# A MUSICALLY AWARE SYSTEM FOR BROWSING AND INTERACTING WITH AUDIO MUSIC COLLECTIONS

**Mohamed Sordo, Gopala K. Koduri, Sertan Şentürk, Sankalp Gulati and Xavier Serra**
[1] Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain.
{mohamed.sordo,gopala.koduri,sertan.senturk,sankalp.gulati,xavier.serra}@upf.edu

## ABSTRACT

In the context of the CompMusic project we are developing methods to automatically describe/annotate audio music recordings pertaining to various music cultures. As a way to demonstrate the usefulness of the methods we are also developing a system to browse and interact with specific audio collections. The system is an online web application that interfaces with all the data gathered (audio, scores plus contextual information) and all the descriptions that are automatically generated with the developed methods. In this paper we present the basic architecture of the proposed system, the types of data sources that it includes, and we mention some of the culture specific issues that we are working on for its development. The system is in a preliminary stage but it shows the potential that MIR technologies can have in browsing and interacting with music collections of various cultures.

## 1. INTRODUCTION

Most music traditions around the world differ in the way their music is produced, used, and understood. This is because each music tradition evolves together with the community that supports and enjoys that music, that is, their music is influenced by their culture, language, geography, and in general by their personality.

The development of computer-based tools for accessing and listening to audio recordings has to consider the cultural context of both the music and the user. Most existing commercial tools, normally referred as audio players, are heavily biased towards Western commercial music. These tools allow users to search and navigate through the music catalogs efficiently by using basic metadata (such as title, album name or artist name) but they lack tools for more complex ways to filter, navigate, and especially to explore the specific musical concepts that characterize a given type of music.

Listening to an audio recording is a significant part of our interaction with a musical work, but the social context, the description of the works, musicians, and of the musical concepts related to the music are relevant information that complement and enrich our musical experience. The developed application enriches the process of searching and listening to music recordings by taking advantage of the results of the CompMusic project [1], thus demonstrating the use of the technologies developed in a practical context.

Next we mention some of the related applications that have been developed and in the followings sections we go over the components of our proposed system.

## 2. BACKGROUND

Most audio players, such as iTunes [1], are aimed at listening to music while providing users limited access to audio metadata. Others, like Amarok [2], also access additional information sources related to the music from Wikipedia. Online music streaming services such as Grooveshark [3] and Spotify [4] demonstrate a similar music experience while hosting large-scale and easily accessible audio music sources with a social layer. An engaging example of social interaction is the capability of posting timed commentaries on the waveform visualization of a music recording in SoundCloud [5].

An interesting web application that provides richer browsing capabilities is Freesound [6]. Freesound is an online repository of free audio samples [2] that has been developed over the past few years within our research group, which also provides contextual metadata such as geotags and audio descriptors. This repository, though, is designed for browsing through sounds samples, not for delivering a music listening experience, which is the aim of our application here. A web application with a functionality and research goals similar to our own is Songle [3]. Songle is an online music service that promotes active listening by computing and visualizing structure, beat, melody and chord related descriptors. Songle also allows the users to edit and correct any mistakes in the automatically extracted features. However this system has been designed to be of relevance for commercial pop music, thus the features extracted from the audio recordings may not be relevant to every music culture (e.g., chords).



[1] http://itunes.apple.com
[2] http://amarok.kde.org
[3] http://grooveshark.com
[4] http://spotify.com
[5] http://soundcloud.com
[6] http://freesound.org

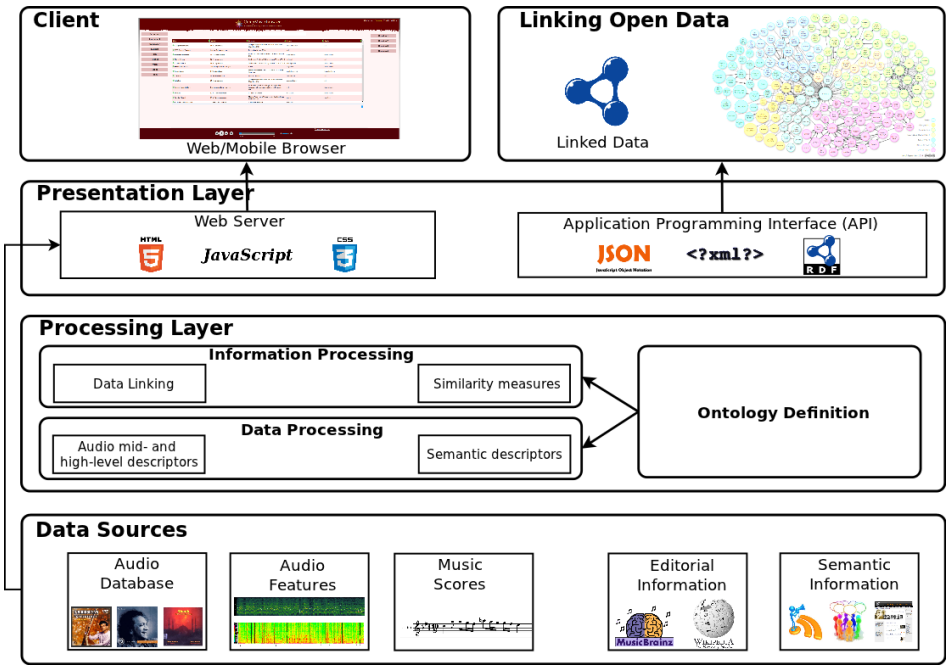


**Figure 1**. Architecture diagram of the proposed system.

## 3. SYSTEM OVERVIEW

The proposed system integrates many types of data and information related to the audio tracks of a music collection and it has an interface that allows a user to navigate through all the information in a musically meaningful way. Figure 1 shows the components of the complete system as we envision it, but only part of it is actually implemented and described here.

### 3.1 Data sources

The system keeps some of the data sources in a database and others are fetch from data repositories accessible through web APIs. Here we briefly describe the different types of data that are integrated and then used by the web application.

#### *3.1.1 Audio recordings*

The first task in building the system has been to gather a representative audio collection of the various music cultures that we are studying (currently Hindustani, Carnatic and Turkish makam). We made a selection from which we could start carrying out research on several musically relevant problems. Experts advised us and we bought around 200 commercial CDs for each repertoire plus we got access to some personal CD collections, gathering more than 300 hours of audio recordings for each collection. The size will grow in time and we aim at reaching 500 hours of audio per collection in the next few years. For the selection of the CDs it was important to choose recordings by recognized and representative artists, with reliable editorial data. The audio tracks that are actually accessible from the system interface are the ones whose metadata is available in MusicBrainz [7] and thus have an identifier that can be used to link the audio tracks with all the other available information. The web interface displays the audio data and it has a simple player to interact with it.

#### *3.1.2 Audio features*

Various low to mid level audio features from the audio recordings of the music collections are extracted and kept together with the audio recordings. Given that we are currently focusing on melodic and rhythmic dimensions of the music we have extracted low level audio features such as: perceptual amplitude, onsets, and predominant pitch [4]. For the Indian music collections, the tonic pitch of the lead performer is also extracted [5]. The Essentia library [6], an audio analysis library developed by our research group that includes most of the common low and mid level feature analysis algorithms, is used to compute these descriptions. We are currently doing research on various culture specific descriptors that will be integrated into Essentia and used in the system as they become available. For example we are working on intonation analysis [7], motivic analysis [8] or rag recognition [9]. The audio features are stored in the system using the YAML format [8].

The web application offers visualization of different audio features, aligned in time with the audio data. In the current version we have adapted the RepoVizz visualization tool [10] and integrated it into the Web interface, as shown in Figure 2. Visualizing the features while listening to the audio adds another dimension to the musical experience.

#### *3.1.3 Music scores*

Each musical culture studied in the CompMusic project uses some form of symbolic representations for their music. In Turkish makam music an extended version of West-

[7] http://musicbrainz.org

[8] http://en.wikipedia.org/wiki/YAML

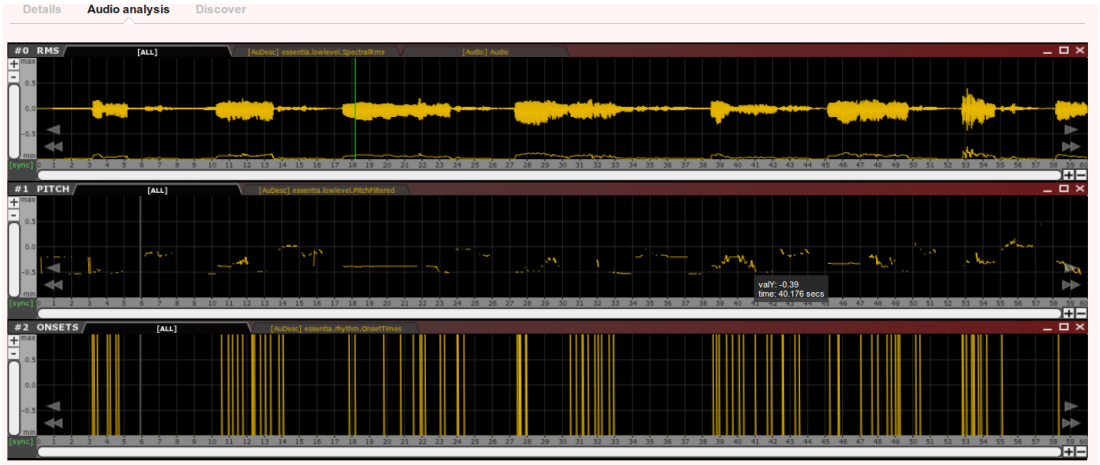


**Figure 2**. Dynamic and interactive visualization of audio features on the system's web interface.

ern classical notation [11] is very much used. Hindustani music uses Bhatkhande notation [12] and Carnatic music uses Dikshitar's notation [13], but these notations are not used much by performers, being mainly used for archival purposes.

To store the scores in machine readable format, we have considered Humdrum [14] or MusicXML [9]. Currently we have 1,700 scores from Turkish makam [15] and their integration into the system is in progress. The specific format to be used has not been decided but most probably will be MusicXML. When not available in a public repository accesible through a web API, the machine readable scores will be stored with the audio recordings and the audio features. The audio recordings and the scores of the same compositions will be linked using MusicBrainz and displayed in a synchronized way.

### 3.1.4 *Editorial metadata*

Every audio recording that we have gathered is accompanied by editorial metadata. Since most audio recordings come from commercial CDs, the editorial metadata comes from the cover or the booklet accompanying the CD. We use MusicBrainz to store and access all this metadata, which includes names of recordings, releases, compositions, composers, performers, and other culture-specific musical concepts. Most of the metadata of the audio recordings obtained was not yet in MusicBrainz, thus we have had to add it ourselves.

MusicBrainz is an open repository of music metadata. It supports all the metadata associated to CDs plus other detailed information about the music. It is designed in such a way that it keeps information about the relations among the previously mentioned musical concepts, thus providing an ontology of music metadata. This metadata is accessible via a web service and it can be easily integrated in the system. However MusicBrainz was designed to support western popular music and it lacks the support for some of our culture specific concepts. We are working closely with the MusicBrainz community to help develop the MusicBrainz framework so that it can better support the the music repertoires we are working on. Table 1 shows the statistics of the three main music collections we have uploaded to MusicBrainz, which is a subset of all the audio recordings that we have gathered, thus we are still in the process of completing them.

| Collection | CDs | Recordings | Performers |
|---|---|---|---|
| Carnatic | 196 | 1001 | 60 |
| Hindustani | 136 | 495 | 64 |
| Turkish-makam | 100 | 1011 | 116 |

**Table 1**. Statistics of the CompMusic collection in MusicBrainz.

There are other information resources than can be used to complement the editorial metadata obtained from MusicBrainz. One such resource is Wikipedia [10], a dynamic and evolving encyclopedia repository of universal knowledge. The complementary information that can be found in Wikipedia, and automatically retrieved with its API, includes artist biographies, description of musical concepts (such as raagas, taalas, makams, etc.) plus other editorial information. All the editorial information is automatically fetched and displayed in the Web interface of our system.

### 3.1.5 *Semantic information*

The previous data sources correspond to information that is either obtained from music editors, communities of experts, or from the audio itself. Another way of obtaining information about our music collections is through user generated content, i.e., information provided by users in a collaborative manner. This includes blogs, album reviews, dedicated websites, social tags, or discussion forums. We have started gathering semantic information from an online dedicated forum of Carnatic music lovers, Rasikas.org [11], in which users engage in many types of discussions, covering most relevant Carnatic music related topics. In our preliminary work [16] we have extracted and analyzed some

[9] http://www.makemusic.com/musicxml

[10] http://wikipedia.org

[11] http://rasikas.org


**Figure 3**. A screenshot of the current user interface of the system.

semantic relations between Carnatic musical concepts. The results of this research has not yet been incorporated into the system.

### 3.2 Processing layer

In this layer we include the modules that process the data sources gathered and obtain the higher level information elements and the other representations that are needed for the different functionalities of the system. This includes the extraction of audio features and the algorithms that will process all the different data sources, in order to get musically meaningful semantic concepts and the distance measures needed to navigate through the different information objects. This is the part of the system that will evolve the most during the course of the CompMusic project. Currently only part of the extraction of audio features is available.

The audio feature extraction algorithms are all integrated into the Essentia library. We will further develop culture specific algorithms to extract melodic and rhythmic characteristics of the different musical repertoires. We are working on the automatic segmentation of the pieces, on the characterization of rhythmic patterns and on the characterization of melodic motives. From these descriptions we should be able to describe the music pieces and their basic music elements, which are very much related to the ragas and talas for Indian music and to the makams and usuls for Turkish music.

In order to go a step further in the description of a musical repertoire we need to elaborate domain specific ontologies. With them we can guide the extraction of the proper semantic concepts and formally represent them together with their relationships. The system can use these ontologies to give the user a musically meaningful interaction with all the available information entities of a given music collection.

The basic mechanism with which we will be able to navigate through the information entities of a given collection is by musically meaningful similarity measures. Examples of these entities can be the actual pieces, a given performer, or specific music elements such as a musical phrase, a rhythmic patterns, or an expressive articulation. Thus we will be able to explore all the musical elements of a musical collection through similarity links.

### 3.3 Presentation layer

The presentation layer of the system includes the interface with users or with other applications. Thus it has a web interface and an API.

#### *3.3.1 Web client*

From the web browser interface the user can access and interact with all the data and information available. In the current version (Figure 3), the main functionality is to access the audio recordings filtered by music concepts which are specific to each culture. These filters are conceived from the near ontological representation of metadata in Musicbrainz, but also from other sources like Wikipedia. The web client uses Ajax [12] calls to retrieve the metadata. Once an audio track is selected we can listen to it while displaying the various audio features (e.g., pitch and onsets) and other musical concepts (e.g., motives) which are extracted from the audio. Future versions of the web client will allow more display and navigation capabilities, including user personalization, session capabilities, annotation and edition of audio features.

#### *3.3.2 Linking open data*

All the information gathered and processed within the system will be published as Linked Data, structured data that can be interlinked with other web resources, and integrated into the Linking Open Data project [17]. The Linking Open Data project is an initiative by the World Wide Web

[12] `http://www.adaptivepath.com/ideas/ajax-new-approach-web-applications`

consortium (W3C [13] ) that encourages web applications to publish their data in a structured way, so that it can be shared and accessed by other web applications. Many of the mentioned data sources in Sections 3.1.4 and 3.1.5 are also part of the Linking Open Data project. This allows the proposed system to access up-to-date data coming from these sources in a structured fashion.

## 4. CONCLUSIONS

In this paper we presented a system for browsing and interacting with audio music collections that is aware of the characteristics of the musical style. It uses knowledge of specific music traditions in order to provide navigation tools that make sense within that musical context. Its main goal is to demonstrate the technologies developed in the CompMusic project in a practical application.

We have presented the basic architecture of the system, the data sources used and we have mentioned some of the research that we are currently doing within the CompMusic project. The system is under active development and it will evolve as we obtain more research results. Specially we should be able to automatically generate more types of higher level semantic information from the available data and we should have more musically meaningful browsing mechanisms.

### Acknowledgments

The CompMusic project is funded by the European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement 267583.

## 5. REFERENCES

[1] X. Serra, "A multicultural approach in music information research," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, Miami, FL, USA, 2011.

[2] V. Akkermans, F. Font, J. Funollet, B. De Jong, G. Roma, S. Togias, and X. Serra, "Freesound 2: An improved platform for sharing audio clips," in *12th International Society for Music Information Retrieval Conference (ISMIR)*, Miami, FL, USA, 2011.

[3] M. Goto, K. Yoshii, H. Fujihara, M. Mauch, and T. Nakano, "Songle: A web service for active music listening improved by user contributions," in *Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR)*, Miami, FL, USA, 2011, pp. 311–316.

[4] J. Salamon and E. Gómez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, pp. 1759–1770, 2012.

[5] J. Salamon, S. Gulati, and X. Serra, "A multipitch approach to tonic identification in indian classical music," in *Proceedings of the 13th International Society for Music Information Retrieval Conference*, Porto, Portugal, 2012.

[6] N. Wack, "Essentia & gaia: Audio analysis and music matching c++ libraries developed by the music technology group," http://mtg.upf.edu/technologies/essentia, Oct. 2011, last accessed Oct. 2011.

[7] G. K. Koduri, J. Serrà, and X. Serra, "Characterization of intonation in carnatic music by parametrizing pitch histograms," in *Proceedings of the 13th International Society for Music Information Retrieval Conference.*, Porto, Portugal, Oct. 2012.

[8] J. C. Ross, T. P. Vinutha, and P. Rao, "Detecting melodic motifs from audio for Hindustani classical music," in *Proceedings of the 13th International Society for Music Information Retrieval Conference*, Porto, Portugal, Oct. 2012.

[9] G. K. Koduri, S. Gulati, P. Rao, and X. Serra, "Raga Recognition based on Pitch Distribution Methods," *Journal of New Music Research*, in press.

[10] O. Mayor, J. Llop, and E. Maestre, "RepoVizz: A Multimodal On-line Database and Browsing Tool for Music Performance Research," in *Proceedings of the 12th International Society for Music Information Retrieval Conference*, 2011.

[11] E. Popescu-Judetz, *Meanings in Turkish Musical Culture*. Istanbul: Pan Yayıncılık, 1996.

[12] V. Bhatkhande, *Hindustani Sangeet Paddhati*. Sangeet Karyalaya, 1934.

[13] S. Dikshitar, *Sangita Sampradaya Pradarsini*. Ettayapuram: na, 1904.

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

[15] K. Karaosmanoğlu, "A Turkish makam music symbolic database for music information retrieval: Symbtr," in *Proceedings of the 13th International Society for Music Information Retrieval Conference*, 2012.

[16] M. Sordo, J. Serrà, G. K. Koduri, and X. Serra, "Extracting semantic information from an online carnatic music forum," in *Proceedings of the 13th International Society for Music Information Retrieval Conference*, Porto, Portugal, 2012.

[17] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data - the story so far," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 5, no. 3, pp. 1–22, 2009.

[13] `http://www.w3.org/`