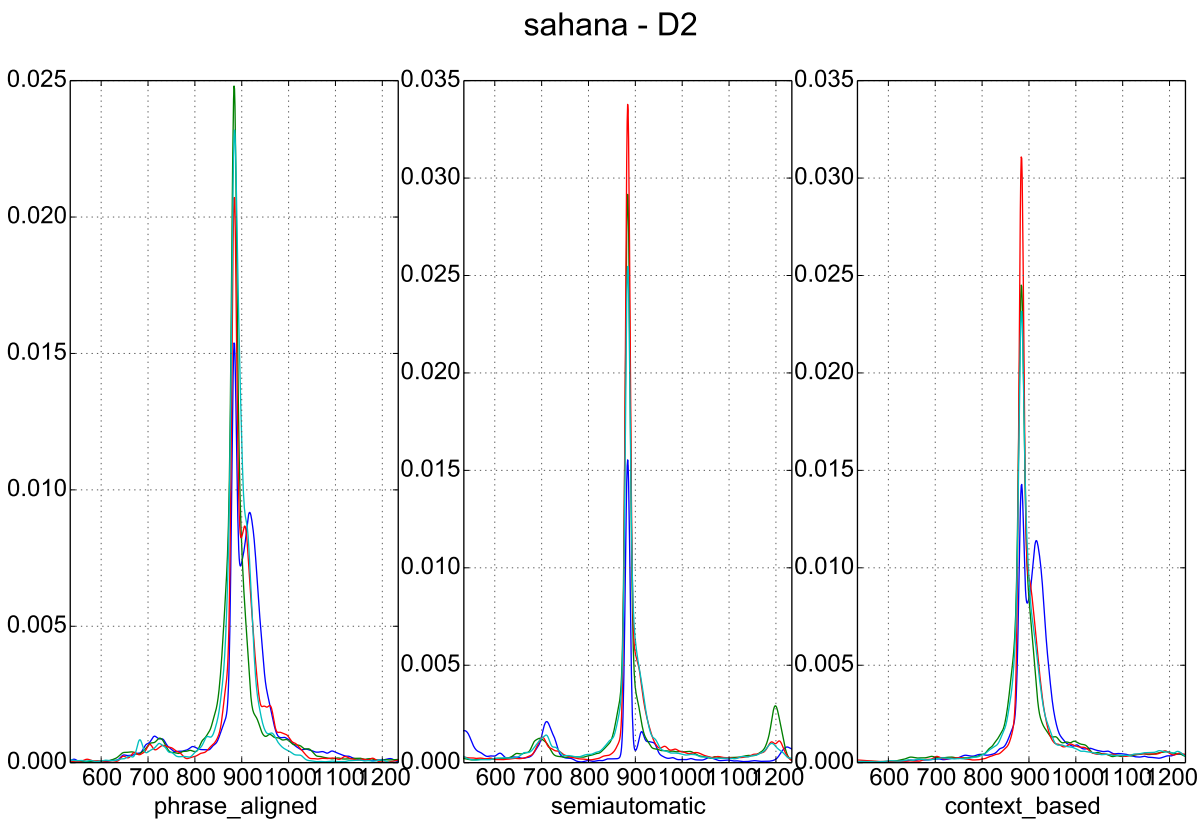


Figure 8.7: Representation for D2 svara in Sahana raaga computed using the three approaches.

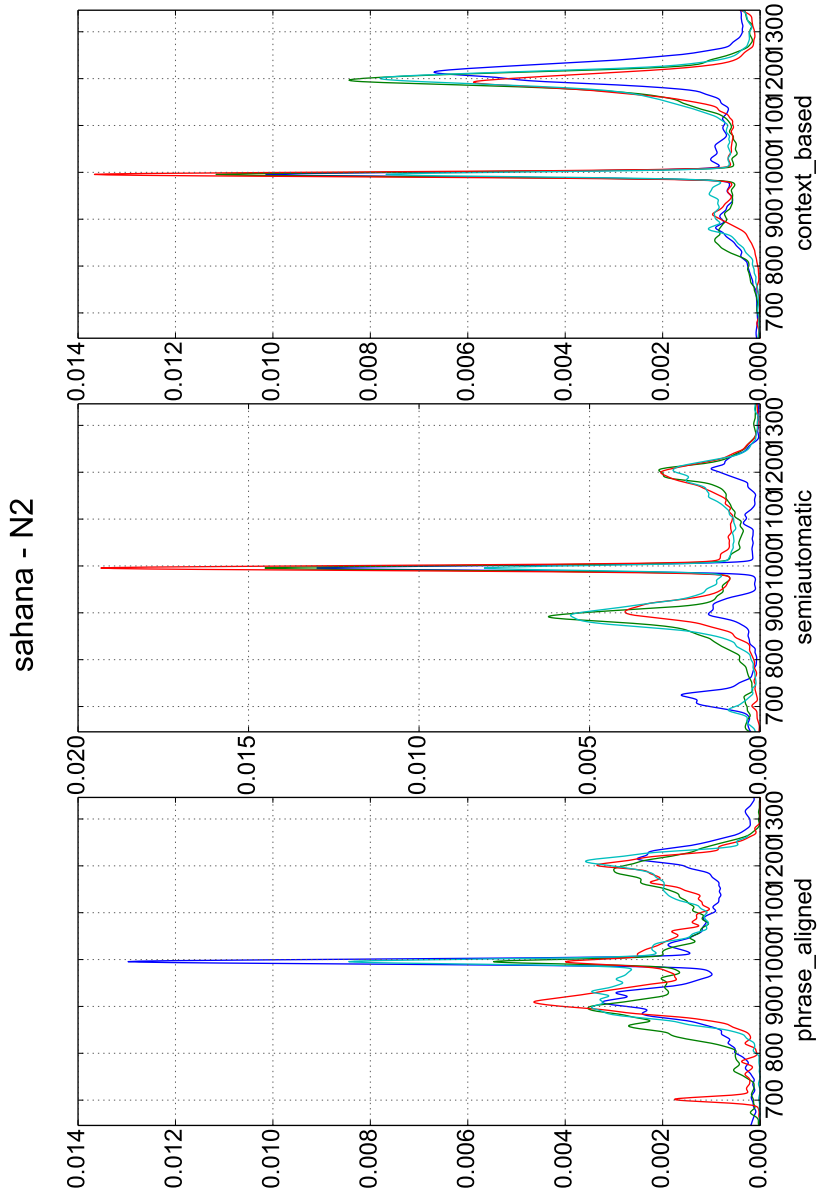


Figure 8.8: Representation for N2 svara in Sahana raaga computed using the three approaches.

fact that the alignment method used relies on the average tempo of the recordings computed from the annotations of the varnam dataset. In order to make the system more self-reliant, there needs to be an initial tempo estimation step similar to Holzapfel et al. (2015).

We have compared the svara representations computed from different approaches qualitatively and quantitatively. It is very encouraging to find that the context-based pitch distributions of svaras which just use the svara information of the raagas perform more or less similar to the approach presented in this chapter which uses scores. Though the lack of diversity in compositions in the varnam dataset did not offer conclusive insights, the results over the kriti dataset show that both the approaches are fairly robust to the variability of svaras across compositions in a given raaga. The varnam dataset helped us in qualitatively analysing the svara representations from the two approaches and comparing them to those obtained using manual annotations. The similarity of the representations and the correspondences therein reinforce that both approaches have been successful in getting close to the ideal representation of svaras.

In the next part, we build ontologies to publish musically meaningful information that can be extracted from this representation. For instance, these can include answering questions such as: i) Is the svara sung steadily? ii) What svaras are part of the movement sung over a given svara? and so on. We discuss ways to incorporate such information in a knowledge-base developing ontology patterns capable of capturing the semantics involved.

The experimental setup including code and data (or links to thereof) are made publicly accessible in a github repository[1].

[1] https://github.com/gopalkoduri/score-aligned-intonation

Part III

A multimodal knowledge-base of Carnatic music

Chapter 9

Ontologies for Indian art music

As part of the CompMusic project (Serra (2011)), we analyze melodic and rhythmic aspects of different music traditions taking advantage of the domain knowledge obtained from musicological texts and experts. This analyses yields musically meaningful information that can be directly related to higher-level musical concepts. Examples of this information include tonic of a given performance (Gulati (2012)), melodic motifs specific to raagas (Gulati et al. (2016a)), intonation characteristics of svaras in different raagas (Part. II), rhythmic events (Srinivasamurthy et al. (2016)) and so on. Further, we have put a major effort in curating an audio collection with well structured metadata that is well-linked within and also to other data sources such as Wikipedia. Our goal is to turn all the aforementioned information coming from different processes and sources into a logically consistent knowledge-base that is publicly accessible.

In chapter 4, we have discussed the aspects of Semantic web domain that help us to realize this objective. The first step in this direction is building ontologies. We discussed how they can help bridge the gap between music information research and disciplines like musicology and music theory. However, the current state of the ontologies for music domain is still limited in its scope to be able to address these issues. There is a need to develop ontologies for concepts and relationships coming from music theory. In the current chapter, we present ontologies which we have developed for Carnatic music addressing these limitations as a first step towards building a multimodal knowledge base. We present what is the first ontology for raaga to the best of our knowledge, and discuss the challenges posed by modeling semantics of its substructures: svara, gamaka and phrases. We also present the top-level ontology developed for Carnatic music, reusing several existing ontologies and defining new vocabularies as necessary.

9.1 Scope of our contributions

An ontology of any concept depends on its purpose and the perspective of knowledge engineer/ontologist, concerning political, cultural, social and philosophical aspects (Bohlman (1999)). An ontology for rāga can stem from different points of view - it's historical evolution, teaching methodologies, role in the society, taxonomies etc. Our choice of design is influenced primarily by its intended use in melodic analysis and data navigation. Therefore, our intent in developing ontologies for Carnatic music concepts is to eventually pave way for a multimodal knowledge-base. We limit ourselves to only those aspects that are essential to this objective. Therefore, it is imperative that these ontologies are not comprehensive of Carnatic music domain in its entirety.

The concept of raaga has evolved through millenia: few aspects like association with time have become less relevant, while a few others like arohana and avarohana have assumed more relevance as the concept is theorized in the last few centuries. The raaga ontology presented in sec. 9.2 represents it as understood in the contemporary context. A complete record of its evolution is beyond the scope of this work. Further, within the contemporary theory of the raaga, we chose to represent those aspects that are deemed relevant for contemporary practice by musicians and musicologists.

The Carnatic music ontology presented in sec. 9.3 builds on top of several ontologies including the raaga ontology and the music ontology (Raimond et al. (2007)). We constrain the scope of this ontology to the intended applications of the resulting knowledge-base[1].. Primarily, it is developed to support systems that facilitate navigation and browsing of music collections, such as Dunya[2] and Saraga[3]. This would require interlinking and computing similarity between diverse music entities. This requires representing a number of music concepts. In the first version of the Carnatic music ontology we present, the concepts included are as follows: raaga, taala, forms, concert, recording, musician, venue and instrument. We reuse several existing ontologies in due course.

[1]The said constraints only help us to limit the scope of the development of ontologies and consequently the knowledge-bases. However, it is to be noted that we have not traded off the completeness in representing a musical concept that we address.

[2]http://dunya.compmusic.upf.edu

[3]https://play.google.com/store/apps/details?id=com.musicmuni.saraga

9.2 Raaga ontology

From our discussion in ch. 2, recall that the fundamental substructures of raaga framework are: svara, gamaka and melodic phrases. In this section, we further elaborate their properties while building ontology patterns which help us in representing those properties and their semantics. Particular emphasis is given to svaras, for they are both central to our work presented in Part. II and to various theories of raaga classification described in musicological texts Ramanathan (2004).

Svaras and their functions

We continue our discussion about svaras from where we left in sec. 2.2. Recall that we mentioned that there are 4 svaras that share their positions with 4 others. This leads to two classes of svaras: natural and vivadi. The latter literally means excluded. From table. 2.1, all except G1, R3, N1 and D3 are considered natural svaras, while those four are classified as vivadi. We begin building our svara ontology using this information[4]. Fig. 9.1 shows part of this ontology. As it evolves, we will explain how the semantics represented in the class hierarchy and relationships come into play.

As aptly put by Viswanathan and Allen (2004), just like various checkers in the game of chess, svaras in rāga have different functions. Certain svaras are said to be more important than the rest. These svaras bring out the mood of the rāga. They are called the *jīva* svaras. The svara which occurs at the beginning of melodic phrases is referred to as *graha* svara. And likewise, *nyāsa* svaras are those svaras which appear at the end of melodic phrases. *Dīrgha* svaras are svaras that are prolonged. A svara that occurs relatively frequently is called *aṁsa* svara, and that which is sparingly used is called *alpa* svara, and so on. Therefore, a given svara in two different raagas can play very different roles. Further, as we learned from our work in previous chapters, svaras also assume their identity from the nature of melodic movements allowed in a raaga. In addition to these roles, we add two more arising from our analyses of svara intonation, and are of relevance to music similarity. These are steady svaras and gamakita svaras. The former refers to svaras sung with gamakas, and the latter without. Fig. 9.2 shows the skeletal definition of raaga class with these relationships.

Svaras also play a major role in several classification schemes of raagas. Certain schemes are said to be outcomes of theorizing the raaga framework, while some of

[4]We present our ontology visually throughout this thesis, unless using syntactic form helps comprehension. When it is required to do so, we use Manchester OWL syntax.

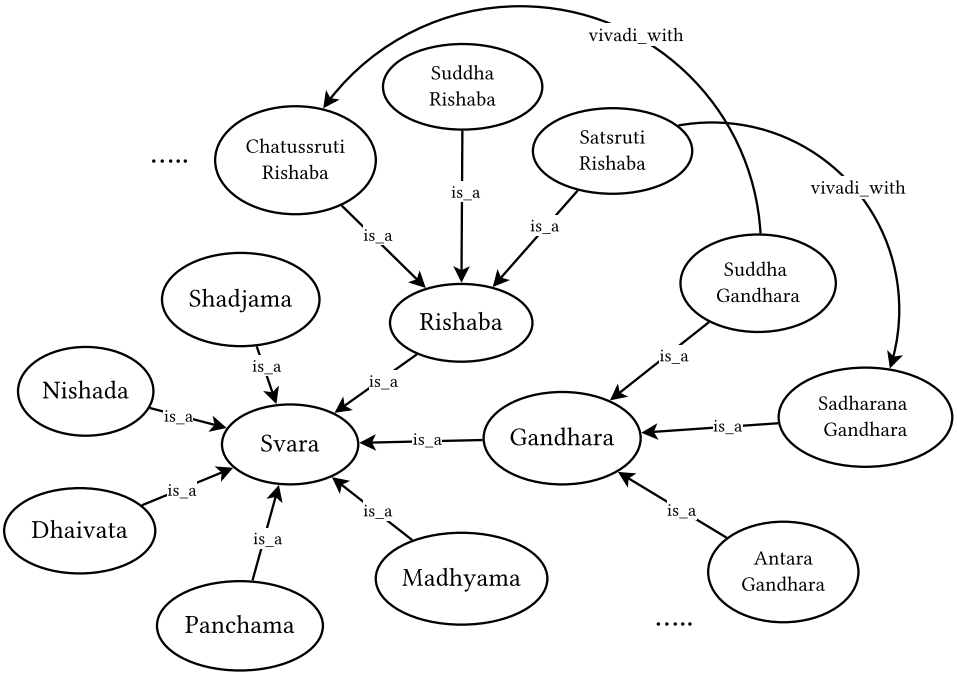


Figure 9.1: A part of the svara ontology showing all the svaras, variants of a couple of svaras, and the relationships between them.

them are a direct consequence of properties of raagas. We begin with introducing a popular framework that organizes raagas based on the constituent svaras. It is called the mēḷakarta system (Sambamoorthy (1998)). According to this system, there are 72 rāgas which are obtained through combinations of the 12 svarastānas with following conditions in place: only one variant of a svara is allowed in a combination, all the svara classes are to be represented, two given combinations differ by at least one svarastāna, and the svaras should be linearly ordered with no vakra[5] pattern. The rāgas thus obtained are called janaka rāgas, literally the parent rāgas. These are further divided into 12 subgroups each with 6 rāgas, based on the svarastānas chosen for five positions of R, G, M, D, N (S, P are invariably present in all parent raagas). The others, called janya/child rāgas are, in theory, derived from them.

However, several janya rāgas pre-date the janaka rāgas by centuries clearly indicating that the mēḷakarta system serves mainly the academic and theoretical pur-

[5]Vakra in Sanskrit literally means twisted. In this context, it means the order of the svaras is twisted/abnormal.

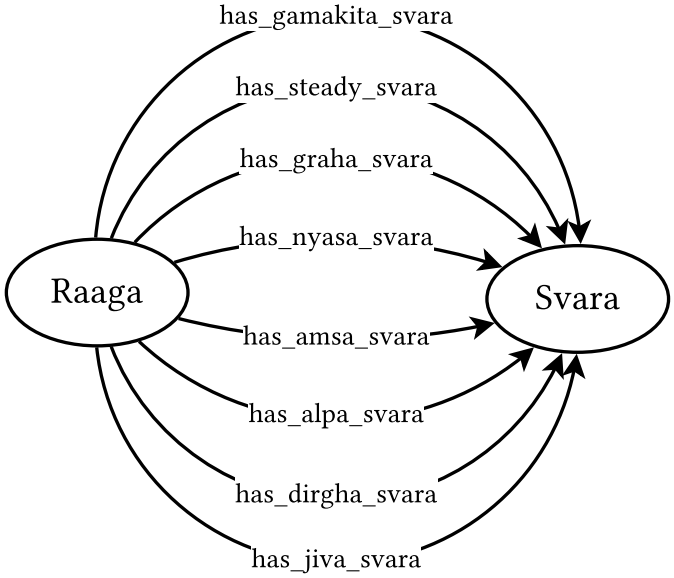


Figure 9.2: A part of the svara ontology showing all the svaras, variants of a couple of svaras, and the relationships between them.

pose of organizing rāgas. This resulted in another classification scheme that divides raagas as scale-based/progression-based (i.e., artificially created scale structure without the defining aspects of a traditional raaga) and phraseology-based raagas (Krishna and Ishwar (2012)).

Progressions

A rāga is typically represented using the ascending (ārōhaṇa) and descending (avarōhaṇa) progressions of the constituent svaras. Order and proximity of the svaras determine their usage in building melodic phrases. The svaras in ascending progression can only be used in melodic phrases which are ascending in nature, and viceversa. This seems to be especially important if the rāga has either differing sets of svaras in the progressions (Eg: Bhairavi rāga in Karṇāṭaka), or there is a vakra pattern in any of them (Eg: Saurāṣṭraṁ rāga in Karṇāṭaka). In the first case, it is imperative that the differing svaras are either used only during ascents or descents. In the latter case, the twisted svara positions allow few transitions which otherwise would not be possible. However, it has been noted that these progressions may not be as relevant to the identity of phraseology-based rāgas (Krishna and Ishwar (2012)).

Based on the characteristics of these progressions and the constituent svaras, there are several other classification schemes of rāgas. We mention a few of them here and refer the readers to extensive musicological texts for more complete list of classifications: Shankar (1983); Bhagyalekshmy (1990). The vakra/krama classification of rāgas is based on the order of svaras in the progressions of the rāga. If there is a deviation or a zig-zag pattern in either of these progressions, the rāga is said to be a vakra rāga. Otherwise, it is a krama rāga. The varjya/sampūrṇa classification of progressions is based on the number of unique svaras they consist of. The sampūrṇa progression have all the 7 svaras. Among the varjya progressions, there can be one, two or three omissions from among the seven svara positions. Based on the number of such omissions, the progression is said to be ṣāḍava, auḍava and svarāntara progression. Different combinations of such ascending and descending progressions give rise to different classes of rāgas.

We now define the Progression class in our ontology. Progression is a melodic sequence which is an ordered list of svaras. OWL in itself does not facilitate modeling order in any data as of yet (Drummond et al. (2006)). To the best of our knowledge, there is no design pattern that represents the order information in a way reasoners can work with. The available alternatives that allow representing sequences are: RDF containers (rdf:Seq in particular), OrderedList ontology[6] and OWL-List ontology (Drummond et al. (2006)). RDF containers do not come with formal semantics which a Description Logic reasoner can make use of[7]. OWL-List ontology is neither maintained nor available. OrderedList ontology makes use of index values of elements to maintain the order in a sequence, which limits the reasoning capabilities over the ontology, but can be used in conjunction with appropriate rules for deducing inferences conditional to a given set of criteria. That said, the rest of the relations over the Progression class in our ontology are shown in fig. 9.3.

We now turn to different kinds of progressions, which are all subclasses of Progression class in hierarchy. Each kind of progression asserts certain criteria on svaras. For instance, all Shadava progressions must contain exactly 5 unique svaras. Some of these criteria can be represented using OWL constructs, and a few others need rules to work with. Listings. 9.2 and 9.2 show the definitions of Sampoorna and Shadava progression classes using cardinality constraints on properties. Audava, Svarantara progressions are also defined likewise. Note that the svaras listed as S, R1 ... D3, N3 are all instances of respective svara classes. That is, R1 is an instance of Suddha Rishaba class and so on. The definition for

[6] http://purl.org/ontology/olo/core#

[7] http://www.w3.org/TR/rdf11-mt/#rdf-containers

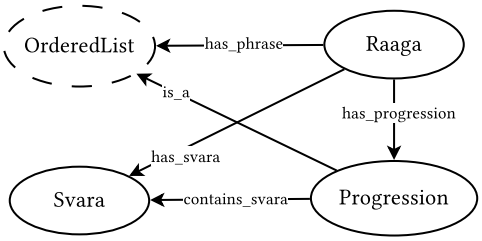


Figure 9.3: Part of the raaga ontology showing Progression class and its relations to other classes.

Sampoorna progression says that it equals to a progression that has at least one variant of each svara. The definition for Shadava progression says that it is a progression that has exactly 6 containsSvara relations with svaras among those listed.

Listing 9.1: Class definition of Sampoorna progression.

```
Progression
and (containsSvara some ({S}))
and (containsSvara some ({R1 , R2 , R3}))
and (containsSvara some ({G1 , G2 , G3}))
and (containsSvara some ({M1 , M2}))
and (containsSvara some ({P}))
and (containsSvara some ({D1 , D2 , D3}))
and (containsSvara some ({N1 , N2 , N3}))
```

Listing 9.2: Class definition of Shadava progression.

```
Progression
and (containsSvara exactly 6 ({S, R1, R2, R3, G1, G2, G3, M1, M2, P
    ,
                  D1, D2, D3, N1, N2, N3}))
```

Phrases and gamakas

Phrases, like progressions, are sequences of svaras. This notion helps in representing scores/notations, however the melodic contours cannot be fit in this definition for reasons we have presented throughout Part. II. Unless the melodic contours are represented using a data model for further abstraction, we do not see any particular advantage in representing them as they are (i.e., a sequence of pitch values) in an ontology. Having said that, it is beneficial to instead link them to their symbolic counterparts which find place in the ontology. Therefore, the

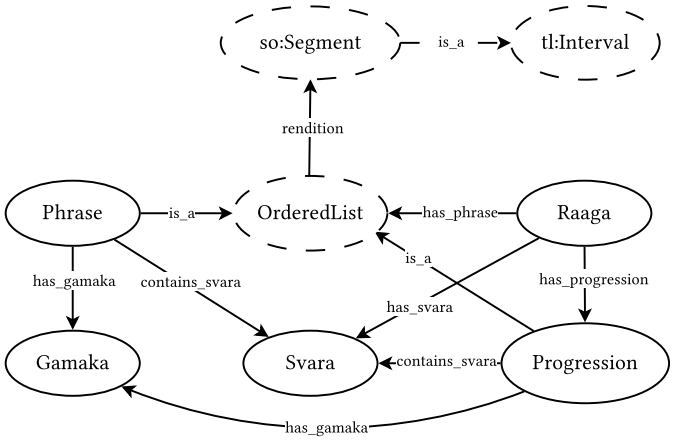


Figure 9.4: Overview of our ontology showing Phrase and Gamaka classes with their relationships.

Phrase class is defined as a subclass of OrderedList, which further can be linked to Segment class in Segment ontology (Fields and Page (2011)). The latter is a subclass of Interval class in Timeline ontology used to express relative/absolute positions and durations over a timeline.

On the other hand, gamakas are not defined as a sequences of svaras. They are melodic movements, whose shape can vary depending on the context, given the rāga and the svaras they are sung with. As of yet, marking gamaka boundaries either in an audio sample or a pitch contour is a difficult task as they are coalesced in the pitch continuum. Hence, one can only assert if the gamaka is used in a certain segment, but not demarcate it precisely. For this reason, we chose to model the gamakas as abstract sequences in which we do not specify the exact nature of gamaka in a direct manner. It is made a part of a longer sequence without an explicit indication of its beginning and ending within the sequence. Further, like in the case of phrases, each sequence is linked to a segment of an audio recording which corresponds to its rendition. Fig. 9.4 shows the corresponding extensions to include phrase and gamaka concepts in our ontology.

Further, there are various ways to classify gamakas. The most accepted classification schemes speak of 10 or 15 different types (Dikshitar et al. (1904); Janakiraman (2008)). However, Krishna and Ishwar (2012) clarifies that only a subset of them are used in contemporary practice and lists them plotting their melodic contours. Our ontology lists the latter as possible individual instances of Gamaka class. Krishna and Ishwar (2012) lists them and discusses each type elaborately.

Data models

One of the main sources of information in our knowledge-base is analyses of audio music recordings. Each analysis results in a compact representation of a music concept conceived as or extracted using a data model. In our analysis of svara intonation, this data model is a normalized pitch histogram that represents a given svara as a probabilistic phenomenon. The information resulting from such models needs to be linked to concerned music concepts in our ontology.

For this, we put forward an ontology that is subsumed by raaga ontology, to facilitate expressing and linking this information. It consists of the Data Model class. This class contains further subclasses each further differing and narrow in their definition as required. Right now though, it has just one subclass called Pitch Distribution. Note that the Data Model class is a general class that is intended to model anything. Therefore, the relation is_about is not restricted to be defined with any specific class. In our case though, this relation links it with the Svara_Manifestation class, which links it with appropriate Svara class and the audio recording from which it is extracted from, using manifest_svara and manifest_in relations defined on it.

Including provenance of such information becomes important for various reasons: i) enhances trust in applications using it, as the origin of the model is made transparent, and ii) most importantly, it also facilitates a systematic and qualitative comparison of different models about the same concept. Therefore, we define a property on Data Model class that links it to a relevant resource. The latter corresponds to frbr:Work[8]. Fig. 9.5 shows the corresponding new classes and their relations linked with the raaga ontology.

Rules

As fig. 9.4 shows, the rāga ontology subsumes sequence and svara ontologies to effect various classification schemes. Within the limits of expressiveness in OWL 2 DL, we have represented a few within the ontology (eg: listings. 9.2 and 9.2). We have earlier mentioned that a few other classification schemes need rules outside the ontology to be effected. This is both in part due to the limitations of OWL 2 DL, and also due to limited support for property chaining among the reasoners. Relations defined using property chain are those where, if a given series of relations exist between class A and class B through other classes, then a new relation is brought into effect between A and B. Such properties are often referred to as 'complex relations', and have scant support in inference procedures.

[8] http://vocab.org/frbr/core

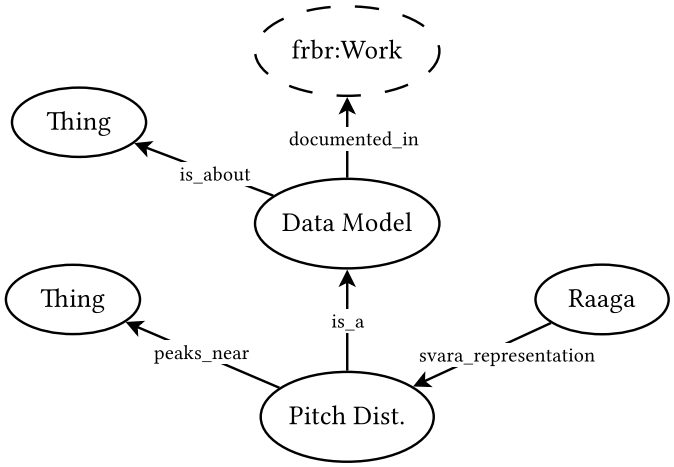


Figure 9.5: Data Model extension to our ontology to express information extracted from audio analyses.

Therefore, a viable alternative is defining rules outside the ontology. These set of rules lead to the necessary classification of raagas from which the resulting facts can be added back to the ontology or knowledge-base. There is more than one way to do this (see ch. 4): Rule Interchange Format (RIF), SPARQL Inferencing Notation (SPIN/SPARQL rules) and Semantic Web Rule Language (SWRL). As the kind of rules one requires are often application and task specific, we do not put efforts into building a comprehensive list of rules. Here, we only demonstrate them using SPARQL rules[9] to perform krama/vakra classification.

As mentioned in sec. 9.2, each rāga has two progressions - ascending and descending. A descent in an ascending progression, or an ascent in the descending progression makes the rāga a vakra rāga. Otherwise, it is classified as a krama rāga. In order to do this using our svara and melodic sequence ontologies, in each given descending progression, we check if a *ol:isFollowedBy* relationship exists between any two consequent svaras among *Shadja, Rishaba, Gandhara, Madhyama, Panchama, Dhaivata* and *Nishada*, in that order. Such relationship would mark an ascent pattern in descending progressions. This order of svaras is reversed and the operation is repeated for checking descent patterns in ascending progressions. Listing 9.2 shows an example SPARQL rule for the classification scheme.

[9]The rule examples are shown using SPARQL query syntax

Listing 9.3: One of the several SPARQL rules used in the classification of rāgas into vakra/krama rāgas.

```
CONSTRUCT {?raaga a :vakraRaaga}
WHERE {
                ?raaga :hasArohana ?progression.
                ?progression olo:has_slot ?slot1.
                ?progression olo:has_slot ?slot2.
                ?slot1 olo:has_item :Gandhara.
                ?slot2 olo:has_item :Rishaba.
                ?slot1 olo:has_index ?index1.
                ?slot2 olo:has_index ?index2.

                FILTER (?index1 > ?index2).
        }
```

Rules can achieve a lot more than just classification. Take the case of constraints over a svara with regards to usage of gamakas. For instance, if the given rāga has R1 and G1 svaras, R1 can not take a gamaka since G1 is very close and it is difficult to sing a gamaka on R1 without touching G2, which would result in violation of rāga's properties[10]. Such conditions can be used in creating new facts on svaras, and even in cross-verifying the svara representations whether the observations from them fall in line with such deduced facts. A similar inference is possible from observing the peak positions of Pitch Distribution class. If it has multiple peaks, it clearly indicates that the svara is sung with gamakas, else it can be classified as a steady svara (see fig. 9.2 to understand how these inferences can be added to knowledge-base).

Summary

We discussed different aspects of the rāga framework and developed ontology patterns to capture their semantics. This version of the ontology clearly laid more emphasis on the role of svaras - ranging from their functions in the raaga to their importance in various classification schemes.

We demonstrated with examples that the description of a musical aspect obtained from the analysis of the audio recordings, interpreted with the help of the ontology, gives insights which otherwise are not explicit or obvious. For instance, the way a particular svara is intoned (say, the nature of semi-tonal oscillation) can

[10] A Karṇāṭaka musician and trainer explains this taking an example from Karṇāṭaka music in this podcast episode: http://raagarasika.podbean.com/2008/09/30/episode-15-featured-raaga-sivaranjani/

be identified by combining the intonation description (obtained from audio data analysis) and the nature of usage of svara in the rāga (obtained from ontology).

Modeling sequences, more specifically gamakas, is not trivial given the limitations on expressiveness of OWL-DL[11] and also due to the variety of temporal variations possible for a given gamaka based on the context. Despite that, the way we currently represent gamaka in the rāga ontology possibly gives way to a symbiotic loop with motif analysis of audio recordings. For a given gamaka, melodic sequences similar to the ones which have the gamaka, obtained using motif analysis, can be used to enhance/reinforce the prevailing representation of the gamaka in the knowledge-base. In turn, the broad variety of sequences thus pooled in for a gamaka in the ontology can potentially help is zeroing in on more concrete form of a gamaka. This further can guide a supervised system to identify more musically meaningful melodic sequences from audio recordings.

As mentioned in sec. 9.1, the scope of this ontology is limited by our goals. However, one can further extend this to overcome such limitations depending on their application needs. These include but not limited to: differences in the definition of rāga based on school/lineage and historical periods, other theory-driven raaga classifications, therapeutic aspects of raaga and so on. In the rest of the chapter, we develop ontologies for music concepts like taala (rhythmic framework), forms and instruments, eventually bringing them together along with raaga ontology resulting in the Carnatic music ontology.

9.3 The Carnatic music ontology

The Carnatic music ontology brings together several concepts in describing their hierarchies and relationships which are relevant for use in navigation and browsing systems in the semantic web context. The core set of sub-ontologies presented in the version discussed here include raaga, taala, forms, performer and work. Besides the raaga ontology, the bulk of Carnatic music ontology is made of taala and forms ontology extensions. For the rest, we depend on existing ontologies and vocabularies, which mainly consist of the music ontology, event ontology, and schema.org vocabularies.

Taala ontology

Taala is the rhythmic framework in Carnatic music. We have introduced the related concepts in sec. 2.2. Following that discussion, our taala ontology comprises

[11] Constructors used for the rāga ontology come from $\mathcal{SROIQ}$ variety of Description Logics.

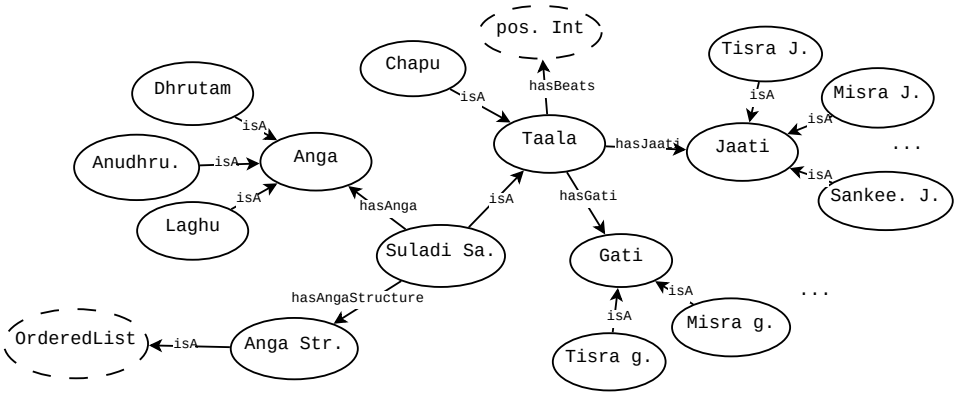


Figure 9.6: Classes and relationships in the Taala ontology.

of the following top-level classes: Taala, Jaati, Gati, Anga and Anga Structure. There are two classes of taalas that are in-vogue in Carnatic art music: Suladi sapta taalas and Chapu taalas. The Suladi Sapta taalas is a collection of seven kinds of anga structures. Where as chapu taalas are represented using the number of beats per cycle without reference to anga structure. As we learned earlier, Jaati and Gati are modifiers to the Suladi sapta taala class which determine the number of beats and its speed respectively. Each of them have five variants. The three Angas: Dhrutam, Anudhrutam and Laghu are listed as subclasses of Anga class. Taala is related to this class by has_anga property. However, the actual sequence of these angas is important in the definition of a taala. Therefore, we define another class which is a subclass of OrderedList, called Anga Structure to facilitate this. Fig. 9.6 shows these classes and their relationships.

Carnatic Forms ontology

We have discussed different types of melodic forms in Carnatic music, and a few of their classification schemes (sec. 2.2). We consolidate those schemes to the following pairs of classes in our ontology: Pure and Applied, Improvisatory and Compositional, Abhyasa gana and Sabha gana, Lyrical and Syllabic, and Freeflow and Rhythmic. In Lyrical and Syllabic classification, the latter refers to those forms which are sung with individual syllables - non-sensical (Like in alapana, where vowels and a few consonants are used) or otherwise (Like in swara kalpana where syllables correspond to Carnatic solfege). Each melodic form can simultaneously belong to more than a class. Fig. 9.7 shows a few of these classes with a couple of forms appropriately classified. For comprehensibility, we have omitted other forms, which can be found in the ontology.

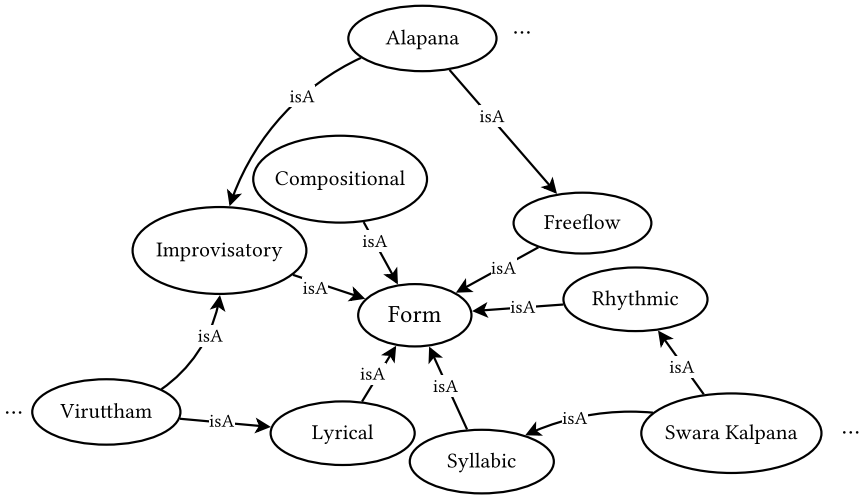


Figure 9.7: Classes and relationships in the Taala ontology.

Performer ontology

We found the existing ontologies to be limited in expressing a seemingly general aspect of music concerts and/or albums, which is to express the role of different artists in the ensemble. It requires expressing that in a given concert, an artist X played an instrument Y in a role Z. The Performance class in the music ontology comes close to expressing this. However, it defines instrument, performer properties directly on the Performance class, which results in a loss of information that conveys who played what instrument. Further, it does not have role information. The performer relation defined in schema.org vocabulary is too generic and falls short of expressing the required information.

We designed an ontology pattern that would facilitate expressing such relations. Fig. 9.8 shows the Performer ontology. The Performer class links to Artist, Instrument and Role classes in defining a given instance of a performer. This pattern enables expressing any possible combination of artist, instrument and role as a performer, which can be linked to an mo:Event such as a co:Concert (defined soon), or a mo:MusicalManifestation such as a mo:Release or a mo:Record. Note that the Performer and Role classes can easily be extended to meet other possible requirements.

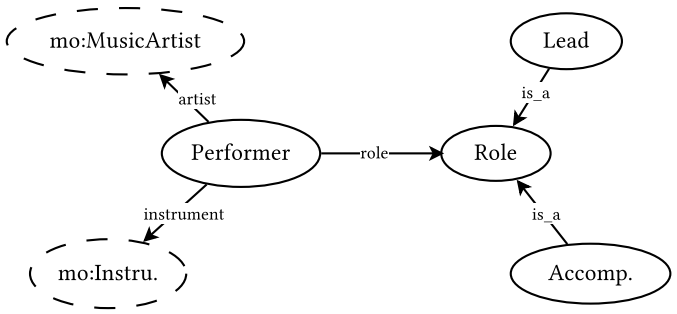


Figure 9.8: Classes and relationships in the Taala ontology.

Concepts imported from other ontologies

Rest of the concepts required to express the knowledge and facts in Carnatic music domain, relevant to our intended applications, are modeled after and linked to those from the existing ontologies. We further define additional relations over them as required. Following are some of the concepts reused from the music ontology: Instrument (defaults to a SKOS representation of MusicBrainz instrument tree), Record, Release, Lyrics, Score, Performance and MusicArtist. Their usage and hierarchy in our ontology is logically consistent with their definitions as given in the music ontology. However, we create new relations among them in order to express the knowledge is Carnatic music domain as succinctly as possible. The Place concept is reused from the schema.org vocabulary to represent concert venue and other places such as artist birth and death place.

Concert becomes the central unit of all musical activity in Carnatic music. So much so, that most commercial releases are basically recordings of concerts. Consequently, our ontology reflects this leaning towards Concert class. It is modeled as a subclass of Event class in the event ontology. It contains a set of Performances, which are modeled as Event-s too. They are essentially renditions in one creative Work or the other. This Work can correspond to any kind of Form, and has Raaga and Taala defined as applicable. If it corresponds to a Compositional Form, the Work is linked to appropriate Notation and mo:Lyrics as well. Each such Performance of a Work has a set of Performer-s indicating which artist played what instrument. Fig. 9.9 shows these concepts and the relations defined among them. Note that we did not reuse mo:MusicalWork in the place of Work class. There are significant differences in the way a creative Work is used and interpreted in western music and Indian art music. For instance, the composition of a work involves creating an indicative notation which is not intended to be performed as is. The notation is used only as a memory aid, unlike a score in western classical/popular

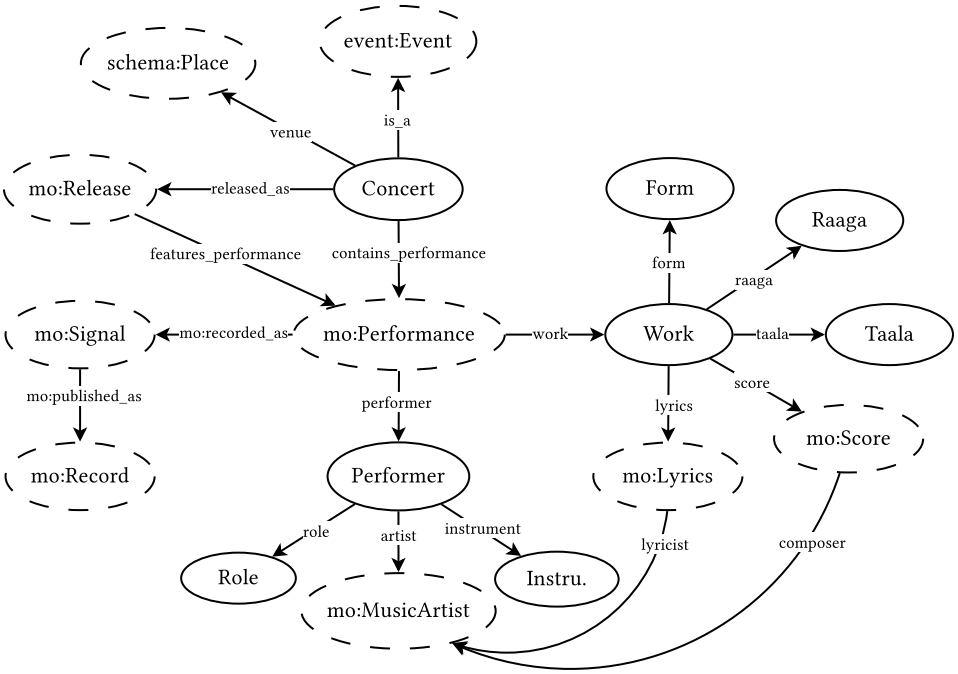


Figure 9.9: The Carnatic music ontology that subsumes Raaga, Taala, Form and Performer ontologies to describe aspects of Carnatic music.

music, where it is intended to be faithfully reproduced.

9.4 Summary & conclusions

In this chapter, we presented the raaga ontology and the Carnatic music ontology which subsumes it besides other extensions that include form, taala and performer. The raaga ontology models the concepts of svara and phrase in Carnatic music to further facilitate various classification schemes that depend of properties of these substructures. OWL 2 DL facilitates expressing a few of these properties in the definitions of the classes, such as Sampoorna Progression. However, a few others such as Vakra Progression had to be defined outside the ontology in one of the rule languages. This is partly due to the limitations of expressiveness of OWL 2 DL, and partly due to the lack of support in reasoners for complex properties such as the ones defined by chaining multiple other properties.

We reuse several concepts and relations from the existing ontologies such as the music ontology in developing the Carnatic music ontology. However, the bulk

of this ontology comes from the definitions of these concepts: raaga, taala, form and performer. The Taala and Form ontology extensions define a hierarchy of concepts and their relations, both coming from music theory (E.g: Improvisatory vs Compositional forms) and an application perspective (E.g: Lyrical vs Syllabic forms). We also define a Work class that is different from mo:MusicalWork owing to the differences in their usage and interpretation in their respective communities. Taking note of the limitations of current ontologies in expressing the association between role (lead, accompanying etc), instrument and artist (person, band or computer) in a performance and/or a recording, we define the Performer class to overcome the same.

The landscape of semantic web technologies is a rapidly changing one. Though on one hand, we had to continually update the ontologies to take advantage of these changes, we believe the advantages offered by explicit semantics far outweigh the efforts needed to maintain the ontologies. Often, these changes bring in improved expressiveness of OWL DL ontology language or better support for OWL DL semantics in the reasoners. However, owing to this, backward compatibility becomes a challenge which to an extent can be handled by enforcing proper versioning of the ontologies. In the following chapters, we make use of these ontologies in providing for a groundtruth of concepts and relations for information extraction from text (ch. 10), and structuring and integrating information from multiple sources (ch. 12).

The ontologies developed as part of this thesis, and the CompMusic project in general, are available as a github repository[12]. The same will be the main web reference for the persistent URLs and the documentation of the ontologies.

[12]https://github.com/gopalkoduri/ontologies

Chapter 10

Concept and relation extraction from unstructured text

In the past decade, domain-independent approaches to information extraction have paved way for its web-scale applications. Adapting them further to acquire knowledge from thematic domains can greatly reduce the need for manual knowledge engineering. This requires understanding how amenable the assertions extracted by such approaches are to ontologization. To this extent, we propose a framework for a comparative extrinsic evaluation of the open information extraction systems. The first part of the framework compares the volume of assertions along different dimensions with an aim to understand their coverage of the domain quantitatively. In the second part, the assertions are evaluated qualitatively by employing them in three of the fundamental tasks of ontologization: entity identification, concept identification and semantic relation extraction. The results from each task are validated against structured content in Wikipedia and/or are manually checked as necessary. The results from the two parts of the framework, when juxtaposed against each other, give us concrete insights into the differences between the performances and the nature of the approaches.

The advent of the semantic web and the linked open data movements have not only resulted in a growing number of community-built structured data sources like Wikidata and DBpedia, but also catalyzed the development of domain-independent approaches for extracting information from unstructured text, further enriching them. Open information extraction (OIE) is one such paradigm that has emerged in the past decade, and has been used to extract assertions from unstructured data

at web-scale with a considerable success (Etzioni and Banko (2008)). Until recently, domain-specific approaches to information extraction from text required manual knowledge engineering as a prerequisite (Sarawagi (2008)). The OIE approaches, however, do not require a pre-specified vocabulary and/or relation-specific input. Therefore, adapting them to information extraction from thematic domains, like Carnatic music, would alleviate the need or manual knowledge engineering.

The process of structuring the assertions extracted from these approaches poses certain challenges. There has been little work so far to identify and address such issues. The advances in OIE, including the recent systems such as NELL[1], are largely directed towards taking advantage of the volume of web data. This meant such systems rely to a good measure on repetitions in the data. In doing so, the recall of the systems is often traded off for a good precision. The adaptation of such systems to acquire knowledge from a given domain is an exciting prospective direction. Soderland et al. (2010) have first attempted to adapt TextRunner system (Etzioni and Banko (2008)) to populate the knowledge-base concerning facts in the domain of football game. One of their findings is that the limited recall of the OIE systems is the major bottleneck in acquiring a good coverage of the relation types.

This chapter is organized as follows. In sec. 10.1, the OIE approaches that we chose to compare are discussed and in sec. 10.2, an overview of the data we work with is presented. In sec. 10.3, we present the framework with various quantitative and qualitative measures for analyzing the quality of assertions extracted, and in sec. 10.4, we demonstrate it on the music domain. Sec. 11.5 concludes the paper with our remarks and future direction of this work.

10.1 Open Information Extraction

Information extraction is the task of obtaining a set of assertions from the natural language text, featuring the entities and the relations of the corresponding domain. The approaches are diverse ranging from those which learn from the labeled training samples for the desired set of target relations, to those which operate in an unsupervised manner. An easy access to large volume of unstructured text on the web has necessitated approaches that scale appropriately to take advantage of this data. Open information extraction aims to extract the assertions from voluminous data without requiring a pre-specified vocabulary or labeled data for relations (Etzioni and Banko (2008)).

[1] http://rtw.ml.cmu.edu/rtw/

ReVerb & OpenIE 4.0 For demonstrating our evaluation framework, we choose two state-of-the-art OIE systems: ReVerb (Fader et al. (2011)) and OpenIE 4.0 (Mausam et al. (2012)), which are shown to have outperformed the earlier systems such as TextRunner, woe^{pos} and woe^{parse} (Wu and Weld (2010)). ReVerb addresses the issue of incoherent and uninformative extractions[2] found with the former systems, by using few syntactic and lexical constraints. OLLIE (Mausam et al. (2012)) is a successor of ReVerb, and includes the noun-mediated relations which are not handled by the latter. It also incorporates the context of the assertions in the form n-ary relations. OpenIE 4.0 employs a similar methodology to that of OLLIE, to retrieve assertions using semantic role labeling, also known as shallow semantic parsing. The implementations for both ReVerb and OpenIE 4.0 are available online[3].

Semantic parsing On the other hand, deep semantic parsing is an active research topic in the natural language processing (NLP) community, which aims to obtain a complete logical form of a given sentence. It is used in applications such as question-answering systems, robotic navigation and further has several direct implications for OIE as it is domain-independent and is shown to be web-scalable (Harrington and Clark (2007)). To our knowledge, there is no existing literature that compares semantic parsing with the likes of ReVerb and OpenIE 4.0. We therefore built an information extraction wrapper around a state-of-the-art semantic parser and compare with the selected OIE systems. What follows is a brief description of this system.

We use Combinatory Categorial Grammar (CCG) (Steedman (2000)) as our grammatical framework to parse natural language sentences to logical representation. CCG is known for its transparency between syntax and semantics, i.e. given the syntactic structure (CCG derivation) of a sentence, a semantic representation can be built deterministically from its derivation. Each word in a sentence is first assigned a CCG category based on its context. Each category represents the syntactic constraints that the word has to satisfy. For example, in Fig. 10.1, the word *plays* is assigned a syntactic category $(S\backslash NP)/NP$ implying that *plays* take a noun (NP) argument on its right, and a noun argument (NP) on its left to form a sentence (S). An equivalent semantic category in terms of a lambda function is constructed from the syntactic category, here $\lambda x.\lambda y.\text{plays}(subj,\ y) \wedge \text{plays}(obj,\ x)$ with *plays* representing the predicate, x and y representing the object (guitar) and subject (John) arguments. CCG defines a set of combinators using which the adjacent categories combine to form

[2]We have used the terms *assertions* and *extractions* analogously.
[3]Available at https://github.com/knowitall/

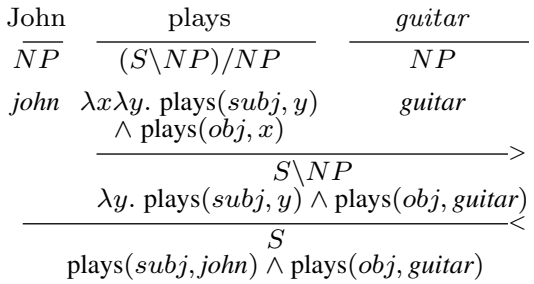


Figure 10.1: An example showing the CCG syntactic and semantic derivation of 'John plays guitar'.

syntactic categories of larger text units like phrases (*e.g:* plays guitar), from there on leading to parsing a whole sentence. Correspondingly, the lambda functions of the categories compose, eventually leading to the semantic representation of the sentence. The advantage with CCG is that the complexity of obtaining a logical representation of a sentence is simplified into the task of assigning categories to words. We use a modified version of Boxer (Bos et al. (2004)) to further convert our sentences of interest to the triple form (subject, relation phrase, object).

10.2 Data

A major challenge in developing technologies for the exploration and the navigation of music repertoires from around the world lies in obtaining and using their cultural context. The vocabulary used for describing and relating the entities (musical concepts, roles of people involved etc) differs to a great extent from music to music. Most commercial platforms have a limited view of such context, resulting in poor navigation and exploration systems that fail to address the cultural diversity of the world. Within the music information research community, there is a growing interest for developing culture-aware approaches to address this problem (Serra (2011)). Such approaches are diverse in terms of the data they work with (audio, metadata and contextual-data) and methodologies they employ (Serra et al. (2013)).

However, to our knowledge, there are no major attempts that use web text, arguably the largest openly available data source. As a first step in this direction, we choose to demonstrate our framework in the music domain. As we know from ch. 2, Indian art music traditions: Carnatic and Hindustani, have a very distinct character, especially when compared to the popular music styles that drive the

music market worldwide. The terminology and the structuring of the knowledge in these music traditions differs substantially from what people are accustomed to, on most commercial platforms (Krishna and Ishwar (2012)). Therefore, we believe they make a suitable yet challenging thematic domain to analyze the quality of the assertions for ontologization.

Text corpus

Our data consists of the plain text obtained from the Wikipedia pages corresponding to the Carnatic and Hindustani music traditions, after removing the tables, listings, figures, infoboxes and other structured content. Pages from the parent categories corresponding to both music traditions are obtained recursively. Text from each page is tokenized to sentences, which are further filtered using the following constraints: a minimum number of 3 words and a maximum of 21 words per sentence, with each word not exceeding 30 characters in length. By trial and error, these constraints are found to reduce the number of malformed and highly complex sentences. In the semantic parsing based system, to address the multiword named entities, we identify consecutive NNPs and merge them into one in a pre-processing step. In ReVerb and OpenIE4.0, they are handled in their respective implementations.

We observed that a majority of the sentences featured pronouns. The resulting assertions only partially contribute to ontologization. For instance, consider the sentence 'She is a composer'. The resulting assertion would be (She, is a, composer). A few such sentences might help us learn that there exists a concept called *composer*. However, such assertions are not useful in identifying entities of the corresponding concept. Therefore, the pronouns in the text from each page are resolved using the deterministic coreference resolution described by Lee et al. (2013)[4]. There were a few false assertions as a result. However, we observed a substantial rise in the recall of the entities in the domain. Table. 10.1 lists the total number of sentences, and the number of assertions extracted using the OIE systems. ReVerb and Open IE 4.0 associate a confidence score with the extracted assertions. We did not however choose to filter them based on this score, as Soderland et al. (2010) advocates that a system with a better recall at the cost of lower precision is actually preferred for knowledge-base population using open information extraction. All the assertions are converted to the triple form.

[4]Available online at http://nlp.stanford.edu/software/dcoref.shtml

Music	#Sentences	#ReVerb	#OpenIE 4.0	#Sem. Parsing
Carnatic	10284	9844	15013	19241
Hindustani	10724	9944	15777	18496

Table 10.1: The number of sentences for each music, and the number of extractions obtained from the OIE systems.

Gold standard

We use the ontologies manually engineered in ch. 9 as a reference to prepare groundtruth for the evaluation in the tasks of concept identification and semantic relation extraction[5]. The concepts and relation-types used in our evaluation framework come from the ontologies, with added synonyms and possible variations. The groundtruth for entities for each concept correspond to the page-titles in the respective subcategory in the Wikipedia (eg: Carnatic_musicians).

10.3 Evaluation framework

In the information extraction literature, different approaches are evaluated by measuring their performances on a set of labeled sentences or by employing human judges (Fader et al. (2011); Mausam et al. (2012)). Our goal, however, is to evaluate them by the usefulness of the assertions extracted. We quantify this using a series of tasks that help in understanding the coverage of the entities and the relation types of a given domain in the extracted assertions, quantitatively and qualitatively. The tasks discussed in the first part of the evaluation compare the volume of the assertions. While in the second part, we validate to what extent the assertions yield to be structured using our ontologies. We then juxtapose and compare the results from both parts of the evaluation.

Quantitative assessment

We study the distribution of the extracted assertions along four different aspects to gain an insight into their coverage of the domain with respect to each of them: **sentences**, **entities**, **relation types** and **concepts**. For the purpose of analyses discussed in this subsection, the subject of a triple is taken for an entity, and

[5] Available at https://github.com/gopalkoduri/ontologies

the object of a triple featuring a subsumption relation phrase[6] (e.g: is a, be etc..) is taken for a concept. For instance, in the triplet (*Tyagaraja*, *is a*, *composer*), *Tyagaraja* is taken for an entity, and as the triple has a subsumption relation type *is a*, *composer* is taken for a concept.

Observations from the distribution of the number of extractions for sentences give a crude perspective of the modularity of the information extraction approach, which is its ability to identify multiple, distinct relations from a given sentence. The distribution of extractions for the entities allows us to gain an overview of the scope of the extracted relations in identifying the entities in the given domain as well as in describing a given entity. The distribution of extractions for relation types allows us to understand the relevance and coverage of an identified relation-type in the domain.

As we will see, a large majority of the assertions from all the OIE systems correspond to the subsumption relation type, often outnumbering the other relation types by orders of magnitude. Therefore, it is important to further analyze this relation type. These relations mainly inform us about the concept membership of the entities. Hence, they assume importance for ontologization as they are resourceful in defining the taxonomy of the given domain. The distribution of extractions for concepts would reveal to what extent the assertions actually carry the required information.

Qualitative assessment

Though the number of relation-types, concepts and entities are representative of the size of the ontology learned and the volume by which it is populated, the numbers alone might be misleading as not all the extractions from a given system are unique and meaningful. The differences between relative performances in quantitative and qualitative analysis expose those systems which overgenerate wrong/redundant relation-types.

Consequently, the tasks discussed in this subsection are complementary to those presented in the former, validating whether the quantitative observations correlate with the performances of OIE systems on various tasks in ontologization. For this, we consider the three fundamental tasks of ontologization: entity identification, concept identification and semantic relation extraction (Petasis and Karkaletsis (2011)).

[6] A subsumption relation phrase is that which indicate a hierarchy in the taxonomy of the domain.

Concept identification. It is the task of identifying the concepts in the domain. Objects from all the triples featuring a subsumption relation phrase are collected. They are disambiguated based on their spellings, mostly automatically using string matching and edit distance measures[7] with minimal manual intervention where necessary. The resulting objects are taken to be the candidate concepts of the given domain. We compare the coverage of these against the concepts[8] in the manually built ontologies.

Entity identification. It concerns with finding the entities of a given domain and assigning each to a concept. The set of subjects from all the triples are considered as candidates entities in the domain. A list of titles of the Wikpedia pages in the domain along with the categories each page belong to, is acquired. The page titles correspond to entities, and the categories are manually mapped to concepts in our ontology. This constitutes the reference with which we compare the results from the two subtasks of entity identification.

For evaluating the first subtask of entity identification, i.e., finding the entities of the given domain, we measure the overlapping (O) and the residual (R) portions of the candidate entities from each system with respect to the reference set. If X is the set of candidate entities and Y is the reference set, O and R are defined as:

$$O(X,Y) = \frac{|X \cap Y|}{|Y|} \tag{10.1}$$
$$R(X,Y) = \frac{|X - Y|}{|X|}$$

The overlap and residual measures are preferred to other standard measures such as precision and recall as there are legitimate entities recognized in the assertions that are not part of the groundtruth. For technical correctness, we chose to evaluate using custom measures that convey the same information. The O measure is the same as recall measure provided that Y has all the possible true positives. But in our case, it is not so and X potentially will have candidates that are true positives but not in Y. Therefore, we avoid terming O a recall measure. On the other hand, R tells us about the proportion of those candidates obtained, which are not found in Y. We employ R with inter-system agreement to draw meaningful conclusions about their results.

The second subtask, concept assignment, is evaluated using two methods. In the first method, we manually build a set of rules over subsumption relation type for each concept. For instance, for an entity to belong to the concept *singers*, it

[7] Available at https://github.com/gopalkoduri/string-matching
[8] The term concept is used analogous to a class in ontologies.

must have either of the words *vocalist* or *singer* in the object of the corresponding triples with subsumption relation type. All the entities satisfying the rules for a given concept are assigned to it.

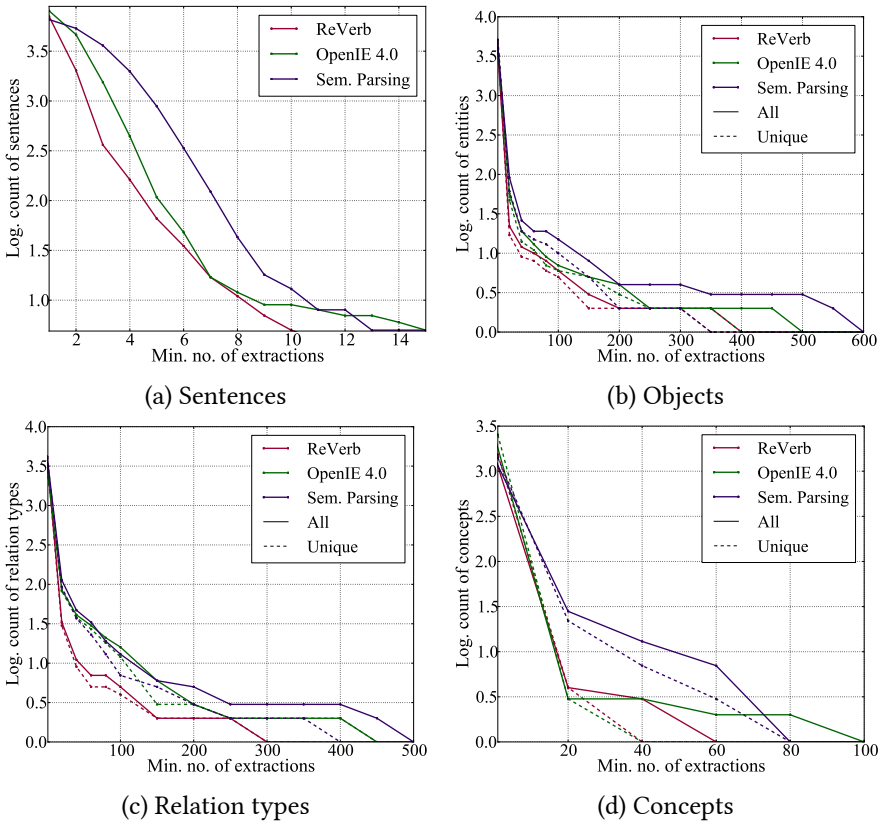
In the second method, a given entity is reduced to be represented by a term vector corresponding to the words from objects of the assertions it is part of. Following this, each concept is initiated with a seedset of entities belonging to it. A given concept is taken to be an abstract super entity, represented by the union of the term vectors of the constituting entities. A bootstrapping mechanism is started, which in a given iteration, finds the closest entity to the given concept and adds it to the seedset, and recomputes its representation. The distance between given two entities corresponds to the cosine similarity between the term vectors transformed using TF-IDF, followed by Latent Semantic Indexing (Řehůřek and Sojka (2010)). Unlike the first method, which is constrained to assertions with subsumption relation type, this method takes advantage of the full spectrum of relation types. Results from the both methods are evaluated using O and R measures from eq. 10.1, where X and Y correspond to the candidate set of entities obtained using one of the methods for a given concept, and the reference set of entities respectively.

Semantic relation extraction. It refers to the relation types other than those which convey concept hierarchies. The assertion shown in Fig. 10.1 is one such example, where *plays* is a relation that connects *person* and *musical instrument* concepts. We compare the OIE systems in this task by two measures: breadth and depth of the identified relation types. Breadth corresponds to the absolute number of valid relation types identified for each concept, and the depth corresponds to the number of assertions for a given relation type that consist of the identified entities. The valid relation types are manually marked from among the relation phrases in the assertions.

10.4 Results and discussion

Quantitative assessment

Figs. 10.2a and 10.3a show the distribution of the number of extracted assertions using each of the OIE systems. Notice that the y-axis is a log scale. Between ReVerb and OpenIE 4.0, the latter seem to perform better, which can be attributed to the noun mediated relations. The semantic parsing based system, however, retrieves substantially more relations per sentence than these two. A tight coupling between syntax and semantics proves to be advantageous in chunking different types of assertions, as well as relating entities far off each other in a given sen-



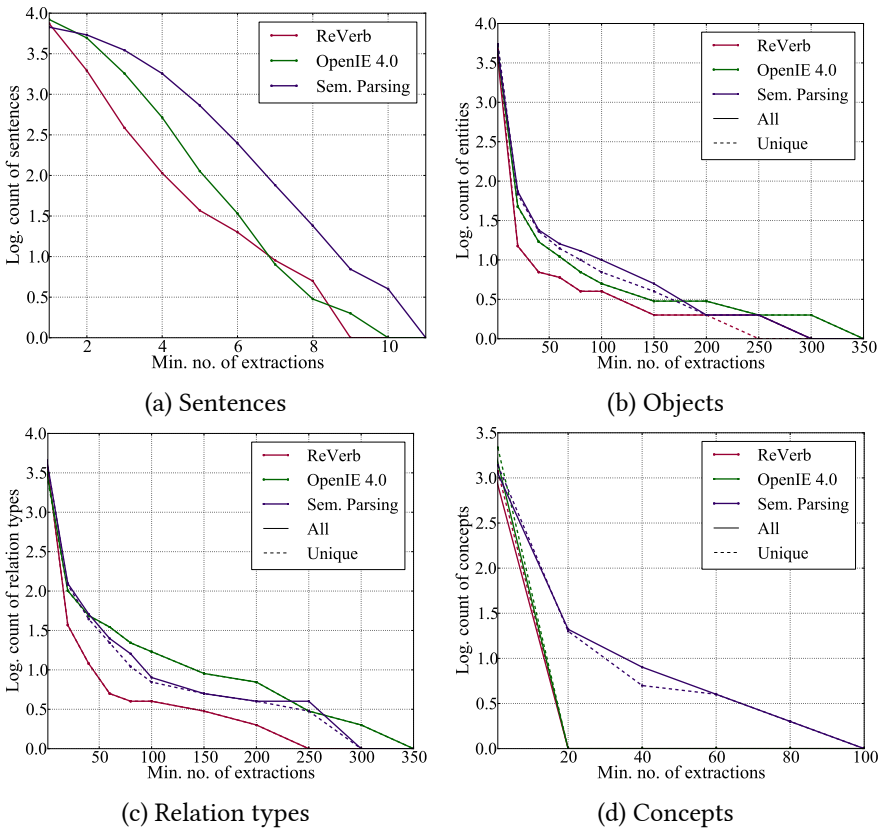
(a) Sentences (b) Objects

(c) Relation types (d) Concepts

Figure 10.2: Distribution of no. of extractions from OIE systems for Carnatic music shown along different aspects. For a given number of extractions on x-axis, the y-axis shows the logarithmic count of the instances within the aspect, which have at least those many extractions.

tence. For instance, in sentences which feature a single subject, but multiple relations (e.g: Chittibabu is a renowned Veena player, born in Kakinada to Ranga Rao and Sundaramma.), it performed thoroughly well compared to others.

Figs. 10.2b and 10.3b show the corresponding distribution for entities. Recall that we defined an entity to be the subject of a triple. The few entities with a disproportionately high number of extractions are usually the pronouns (despite resolving most of them), followed by musical terms. The semantic parsing based system retrieves slightly more number of assertions per entity compared to OpenIE 4.0, which in turn performs better than ReVerb. We observed some redundancy in assertions for a given entity, which is beneficial as this can be used as a measure of



(a) Sentences (b) Objects

(c) Relation types (d) Concepts

Figure 10.3: Distribution of no. of extractions from OIE systems for Hindustani music shown along different aspects. For a given number of extractions on x-axis, the y-axis shows the logarithmic count of the instances within the aspect, which have at least those many extractions.

confidence in asserting the corresponding relation. In order to analyze this, we have also plotted the distributions of number of unique extractions for entities (shown in dashed lines in the figures). We can observe that the semantic parsing based system retrieves substantially more number of redundant assertions per entity compared to the other two.

Figs. 10.2c and 10.3c show the distribution of the number of extracted assertions for the relation types. The results for semantic parsing based system and OpenIE 4.0 are more or less the same, both of which are substantially better than ReVerb. The redundancy in assertions, seen as the difference between the distributions shown by solid and dashed lines, is not as pronounced as it is for entities. The decline in the total number of relation-types as the number of extractions go higher, is less steep than in the case of entities. Unless the vocabulary in the domain is itself limited, this may indicate a slightly better coverage of relation types compared to that of the entities in the domain.

Figs. 10.2d and 10.3d show the distribution of the number of extracted assertions for the concepts. Recall that a concept is defined to be the object of a triple with subsumption relation phrase. The difference between the semantic parsing based system and the other two is quite marked, with the former retrieving more assertions per concept. For Hindustani music, the coverage of concepts in the assertions of ReVerb and OpenIE 4.0 is very low, with no concept having more than 20 assertions.

To summarize, the results indicate that the semantic parsing based system has a better coverage of entities, concepts and relation types of the domain than OpenIE 4.0, which is followed by ReVerb. It also retrieves more assertions per sentence compared to the other two. The results of quantitative assessment for Carnatic and Hindustani music correlate with each other, which shows that results are consistent. Note that this subsection has only provided the quantitative information, which by no means is complete in itself. The results we discuss in the following subsection complement these providing qualitative observations along the same dimensions (i.e., entities, concepts and relation types).

Qualitative assessment

In this subsection, we present the results for various tasks in the ontologization of Indian art music domain: concept identification, entity identification and semantic relation extraction.

Concept identification. Recall that we define candidate concepts to be the collection of objects from the triples with subsumption relation type. We map these

Music	#Ontology	#ReVerb	#OpenIE 4.0	#Sem. Parsing
Carnatic	53	4	4	22
Hindustani	55	1	2	9

Table 10.2: The number of concepts in the ontologies for each music, and those mapped from the assertions of the OIE systems.

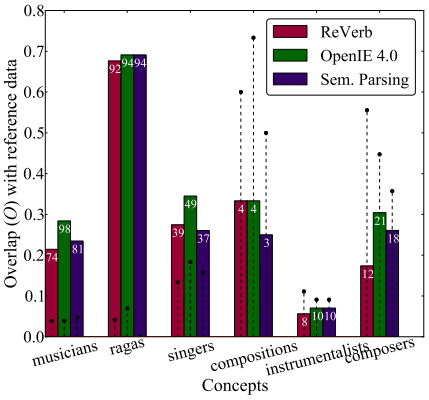
Music	#Reference	#ReVerb	#OpenIE 4.0	#Sem. Parsing
Carnatic	618	349	364	364
Hindustani	697	396	410	399

Table 10.3: The number of entities in the reference data for each music, and those identified using the OIE systems.

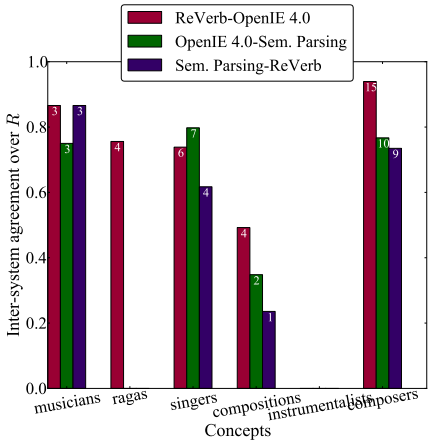
to the concepts in the ontologies as described in sec. 10.3. Table. 10.2 shows the number of concepts in the ontolgies for each music, and the number of concepts mapped from the assertions of the OIE systems. These results are consistent with our earlier observations of the results shown in fig. 10.2d.

Entity identification. The first subtask in this part is to find the entities in the domain. The candidate entities from each OIE system are defined to be the collection of subjects from all its triples. Table. 10.3 shows the total number of entities in the reference data taken from Wikipedia for each music, and the number of entities in the intersubsection of these with the candidate entities of each OIE system. There is no marked difference between the results, with nearly all the systems having about 60% of the entities from reference data in their assertions. However, it is observed that these correspond to only about 7% of all the possible candidate entities.

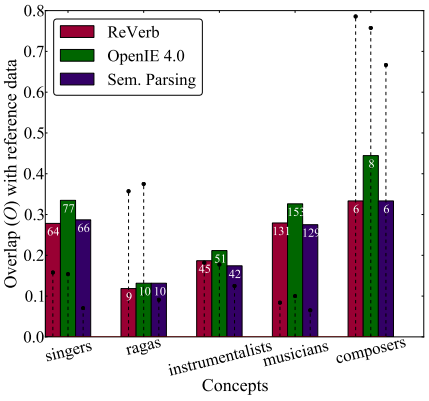
For the second subtask of entity identification, i.e., assigning entities to concepts, we have considered those concepts from our ontology for which there is a corresponding category on Wikipedia, each having at least 20 pages. This was done to avoid manual labeling of entities. We found 5 such concepts for Hindustani music: musicians, singers, instrumentalists, composers and ragas. For Carnatic music, in addition to these, we have found another concept: compositions. As discussed, we evaluate this task using two methods: rule-based and bootstrapping-based method.



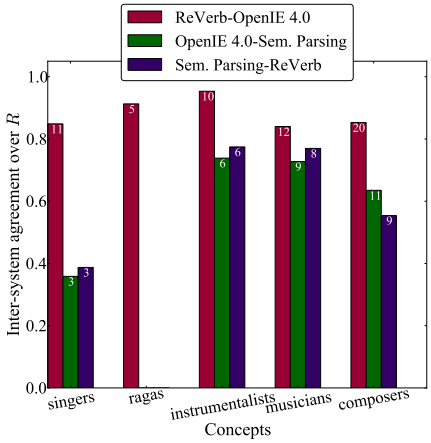
(a) Overlap with reference data.



(b) Inter-system agreement for residual entity cadidates.



(c) Overlap with reference data.



(d) Inter-system agreement for residual entity cadidates.

Figure 10.4: Results for rule-based concept assignment of entities identified in Carnatic (top) and Hindustani (bottom) music.

Figs. 10.4a and 10.4c show the overlap (O, see eq. 10.1) on rule-based concept assignment for entities found in Carnatic and Hindustani music using the OIE systems. The stem plots in the figures show residual portion of entities (R). The most notable performances are seen for the raga concept in Carnatic music. This can be attributed to two specific reasons: most Carnatic ragas on Wikipedia are described using a template, and the terminology consists mainly of Sanskrit terms which set them apart from the rest (mainly people). On the other hand, the description for Hindustani ragas varied a lot from raga to raga, and often the Sanskrit terms are inconsistently romanized making it hard for OIE systems to retrieve meaningful assertions. In theory, there is a template for almost every category, but there is a lot of variability, such as this, in describing the corresponding entities, except in the case of Carnatic ragas. OpenIE 4.0 seems to perform slightly better in terms of overlap, compared to the semantic parsing based system and ReVerb.

It is noteworthy to observe that residual entity candidates are consistently less in number for the semantic parsing based system. As mentioned earlier, there are two possibilities with them: they can be either false positives, or true positives which are not found in the reference data. In most cases, they are observed to be false positives. However, there are also a few of the latter. In order to understand them further, we have plotted the inter-system agreement in figs. 10.4b and 10.4d, which is given by the cosine similarity between R of different systems. ReVerb and OpenIE 4.0 agree with each other consistently higher over many of the concepts. We have observed that the cases where two or more systems agree on the candidature of a given entity, it is highly probable that the entity actually belongs to the concept. All the figures also show absolute numbers to put into perspective the proportion of residual entities where the systems agree with each other.

The second method for the evaluation of concept assignment employs bootstrapping as discussed in sec. 10.3. This process involves selection of a seedset and determining the number of bootstrapping iterations. For the sake of brevity, we have set the size of seedset to be the same for all the concepts, which is 3. The entities in the seedset are randomly chosen from the ones among the reference data taken from Wikipedia. However, as the bootstrapping process itself can be sensitive to the initial selection of the entities in the seedset, the whole process is repeated 5 times with randomly chosen seedsets. The bootstrapping method is terminated once the size of seedset reaches that of the corresponding concept in the reference data. After every 5 instances added during the process, we measure the overlap (O) and residual (R) portions of the seedset with respect to the reference data. Figs. 10.5 and 10.6 show their mean over 5 runs.

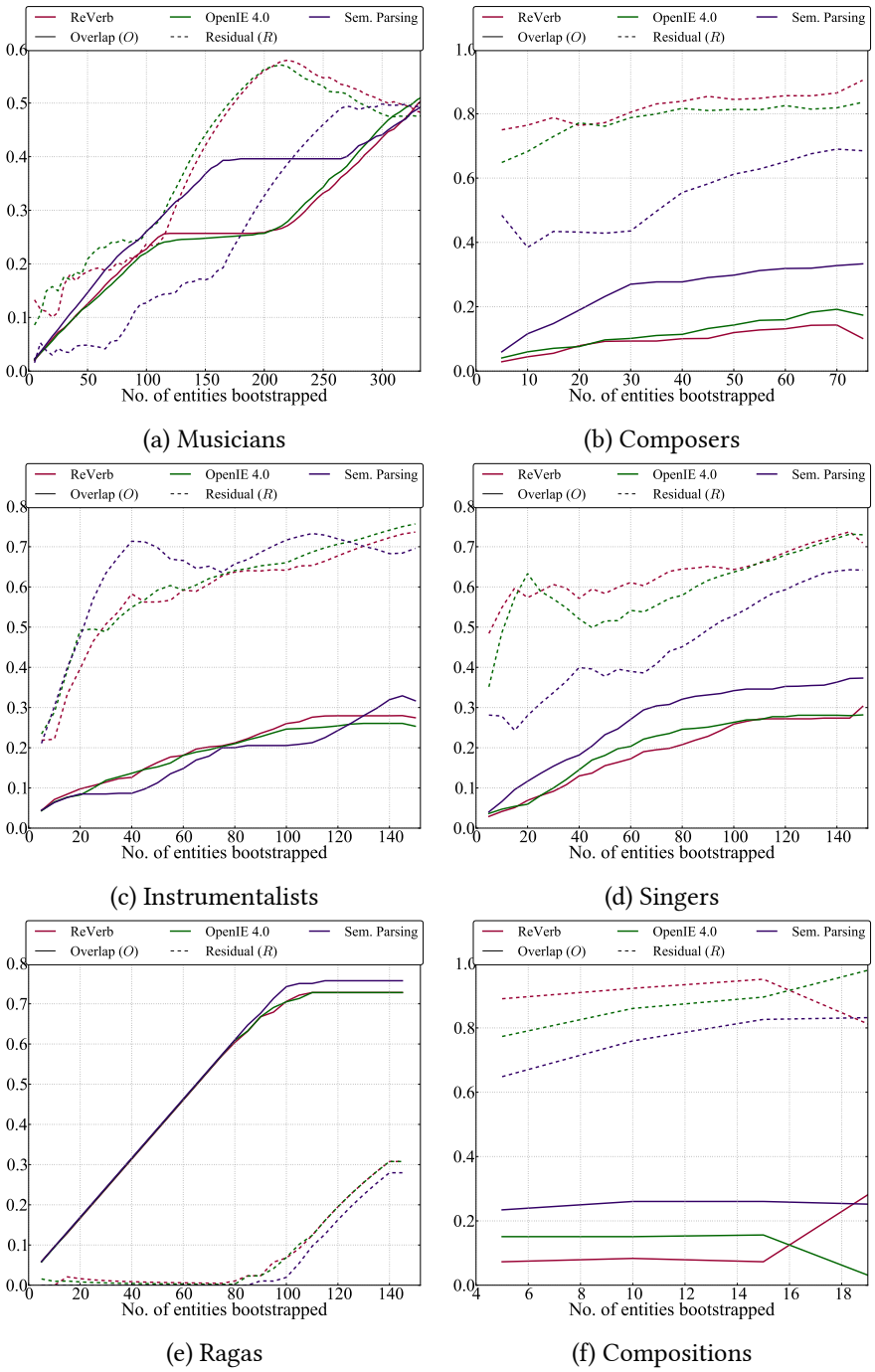
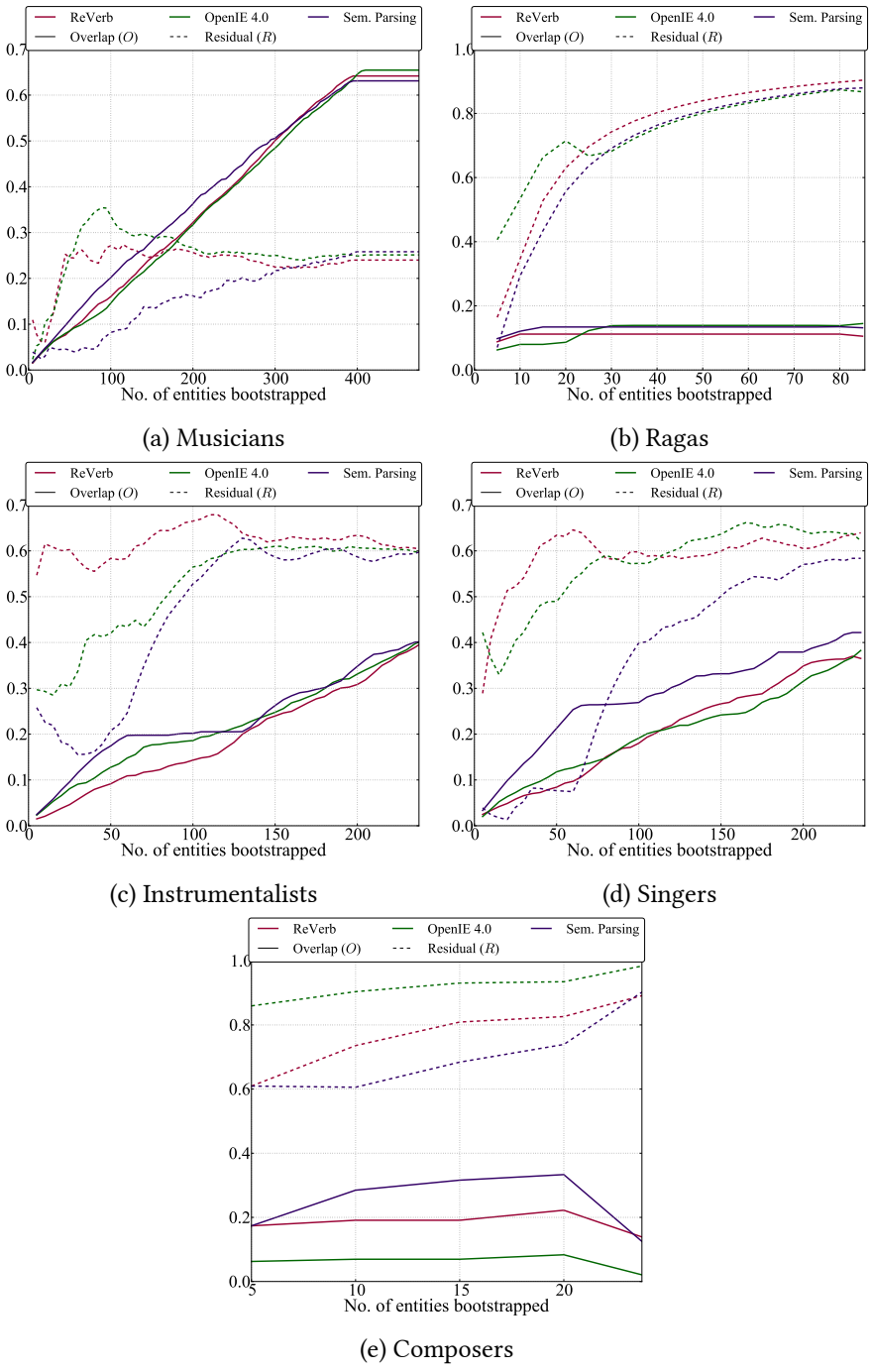


Figure 10.5: Results for bootstrapping-based concept assignment of entities identified in Carnatic music



(a) Musicians

(b) Ragas

(c) Instrumentalists

(d) Singers

(e) Composers

Figure 10.6: Results for bootstrapping-based concept assignment of entities identified in Hindustani music. The results for hindustani composers were not included due to space constraints.

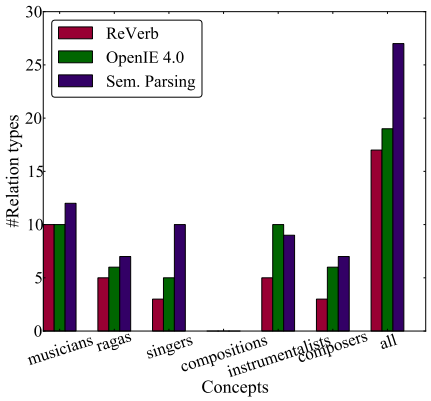
In most categories and for all the three systems, it can be seen that R grows quickly over iterations, making the residual portion the majority among the candidate entities, which brings the precision down. The semantic parsing based system consistently outperforms the other two methods, both in terms of having higher O, and lower R. Between ReVerb and OpenIE 4.0, there is no substantial difference in terms of O. For Carnatic singer and instrumentalist categories, however, the latter results in a lower R, and a slightly higher O compared to the former. These results partly contrast with those obtained for the rule-based concept assignment (fig. 10.4). However, remember that the coverage of concepts in the domain is observed to be remarkably better in the case of the semantic parsing based system (fig. 10.2d). As the bootstrapping method uses the objects from the triples (which are the candidate concepts), it seems logical that the semantic parsing based system has performed substantially better than the other two.

Semantic relation extraction. For the purpose of this task, the subsumption relation types and also those relation phrases which do not have a consistent meaning across the assertions were discarded. Then, following the procedure discussed in sec.10.3, we have marked the valid relation types for each concept, and obtained the corresponding assertions featuring the entities in the domain. Fig. 10.7 shows the results, for both Carnatic and Hindustani music.

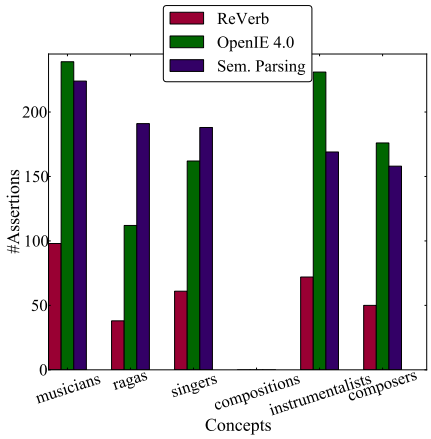
In terms of the breadth of the valid relation types (see sec. 10.3), the semantic parsing based system performs better than OpenIE 4.0, which in most cases fares better than ReVerb. However, in terms of the depth of relation types, both the semantic parsing based system and OpenIE 4.0 perform competitively, with the former scoring high in a few categories while the latter in few others. Compared to these two, ReVerb scores substantially less in this task. These results can be attributed to two reasons: the former two handle noun-mediated relations, whereas ReVerb does not, the average number of assertions for a noun-mediated relation type is observed to be usually higher than the verb-mediated relations. The second reason can be specific to the domain, where the relations with musical concepts are mostly noun-mediated (e.g: Abhogi, is a raga in, Carnatic music). Notice that there is a strong concurrence between these results and those shown in fig. 10.2c, with a noticeable correlation in the differences between the performance of the OIE systems.

10.5 Summary & conclusions

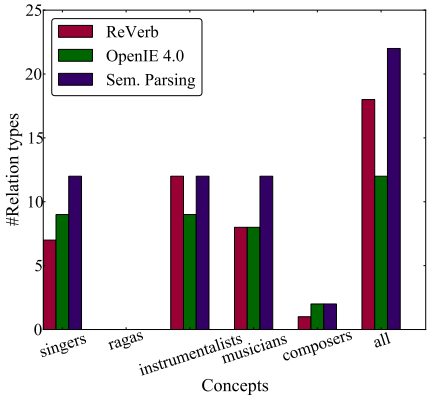
Domains like Indian art music often lack the advantages of the scale of data available. The OIE systems rely to a great extent on repetition in the data which is a



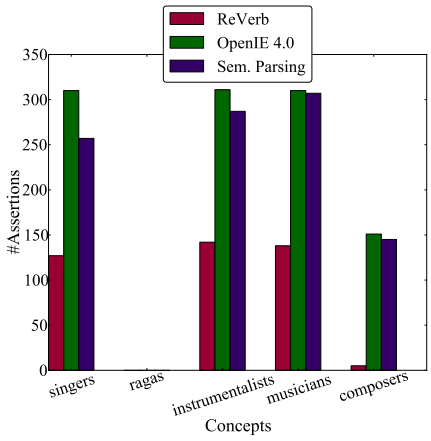
(a) Carnatic music: No. of valid relation types



(b) Carnatic music: No. of corresponding assertions



(c) Hindustani music: No. of valid relation types



(d) Hindustani music: No. of corresponding assertions

Figure 10.7: Semantic relation extraction task: The number of valid relation types marked for each concept, and the number of corresponding assertions that include the entities in the domain.

natural consequence of the scale of the web. As a result, OIE systems perform with very low recall over domains with limited data. In this context, it becomes important to bring out the differences between OIE systems in terms of their performance over such domains.

In this chapter, we have presented a framework for comparative evaluation of OIE systems for ontologization of thematic domains like Carnatic music. We have demonstrated it using three OIE systems in ontologizing the Indian art music domain. The results lead us to better understand the behavior of the systems from different perspectives, which can be used to guide the work in adapting OIE to thematic domains. The source code for the framework, links to various software components used in the demonstration, the ontologies and the data are made available online[9]. As we have seen, the results from the quantitative and the qualitative evaluation have a strong agreement with each other. The results indicate that the semantic parsing based system has performed comparatively better than the other two systems, and has certain desirable advantages over the other two.

A particular limitation of the framework that is of concern is the availability of groudtruth for qualitative evaluation. The current framework hinges on the structured content in Wikipedia/DBpedia to this extent. However, as our case study with Indian art music shows, they lack finer segregation of concepts (compared to our ontologies). The measures proposed in eq. 10.1, used in conjunction with inter-system agreement and various other observations over the data, helped us to overcome this limitation to a certain extent.

In ch. 12, we consolidate the information extracted using the three OIE systems to be published as a knowledgebase using our ontologies. Further, we combine this with metadata and intonation description extracted from the audio data. In the next chapter, we propose a methodology to use natural language text in bringing the salient aspects of a given music, which contrast it with other music genres/traditions. Further, as part of its evaluation, we propose a salience-aware semantic distance to use the acquired knowledge in a recommendation task.

The code, data and results constituting the experiments reported in this paper can be accessed at this github repository[10].

[9] https://github.com/gopalkoduri/openie_eval
[10] https://github.com/gopalkoduri/nerpari

Chapter 11

Quantifying the Salience of Musical Characteristics From Unstructured Text

Music is a discerning window to the rich diversity of the world. We hypothesize that identifying the differences between music from different cultures will lead to richer information models representative of them. Using five music styles, this paper presents a novel approach to bring out the saliences of a given music by rank-ordering its characteristics by relevance analyzing a natural language text corpus. The results agree with the cultural reality reflecting the diverse nature of the music styles. Further, to gather insights into the usefulness of this knowledge, an extrinsic comparative evaluation is performed. Similarities between entities in each music style are computed based on a salience-aware semantic distance proposed using the knowledge acquired. These are compared with the similarities computed using an existing linked-data based distance measure. A sizable overlap accompanied by an analysis of experts' preferences over the non-overlapping portions indicate that the knowledge acquired using our approach is indeed musically meaningful and is further complementary in nature to the existing structured information.

Music traditions from around the world share a few common characteristics. Yet, they differ substantially when viewed within their geographical and cultural context (Serra (2011)). Even among the seemingly usual characteristics, such as the musical concepts (melody, rhythm, ...) and the people involved in making the music (performers, composers, ...), their relevance and role vary from music to music. Consider the role of dance in different music styles. In Flamenco, it be-

comes an integral part of the music and is therefore seen as an important aspect of the music itself. Whereas in Jazz, it is not as closely associated.

Most commercial music platforms are agnostic to such differing characteristics of music, which inhibits them from scaling their recommendation services to meet the cultural diversity. To a certain extent, collaborative filtering techniques (Sarwar et al. (2001)) and context-based recommendation systems (Knees and Schedl (2013)) implicitly avail such information latent in listener activities and the community provided data such as tags. However, to our knowledge, there are no known approaches that explicitly incorporate the relevance of different musical characteristics.

We formally define the problem of quantifying the relevance or salience of characteristics of a given music as follows. E is a set of entities that make up the music, which includes its entire vocabulary. C is a set of its characteristics. Any given entity can posses more than a characteristic. C_k is a set of entities that share a characteristic, c_k. Entities include names of scales, chords, raagas, rhythm cycles, people and so on. An example for a characteristic is composing (c_k). All the entities who possess this characteristic constitute a set, C_k.

$$\begin{aligned} E &= \{e_i \mid e_i \text{ is an entity}\} \\ C &= \{c_i \mid c_i \text{ is a characteristic}\} \\ C_k &= \{e_i \mid e_i \text{ has a characteristic } c_k\} \end{aligned} \tag{11.1}$$

The first part of this paper presents our system, called *Vichakshana*[1], for quantifying the salience of the characteristics (C) of a given music and rank-ordering them, thus bringing out the most defining aspects of each music. Using the scores of C, we then propose a salience-aware semantic distance ($SASD$) to discover the related entities of a query entity. In the second part of the paper, We use an evaluation methodology to compare the results of a recommendation system[2] using $SASD$ with a linked data based recommendation system (Passant and Decker (2010)). Our primary intention is to understand the common and complementary aspects between the knowledge available as linked open data and the information our approach extracts.

The remainder of the paper is organized as follows. In 11.1, we describe the data we work with. Secs. 11.2 & 11.3 present our approach with details of its application on different music styles, and the $SASD$. In sec. 11.4, we present the

[1] *Vichakshana*, in Telugu language, means [wise] discretion.

[2] In this paper, we use the term *recommendation system* loosely to mean any system that can be used in retrieving related entities.

Music	Pages (E)	Categories (C)	Words
Baroque	2439	2476	901243
Carnatic	618	631	251533
Flamenco	322	1113	100854
Hindustani	697	492	317241
Jazz	21566	14500	5797726

Table 11.1: Details of the text-corpus taken from Wikipedia.

evaluation methodology and an extrinsic comparative analysis of the recommendation system built using $SASD$. In sec. 11.5, we conclude with a summary of the paper, the current work in progress and possible future directions.

11.1 Data

The natural language descriptions are a rich source of data about a given music. The web is voluminous in this sense, but also very noisy: with varying spellings, scholarly value etc. As the impact of such noise on the results is difficult to keep track of, we chose to present the results of our approach using text corpus extracted from the Wikipedia. Further, for our work, we need to acquire the characteristics of a given music. Automatically detecting them is part of the research on ontologization at the intersection of information extraction and knowledge engineering domains, which is a challenge in itself. These characteristics often directly correspond to the subsumption hierarchies and the class memberships in ontologies (Petasis and Karkaletsis (2011)). In this paper, we address the issue of rank-ordering the characteristics based on their salience. Therefore, in order to avoid digression from the problem being addressed, we rely on Wikipedia for obtaining the characteristics which roughly correspond to the categories each page is associated with. We keep only the plain text from the pages removing other structured information such as hyperlinks, info-boxes and tables.

We have selected five different music styles to work with: two Indian art music traditions (Carnatic and Hindustani), Baroque, Flamenco and Jazz, which together constitute a diverse set of music styles. Table. 11.1 shows the number of pages, categories (which correspond to E, C respectively), and the cumulative number of words across all the pages for each music style. There are as many contrasting features between them as there are similarities. Baroque is an Eu-

ropean classical music tradition which is no longer active, while Carnatic and Hindustani are actively practiced Indian art music traditions. Flamenco and Jazz are active popular music styles with distinct characteristics.

11.2 *Vichakshana*

A given entity in a music can be characterized by the references to other related entities in its description. In a way, such references can be understood to *explain* the given entity. Analysis of the structure of a network of references combined with the characteristics of each entity would yield us certain insight into the nature of the music. This is the intuition that our system, *Vichakshana*, builds upon.

The process broadly consists of three steps: entity linking, entity ranking and salience computation. The first step involves identifying the references to other entities from the content of a given page. This is performed using the DBpedia spotlight[3], which uses a combination of language-dependent and -independent approaches to contextual phrase spotting and disambiguation (Daiber et al. (2013)). A weighted directed graph (G) is created with the entities as nodes and the references as edges. The weight of an edge (w_{e_i,e_j}) is defined as follows:

$$w_{e_i,e_j} = \frac{n_{e_i,e_j}}{\Sigma_k n_{e_i,e_k}} \tag{11.2}$$

where n_{e_i,e_j} is the number of references from e_i to e_j. We have observed that the link structure in the graphs thus obtained is very sparse. Therefore, the references to entities which are outside the set of E are eliminated. Table. 11.2 shows topology of all the graphs.

In order to compute the salience score for a given C_i, we require a measure for the relevance of the constituting entities in the given music. Pagerank is a widely used algorithm to compute the relevance of a node in a hyperlink graph (Page et al. (1999)). Intuitively, it is an iterative algorithm in which nodes acquire relevance from their incoming edges. A reference from a node with a high pagerank to another node contributes positively to the relevance score of the latter. In this sense, it can also be understood as a variant of the eigenvector centrality (Newman (2010)). We use a slightly modified version of the original pagerank algorithm to use edge weights in propagating the score of a given node to its

[3]We use a locally deployed version of DBpedia spotlight with the statistical backend, available openly at https://github.com/dbpedia-spotlight/dbpedia-spotlight

Graph	Nodes	Edges	Density	Avg. Clust.	Avg. Deg.
Baroque (I)	14278	44809	0.0002	0.002	3.14
Baroque	2059	7118	0.0017	0.018	3.46
Carnatic (I)	4524	12952	0.0006	0.003	2.86
Carnatic	602	3291	0.0091	0.03	5.47
Flamenco (I)	2671	5459	0.0008	0.004	2.04
Flamenco	312	846	0.0087	0.027	2.71
Hindustani (I)	7011	17754	0.0004	0.002	2.53
Hindustani	681	3774	0.0081	0.027	5.54
Jazz (I)	87918	381514	0.0	0.004	4.34
Jazz	17650	119107	0.0004	0.019	6.75

Table 11.2: Topology of the graphs obtained on entity linking, before and after the references to entities outside E are eliminated. Rows with '(I)' denote the former.

neighbors. Eq. 11.3 describes the corresponding computations.

$$\begin{aligned} A_{e_i,e_j} &= w_{e_i,e_j} \\ D_{e_i,e_i} &= max(e_i^{out}, 1) \\ P &= D(D - \alpha A)^{-1}\beta \end{aligned} \tag{11.3}$$

where A is the adjacency matrix corresponding to the graph G, D is the diagonal matrix with the diagonal elements set to the out degree of the corresponding node (e_i^{out}), P is the resulting pagerank values of all the nodes. α is an activation constant set to 0.85 in our analysis, and β is an array of additive constants which are all set to 1. For more explanation on pagerank and the constants, we refer the reader to (Newman (2010); Page et al. (1999)).

Given a C_k, a naive and simple salience score can be the mean of pagerank scores of all the constituting entities. Remember that an entity can have multiple characteristics. A simple scoring method, such as this one, would imply that the pagerank score of a given entity equally contributes to the salience score of every C_k it belongs to. However, it is desirable that an entity contributes more to those characteristics which are more specific to it. As our data does not contain this

Baroque	Carnatic	Flamenco	Hindustani	Jazz
Anglican saints, Organ improvisers	Carnatic music terminology	Spanish dances, Spanish folk music	Carnatic music	African-American music, Western swing ...
Music in Leipzig, Thomaskantors	People from Tiruvarur district	People from Algeciras, Spanish people of Portuguese descent	Formal sections in music analysis	Burials at Woodlawn Cemetery (Bronx)
Composers for cello	Indian classical music	Andalusian music	Indian classical music	Rec labels- established in 1916, disestablished in 1940
Harpsichord, Keyboard instruments	Ragas	Andalusian music, Flamenco styles, Spanish music, Vocal music	String instruments	Presidential Medal of Freedom recipients
Music catalogues	Chennai culture	Latin jazz musicians, Spanish guitarists	Hand drums	Bass (sound)
Baroque instruments, Necked bowl lutes ...	Carnatic music	1950 births	Ragas	American jazz
Collected editions of classical composers, Johann Sebastian Bach	Indian Vaishnavites, Kannada people	Romani guitarists	Bangladeshi-, Hindustani-, Pakistani-musical instru.	ABC Records artists
1685 births	Dvaita, Indian philosophers ...	Spanish musicians	Culture of- Bihar, Uttar Pradesh ...	Amplified instruments, Bass guitars ...
People from Halle (Saale)	Carnatic Ragas	People from Córdoba, Andalusia	Necked bowl lutes, String instruments with sympathetic strings ...	Jazz ensembles
German Lutherans	Hand drums, Pitched percussion ...	Male ballet dancers, Spanish dancers	Hindustani music	American Buddhists, Converts to Buddhism ...
English people of German descent ...	Telugu people	Cancer deaths in Spain	Sitars	EMI
Cantatas by Johann Sebastian Bach	Hindu monarchs, Maharajas of Travancore ...	1958 births	Carnatic music terminology	Jazz instruments
Medieval music	Bhakti movement	People from Cadiz	Music schools in India, Vocal gharanas	Jazz genres
Composers for violin	1680 deaths, Hindu poets, History of Andhra Pradesh, Telugu poets	2004 deaths	Carnatic Ragas	Pablo Records artists
1750 deaths	People from Thanjavur district	Flamenco groups	Dark Horse Records artists, Grammy Award-winning artists	Companies based in California, Labels distributed by UMG ...

Table 11.3: Top 15 characteristics ordered by their salience to different music styles. Note that as Carnatic and Hindustani share a large portion of musical terminology which are categorized into Carnatic music on Wikipedia, we see many Carnatic music characteristics for Hindustani music.

information, we hypothesize that the fewer the number of other entities which share a characteristic with the given entity, the more specific it is. Formally, if each entity (e_i) represents a document with the characteristics (c_i) as the terms, the inverse document frequency of a c_k with respect to E would yield us a measure that can be used to weigh the pagerank score of a given entity in computing the saliences of all C_k it is associated with. Eq. 11.4 describes this process in detail along with the steps for computing the salience score of a C_k (given by S_{C_k}) from the pagerank values of its entities.

$$\text{idf}(c_k) = \log \frac{|E|}{1 + |\{e_i \in C_k\}|} \tag{11.4}$$
$$S_{C_k} = \frac{1}{|C_k|} \sum_{e_i \in C_k} P(e_i) \times \text{idf}(c_k)$$

This gives us a list of characteristics of a music ordered by their salience. We have observed that several characteristics have a considerable overlap between them. For instance, the characteristics *Music festivals in India* and *Carnatic music festivals in India* have more or less the same set of entities, with respect to Carnatic music. We consider them redundant even though semantically one is a more specific form of the other. As we will see in sec. 11.3, this is undesirable for applications using these salience scores. We handle such cases by *merging* them and assigning a common rank to each such group. This is performed using an undirected weighted graph constructed with the characteristics (C) as nodes. The weights of edges correspond to the cosine similarity between the corresponding sets of entities, C_i and C_j. Those edges with a weight less than 0.5 are filtered out, and then the closely related communities are identified using the Louvain method (Blondel et al. (2008)). Each such community represents a group of characteristics which have a great overlap between the corresponding entities, and is assigned a common rank based on the new salience score recomputed using eq. 11.4 considering each community as a single characteristic.

It is also observed that the weights from Eq. 11.4 inadvertently resulted in a high rank for characteristics that are relevant to a musician, but not to the given music in general. For instance, if a very popular musician also happens to be a politician, the political characteristics are ranked high even though they are irrelevant to the music. However, if there is a certain regularity to such associations (eg: more musicians are also politicians), it is dsizeesirable to incorporate and rank those characteristics. Towards this extent, we constrain the ranked characteristics to a set of those which have at least a minimum number of entities linked to them. A high threshold includes the risk of discarding meaningful characteristics,

while not employing such threshold would result in spurious ranking. Merging overlapping characteristics minimizes the impact of irrelevant ones provided they are not associated to a musical entity by chance. We have empirically chosen the value for the threshold to be three.

Table. 11.3 shows the top few characteristics of each music style, ordered by their salience to the music. The similarities and contrasts between the music styles are quite apparent. In Baroque, Carnatic and Flamenco, various groups of people identified by their region/language occupy a prominent place, while in Hindustani and Jazz, such groups are relatively less prominent. This might be due to the fact that the latter two are spread out over larger regions than the former three. In Carnatic and Hindustani, the terminology and musical concepts turn out to be more relevant than in other music styles. In Jazz and Flamenco, the salience of record labels is quite high whereas in other music styles, it is almost non-existent or very less. This can be due to the fact that the primary medium by which Carnatic/Hindustani music reaches people is a concert. In the case of Baroque, this is because it is no longer an active music tradition. The results also highlight the distinct features of each music. The prominence of religion in Carnatic music, dance in Flamenco, and Gharanas/schools in Hindustani music is noticeable and each contrasts with other music styles.

A direct objective evaluation of these results in not feasible as it is impractical to obtain a consensus on a highly subjective notion such as the relevance or salience of something/ somebody in a music. Therefore, we present an extrinsic evaluation of the results using the task of music recommendation. We use the salience scores of the characteristics of a given music to relate the entities using a distance measure, and compare the results with a recommendation system that feeds on linked-data.

11.3 Salience-aware semantic distance

Using the graph (G) and salience scores (S), we propose a salience-aware semantic distance (SASD). It is a weighted average of three parts. The first and prominent part is a function of salience scores. The second part is a function of length of the shortest path between the two nodes in G, while the third part is a function of their cocitation index, which is the number of other nodes in G that point to both the given nodes. Eq. 11.5 formally defines the three parts and the sum. The values of all the parts of the distance and the weighted sum range

between 0 (nearest) and 1 (farthest).

$$
\begin{aligned}
S_{e_i,e_j} &= \{S_{C_k} \mid e_i \in C_k \text{ and } e_j \in C_k\} \qquad (11.5)\\
A'_{e_i,e_j} &= \begin{cases} 1 & \text{if } A_{e_i,e_j} > 0 \\ 0 & \text{otherwise .} \end{cases} \\
D1 &= \frac{1}{1 + |S_{e_i,e_j}| + mean(S_{e_i,e_j})} \\
D2 &= \frac{p^2_{e_i,e_j}}{1 + p^2_{e_i,e_j}} \\
D3 &= \frac{\Sigma_k A'_{e_k,e_i} A'_{e_k,e_j}}{\sqrt{\left|\Sigma A'_{e_k,e_i}\right| \times \left|\Sigma A'_{e_k,e_j}\right|}} \\
SASD_{e_i,e_j} &= \left(\frac{1}{2}\right) D1 + \left(\frac{1}{4}\right)(D2 + D3)
\end{aligned}
$$

where S_{e_i,e_j} corresponds to the salience scores of the common characteristics of e_i and e_j, A' corresponds to the adjacency matrix of the unweighted graph equivalent of G, and p_{e_i,e_j} is the length of the shortest path between e_i and e_j in G. The first part of the distance is weighed more making the role of knowledge extracted using *Vichakshana* more pronounced in relating the entities. The role of the other two parts of the distance is often limited to further sort the list of related entities obtained using the first part.

11.4 Evaluation

Methodology

Our evaluation methodology is primarily intended to streamline the two stages of the objective and the subjective forms of evaluations for comparing the recommendation systems. The first stage corresponds to an objective comparison of the results over three measures: yield (Y), overlap (O) and rank-correlation (RC) within the overlap. The former two measures are defined as follows:

$$
\begin{aligned}
Y^I &= \frac{\left|\{e_k \mid \left|R^I_{e_k}\right| > 0\}\right|}{|E|} \qquad (11.6)\\
O_{e_k} &= \frac{|R^I_{e_k} \cap R^J_{e_k}|}{\max(|R^I_{e_k}|, |R^J_{e_k}|)}
\end{aligned}
$$

where $R^I_{e_k}$ denotes an ordered-list of recommendations for a given entity e_k generated using an approach I. Therefore, Y^I is the proportion of entities which have non-empty set of recommendations using approach I. O_{e_k} is the proportion of the common set of entities in $R^I_{e_k}$ and $R^J_{e_k}$. For measuring rank-correlation, we use Kendall's Tau, which is preferred over other standard alternatives as it is known to be robust for smaller sample sizes. We use the Tau-b variant which accounts for tied pairs (Kendall and Gibbons (1990)).

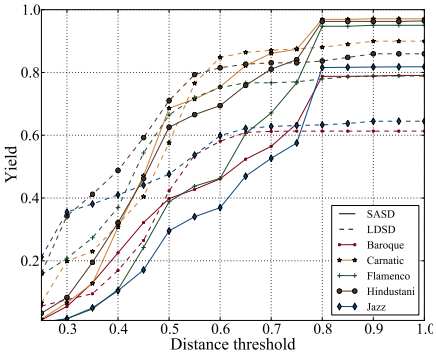
The set E is divided into three different sets based on our analysis in this stage. The first one (E_1) is the set of entities for which $O \geq \frac{1}{3}$ and RC is greater than the median of all values. This set corresponds to those entities where both approaches broadly agree with each other. The second set (E_2) consists of those entities where $O \geq \frac{1}{3}$ and RC is less than the median of all values. The last set of entities (E_3) is where $O < \frac{1}{3}$.

In the second stage, which is a subjective form of evaluation, the music experts (mostly practicing musicians) record their feedback for questionnaires based on the latter two sets of entities. The one based on E_2 has, for each query entity, two rank-ordered lists with exactly the same set of entities (i.e., the overlapping portion). The experts are asked to pick the better one. The motive behind this is to understand whether one system is preferable to the other in ranking the entities. The questionnaire based on E_3 also has two lists for each query entity, but this time without an emphasis on the order of the items. The experts are asked to pick the entities in each list that are the most relevant to the query entity, and also the overall better list. Evidently, the motive here is to understand which of the approaches produces more appropriate recommendations. An analysis of their opinions would let us know whether a particular approach is preferred over the other, and can further be used to investigate why.

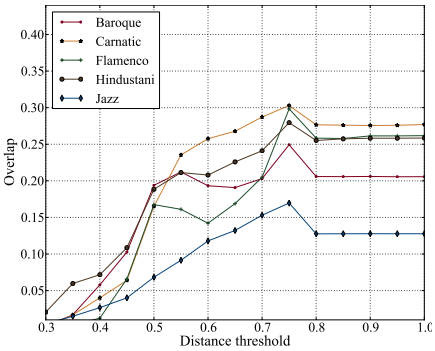
Results & discussion

As mentioned earlier, our approach borders on the unstructured contextual-data based approaches and the linked-data based approaches. More specifically, SASD is used to relate entities with the knowledge extracted from unstructured data using *Vichakshana*. This is unlike the other contextual-data based approaches which use community-generated data such as social networks, tags and user behavior, where the latent information in the social dynamics plays a significant role in their outcome. Therefore, we choose to compare our approach with others that build upon similar data sources to ensure a fair evaluation.

We compare the results from our approach with DBrec system (Passant and Decker



(a) The yield (Y) for both the systems shown for different music styles.



(b) The thick-lines show the mean overlap between R^{sasd} and R^{ldsd} for different music styles. The dotted lines shows the standard deviation.

Figure 11.1: Results for the analysis of overlap between the two recommendation systems. X-axis in both the figures denote the distance threshold beyond which two entities are considered unrelated.

(2010)), which is based on DBpedia[4]. As it is shown to perform comparably well with other context-based approaches that build on diverse sources of data, this comparison helps us to put the results of our system in perspective with both the linked-data based and the other context-based systems. However, note that our system uses only the salience scores and the entity references, but not the structured data from Wikipedia.

For all the experiments hence forth, the size of the recommendations corresponds to ten[5]. Fig. 11.1a shows Y^{sasd} and Y^{ldsd}, and fig. 11.1b shows the mean overlap between R^{sasd} and R^{ldsd} for different distance thresholds. Y^{ldsd} steeply rises until 0.6 and saturates, indicating that the practical limit for LDSD between two entities is 0.6, where as it is 0.8 for SASD. In line with this, the mean overlap in fig. 11.1b rises until a distance threshold of 0.75, where the overlap for all the music styles between the two systems is the maximum. Following that, it slightly drops, which must be a consequence of the gain in Y^{sasd} compared to Y^{ldsd} as shown in fig. 11.1a. We can deduce that there is a sizable overlap between the recommendations of the two systems. However, which of the non-overlapping

[4]DBpedia collects the structured content from Wikipedia.

[5]Results for other sizes of the recommendations show a similar behavior. For the sake of brevity, we report on recommendation sets of size ten.

Mus	Pos	Mean	Std	Neg	Mean	Std
Baroque	59%	0.47	0.25	41%	-0.32	0.3
Carnatic	56%	0.47	0.26	44%	-0.29	0.26
Flamenco	69%	0.41	0.22	31%	-0.2	0.29
Hindustani	65%	0.49	0.26	35%	-0.23	0.24
Jazz	55%	0.47	0.25	45%	-0.27	0.28
Avg	60%	0.46	0.24	39%	-0.26	0.27

Table 11.4: Results for rank-correlation between the two approaches, showing the % of entities with positive and negative rank-correlation, along with their mean and standard deviation.

Music	E_1	E_2	E_3
Baroque	17%	12%	31%
Carnatic	27%	21%	39%
Flamenco	31%	14%	29%
Hindustani	28%	14%	39%
Jazz	11%	8%	34%

Table 11.5: Proportions of the E_1, E_2 and E_3 across all the music styles.

recommendations are more meaningful is an issue we will have to address in the subjective evaluation.

The rank-correlation (RC) is analyzed between those $R_{e_k}^{sasd}$ and $R_{e_k}^{ldsd}$ which have an overlap of 0.3, with a distance threshold of 0.75 (which is roughly the same as the configuration corresponding to the highest overlap from fig. 11.1b). Note that RC ranges from -1 (complete disagreement) to 1 (perfect agreement). We consider a value of zero to be a negative correlation. Table. 11.4 shows the results. We observe consistently more positive correlations across all styles of music. Further, the mean of the positive correlations indicates a strong agreement between the recommendation systems. Based on the analysis so far, we divide the entities into three sets as discussed in sec. 11.4. Table. 11.5 shows the proportions of the three sets for different music styles.

For the subjective evaluation with music experts to further understand E_2 and

E_3, we randomly sampled 20 entities from Carnatic music[6] ensuring that there is equal coverage of popular and less popular entities, as well as E_2 and E_3. Note that these entities comprise of not just artists and songs, but any musical entity (eg: a place). The measure of popularity of an entity is its PageRank value in the graph G.

A total of 424 responses were recorded from 10 Carnatic music experts[7], all of whom are practicing musicians with a median age of 25. Table. 11.6 shows the aggregate results for questionnaires based on E_2 and E_3. The overall results do not seem to indicate a strong preference to one system or the other. However, it is evident that the responses concerning different entities are very divided. There are certain interesting observations from the responses over E_3. The number of cases in E_3 where DBrec system is preferred seems slightly higher. Yet, the number of cases where more entities are specifically marked relevant to the given query entity is higher for $SASD$.

In order to further understand this phenomenon, we have gone through the recommendations from the two systems and the responses recorded for each entity. Consider the case of *Kshetrayya*, a composer in E_3. The list of recommendations using $SASD$ is dominated by other composers sharing some characteristics (like geographic location, language etc). Those from DBrec system ranks the performers who often sing his compositions higher than the fellow composers. This resulted in more experts preferring DBrec system. However, the number of recommendations explicitly marked as relevant are marginally higher for $SASD$. Another example is the case of *M. Balamuralikrishna*, a performer. The recommendations from $SASD$ do not include his teacher whereas those from DBrec system do. This is a result of a low recall in entity linking using DBpedia Spotlight. While the ratio of experts preferring DBrec system to SASD is 7:2, the corresponding ratio of absolute number of entities selected as relevant is 7:6. There are several other cases like these. This trend clearly indicates that though our system missed few important relations between entities (such as those between different types of entities), the recommendations made are still very relevant, often times even more than the recommendations from DBrec system.

More formally, for a given query entity, $SASD$ is inherently biased to select and rank higher other entities which are of the same type (Eg: Composer). It is also highly sensitive to recall in entity linking. On the other hand, DBrec system has

[6] Carnatic music was chosen for the subjective evaluation as the authors have better access to its community compared to other music styles.

[7] By experts, we mean those who are thoroughly well-versed with the domain and are learned-artists.

	E_2				E_3			
	SA.	LD.	Both	None	SA.	LD.	Both	None
Overall preference	30%	30%	10%	30%	40%	50%	10%	0%
Entities specifically marked as relevant	n/a	n/a	n/a	n/a	50%	30%	20%	0%

Table 11.6: Results of the subjective evaluation of the two recommendation systems. The first row of results show the % of query entities where a particular recommender system is more favored. The second row shows the % of query entities where more number of entities in the corresponding recommendation list are marked as specifically relevant to the query entity.

access to a richer link structure that spans different entity types.

When we view the results again in the light of this stark difference in the nature of information both the systems had access to, the results seem to clearly indicate that the information extracted using *Vichakshana* is meaningful in itself. Further, it is complementary to the existing structured content on Wikipedia which DBrec depends on, and is useful for improving the music recommendations, both in terms of better ranking and more importantly, finding relevant content with an emphasis on the distinct characteristics of a given domain.

11.5 Conclusions

We have presented and formally defined the idea of quantifying the salience of characteristics of a music, and how it leads us to extracting culture-specific information about a music using natural language text. We have shown that the performance of a recommendation system built using the information extracted is comparable to that of the linked-data based recommendations. The main contributions of the paper are as follows: i) A novel approach that quantifies the salience of characteristics of a music, ii) A salience-aware semantic distance that builds upon the knowledge extracted, iii) Open-source python implementations of *Vichakshana* and SASD[8].

[8] Available at https://github.com/gopalkoduri/vichakshana-public

Chapter 12

Knowledge-base population: Structuring and interlinking the data sources

This is the concluding chapter in which we describe the processes for merging the information from various data sources and publish them using our ontologies. This procedure involves three steps. In the first step, we publish the editorial metadata of commercial releases that are part of the Carnatic corpus of the CompMusic collection, using our ontologies and the music ontology amongst others. In the second step, we compute the context-based pitch distributions for svaras in different raagas using the audio data and link them to the recordings in the KB. In the third step, we link the assertions retrieved from the open information extraction systems to appropriate concepts and relationships in the KB. Each of these steps are presented in the following sections, mentioning their quantitative contribution to the KB. We then conclude with the main challenges that we encountered in accomplishing our goals initially stated, and layout future course of research to take this work further.

12.1 Editorial metadata

Statistics of the Carnatic corpus are shown in fig. 1.2. All the metadata is uploaded and structured on Musicbrainz[1]. We use the URLs originating on Musicbrainz as unique identifiers for most entities. For others which do not have a URI/unique

[1] Available at https://musicbrainz.org/user/compmusic/collections

identity on Musicbrainz, like raagas and taalas, we use the URLs minted in Dunya. One of the challenges in structuring this data had been the common occurrence of numerous variations in the spellings of entity names. We automated merging of such aliases partly, using a combination of string matching algorithms[2]. In the next step, we create a wrapper on Dunya API that maps its database schema to concepts and relations in our ontologies. Following this, these concepts are further mapped to DBpedia using DBpedia spotlight. This completes publishing of the editorial metadata and its interlinking with Musicbrainz and DBpedia.

We observe that keeping concepts and relations that are part of music creation workflows adds a certain overhead in adding information to the KB. For KBs that are specific to applications that are not concerned about such information, this overhead does not add any value. For instance, using the current version of our ontology as described in the thesis, the relation between a mo:Record and a mo:Release consists of a chain of at least 3 relations that connect 4 concepts inclusive of them. The intermediary concepts include mo:Performance and mo:Signal, which in our case do not have an identifier and hence had to be represented using a blank node. For queries, and even inferences, this adds to the complexity and hence we believe custom relations defined directly between the relevant concepts are more preferred. However, we keep the status quo for two reasons: i) This KB is not being tailored to a specific use, ii) The scale of the KB is small.

Svara information in raagas will be of particular interest to the possible applications of the resulting knowledge-base. Hence, we also crawled semi-structured data from the web concerning raagas such as this[3] listing of raagas. We then publish this information using our ontology and link it with entities from editorial metadata. This brings in valuable information about a raaga such as its arohana and avarohana progressions, therefore the svaras it has, whether it is a melakartha raaga or a janya raaga and so on. Together, the editorial metadata and the information extracted from semi-structured data account for 83000 triples in the KB (73670 and 9333 triples respectively) as of the date of writing this thesis[4]. Listing. 12.1 shows Hanumatodi, an entity belonging to Raaga class, as expressed using our ontology.

[2] Available at https://github.com/gopalkoduri/string-matching

[3] http://www.nerur.com/music/ragalist.php

[4] As this KB is an evolving resource, we refer the reader to a companion page for up-to-date examples, details and statistics of the KB - https://github.com/gopalkoduri/knowledge-bases

Listing 12.1: Part of Hanumatodi as expressed using our ontology.

```
<http://dunya.compmusic.upf.edu/carnatic/raaga/a9413dff-91d1-4e29-
    ad92-c04019dce5b8> a co:Melakarta_Raaga ;
  co:has_arohana [ a olo:OrderedList ;
    ns1:creator "semi-structured-nerur"^^xsd:string ;
    olo:length 2 ;
    olo:slot [ olo:index 5 ;
    olo:item co:P ],
    [ olo:index 2 ;
    olo:item co:R1 ],
    [ olo:index 1 ;
    olo:item co:S ],
    [ olo:index 3 ;
    olo:item co:G2 ],
    [ olo:index 7 ;
    olo:item co:N2 ],
    [ olo:index 4 ;
    olo:item co:M1 ],
    [ olo:index 6 ;
    olo:item co:D1 ] ] ;

    co:has_svara co:D1,
                 co:G2,
                 co:M1,
                 co:N2,
                 co:P,
                 co:R1,
                 co:S ;

    co:mela_number 8 .
```

Besides the assertions that we explicitly published, the KB can also be used to make inferences resulting from definitions of other concepts or a set of rules (see sec. 9.2). For instance, our definitions of Sampoorna_Progression and Sampoorna_Sampoorna raaga further trigger Hanumathodi's classification into Sampoorna_Sampoorna raaga. They together conclude that, any raaga that has at least one variant of each top-level svara class (such as Shadjama, Rishaba and so on) in both its arohana and avarohana, is a Sampoorna_Sampoorna raaga. Let us now consider the case of Dharbaru raaga. In this case, it is a rule outside the ontology that results in its classification as a Vakra raaga. Therefore, we find several such inferred statements about the domain in this KB.

12.2 Intonation description

We compute the context-based pitch distributions of svaras for all the recordings in corpus for which we have raaga and svara information. From the normalized distributions, we identify peaks (see sec. 5.5). We use the Data Model class to

link this information to Record class. This added knowledge of svaras facilitates computation of similarity between recordings and raagas, by extension.

Not all the svaras are equally important in a raaga, as we mentioned earlier. A few assume greater role by virtue of their functionality. This information can be availed from the KB and then used in weighing the role of different svaras when computing similarities. Further, as in the case of editorial metadata, rules can be applied to draw inferences. For instance, if a given distribution has multiple peaks, it can be inferred that the corresponding svara is sung as a movement that spans an interval corresponding to the maximum difference in the peak locations. Whereas those which have just one peak marked can be inferred to be sung as steady svaras.

In the Carnatic corpus, there are 2338 recordings in 161 raagas for which we have all the necessary data to compute the intonation description. Information extracted from these result in about 70000 triples being added to the KB. These do not include the inferences. The Pitch_Distribution class has information about peaks described in RDF Literals (integers) in our current version of the ontology. We use rules and SPARQL queries over these to make observations and infer further valid facts to be added to the KB.

12.3 Structured information from natural language text

As we already know from ch. 10, the precision and recall of different open information extraction systems in identifying concepts, objects (instances) and relations from natural language text are not at a desirable level. However, if we restrict the relation-types to a known set, we observed that the resulting extractions, though fewer in number, come with better precision. These relations include the hierarchical relation (is a) and biographical relations (brother of, born in etc). We further link most of the concepts and the objects to those on DBpedia, which links them to the relevant structured information published elsewhere.

12.4 Possible courses of future work

There are two main lines of work in this thesis that can be further extended. The first one concerns the intonation description from audio music recordings. In this thesis, we have explored aggregate approaches that discard the temporal information. Though the resulting description is musically meaningful, it lacks the granularity that becomes necessary in more sophisticated systems such as those studying variations between artists performing the same raaga. This line

of work can benefit from both the current thesis and the work concerning melodic pattern analysis in the CompMusic project (see **?**, and the references therein).

The other line of work which we believe can further be extended from this thesis is combining ontology development with natural language text processing. For thematic domains where the natural language text is a scant resource, OIE systems are shown to have a very limited success. Of the ones that are evaluated, semantic parsing based system is the most promising and we believe this requires further exploration. Its main advantage is in extracting a greater number of assertions per natural language statement, compared to other competing systems. This brings in the much required redundancy in the assertions, although to a limited extent. A particular advantage of thematic domains is that it mostly requires a smaller vocabulary (hierarchy of concepts) required to describe it. This can be used in our favor where vocabularies manually engineered can be combined with a semantic parsing based system in building a knowledge-base.

Improving and revising the ontologies and knowledge-bases is something which we will continue to do post this thesis duration. In connection to this, and besides the possible extensions to our work which we already mentioned, another important line of work in MIR that complements this, is research and development of applications that take advantage of the knowledge-bases and ontologies. We have been working on three applications that use parts of the information in the KB. We discuss them briefly in Appendix. XX.

12.5 Summary & conclusions

We have discussed and documented our research efforts in building a multi-modal knowledge-base for the specific case of Carnatic music. Recall that there are three primary objectives to our work (sec, 1.2), for each of which we now review the status.

- *Extract a musically meaningful representation of svara intonation in Carnatic music, analyzing audio music recordings.*

  We have consolidated a representational model for svara that expands the scope of the current note model in use by addressing the notion of variability in svaras. We have presented two approaches that partially (ch. 6) and fully (ch. 8) exploit scores to describe pitch content in the audio music recordings, using our model. These approaches are shown to perform significantly better than the state-of-the-art parametrization of svara distributions, by a good margin in a raaga classification task.

We have qualitatively compared the svara representations computed from these approaches to those obtained using manually annotated data (ch. 7). For most svaras, there are clear correspondences between svara representations from different approaches. These include the overall shape of the distribution, the peak positions and their relative importance indicated by the peak amplitude. This clearly indicates that our approaches have been successful to a good measure in obtaining representations of svaras which can further be used in retrieving musically meaningful information.

- *Extract music concepts and the corresponding individual entities along with relevant relations from natural language text.*

  We have presented a framework for comparative evaluation of OIE systems for ontologization of thematic domains. We have demonstrated it using three OIE systems in ontologizing the Indian art music domain. The results lead us to better understand the behavior of the systems from different perspectives, which can be used to guide the work in adapting OIE to thematic domains. The main challenge with these domains is the lack of large scale data. The OIE systems rely to a great extent on repetition in the data which is a natural consequence of the scale of the web. As a result, OIE systems perform with very low recall over domains with limited data. In this context, it has become important to bring out the differences between OIE systems in terms of their performance over such domains.

  We have shown that the semantic parsing based system performs better compared to the other two that are discussed. Owing to the fact that there are very few repetitions in the data we could gather, we had to rely on inter-system agreement in concept and relation extraction for determining precision. Linking concepts and entities has been relatively less hassle-free compared to mapping relations extracted with properties in ontologies. Due to the possible variations in expressing a relation, the number of useful relation extractions that can be mapped is also scanty. For instance, all the following sentences express the same information: i) Tyagaraja is born in Thiruvarur, ii) Tyagaraja is from Thiruvarur, iii) Tyagaraja's native place is Thiruvarur. Active and passive variations of these sentences further complicate the consolidation of relation types and their mapping to ontology.

- *Build an ontology to structure information extracted from audio music recordings and natural language text.*

  We have presented the raaga ontology and the Carnatic music ontology which subsumes it besides the other extensions that include form, taala and

performer. The raaga ontology models the concepts of svara and phrase in Carnatic music to further facilitate various classification schemes that depend of properties of these substructures. We have outlined the limitations of OWL DL languages in expressing sequence information across these ontologies, which are overcome using rules alongside them.

In this chapter, we presented the use of these ontologies in building a knowledgebase encompassing information extracted from different sources, and the challenges involved in doing so. We have also elucidated the new possibilities in computing similarities between different entities taking advantage of the enriched information.

The knowledge-bases resulting from our work can be accessed at this github repository[5]. This will remain the main reference for these data dumps which will be version in conjunction with the ontologies. We will also keep the companion page[6] for this thesis updated with all the resources related to this thesis.

[5]https://github.com/gopalkoduri/knowledge-bases
[6]http://compmusic.upf.edu/node/333

Appendix A

Supplementary content for Part II

A.1 Additional plots and data for peak detection

The following plots show the impact of varying different parameters on different peak detection methods reported in sec. 5.5.

Parameter	Range (step size)
Kernel size for Gaussian filter	5 to 15 (2)
Intervallic constraint (I_C)	30 to 100 cents (10)
Peak amplitude threshold (A_T)	$1.0 \cdot 10^{-5}$ to $1.0 \cdot 10^{-4}$ ($1.0 \cdot 10^{-5}$)
Valley depth threshold (D_T)	$1.0 \cdot 10^{-5}$ to $1.0 \cdot 10^{-4}$ ($1.0 \cdot 10^{-5}$)

Table A.1: Range of values of each parameter over which grid search is performed to obtain the best combination of parameters.

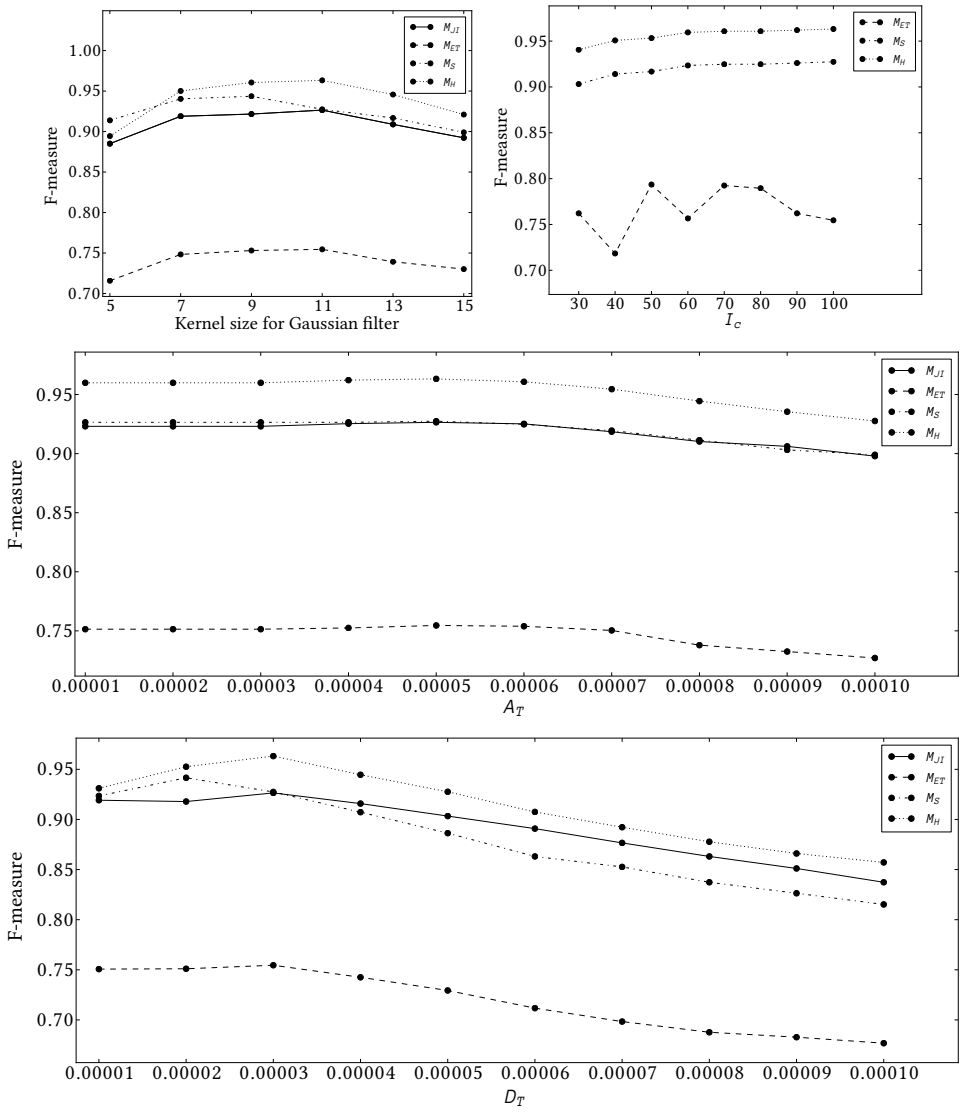


Figure A.1: Impact of varying each parameter on the four peak detection methods. Y-axis indicates values of f-measure. and X-axis indicates label and corresponding values for each parameter.

A.2 Decision table resulting from raaga classification

Following is a part of the decision tree model obtained in raaga classification test reported in sec. 6.3. It clearly indicates the prominence of the features that we introduced in intonation description, especially in the deeper sections of the tree. Notice how variance, kurtosis, variance, skew and mean are influential at various points.

```
amplitude_813 > -0.61524
|   amplitude_386 <= -1.207681
|   |   variance_498 <= -2.339884
|   |   |   variance_996 <= -1.355139: Varali (12.0)
|   |   |   variance_996 > -1.355139: Sanmukhapriya (12.0)
|   |   variance_498 > -2.339884
|   |   |   amplitude_111 <= -0.535038
|   |   |   |   kurtosis_813 <= 1.616033: Husseni (8.0)
|   |   |   |   kurtosis_813 > 1.616033
|   |   |   |   |   position_813 <= 1.617055: Husseni (3.0)
|   |   |   |   |   position_813 > 1.617055
|   |   |   |   |   |   skew2_701 <= 0.042324: Mukhari (7.0)
|   |   |   |   |   |   skew2_701 > 0.042324
|   |   |   |   |   |   |   amplitude_0 <= -0.708484: Mukhari (4.0/1.0)
|   |   |   |   |   |   |   amplitude_0 > -0.708484
|   |   |   |   |   |   |   |   mean_315 <= 1.307345: Bhairavi (11.0)
|   |   |   |   |   |   |   |   mean_315 > 1.307345: Mukhari (3.0/1.0)
|   |   |   amplitude_111 > -0.535038
|   |   |   |   amplitude_203 <= -1.935112
|   |   |   |   |   mean_813 <= 1.621428: Todi (12.0)
|   |   |   |   |   mean_813 > 1.621428: Dhanyasi (12.0)
|   |   |   |   amplitude_203 > -1.935112: Sindhubhairavi (12.0)
|   amplitude_386 > -1.207681
|   |   amplitude_498 <= -2.339884: Kamavardani/Pantuvarali (12.0)
|   |   amplitude_498 > -2.339884
|   |   |   kurtosis_498 <= 0.415956
|   |   |   |   variance_386 <= 0.826298: Mayamalavagaula (3.0)
|   |   |   |   variance_386 > 0.826298: Saveri (12.0)
|   |   |   kurtosis_498 > 0.415956: Mayamalavagaula (9.0)
```

A.3 Applications

Dunya

Dunya comprises the music corpora and related software tools that have been developed as part of the CompMusic project. 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 web interface available online[1] allows researchers and users to browse and navigate through the collections using criteria originating in metadata, cultural information, audio features and musically meaningful information extracted from the audio music recordings. The recording and raaga similarities are computed using intonation descriptions developed in this thesis.

[1] http://dunya.compmusic.upf.edu

Saraga

Saraga is an android application that provides an enriched listening atmosphere over a collection of Carnatic and Hindustani music. It allows Indian art music connoisseurs and casual listeners to navigate, discover and listen to these music traditions using familiar, relevant and culturally grounded concepts. Sarāga includes inclusive designing of innovative visualizations and inter and intra-song navigation patterns that present musically rich information to the user on a limited screen estate such as mobiles. These time synchronized visualizations of musically relevant facets such as melodic patterns, samas locations and sections provides a user with better understanding and appreciation of these music traditions. This application contains the svara distribution plots presented in a way users understand the relative distribution of pitches in a given recording. This application is available for use on Google playstore[2].

Riyaz

Riyaz is an android application that aims to facilitate music learning for beginner to intermediate level music students by making their practice (riyaz) sessions more efficient. This application includes music technologies that employ perceptually relevant models to automatically evaluate how well a student is singing compared to a reference music lesson. Students get a fine grained feedback on their singing. In this application, we intend to incorporate the intonation description developed in this thesis as a way to compare the user sung snippets to those from a known reference. This application is available for use on Google playstore[3].

[2]https://play.google.com/store/apps/details?id=com.musicmuni.saraga
[3]https://play.google.com/store/apps/details?id=com.musicmuni.riyaz

Bibliography

Each reference indicates the pages where it appears.

Sareh Aghaei. Evolution of the World Wide Web : From Web 1.0 to Web 4.0. International journal of Web & Semantic Technology, 3(1):1–10, 2012. ISSN 09762280. doi: 10.5121/ijwest.2012.3101. 58

Vincent Akkermans, Joan Serrà, and Perfecto Herrera. Shape-based spectral contrast descriptor. In Sound and Music Computing, number July, pages 23–25, 2009. 84, 85

Franz Baader. The Description Logic Handbook. Cambridge University Press, Cambridge, 2007. ISBN 9780511711787. doi: 10.1017/CBO9780511711787. 68

Sandeep Bagchee. NAD Understanding Raga Music. Business Publications Inc, 1998. ISBN 81-86982-07-8. 27

L.L. Balkwill and W.F. Thompson. A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. Music Perception, 17 (1):43–64, 1999. ISSN 0730-7829. 32, 33

Dana H Ballard. GENERALIZING THE HOUGH TRANSFORM TO DETECT ARBITRARY SHAPES. Pattern Recognntton, 11(11):111–122, 1122. 126

Shreyas Belle, Rushikesh Joshi, and Preeti Rao. Raga Identification by using Swara Intonation. Journal of ITC Sangeet Research Academy, 23, 2009. 31, 40, 82

Ashwin Bellur, Vignesh Ishwar, Xavier Serra, and Hema A Murthy. A Knowledge Based Signal Processsing Approach to Tonic Identification in Indian Classical Music. In 2nd CompMusic Workshop, pages 113–118, Istanbul, 2012. 29

Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. Scientific American, 284(5):34–43, 2001. 57, 58, 68

Tim Berners-lee, Yuhsin Chen, Lydia Chilton, Dan Connolly, Ruth Dhanaraj, James Hollenbach, Adam Lerer, and David Sheets. Tabulator: Exploring and Analyzing linked data on the Semantic Web. Swui, 2006(i):16, 2006. doi: 10.1.1.97.950. 65

S. Bhagyalekshmy. Ragas in Carnatic Music. CBH Publications, Nagercoil, 8 edition, 1990. ISBN 8185381127. 19, 144

Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. DBpedia: A Nucleus for a Web of Open Data. In ISWC, pages 722–735, 2007. 71, 74

Christian Bizer, T Heath, and T Berners-Lee. Linked data-the story so far. International Journal on Semantic Web and Information Systems, 5(3):1–22, 2009a. 57, 65

Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. DBpedia - A crystallization point for the Web of Data. Web Semantics: Science, Services and Agents on the World Wide Web, 7(3):154–165, sep 2009b. ISSN 15708268. doi: 10.1016/j.websem.2009.07.002. 74

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10):P10008, oct 2008. ISSN 1742-5468. doi: 10.1088/1742-5468/2008/10/P10008. 183

PV Bohlman. Ontologies of Music. In Nicholas Cook and Mark Everist, editors, Rethinking music. Oxford University Press, 1999. ISBN 0-19-879004-1. 140

Johan Bos, Stephen Clark, Mark Steedman, James R. Curran, and Julia Hockenmaier. Wide-coverage semantic representations from a CCG parser. In COLING, pages 1240–1246, Morristown, NJ, USA, 2004. Association for Computational Linguistics. doi: 10.3115/1220355.1220535. 160

Bar\i\cs Barıs Bozkurt, Ozan Yarman, M. Kemal Karaosmanoğlu, Can Akkoc, M Kemal Karaosmano\uglu, and Can Akkoç. Weighing Diverse Theoretical Models on Turkish Maqam Music Against Pitch Measurements: A Comparison of Peaks Automatically Derived from Frequency Histograms with Proposed Scale Tones. Journal of New Music Research, 38(1):45–70, mar 2009. ISSN 0929-8215. doi: 10.1080/09298210903147673. 88

Barış Bozkurt. Pitch Histogram based analysis of Makam Music in Turkey. In Proc. Les corpus de l'oralité, 2011. 36, 82

C Cannam, C Landone, and M Sandler. Sonic Visualiser: An Open Source Application for Viewing, Analysing, and Annotating Music Audio Files. In Proceedings of the ACM Multimedia 2010 International Conference, pages 1467–1468,

Firenze, Italy, oct 2010a. 117

Chris Cannam, Michael O. Jewell, Christophe Rhodes, Mark Sandler, and Mark D'Inverno. Linked Data And You: Bringing music research software into the Semantic Web. Journal of New Music Research, 39(4):313–325, 2010b. ISSN 0929-8215. doi: 10.1080/09298215.2010.522715. 75, 76

Ò Celma. Foafing the Music Bridging the semantic gap in music recommendation. In 5th International Semantic Web Conference (ISWC), pages 927–934, Athens, GA, USA, 2006. 68

Oscar Celma. Music recommendation. In Music Recommendation and Discovery, pages 43–85. Springer, 2010. 3

J Chakravorty, B Mukherjee, and Ashok Kumar Datta. Some Studies On Machine Recognition of Ragas In Indian Classical Music. Journal of Acoustic Society of India, Vol. XVII(3&4):1–4, 1989. 35

Parag Chordia and Alex Rae. Raag recognition using pitch-class and pitch-class dyad distributions. In International Conference on Music Information Retrieval, pages 431–436, 2007. 20, 36, 37, 82

Parag Chordia and Alex Rae. Understanding emotion in raag: An empirical study of listener responses. In Computer Music Modeling and Retrieval, pages 110–124. Springer, 2009. 32, 33

Parag Chordia, Sertan \cSentürk, and Sertan Şentürk. Joint Recognition of Raag and Tonic in North Indian Music. Journal of New Music Research, 37(3):82–98, 2013. doi: 10.1162/COMJ. 29, 82

Martin R L Clayton. Time in Indian Music: Rhythm, Metre and Form in North Indian Rag Performance. Oxford University Press, 2000. 23, 27

Arshia Cont. A coupled duration-focused architecture for real-time music-to-score alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(6):974–987, 2010. 125

Joachim Daiber, Max Jakob, Chris Hokamp, and PN Mendes. Improving efficiency and accuracy in multilingual entity extraction. In International Conference on Semantic Systems, pages 3–6, 2013. ISBN 9781450319720. 180

Dipanjan Das and Monojit Choudhury. Finite State Models for Generation of Hindustani Classical Music. In Frontiers of Research on Speech and Music, 2005. 36

Ashok Kumar Datta, R. Sengupta, Nityananda Dey, Dipali Nag, and A. Mukerjee. Objective Analysis of the Interval Boundaries and Swara-Shruti relations in Hindustani vocal music from actual performances. Journal of ITC Sangeet Research Academy, 2006. 29, 30

Alain de Cheveigné, Hideki Kawahara, Alain de Cheveigné, and Hideki Kawahara. YIN, a fundamental frequency estimator for speech and music. The Journal of the Acoustical Society of America, 111(4):1917, 2002. ISSN 00014966. doi: 10.1121/1.1458024. 41, 86

A Victor Devadoss and S. Asservatham. A Comparative Study between Properties of Carnatic Raagas and Emotions of Devotional Songs Using Bidirectional Associative Memories (BAM). International Journal of Engineering Research & Technology, 2(6):1380–1385, 2013. 32

Pranay Dighe, Harish Karnick, and Bhiksha Raj. Swara Histogram Based Structural Analysis and Identification of Indian Classical Ragas. Proc. of the 14th International Society for Music Information Retrieval Conference, 2013. 36

Subbarama Dikshitar, P. P. Narayanaswami, and Vidya Jayaraman. Sangita Sampradaya Pradarsini. Online, 1904. 18, 21, 27, 146

Simon Dixon and Gerhard Widmer. Match: A music alignment tool chest. In International Society for Music Information Retrieval Conference, pages 492–497, 2005. 125

N. Drummond, A.L. Rector, R. Stevens, G. Moulton, M. Horridge, H Wang, and J. Seidenberg. Putting OWL in order: Patterns for sequences in OWL. In 2nd OWL Experiences and Directions Workshop, International Semantic Web Conference, pages 1–14, 2006. 144

Shrey Dutta and Hema A. Murthy. DISCOVERING TYPICAL MOTIFS OF A RAGA FROM ONE-LINERS OF SONGS IN CARNATIC MUSIC. In International Society for Music Information Retrieval, pages 397–402, 2014. 31

Fredo Erxleben, Michael Günther, Markus Krötzsch, Julian Mendez, and Denny Vrandečić. Introducing wikidata to the linked data web. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8796:50–65, 2014. ISSN 16113349. doi: 10.1007/978-3-319-11964-9. 74

Oren Etzioni and Michele Banko. Open Information Extraction from the Web. Communications of the ACM, 51(12):68–74, 2008. 158

Anthony Fader, Stephen Soderland, and Oren Etzioni. Identifying Relations for Open Information Extraction. In Empirical Methods in Natural Language Processing, 2011. 159, 162

Gyorgy Fazekas and Mark B Sandler. The Studio Ontology Framework. In ISMIR, number Ismir, pages 471–476, 2011. ISBN 9780615548654. 70

György Fazekas, Yves Raimond, Kurt Jacobson, and Mark Sandler. An Overview of Semantic Web Activities in the OMRAS2 Project. Journal of New Music Research, 39(4):295–311, dec 2010. ISSN 0929-8215. doi: 10.1080/09298215.

2010.536555. xxv, 69, 75

Christiane Fellbaum. WordNet. Wiley Online Library, 1998. 4

Alfio Ferrara, Luca a. Ludovico, Stefano Montanelli, Silvana Castano, and Goffredo Haus. A Semantic Web ontology for context-based classification and retrieval of music resources. ACM Transactions on Multimedia Computing, Communications, and Applications, 2(3):177–198, aug 2006. ISSN 15516857. doi: 10.1145/1152149.1152151. 68, 72

Ben Fields and K Page. The segment ontology: Bridging music-generic and domain-specific. In International Conference on Multimedia and Expo, pages 1–6, 2011. 146

Christian Fremerey, Meinhard Müller, and Michael Clausen. Handling repeats and jumps in score-performance synchronization. In International Society for Music Information Retrieval Conference, pages 243–248, 2010. 125

Takuya Fujishima. Realtime chord recognition of musical sound: A system using common lisp music. In Proc ICMC 1999, volume 9, pages 464–467, 1999. 124

Joe Futrelle and J Stephen Downie. Interdisciplinary Communities and Research Issues in Music Information Retrieval. Library and Information Science, pages 215–221, 2002. ISSN 0929-8215. doi: 10.1076/jnmr.32.2.121.16740. 3

R. Garcia and Òscar Celma. Semantic integration and retrieval of multimedia metadata. In Proceedings of 4rd International Semantic Web Conference. Knowledge Markup and Semantic Annotation Workshop, Galway, Ireland, pages 1–12. Citeseer, 2005. 68, 72

Ali C. Gedik and Bariş Barış Bozkurt. Pitch-frequency histogram-based music information retrieval for Turkish music. Signal Processing, 90(4):1049–1063, apr 2010. ISSN 01651684. doi: 10.1016/j.sigpro.2009.06.017. 36, 82, 124

Emilia Gómez. Tonal Description of Music Audio Signals. PhD thesis, Universitat Pompeu Fabra, 2006. 124

Karen F. Gracy, Marcia Lei Zeng, and Laurence Skirvin. Exploring methods to improve access to music resources by aligning library data with linked data: A report of methodologies and preliminary findings. Journal of the American Society for Information Science and Technology, 64(10):2078–2099, 2013. ISSN 15322882. doi: 10.1002/asi.22914. 70

Sankalp Gulati. A Tonic Identification Approach for Indian Art Music. Masters' thesis, Universitat Pompeu Fabra, 2012. 29, 42, 86, 139

Sankalp Gulati, Ashwin Bellur, Justin Salamon, Ranjani H. G., Vignesh Ishwar, Hema A Murthy, and Xavier Serra. Automatic Tonic Identification in Indian Art Music: Approaches and Evaluation. Journal of New Music Research, 43 (01):55–71, 2014. doi: 10.1080/09298215.2013.875042. 29

Sankalp Gulati, Joan Serra, Vignesh Ishwar, and Xavier Serra. Mining melodic patterns in large audio collections of Indian art music. Proceedings - 10th International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2014, pages 264–271, 2015. doi: 10.1109/SITIS.2014.73. 32

Sankalp Gulati, Joan Serr, Vignesh Ishwar, and Xavier Serra. DISCOVERING R AGA MOTIFS BY CHARACTERIZING COMMUNITIES IN NETWORKS OF MELODIC PATTERNS Music Technology Group , Universitat Pompeu Fabra , Barcelona , Spain Telefonica Research , Barcelona , Spain. Icassp 2016, pages 286–290, 2016a. 32, 139

Sankalp Gulati, Joan Serr, and Xavier Serra. PHRASE-BASED R AGA RECOGNITION USING VECTOR SPACE MODELING. Icassp 2016, pages 66–70, 2016b. 11, 32

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The WEKA data mining software: an update. SIGKDD Explor. Newsl., 11(1):10–18, 2009. ISSN 1931-0145. doi: 10.1145/1656274.1656278. 85, 86, 93

Brian Harrington and Stephen Clark. Asknet: Automated semantic knowledge network. In AAAI, pages 889–894, 2007. 159

C Harte, M Sandler, S Abdallah, and E Gómez. Symbolic representation of musical chords: A proposed syntax for text annotations. In Proc ISMIR, volume 56, pages 66–71, 2005. ISBN 0955117909. 73

Andre Holzapfel, Umut Simsekli, Sertan Senturk, and Ali Taylan Cemgil. Section-level modeling of musical audio for linking performances to scores in Turkish makam music. In ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, volume 2015-Augus, pages 141–145, Brisbane, Australia, 2015. IEEE. ISBN 9781467369978. doi: 10.1109/ICASSP.2015.7177948. 125, 136

Vignesh Ishwar, Shrey Dutta, Ashwin Bellur, and HA Murthy. MOTIF SPOTTING IN AN ALAPANA IN CARNATIC MUSIC. In ISMIR 2013, Curitiba, Brazil, 2013. 31

Kurt Jacobson. Connections in Music. PhD thesis, Queen Mary University of London, 2011. 70, 73

Kurt Jacobson and Yves Raimond. An Ecosystem for Transparent Music Similarity in an Open World. In George Tzanetakis and Keiji Hirata, editors, Information Retrieval, number Ismir, pages 33–38, 2009. 70

N. A. Jairazbhoy. The raags of North Indian music, their structure and evolution. Faber, 1971. 19

S. R. Janakiraman. Essentials of Musicology in South Indian Music. The Indian

Music Publishing House, 2008. 19, 23, 146

Dan-Ning Jiang, Lie Lu, Hong-jiang Zhang, Jian-Hua Tao, and Lian-hong Cai. Music type classification by spectral contrast feature. In IEEE International Conference on Multimedia and Expo, volume 1, pages 113–116. IEEE, 2002. ISBN 0780373049. 84

Cyril Joder, Slim Essid, and Senior Member. A Conditional Random Field Framework for Robust and Scalable Audio-to-Score Matching. 19(8):1–13, 2010. doi: 10.1109/TASL.2011.2134092. 125

Divya Mansingh Kaul. Hindustani and Persio-Arabian Music: An Indepth, Comparative Study. Kanishka Publishers, Distributors, New Delhi, first edition, 2007. ISBN 81-7391-923-2. 18

Maurice George Kendall and Jean Dickinson Gibbons. Rank Correlation Methods. E. Arnold, 5th edition, 1990. ISBN 978-0195208375. 186

HG Kim, Nicolas Moreau, and Thomas Sikora. MPEG-7 audio and beyond: Audio content indexing and retrieval. John Wiley & Sons, 2006. ISBN 047009334X. 85

Peter Knees and Markus Schedl. A Survey of Music Similarity and Recommendation from Music Context Data. ACM Trans. Multimedia Comput. Commun. Appl., 10(1):2:1–2:21, dec 2013. ISSN 1551-6857. doi: 10.1145/2542205.2542206. 178

Georgi Kobilarov, Tom Scott, Yves Raimond, Silver Oliver, Chris Sizemore, Michael Smethurst, Christian Bizer, and Robert Lee. Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections. In Proceedings of the European Semantic Web Conference, pages 723–737, 2009. 70, 74

Gopala Krishna Koduri and Bipin Indurkhya. A Behavioral Study of Emotions in South Indian Classical Music and its Implications in Music Recommendation Systems. In SAPMIA, ACM Multimedia, pages 55–60, 2010. 32, 33

Gopala Krishna Koduri, Sankalp Gulati, Preeti Rao, and Xavier Serra. Rāga Recognition based on Pitch Distribution Methods. Journal of New Music Research, 41(4):337–350, 2012. doi: 10.1080/09298215.2012.735246. 95

T M Krishna. A Southern Music. HarperCollins Publishers India, 2013. 25

T M Krishna and Vignesh Ishwar. Karṇāṭik Music : Svara, Gamaka, Phraseology And Rāga Identity. In 2nd CompMusic Workshop, pages 12–18, 2012. 20, 21, 110, 116, 124, 143, 146, 161

Arvindh Krishnaswamy. On the twelve basic intervals in South Indian classical music. In Audio Engineering Society Convention, number ii, page 5903, 2003. 30, 56, 82, 124

Arvindh Krishnaswamy. Melodic atoms for transcribing carnatic music. In International Conference on Music Information Retrieval, pages 1–4, 2004. xxii, 20, 30, 39, 102

Vijay Kumar, Harit Pandya, and C. V. Jawahar. Identifying ragas in Indian music. Proceedings - International Conference on Pattern Recognition, 2014(August): 767–772, 2014. ISSN 10514651. doi: 10.1109/ICPR.2014.142. 37

Heeyoung Lee, Angel Chang, Yves Peirsman, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. Deterministic Coreference Resolution Based on Entity-Centric, Precision-Ranked Rules. Computational Linguistics, 39(4):885–916, dec 2013. ISSN 0891-2017. doi: 10.1162/COLI_a_00152. 161

Mark Levy. Intonation in North Indian Music. Biblia Implex Pvt. Ltd, New Delhi, 1982. 30, 56, 81

A Maezawa, H G Okuno, T Ogata, and M Goto. Polyphonic audio-to-score alignment based on Bayesian Latent Harmonic Allocation Hidden Markov Model. In IEEE International Conference on Acoustics, Speech and Signal Processing, pages 185–188, may 2011. doi: 10.1109/ICASSP.2011.5946371. 125

Farzaneh Mahdisoltani, Joanna Biega, and Fabian Suchanek. Yago3: A knowledge base from multilingual wikipedias. In 7th Biennial Conference on Innovative Data Systems Research. CIDR Conference, 2014. 74

Mausam, Michael Schmitz, Robert Bart, Stephen Soderland, and Oren Etzioni. Open Language Learning for Information Extraction. In Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2012. 159, 162

Wim Van Der Meer and Suvarnalata Rao. MICROTONALITY IN INDIAN MUSIC: MYTH OR REALITY ? In FRSM, 2009. 28, 29, 31

PN Mendes and Max Jakob. DBpedia spotlight: shedding light on the web of documents. In International Conference on Semantic Systems, pages 1–8, 2011. ISBN 9781450306218. 75

M. Narmada. Indian Music and Sancharas in Raagas. Somnath Dhall, Sanjay Prakashan, Delhi, 2001. ISBN 81-7453-044-4. 18, 27

Mark Newman. Networks. Oxford University Press, mar 2010. ISBN 9780199206650. doi: 10.1093/acprof:oso/9780199206650.001.0001. 180, 181

Bernhard Niedermayer. Accurate Audio-to-Score Alignment – Data Acquisition in the Context of Computational Musicology. PhD thesis, Johannes Kepler Universität, 2012. 125

Sergio Oramas, Vito Claudio Ostuni, Tommaso Di Noia, Xavier Serra, and Eugenio Di Sciascio. Sound and Music Recommendation with Knowledge Graphs. ACM Trans. Intell. Syst. Technol., 9(4):1–21, 2015. 71, 74

Tim O'Reilly. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. Communications & Strategies, 1(First Quarter): 17, 2007. ISSN 11578637. doi: 10.2139/ssrn.1008839. 58

L Page, S Brin, R Motwani, and T Winograd. The PageRank citation ranking: bringing order to the web. Technical report, Stanford InfoLab, 1999. 180, 181

G Palshikar. Simple algorithms for peak detection in time-series. In International Conference on Advanced Data Analysis, Business Analytics and Intelligence, pages 1–13, 2009. 88, 89

Gaurav Pandey, Chaitanya Mishra, and Paul Ipe. Tansen: A system for automatic raga identification. In Indian International Conference on Artificial Intelligence, pages 1350–1363, 2003. 36, 37, 38

Alexandre Passant. Measuring Semantic Distance on Linking Data and Using it for Resources Recommendations. In AAAI, pages 93–98, 2010a. 75

Alexandre Passant. Dbrec—music recommendations using DBpedia. In International Semantic Web Conference, pages 209–224, 2010b. 75

Alexandre Passant and Stefan Decker. Hey! ho! let's go! explanatory music recommendations with dbrec. The Semantic Web: Research and Applications, 1380(2):411–415, 2010. 178, 186

Georgios Petasis and Vangelis Karkaletsis. Ontology population and enrichment: State of the art. In Knowledge-driven multimedia information extraction and ontology evolution, pages 134–166. Springer-Verlag, 2011. 163, 179

A Porter, M Sordo, and Xavier Serra. Dunya: A System for Browsing Audio Music Collections Exploiting Cultural Context. In ISMIR, pages 101–106, Curitiba, Brazil, 2013. 8

Yves Raimond. A Distributed Music Information System. PhD thesis, University of London, 2008. 69, 72, 73

Yves Raimond, Samer Abdallah, Mark Sandler, and Frederick Giasson. The Music Ontology. In ISMIR, pages 1–6, 2007. 69, 140

Yves Raimond, Christopher Sutton, and Mark Sandler. Interlinking Music-Related Data on the Web. IEEE Multimedia, 16(2):52–63, apr 2009. ISSN 1070-986X. doi: 10.1109/MMUL.2009.29. 75

C. V. Raman. The Indian musical drums. Journal of Mathematical Sciences, 1(3): 179–188, 1934. ISSN 0253-4142. 83

Hema Ramanathan. Ragalaksanasangraha (Collection of Raga Descriptions). N. Ramanathan, Chennai, 2004. 27, 141

N. Ramanathan. Sruti - Its Understanding In The Ancient, Medieval And Modern Periods. Journal of the Indian Musicological Society, 12:31–37, 1981. 28, 29

Adya Rangacharya. The Natyasastra. Munshiram Manoharlal Publishers, 2010. 28

H.G. Ranjani, S. Arthi, and T.V. Sreenivas. Carnatic music analysis: Shadja, swara identification and rAga verification in AlApana using stochastic models. In Applications of Signal Processing to Audio and Acoustics (WASPAA), IEEE Workshop, pages 29–32, 2011. ISBN 9781457706936. 29, 41

Preeti Rao, Joe Cheri Ross, Kaustuv Kanti Ganguli, Vedhas Pandit, Vignesh Ishwar, Ashwin Bellur, and Hema a. Murthy. Classification of Melodic Motifs in Raga Music with Time-series Matching. Journal of New Music Research, 43 (November 2014):115–131, 2014. ISSN 0929-8215. doi: 10.1080/09298215.2013.873470. 31

Suvarnalata Rao. SHRUTI IN CONTEMPORARY HINDUSTANI MUSIC. In FRSM, pages 110–121, 2004. 28, 29, 30

Suvarnalata Rao and Wim van der Meer. The Construction, Reconstruction and Deconstruction of Shruti. In J. Bor, editor, Hindustani Music - Thirteenth to Twentieth Centuries, pages 673–696. Codarts & Manohar, New Delhi, 2010. 29

T. K. Govinda Rao. Compositions of Tyagaraja. Ganamandir Publications, Chennai, 1995. 13

T. K. Govinda Rao. Compositions of Muddusvami Dikshitar. Ganamandir Publications, Chennai, 1997a. 13

T. K. Govinda Rao. Compositions of Syama Sastri. Ganamandir Publications, Chennai, 1997b. 13

T. K. Govinda Rao. Varṇasāgaraṁ. Ganamandir Publications, Chennai, 2006. 116

JC Ross, TP Vinutha, and Preeti Rao. Detecting melodic motifs from audio for Hindustani classical music. In ISMIR, pages 193–198, 2012. 31

H Sahasrabuddhe and R Upadhye. On the computational model of raag music of india. In Workshop on AI and Music: European Conference on AI, 1992. 36

Justin Salamon and Emilia Gomez. Melody Extraction From Polyphonic Music Signals Using Pitch Contour Characteristics. IEEE Transactions on Audio, Speech, and Language Processing, 20(6):1759–1770, aug 2012. ISSN 1558-7916. doi: 10.1109/TASL.2012.2188515. 41

Justin Salamon, Sankalp Gulati, and Xavier Serra. A Multipitch Approach to Tonic Identification in Indian Classical Music. In ISMIR, number Ismir, pages 499–504, Porto, 2012. 86, 126

P. Sambamoorthy. South Indian Music (6 Volumes). The Indian Music Publishing House, 1998. 19, 27, 142

Sunita Sarawagi. Information Extraction. Foundations and Trends in Databases,

1(3):261–377, 2008. doi: 10.1561/1500000003. 158

Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In World Wide Web, pages 285–295. ACM, 2001. 178

Sertan Senturk, André Holzapfel, and Xavier Serra. Linking Scores and Audio Recordings in Makam Music of Turkey. Journal of New Music Research, 8215 (November):35–53, 2014. ISSN 17445027. doi: 10.1080/09298215.2013.864681. 126

Joan Serrà, Gopala Krishna Koduri, Marius Miron, and Xavier Serra. Assessing the tuning of sung indian classical music. In International Conference on Music Information Retrieval, pages 263–268, 2011. ISBN 9780615548654. 30, 31

X. Serra, M. Magas, E. Benetos, M. Chudy, S. Dixon, A. Flexer, E. Gómez, F. Gouyon, P. Herrera, S. Jordà, O. Paytuvi, G. Peeters, J. Schlüter, H. Vinet, and G. Widmer. Roadmap for Music Information Research. 2013. ISBN 9782954035116. 3, 67, 77, 160

Xavier Serra. A Multicultural Approach in Music Information Research. In ISMIR, pages 151–156, 2011. 5, 8, 42, 56, 67, 77, 91, 116, 139, 160, 177

Xavier Serra. Data gathering for a culture specific approach in MIR. In Proceedings of the 21st international conference companion on World Wide Web - WWW '12 Companion, page 867, New York, New York, USA, 2012. ACM Press. ISBN 9781450312301. doi: 10.1145/2187980.2188216. 42

Vidya Shankar. The art and science of Carnatic music. Music Academy Madras, Chennai, 1983. 19, 21, 23, 88, 91, 144

P Sharma and K Vatsayan. Brihaddeshi of Sri Matanga Muni, 1992. 20

Surendra Shetty and K. K. Achary. Raga Mining of Indian Music by Extracting Arohana-Avarohana Pattern. International Journal of Recent Trends in Engineering, 1(1), 2009. 36

M Sinith and K Rajeev. Hidden markov model based recognition of musical pattern in south Indian classical music. In IEEE International Conference on Signal and Image Processing, Hubli, India, 2006. 36

Malcolm Slaney. Auditory toolbox. Technical report, 1998. 85

Stephen Soderland, Brendan Roof, Bo Qin, Shi Xu, and O Etzioni. Adapting open information extraction to domain-specific relations. AI Magazine, pages 93–102, 2010. 158, 161

Yading Song, Simon Dixon, and Marcus Pearce. A survey of music recommendation systems and future perspectives. In 9th International Symposium on Computer Music Modeling and Retrieval, 2012. 4

Mohamed Sordo, Joan Serrà, Gopala Krishna Koduri, and Xavier Serra. A Method For Extracting Semantic Information From On-line Art Music Discussion Forums. In 2nd CompMusic Workshop, pages 55–60, 2012. 71

Lucia Specia, Enrico Motta, Enrico Franconi, Michael Kifer, and Wolfgang May. Integrating Folksonomies with the Semantic Web. Lecture Notes in Computer Science -The Semantic Web: Research and Applications, 4519(September 2006): 624–639, 2007. ISSN 0302-9743. doi: 10.1007/978-3-540-72667-8. 71

Rajeswari Sridhar and TV Geetha. Raga Identification of Carnatic music for music Information Retrieval. International Journal of Recent trends in Engineering, 1(1):1–4, 2009. 36, 37

Ajay Srinivasamurthy, André Holzapfel, and Xavier Serra. In Search of Automatic Rhythm Analysis Methods for Turkish and Indian Art Music. Journal of New Music Research, 43(1):94–114, jan 2014a. ISSN 0929-8215. doi: 10.1080/09298215.2013.879902. 23

Ajay Srinivasamurthy, Gopala Krishna Koduri, Sankalp Gulati, Vignesh Ishwar, and Xavier Serra. Corpora for Music Information Research in Indian Art Music. In International Computer Music Conference/Sound and Music Computing Conference, pages 1029–1036, Athens, Greece, 2014b. 8

Ajay Srinivasamurthy, Andre Holzapfel, Ali Taylan Cemgil, and Xavier Serra. A generalized Bayesian model for tracking long metrical cycles in acoustic music signals. In 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016), pages 76–80, Shanghai, China, 2016. IEEE, IEEE. 139

V Sriram. The Devadasi and the Saint: The Life and Times of Bangalore Nagarathnamma. East West Books (Madras), 2007. 18

S Staab and R Studer. Handbook on ontologies. 2009. ISBN 9783540664413. 68

Mark Steedman. The Syntactic Process. MIT Press, Cambridge, MA, USA, 2000. ISBN 0-262-19420-1. 159

Robert Stevens, Carole A Goble, and Sean Bechhofer. Ontology-based knowledge representation for bioinformatics. Briefings in Bioinformatics, 1(4):398–414, 2000. doi: 10.1093/bib/1.4.398. 4, 68

M Subramanian. Analysis of Gamakams of Carnatic Music using the Computer. Technical Report 1, 2002. 101

M Subramanian. Carnatic Ragam Thodi – Pitch Analysis of Notes and Gamakams. Journal of the Sangeet Natak Akademi, XLI(1):3–28, 2007. 31, 82, 101, 102, 124

D Swathi. Analysis of Carnatic Music : A Signal Processing Perspective. Unpublished work – the study is rather preliminary, IIT Madras, 2009. 30, 81

George Tzanetakis and Perry Cook. Musical genre classification of audio signals.

IEEE Transactions on speech and audio processing, 10(5):293–302, 2002. 84

Mike Uschold and Michael Gruninger. Ontologies : Principles , Methods and Applications. Knowledge Engineering Review, 11(2):69, 1996. 64

R. Vedavalli. Varnam - the mother of manodharma sangeetam (Part II). Sruti, pages 59–62, feb 2013a. 116

R. Vedavalli. Varnam - the mother of manodharma sangeetam (Part I). Sruti, pages 61–63, jan 2013b. 116

K.H. Veltman. Towards a Semantic web for Culture. Journal of Digital Information, 4(4), 2004. 77

T. Viswanathan and Matthew Harp Allen. Music in South India. Oxford University Press, 2004. 11, 19, 27, 104, 141

Jun Wang, Xiaoou Chen, Yajie Hu, and Tao Feng. Predicting High-level Music Semantics using Social Tags via Ontology-based Reasoning. In ismir2010.ismir.net, number Ismir, pages 405–410, 2010. 71, 74

Alicja Wieczorkowska, Ashok Kumar Datta, R. Sengupta, Nityananda Dey, and Bhaswati Mukherjee. On Search for Emotion in Hindusthani Vocal Music. Advances in Music Information Retrieval, pages 285–304, 2010. 32

Thomas Wilmering, György Fazekas, and Mark B. Sandler. The Audio Effects Ontology. Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR-2013), 2013. 70

IH Witten and E Frank. Data Mining: Practical machine learning tools and techniques, volume 36. ACM, New York, NY, USA, 2005. doi: 10.1145/2020976.2021004. 93

F Wu and DS Weld. Open information extraction using Wikipedia. In Association for Computational Linguistics, number July, pages 118–127, 2010. 159

Xian Wu, Lei Zhang, and Yong Yu. Exploring social annotations for the semantic web. Proceedings of the 15th international conference on World Wide Web - WWW '06, page 417, 2006. doi: 10.1145/1135777.1135839. 71

R Řehůřek and Petr Sojka. Software framework for topic modelling with large corpora. In Workshop on New Challenges for NLP Frameworks, LREC, pages 45–50, Valletta, Malta, may 2010. ELRA. 165

Sertan Şentürk, Sankalp Gulati, and Xavier Serra. Towards alignment of score and audio recordings of Ottoman-Turkish makam music. In International Workshop on Folk Music Analysis, pages 57–60, Istanbul, Turkey, 2014. Computer Engineering Department, Bo\u{g}aziçi University. 126