

Culture-aware approaches to modeling and description of intonation using multimodal data

Gopala Krishna Koduri

Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain.
gopala.koduri@upf.edu

Abstract. Computational approaches that conform to the cultural context are of paramount importance in music information research. The current state-of-the-art has a limited view of such context, which manifests in our ontologies, data-, cognition- and interaction-models that are biased to the market-driven popular music. In a step towards addressing this, the thesis draws upon multimodal data sources concerning art music traditions, extracting culturally relevant and musically meaningful information about melodic intervals from each of them and structuring it with formal knowledge representations. As part of this, we propose novel approaches to describe intonation in audio music recordings and to use and adapt the semantic web infrastructure to complement this with the knowledge extracted from text data. Due to the complementary nature of the data sources, structuring and linking the extracted information results in a symbiosis mutually enriching their information. Over this multimodal knowledge base, we propose similarity measures for the discovery of musical entities, yielding a culturally-sound navigation space.

1 Problem

Music traditions from around the world share a few common characteristics. Yet, they differ substantially when viewed within their geographical and cultural context [1]. Consider the fundamental melodic unit. In Indian art music, the definition of this unit, called *svara*, constitutes a frequency with an allowed space around it owing to its shared context with the neighboring *svaras*. This space is further constrained by certain motifs as allowed/disallowed by the melodic framework it belongs to, called *rāga* [2]. This formulation differs to a great extent compared to the definitions of analogous concepts in other music traditions like notes in *Makam* music of Turkey or Western classical music. Indeed, the similarity/relation¹ between *rāgas* is dictated by terms different to those concerning *makams* or scales.

The current state of the music information research (MIR) and most commercial music platforms are largely agnostic to such differences, which inhibit them from adapting their information models to meet the cultural diversity. The need for integrating and using cultural context in MIR has been emphasized time and again [3, 1]. In part, this is to address an increasing bias towards the market-driven popular music in the information models developed and used by the research and the development communities. This is

¹ The terms similarity and relation are used synonymously henceforth.

reflected even in the tasks as basic as music browsing: most of the popular music library services² structure the content using the severely limited artist-album-track schema. For music information research and technologies to evolve beyond their current limitations in this regard, it is imperative to rely on the cultural context of the music which broadens their purview and further opens up new avenues for research [1].

Within this context, the progress in semantic web technologies over the last decade assumes a significant relevance for MIR. The linked open data movement has resulted in an increasing number of structured data sources for music including Freebase, DBpedia and LinkedBrainz. This has spawned the research and development of music applications built around such resources, some of which are reviewed in sec. 2. On the other hand, development of expressive description logics [4] has resulted in ontology languages that can be used to represent complex musical concepts such as the melodic sequences/patterns [5]. In summary, the current state of the semantic web technologies turns the problem into a tenable one. Addressing it further helps in expanding the scope of linked open data resources in the music domain by combining the information extracted from different modalities of data, and testing the limitations of the ontology languages in structuring the music domain knowledge.

In a step to address the problem described, we propose: a) developing/adapting methodologies to extract musically meaningful information about the melodic units from different modalities of data (audio music recordings and text), b) structuring this information to create an integrated multimodal knowledge base, and c) developing similarity/relatedness measures over this resource to relate the entities to facilitate exploration and navigation that conforms to the cultural context of the music.

2 State of the art

Within the context-based³ MIR, there are broadly two classes of approaches to music similarity and recommendation: those which use the knowledge inherent in the contextual data in an implicit manner, and those which take advantage of the semantic web technologies. The former draws upon data sources such as tags, lyrics and unstructured text for applications ranging from playlist generation and auto tagging, to music recommendation, search engines and interfaces (see [6] for a recent review of the related work).

The latter class of approaches is a relatively recent direction in MIR, which can broadly be understood to address information management, music similarity and recommendation. Raimond developed the Music Ontology, which builds on top of several existing ontologies, defining generic music-specific vocabulary [7]. It allows describing music creation workflows, temporal events and editorial metadata. Recently, Oramas proposed a methodology to structure the social media content using the community-generated data such as tags, with formal knowledge representations [8].

Foafing-the-music [9] uses multiple data sources such as contextual-data from RSS feeds, content-based features and user profiles, and links them by describing the data

² such as iTunes, Amarak, Windows media player etc.

³ *Context* in MIR refers to non-audio data about music such as lyrics, community-generated data etc.

using a common ontology. This information is further used for music recommendation. Jacobson *et al* [10] show that the network structure of artists on Myspace correspond to the genres. This data was published using music ontology [7] to be used by other music applications. DBrec system [11] uses a linked-data based semantic distance to relate musical entities, based on the link structure of the music-related resources on DBpedia, resulting in explanatory recommendations.

In the content-based⁴ MIR, there is a steadily growing interest in incorporating the musicological knowledge in the information retrieval methodologies. Srinivasamurthy *et al* [12] review several state-of-the-art approaches to rhythm description and evaluate them on Indian art music traditions and Makam music of Turkey. Gulati *et al* [13] evaluate various tonic identification approaches over Indian art music traditions using diverse datasets under different conditions of the data (E.g., duration of the input audio clip, presence/absence of metadata etc). Rao *et al* [14] propose methods to classify melodic patterns in Indian art music and evaluate them using retrieval experiments.

In summary, the context- and content-based approaches are advancing to incorporate culture-specific information. However, to fully realize the potential benefits in this direction, we believe that a) the information extracted from different modalities must be used to complement and enrich each other and b) the approaches themselves must have an integrated access to the information extracted from one another.

3 Proposed Approach

3.1 Methodology

Creating a multimodal knowledge base for a music involves content-based analyses: melodic (tonic identification, motif detection, intonation description and so on) and rhythmic (downbeat estimation, cycle tracking and so on), and context-based analyses. We choose two of these tasks, one from each, to be addressed in this thesis: intonation description, and using the linked data cloud and adapting open information extraction (Open IE) for knowledge base population of a given music. Fig. 1 shows an overview of our approach, which is detailed in the rest of this section.

Intonation: It is a fundamental music concept that is relevant to most melodic music traditions. It is characteristic of the melodic framework and key to the musical expression of the artist. Describing intonation is of importance to several music information retrieval tasks. The research questions that we address are: *How can we computationally model the concept of intonation? How concisely can we describe the intonation of notes/svaras in a given music recording? In what ways is this description useful to relate the musical entities? How do we adapt the method to a given music tradition?*

Open IE for Knowledge Base population: In the past decade, domain-independent approaches to information extraction have paved way for its web-scale applications. Open IE is one such paradigm that has been used to extract assertions from unstructured data at web-scale with a considerable success [15]. Until recently, domain-specific approaches

⁴ Content in MIR is analogous to audio data.

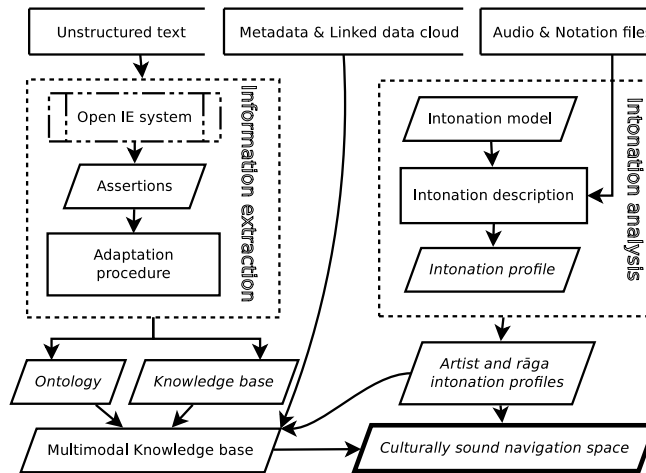


Fig. 1. An overview of the proposed approach.

to information extraction from text required manual knowledge engineering as a prerequisite [16]. The Open IE approaches, however, do not require a pre-specified vocabulary and/or relation-specific input. Therefore, adapting them to information extraction from thematic domains would alleviate the need for manual knowledge engineering.

The process of channeling the assertions extracted from these approaches into a coherent knowledge base poses certain challenges. There has been little work so far to identify and address such issues. The advances in Open IE, including the recent systems such as NELL⁵, are largely directed towards taking advantage of the volume of web 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. The research questions addressed in this direction are: *How do we process the assertions from an open information extraction system to be used for knowledge base population? Given a thematic domain, how do we adapt open IE systems?*

Multimodal Knowledge Base: Let us consider an example to understand the potential benefits of structuring and combining information obtained from content- and context-based analyses. Consider two artists who are said to be similar. For a listener to appreciate/understand the similarity, it is necessary to quantify this relation (how similar are they?) and provide a meaningful description (why are they similar?).

Structuring the information using ontologies and combining the different sources using the linked data framework reveals non-trivial, deeper relations between entities. The content-based analyses give us an objective view of such relations between entities (E.g., Are the two artists singing a particular phrase in a similar way? Is there a lot of variation in how they intone svaras?). Let us say that the two artists in our example are similar by this analyses. The context-based analyses reveal information that is not part of the audio content: biographical details, social connections, popularity etc. This information gives us the possible reasons for their similarity (E.g., They received train-

⁵ <http://rtw.ml.cmu.edu/rtw/>

ing from the same teacher, or they belong to same lineage, etc.). The research questions that we address towards achieving this are: *How do we represent the complex musical concepts related to intonation in an ontology? What are the ways to combine the information extracted from content- and context-based approaches? How do we model culturally-relevant similarity measures over the multimodal knowledge base?*

3.2 Data repertoires

The two art music traditions of the Indian subcontinent - Carnatic and Hindustani - are among the oldest yet actively practiced. With a thriving community of scholars, there are thorough musicological and cultural studies about them. The terminology and the structuring of the knowledge in these music traditions differ substantially from the information models in vogue in the MIR community [2]. Therefore, they make a suitable and challenging test-ground for developing culture-aware information models. The data for our work include audio, metadata and community-generated text data. The audio collection is carefully curated with experts' inputs and the corresponding metadata is stored on MusicBrainz, whose schema is accordingly extended to incorporate the new concepts and relation-types. Table. 1 shows statistics of the current snapshots of Carnatic and Hindustani music collections⁶. The community-generated data includes Wikipedia pages, online discussion forums and web pages concerning these music traditions.

	Releases	Recordings	Hours	Artists	Rāgas
Carnatic	248	1650	346	233	246
Hindustani	233	1096	300	360	176

Table 1. Statistics of the Carnatic and Hindustani music collections.

4 Evaluation

We quantitatively evaluate the different parts of the thesis. As intonation is a characteristic of the rāga, we use rāga classification task for a comparative evaluation with the state-of-the-art. We assess the changes we effect to an Open IE system by comparing the results to that of the original system over a text corpus labeled with the desired concepts and the relation-types. Further we also perform an extrinsic evaluation on the tasks of knowledge base population, using a gold standard consisting of manually engineered ontologies⁷ [5].

On the other hand, the similarity measures developed on the multimodal knowledge base will be implemented and tested within a music discovery platform, Dunya⁸ [17]. It is an online culture-aware navigation and discovery system being developed as part of the CompMusic project [1] to which this thesis will contribute. The similarity computed between the musical entities will be used for generating recommendations to a query entity which the user is interested in. Users of this system will have access to both the

⁶ For further information, see <http://compmusic.upf.edu/corpora>

⁷ Available at <https://github.com/gopalkoduri/ontologies>

⁸ Available at <http://dunya.compmusic.upf.edu/>

baseline (state-of-the-art) music similarity algorithms and the ones we develop. We use the logged activity of the users to analyze their subjective preferences and evaluate the similarity measures.

5 Results

Intonation: The information model commonly used in MIR for melodic notes and intervals identifies notes as a set of discrete points on frequency spectrum separated by certain intervals. One important limitation of this model is that it cannot account for the variability in a given note/svara. Therefore, we proposed a statistical model of notes, that defines the notes as probabilistic phenomenon on frequency spectrum. We have so far proposed three different approaches based on this model which are discussed in detail in [18]. In the first method, we compute the characteristics (such as mean, standard deviation, skewness etc.) of distribution of pitch values around each interval in the pitch-histogram. The improvement over the state-of-the-art which uses just the interval locations and their frequency of occurrence, is not significant. However, using feature selection task, we discovered that they do carry useful information to distinguish rāgas. This aggregate approach discards the contextual information of pitches: mainly the melodic & temporal contexts. Our second approach addresses this by obtaining this contextual information to a limited extent using a moving window as detailed in [18]. This resulted in an intonation description that distinguished rāgas better. Table. 2 shows the results for rāga classification task using the two methods.

	Naive Bayes	3-Nearest Neighbours	SVM	Random forest	Logistic regression	Multilayer Perceptron
Histogram	78.26	78.46	71.79	81.16	78.61	78.78
Cont. info	82.63	82.83	79.90	82.69	81.11	82.17

Table 2. Accuracies obtained using different classifiers in the rāga classification experiment using the Histogram and Contextual-information based approaches for intonation description.

To further improve it qualitatively, we have semi-automatically annotated few songs that are sung to notation without much improvisation, with svara labels⁹. The distribution for each svara is obtained by constructing a histogram of pitch values from its contours. These are verified by domain experts to be representative and meaningful of their respective rāgas. Therefore, our current direction of work involves using the notation alongside the audio to automate the labeling of melodic contours with svaras. We then use the svara contours and the aggregate distribution of all the corresponding pitch values in obtaining a more accurate intonation description (see fig. 1).

OpenIE for Knowledge Base population: In order to validate our hypothesis of diversity in music quantitatively, we have developed an approach that rank-orders the characteristics of a given music based on their salience using a text corpus. The approach uses cross references between the musical entities (E.g., individual instances of *Composers*) in their respective text descriptions to compute their relevance in the music. Then, the

⁹ This dataset is openly available at <http://compmusic.upf.edu/carnatic-varnam-dataset>

different characteristics of the music (E.g., Composers) are assigned a rank based on the importance of the constituent entities. We employed this approach on the Indian art music traditions, Baroque, Flamenco and Jazz. The results support our hypothesis showing the distinct features of each music differentiating it from the rest, such as the prominence of religion in Carnatic music, dance in Flamenco, and gharanas/schools in Hindustani music, and so on. We have further developed a rudimentary distance measure using this knowledge to compute the similarity between various musical entities within a given music. The work in progress involves structuring this knowledge using our ontologies and improving the distance measure.

On the other hand, as a first step in the direction of adapting Open IE systems to knowledge base population, we devised a comprehensive framework for a comparative evaluation. The first part of the framework compares the volume of extracted assertions along different aspects: sentences, entities, relation-types and concepts/classes, with an aim to understand the coverage of the domain quantitatively. In the second part of the framework, the assertions are used in three fundamental tasks of knowledge base population: entity identification, concept identification, and semantic relation extraction. The framework is demonstrated using Indian art music domain, and the results from each task are validated against structured content in Wikipedia and/or using manually engineered ontologies. The results from the two parts of the framework, when juxtaposed against each other, gave us concrete insights into the differences between the performances and the nature of the approaches. The results indicate that the approach based on deep semantic parsing [19, 20] performs better compared to the ones based on shallow semantic parsing (or semantic role labeling)¹⁰ [21] and dependency parsing [22]. Our current line of work in this direction involves adapting the deep semantic parsing based approach to music domain, specifically Indian art music.

6 Conclusions and Future work

We have discussed the need for culture-aware approaches in music information research. Towards addressing this, we have identified the specific issues that will be dealt with in the thesis: intonation description and knowledge base population of music domain using Open IE. We presented the methodology and the results from our work so far. Both the branches (information extraction and intonation analysis) shown in the flowchart in Fig. 1 are far into development, and the results are reaffirming of our hypothesis that using cultural-context can lead to better music similarity measures. Our future work involves:

- Using the ontologies to structure and combine the information obtained from audio and text sources. Two major challenges involved are: disambiguation of concepts and relation-types, and linking the information obtained from different modalities.
- Developing similarity measures that conform to the cultural context of the music.

Acknowledgments: This thesis is supervised by Dr. Xavier Serra, and was partly funded by the European Research Council under the European Union's Seventh Framework Program, as part of the CompMusic project (ERC grant agreement 267583).

¹⁰ We used the implementation available at <https://github.com/knowitall/openie>

References

- [1] Serra, X.: A Multicultural Approach in Music Information Research. In: ISMIR. (2011) 151--156
- [2] Krishna, T.M., Ishwar, V.: Karnāṭik Music : Svāra, Gamaka, Phraseology And Rāga Identity. In: 2nd CompMusic Workshop. (2012) 12--18
- [3] Serra, X., Magas, M., Benetos, E., Chudy, M., Dixon, S., Flexer, A., Gómez, E., Gouyon, F., Herrera, P., Jordà, S., Paytūvi, O., Peeters, G., Schlüter, J., Vinet, H., Widmer, G.: Roadmap for Music Information Research. (2013)
- [4] Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F., eds.: The Description Logic Handbook. Cambridge University Press, Cambridge (2007)
- [5] Koduri, G.K., Serra, X.: A knowledge-based approach to computational analysis of melody in Indian art music. In: International Workshop on Semantic Music and Media colocated with International Semantic Web Conference. (2013) 1--10
- [6] Knees, P., Schedl, M.: A Survey of Music Similarity and Recommendation from Music Context Data. *ACM Trans. Multimedia Comput. Commun. Appl.* **10**(1) (December 2013) 2:1--2:21
- [7] Raimond, Y.: A Distributed Music Information System. PhD thesis, University of London (2008)
- [8] Oramas, S.: Harvesting and Structuring Social Data in Music Information Retrieval. In: Extended Semantic Web Conference (ESWC 2014), Crete, Springer, Springer (2014)
- [9] Celma, O.: Foafing the Music Bridging the semantic gap in music recommendation. In: International Semantic Web Conference, Athens, GA, USA (2006) 927--934
- [10] Jacobson, K., Sandler, M., Fields, B.: Using Audio Analysis and Network Structure to Identify Communities in On-Line Social Networks of Artists. In: ISMIR. (2008) 269--274
- [11] Passant, A., Decker, S.: Hey! ho! let's go! explanatory music recommendations with dbrec. *The Semantic Web: Research and Applications* **1380**(2) (2010) 411--415
- [12] Srinivasamurthy, A., Holzapfel, A., Serra, X.: In Search of Automatic Rhythm Analysis Methods for Turkish and Indian Art Music. *Journal of New Music Research* **43**(1) (January 2014) 94--114
- [13] Gulati, S., Bellur, A., Salamon, J., H. G., R., Ishwar, V., Murthy, H.A., Serra, X.: Automatic Tonic Identification in Indian Art Music: Approaches and Evaluation. *Journal of New Music Research* **43**(01) (2014) 55--71
- [14] Rao, P., Ross, J.C., Ganguli, K.K., Pandit, V., Ishwar, V., Bellur, A., Murthy, H.a.: Classification of Melodic Motifs in Raga Music with Time-series Matching. *Journal of New Music Research* **43**(1) (January 2014) 115--131
- [15] Etzioni, O., Banko, M.: Open Information Extraction from the Web. *Communications of the ACM* **51**(12) (2008) 68--74
- [16] Sarawagi, S.: Information Extraction. *Foundations and Trends in Databases* **1**(3) (2008) 261--377
- [17] Porter, A., Sordo, M., Serra, X.: Dunya: A System for Browsing Audio Music Collections Exploiting Cultural Context. In: ISMIR, Curitiba, Brazil (2013) 101--106
- [18] Koduri, G.K., Ishwar, V., Serrà, J., Serra, X.: Intonation Analysis of Rāgas in Carnatic Music. *Journal of New Music Research* **43**(01) (January 2014) 72--93
- [19] Harrington, B., Clark, S.: Asknet: Automated semantic knowledge network. In: AAAI. (2007) 889--894
- [20] Steedman, M.: *The Syntactic Process*. MIT Press, Cambridge, MA, USA (2000)
- [21] Mausam, Schmitz, M., Bart, R., Soderland, S., Etzioni, O.: Open Language Learning for Information Extraction. In: Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. (2012)
- [22] Fader, A., Soderland, S., Etzioni, O.: Identifying Relations for Open Information Extraction. In: Empirical Methods in Natural Language Processing. (2011)