Journal of Computer and System Sciences 78 (2012) 1219–1231

Contents lists available at SciVerse ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Semantic annotation of digital music

Fazilatur Rahman *, Jawed Siddiqi

Sheffield Hallam University, Faculty of Arts, Computing, Engineering and Sciences, Sheffield S1 2NU, UK

A R T I C L E I N F O

Article history:
Received 25 March 2011
Received in revised form 22 July 2011
Accepted 17 October 2011
Available online 28 October 2011

Keywords:
Semantic
Ontology
Music search
Retrieval
Annotation
MPEG-7

A B S T R A C T

In recent times, digital music items on the internet have been evolving in a vast information space where consumers try to find/locate the piece of music of their choice by means of search engines. The current trend of searching for music by means of music consumers' keywords/tags is unable to provide satisfactory search results. It is argued that search and retrieval of music can be significantly improved provided end-users' tags are associated with semantic information in terms of acoustic metadata – the latter being easy to extract automatically from digital music items. This paper presents a lightweight ontology that will enable music producers to annotate music against MPEG-7 description (with its acoustic metadata) and the generated annotation may in turn be used to deliver meaningful search results. Several potential multimedia ontologies have been explored and a music annotation ontology, named ***mpeg-7Music***, has been designed so that it can be used as a backbone for annotating music items.

1. Introduction

Search engines such as Google and Yahoo provide access to music resources available on the World Wide Web (WWW) by producing results based on keywords and therefore do not take into account the type of users searching for the music items. But, music information retrieval needs to be tailored to fit the tastes and needs of individual listeners [34]. Traditionally, music files are organized using song title, singer and other textual tags (known as syntactic metadata) but such techniques are insufficient for modern users [21] to find their music items of choice. Due to increasing number of music collections traditional browsing of folder hierarchies or search by title and album name tend to be insufficient and as a result finding novel ways of music resource organization has become a research issue [25].

The **general search problem** of digital audio is a huge challenge due to the lack of contextual information available to guide the search. The problem with traditional search engines is the lack of semantics or meanings; they can only find pages that contain the chosen key/search/content word in the text, as a result finding relevant information sometimes becomes impossible. Major search engine vendors as well as scientists predict the future of search engines lies in its semantic capability [27]. But they all have different opinions on how the semantics (or meanings) should be incorporated on the search query as well as the content itself.

Derivation of semantic interpretation from different relationships of acoustic properties has been tried by researchers e.g. in [57]. But, simple mapping or interpretation of acoustic relationship is not enough to be used as a means for effective music search and retrieval as the efficiency of search and retrieval is dependent on the structured and meaningful mapping of the machine level statistics. The structured representation of machine level acoustic information of music (standardized by MPEG-7 [37]) is not understood by most of the ordinary music listeners, though those perceptual feature based representation could lead to finding music easily. So, it would be very useful if we could enrich the semantic interpretation of

* Corresponding author.
E-mail addresses: f.rahman@shu.ac.uk (F. Rahman), j.i.siddiqi@shu.ac.uk (J. Siddiqi).

0022-0000/$ – see front matter
doi:10.1016/j.jcss.2011.10.010

musical pieces with those acoustic/perceptual features following standard representational techniques. At present, ordinary users' metadata vocabulary to describe music is quite unstructured and is not defined with any standard interpretation mechanism that can be linked with acoustic metadata that can be automatically derived from music audio. This paper presents a structured semantic metadata that creates a mapping between music consumers' metadata and the underlying music's acoustic metadata in order to enable music producers to annotate music in a meaningful way which we term as 'semantic annotation of music'.

So, in essence, the motivation of this research arises from three different areas of research. Firstly, derivation of semantic interpretation from acoustic properties of music [57] rather than the current state which only considers simple mapping of acoustic metadata to consumers' tags. Such simple mapping is not suitable for semantic search because it provides no structured mapping of acoustic metadata to the consumers' tags. Secondly, the Motion Picture Expert Group (MPEG) initiative to create a set of standard syntactic vocabulary for multimedia content enabling search and retrieval of the content is known as the MPEG-7 standard. The MPEG-7 Audio part defines syntactic metadata to represent acoustic properties of the audio content. Music audio described using MPEG-7 Audio descriptors could lead to automatic search and retrieval of the music materials. But, MPEG-7 Audio descriptors based acoustic metadata is understood neither by music producers nor by consumers i.e. ordinary listeners. Thirdly, semantic web initiative provides standard tools and techniques to represent contents in way that are both human understandable and machine process-able but do not stipulate how to annotate/categorize music using structured metadata [12].

This paper will utilize standard acoustic feature based audio description schemes (MPEG-7 Audio) to fulfill the purpose of music annotation by music producers to enable effective retrieval of musical resources by music consumers in general. In section two, we will provide the background and related research fields where this work created an impact. Section three will detail the structured semantic metadata for music annotation (the ***mpeg-7Music*** ontology) and a summary of the evaluation of this ontology will be presented in section four. Finally, in section five, we conclude briefly with further thoughts and reflection on the work presented in this paper.

2. Background

2.1. Search and retrieval of music on the WWW

Search engines generally consider semantic search to be implemented involving the analysis of the textual query only, based on the assumption that users are unable to type much more than a simple keyword for searching contents. So search engines (specifically concerning multimedia) are faced with the challenge to derive meaningful outcome in the search results. This leads to the challenge of how to represent and index the multimedia content for efficient retrieval [11]. In order to provide personalized and context aware access to content (mostly digital multimedia contents) collected from different heterogeneous disjoint sources requires an understanding of the content as well as users using them [1] and how to identify content data from music resources and create an understanding for the music listeners who search for them.

Location of implanting semantics with the information content may be either applied during categorization/annotation or during searching the content [6]. If semantic resources are embedded in the web pages themselves then search engine architecture will apply semantic reasoning and then create semantic index for semantic retrieval as well as including traditional keyword index based retrieval. Such a view relies on semantic annotation to associate resources with concepts or instances from structured domain knowledge (semantic vocabulary) and refers to the process that uses semantic annotation algorithm to associate concepts or instances from domain knowledge, annotates documents creating domain resource repository, and generates semantic index repository e.g. in [58]. But, at present, there exists no standard compliant conceptual metadata vocabulary (or domain ontologies) for annotating diverse types of musical objects that can utilize such semantic expansion search. On the other hand semantic resources may be created on the fly by search engines which deploy algorithms to analyze the semantic information; examples of such implementations are Powerset, Cognition, Lexxe etc. [38]. Google sees semantic search technology as part of the algorithmic mix, not as a replacement to its traditional keyword-analysis approach [43]; but, creating semantic resources e.g. creating knowledge of languages on which search algorithms will work is still considered as an expensive endeavor.

The task of content categorization becomes a distinct phase in the life cycle of content starting from content creation/publication phase to content consumption (search/retrieval) phase. Either content categorization is done at the creation stage or at the consumption stage in order to enable semantic search and retrieval; content categorization relies on generating semantic knowledge associated with the resources content. Based on this observation, it can be assumed that the idea of semantic search capabilities are to be used to complement and enhance search technologies by mapping the terms in user's search query to relevant semantic knowledge entities [20]. At present search engine providers are using various algorithms to enhance search experience. Most algorithms used by search engines work with the keywords in the search query and therefore for this reason this paper aims to design a semantic structured vocabulary to annotate musical objects rather than designing a search algorithm.

Searching for music is done by anyone from ordinary listeners to music experts. An untrained music listener not only just enjoys the music but is also aware of the change of key, repetitions and resolution. But, during search for a musical item ordinary listeners need to describe music in the form of keywords. This implies more information to be provided to ordinary listeners about music than that available to them at present [52]. As many more music collections are made available online more often users wish to retrieve music from an audio collection given a query, representing a portion of the music that is either sung/hummed, played, or otherwise encoded using text [53]. Forming the search query using free text or hummed audio (that is expected to retrieve the intended piece of music) leaves ordinary music fans with unsatisfactory search results. As a result, the huge proliferation of digital music in the form of audio resources on the web presents new challenges for search engines about how to incorporate search by musical content [47] for consumers.

Existing methods to search music may be classified into two broad categories based on the type of query allowed by the search engines: first, is the audio query that either contains sung/hummed audio segments or musical scores; and the second one is the textual query that contains keywords.

Depending on how audio queries are expressed and analyzed, audio based search queries may further be classified as musical score based (MIDI score input) [24] and musical metadata based (generated from humming input) [36]. But, such systems rely mostly on the user's ability to present query in the required format that is humming and clearly not all users are good at humming. Search methods therefore not surprisingly are largely based on textual queries. These are processed usually on the basis of subjective metadata like title, album, artist/singer name, genre, style etc. Approaches used by major search engines and commercial music sellers fall under this type. Others attempt to enable processing search query using natural language processing (NLP) techniques covering sound mood, cultural context, user profile and user defined metadata [5,10,33] etc. But, these systems suffer from several drawbacks such as how to generate the metadata (automatically or semi automatically) and filter noisy metadata as well as coverage of a large number of ordinary users.

Audio search and specifically searching for a song is inherently different than searching for any other multimedia resources on the web [4]. For example, search engines can display the top matches to users' search criteria almost immediately provided users keyword matches with song title, album or artist's name. However, in contrast, if the users' query is based on the concept presented by perceptual feature (e.g. timbre, melody or tempo) of the underlying sound or by musical property then how can it be ensured that this musical piece contains enough information for the search engine to locate the intended song by the user. Moreover, they are also unable to handle other keywords such as 'bright', 'sharp' describing tonal aspect of a musical piece; or keywords such as 'rising', 'falling' describing melody of the musical content.

In the context of Music Information Retrieval (MIR) ordinary listeners (who are mainly interested about personalized search and retrieval of music) are potentially one of the main beneficiaries of MIR systems and performance improvement of such a system depends on finding the method of music search that could lead to better understanding of how ordinary listeners (and/or music professionals) interpret music and what they expect from music searches [9]. The focus of this paper is on publishing the musical contents by music producers who can benefit from using the structured vocabulary to categorize the musical content. These categorized musical contents processed by search algorithms will in turn bring satisfactory results to the music consumers.

2.2. Tagging, ontology, annotation and audio descriptors

Tags are generally, free text labels applied to musical content; typically applied by the publisher/producer or the consumer of the musical item. Usually tags are unstructured without any vocabulary limit and provide a channel for narrative and social interaction.[1] Social tags are sets of individual tags; often known as Folksonomy generated from a user-created bottom-up categorical structure with an emergent thesaurus. Comparison of different tag collection approaches [55] shows pros and cons of five music tag collection approaches. Among them surveys, social tags, game-based tagging relies on human participation, and as a result is expensive. The other two approaches – text mining and auto tagging rely on automatic methods requiring less human involvement but suffer from the need of computationally intensive training data. Besides, songs that are not annotated cannot be retrieved and result in what is termed as the cold start problem. Such problem is caused by *popularity bias* i.e. popular songs tend to be annotated more thoroughly than unpopular songs. Sparse or inadequate tags result in cold start, obscure content, ambiguity as well as leading to low or no covering of certain style of songs. As a result, people interested in those less represented music styles will not be able to contribute equally with those compared to those that are heavily represented.

There also exists game based music tagging approaches with purposely built games to tag at the phrase/clip level, solve problem of obscure tags resulting in the potentially high tagging rates [3]. But, a gaming approach needs to overcome the challenge of superficial tags applied by non-fans and collecting weak labels. Another approach to collect tags could be hiring experts for survey or to hand label content – that could result in consistent, strong labelling with fixed structured vocabulary. But such human-labour intensive approach does suffer from the small pre-determined vocabulary and doesn't scale to the long-tail tagging. Besides it is very difficult to construct widely accepted taxonomy. Auto tagging is another idea that uses content analysis to automatically apply tags acquired from other sources (social tags, games, web crawling) can be 'learned'. New music or unpopular music can be auto tagged with the 'learned' tags and can scale to the long tail. Generally, an auto tagging system relies on audio feature extraction modules and then matches the extracted features with pre-labelled example data to categorize the input item. In essence these systems create standard models of classification from already annotated items. So, we chose to enhance current approaches of music tagging by creating a knowledge based

[1] http://wikis.sun.com/display/SocTagsMIR/Social+Tags+and+Music+Information+Retrieval.

representation of musical content. The proposed structured annotation vocabulary will contain implicit association with low level acoustic properties of musical sound so that ordinary users will be able to tag the musical object without being aware of complex scientific representation of low level features. Moreover, low level acoustic features are easy to extract using automated tools.

At present there is no single metadata standard available for music covering all the possible requirements for aforementioned applications of music. The available music metadata schemas and vocabularies are either focused on interoperability or designed as a standard from the start [13]. The ID3 [41] tag contains the artist name, song title and genre to be embedded within the audio file and has got a wide spread use in music players and devices e.g. iTunes, iPod and Winamp. These are general purpose syntactic tags and are designed to organize music items for commercial exploitation of music rather than supporting music consumers' search query and do not enable music producers to annotate music with content based semantic tags.

The Kanzaki Music Vocabulary [31] or Kanzaki ontology presents concepts to describe classical music and performances in order to distinguish musical works from performance events or works from performer in the ensemble defining classes for musical works, events, instruments, performers and relationships between them. Based on this vocabulary, the Music Ontology [45] deals with music-related information on the Semantic Web, including editorial, cultural and acoustic information and has been developed as a ground for more domain-specific knowledge representation. The Music Ontology[2] covers mainly the applications related to editorial musical information (albums and tracks), event related concept concerning work flow involving the composition of a musical work, an arrangement of this work, a performance of this arrangement and a recording of a performance. It enables to create semantic metadata to incorporate several levels of musical information including editorial information, event related concepts (e.g. arrangement and recording performance) and musicological information (e.g. which key was played at a certain time by a person playing the instrument). But, the set of semantic concepts provided by Music Ontology do not cover all layers of information as contained by a musical object that can be utilized both for annotation and retrieval.

Yet, Music Ontology imports the time line ontology to cover both acoustic signal level aspects and universal physical time may be further extended to accommodate perceptual features associated with the timeline. To accommodate both Physical time and MPEG-7 media time in relation to time instant and duration aspects of audio files the proposed ontology has borrowed time line concepts by adding two disjoint subclasses (MediaTime and PhysicalTime) to it and appropriate properties have also been created. Additional details will be presented in section three.

Annotation means adding information to the existing content or resources without changing the original. These annotations are meant to be share-able over diverse network, domain and/or users. Annotations are additional data/information that is tied to the content/resource in question; that represent information about the resources and arises from the interaction between the resource and its user. These characteristics serve to distinguish annotations from the general category of "just additional data". The semantic web initiative by World Wide Web Consortium (W3C) has established standards to make resource content automatically process-able by machines as well as human-readable. Such capabilities would be valuable for sharing thoughts and knowledge. Machine-understandable annotations will enhance intelligent search and retrieval and for that it was required to pay attention to widely accepted knowledge representation techniques to create semantic metadata [42]. There are multiple definitions of annotations in the related literature depending how annotations are created, how they are shared and utilized or as an enabler of machine readable meaningful metadata. However, annotated music items usually provide information that is explicitly related to a music item and such annotated information will formally represent the semantics of the metadata with reference to ontology; the task of semantic annotation is performed by tagging ontology class instance data and mapping it into ontology classes [46].

The ABC ontology [29] actually forms the foundation of MPEG-7 data model to further extend any MPEG-7 ontology. It creates the basic concepts of how the class and property hierarchies and semantic definitions should be derived from MPEG-7 Descriptors [35] and Description Schemes (MPEG-7 Part 5, 2003). The approach is to describe a core subset of multimedia content entities (StillRegion, AudioSegment, VideoSegment, AVSegments etc.) as subclass of the top level MultimediaContent class which was derived from the Resource class. Visual features and properties such as Colour, Texture, Motion and Shape are only applicable to visual entities only; but these are not relevant for digital items containing solely musical audio. The proposed ontology has created the MusicalSegment class by extending the AudioSegment class of ABC ontology.

The Visual Ontology proposed by [28] extends visual feature Colour (Colour class from ABC ontology) to annotate visual resources for efficient retrieval. This visual ontology creates subclass of *Colour* to represent dark and light colour and provides associated property definitions. In a similar fashion the proposed ontology has created the MusicalExpression class and its subclasses – TimbralExpression/MelodicExpressions with appropriate property definitions that connect MPEG-7 timbral and melodic descriptors with semantic concepts.

The OWL-DL representation of full MPEG-7 MDS as proposed in [54] is aimed to serve as an upper level ontology for the ontologies developed to capture MPEG-7 semantics. Their approach was focused towards developing an upper level ontology that forms the OWL based representation of MPEG-7 data types into OWL data type properties; creating OWL class for each MPEG-7 complex types and for every simple attribute of the complex type a datatype property OWL data type

[2] http://musicontology.com/#sec-evolution.

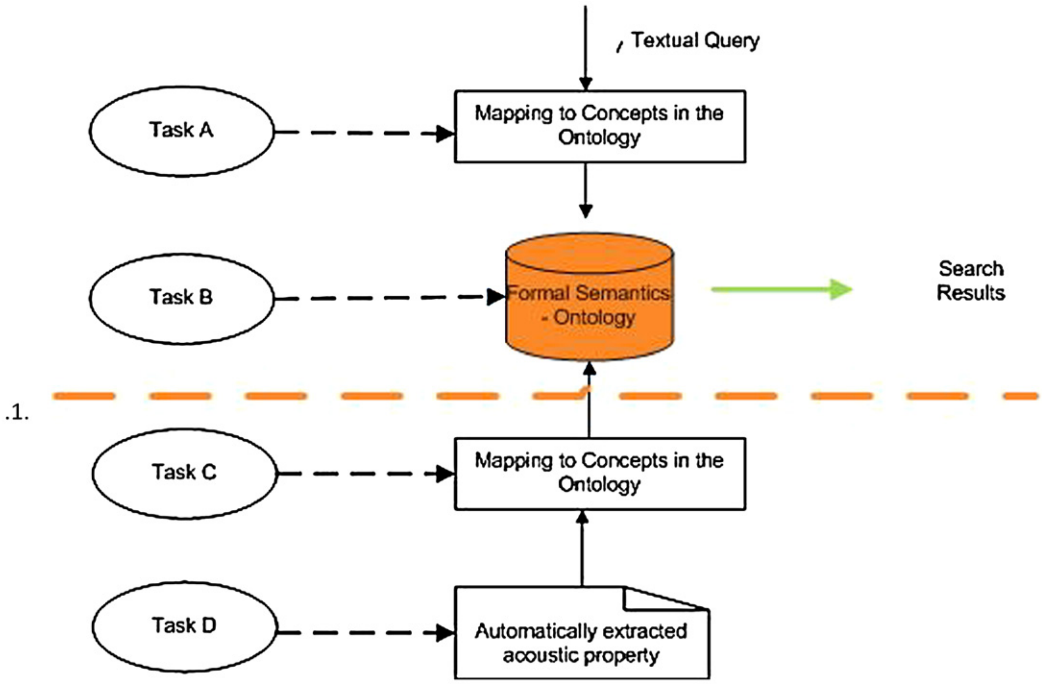

Fig. 1. Requirements within an operational framework for Semantic Search and Retrieval.

property construct and complex attributes are represented as OWL object properties, which relate class instances. No doubt that, development of such an ontology will ease the effort of creating new MPEG-7 ontologies for different application domain but it does require the development of mapping of domain ontology concepts to its upper ontology classes to create annotation for further retrieval of search results, the upper ontology is based on a one-to-one mapping for further extension.

In MPEG-7 (Audio) Part 4 there are seventeen low-level audio descriptors that support general audio description. MPEG-7 **Melody** description tools include *MelodyCountour DS* (for extremely terse melody) and *MelodySequence DS* (for verbose complete melody) for monophonic melodic information. **Timbre** descriptors are aimed at describing perceptual features of instrument sounds. The MPEG-7 Timbre descriptors relate to notions such as attack, brightness or richness of sound. In the timbre descriptors, the two widely used classes of musical instrument sounds have been detailed – the harmonic (sustained – coherent) and percussive (non-sustained) sounds. The timbre description tool combines several low level descriptors for harmonic and percussive sounds. Therefore the proposed music annotation ontology is required to generate appropriate concepts and properties to represent different timbral and melodic class concepts that enable the music producers to derive implicit association with MPEG-7 low level descriptors.

In summary the current state concerning MPEG-7 ontologies is that they do not specify explicitly the representation for audio features. To bridge this gap, the research presented in this paper is novel because it has created the extension of the audio feature ontology for the purpose of annotation of music resources. It is actually a semantic annotation ontology for digital music based on a set of requirements guided by the Common Multimedia Ontology Framework requirements [48,19].

2.3. Requirements of semantic search of music

It is evident that to enhance the user experience with semantic capability will require building structured semantics and knowledge that will create the foundation for semantic search. Search and query framework for human oriented Music Information Retrieval (MIR) system [5] may be designed depending on structured vocabulary e.g. ontological semantics. Use of ontological semantics in different domains is not new [8]. The proposed music annotation vocabulary (providing structured semantics) in this paper is aimed to support this type of MIR system that provides support for textual query.

Incorporation of semantic information as a structured vocabulary in the Search and Query framework is aimed to support integration of information about artist, genre, year, lyrics, as well as automatically extracted acoustic properties like loudness, tempo, and instrumentation by using semantic ontology as the basis for query processing to extract relations between these high level semantic concepts that will allow to satisfy semantic queries involving faster/slower, cheer-up/calm-down, mix of calm-down and cheer-up music etc. The proposed vocabulary falls in the category of formal semantics [30] and will be discussed in detail in section three. Presently we specify the key requirements (they are italicized in the text) that semantic search frameworks need to be fulfill in Fig. 1 and as follows:

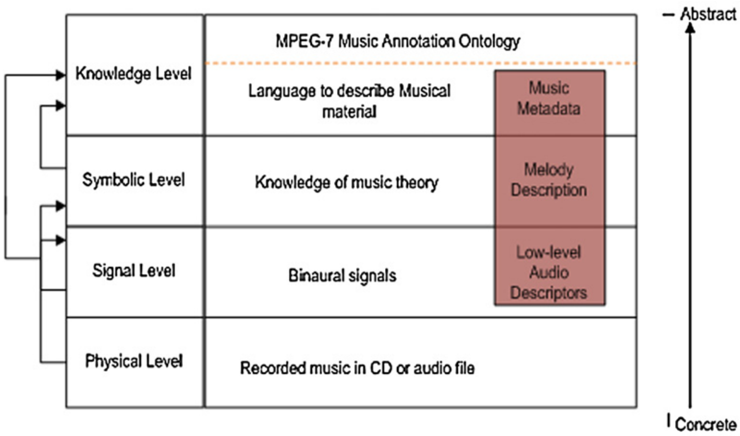

Fig. 2. Representation levels of Music Information-adapted from [56].

- Choose semantic metadata so that *semantic metadata matches the text query* (Task B).
- *Map the textual query to semantic concepts and relations* (Task A).
- *Relate the query to the content itself,* automatic extraction of acoustic property requires us to decide what and how the features of the musical content will be extracted (Task D).
- *Map acoustical property to the semantic metadata structure* to provide search results (Task C).

3. MPEG-7 music annotation ontology

3.1. Representation level of music information

There are four types of music representation found in the existing literature and these types are organized in levels associated to the various type of information conveyed [56]. Fig. 2 depicts four levels of musical information associated with digital music. The lowest level physical layer can be characterized by the acoustic pressure as a function of space and time i.e. the acoustic pressure signals at the level of both eardrums, which characterize information input into the auditory system as played back from digital music files.

The signal level representation specifies a content-unaware representation to transmit any sound (both musical/non-musical/audible and even non-audible signals). This signal level representation of music can be automatically extracted from music file using MPEG-7 feature extractor tools see [14].

Music consumers with little or no understanding of musicology/music theory might be able to use simple phrases [32] such as: rising, falling etc. to describe the musical item and may not be able to properly classify the part or whole music according to the melody [51]. Similarly, timbral (tonal aspect) features of music may be named using everyday words like having gentle, metallic, hard sound characteristics by the listeners [50]. But, these musicological features are understood in the domain of music publishers/producers who are able to classify correctly a musical item using melodic and timbral dimensions. Besides, the MPEG-7 Audio standard provides description schemes based on signal-level audio descriptors that represent musical timbre and melody and these MPEG-7 low-level descriptors automatically generated from music files (using freeware tools e.g. [14]) may be mapped to symbolic level concepts (timbre and melody).

The symbolic representation describes *content-aware* events with regard to the formal concepts of music theory and accounts for discrete events, both in time and in possible event states (e.g. melody shape and motion, musical timbre) with reference to listeners' knowledge of music theory. For instance, musical timbre may be described as multidimensional perceptual attributes that comprise relevant information that conveys the identity of the sound in a musical context. Typically spectral, temporal attributes and acoustical parameter of sound are used to determine timbre such as spectral envelope, spectral centroid, temporal envelope and spectral flux respectively [18]. These in turn create a relation to verbal attributes used to describe timbre perception of complex sounds e.g. smooth-rough or light-dark by correlating verbal attributes with one or more perceptual dimensions: "dry" correlated with the log of the attack time dimension; "round" correlated with the spectral centroid; "brilliant/bright" correlated with spectral centroid while "metallic" was correlated with three perceptual dimensions.

The knowledge level representation is associated with appropriate language structures for describing musical phenomena. We aimed to create a meaningful representation of music metadata using conceptual representation languages specified by W3C e.g. Web Ontology Language (OWL). Using highly interoperable and semantically rich languages musical dimensions may be modeled to create knowledge level representation. But this would first require establishing mapping among the different levels.

Knowledge level representation may be achieved from both signal and symbolic levels (as shown using arrows in Fig. 2). Signal level information that specifies a set of audio descriptors may be mapped to represent objective representations of

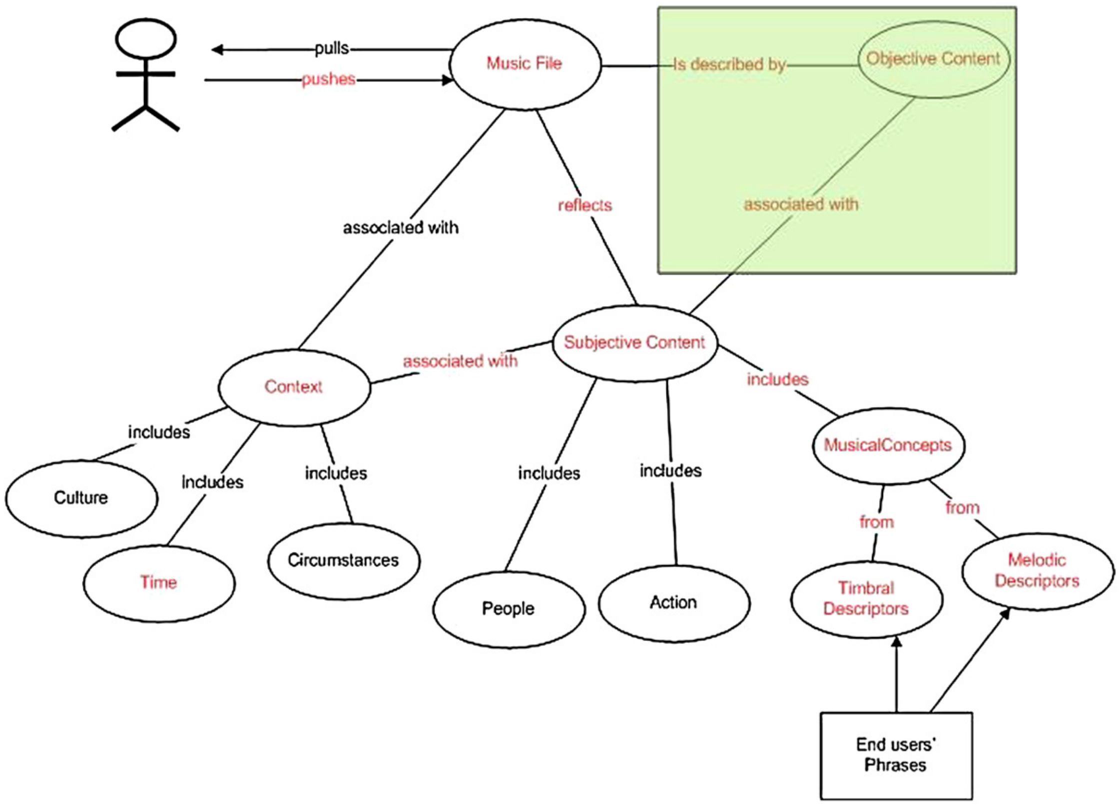

Fig. 3. Use scenario of the music annotation ontology.

music items. Symbolic level Melody and Timbre Description schemas may be conceptually modeled to provide subjective representation of musical items. So, knowledge level representation of music material provides both objective descriptions (mapped from MPEG-7 low level audio features related to timbral and melodic description schema) and subjective descriptions (features modeled from musical dimensions e.g. timbre and melody). As a result, knowledge level representation can contain objective and subjective characteristics that can be created from combinations of signal and symbolic level information respectively.

The limitation of such layered representation is its inadequacy to reflect the structural complexity of the intermediary levels and to model the way higher-level information is structured. To overcome the problem of conversion between various representation levels and to adapt the different level of information to support technical applications such as search and retrieval of music we have proposed here to create a semantic vocabulary to support the process of annotation of music files by music producers. This vocabulary has been designed to provide a set of formal concepts and properties as an abstraction model for signal and symbolic level information.

3.2. Use scenarios of music annotation ontology

Music producers require information to annotate a music file that *is described by* the objective content (MPEG-7 elements) as well as *reflecting* the subjective content within it. Music producers are not best placed to suggest the subjective content; they require music listeners to provide such a subjective content we have chosen to denote the subjective content via a vocabulary of musical concepts selected from a survey conducted at MIT media lab with a view to collecting phrases used by ordinary listeners [50] and online music teaching module [51].

In the proposed ontology (***mpeg-7Music***) the two types of content information – both objective and subjective content of the music files are explicitly defined and are joined together by data type and object properties. Subjective content is also associated with contextual information that *includes* Time. The most important contribution of this ontology is the richness of the relationship that a music file expresses as shown in Fig 3. We have not only involved both the objective (from mpeg-7) and subjective content from established surveys but also the context in which music consumers listen to music; this context incorporates the circumstances and the time in which music is listened to. Specifically, to illustrate the feasibility of this rich and powerful relationship we have included Musical Concepts *from* Timbral and Melodic Descriptors those of which in turn are ***characterized by*** concepts created specially ***to denote*** musical concepts under this ontology.

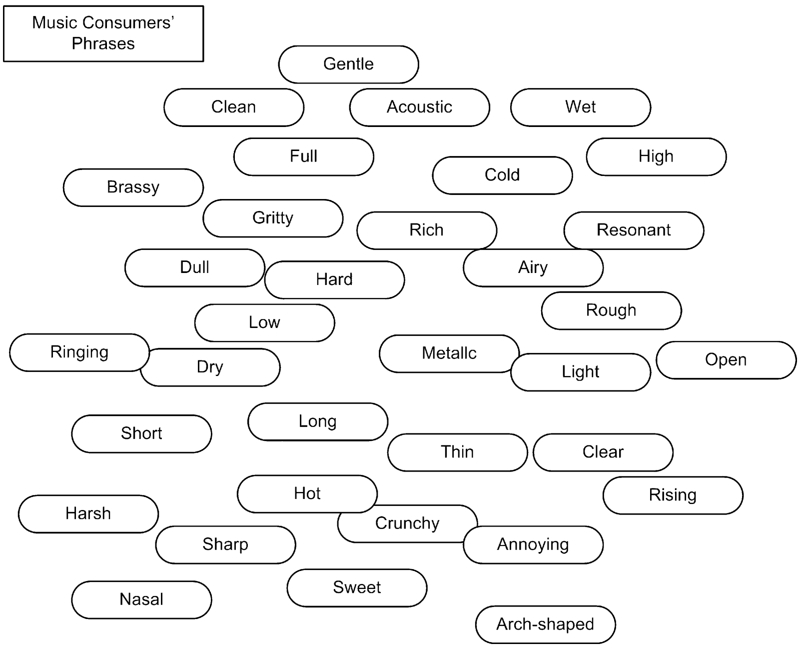

Fig. 4. End users' phrases for musical concepts of the music annotation ontology.

Fig. 3 shows that users annotate their music files with phrases associated with timbral descriptors as well as melodic descriptors but more significantly they are required to know about the objective content of music's acoustic data. The ***mpeg-7Music*** ontology will automatically associate objective data with musical concepts and the user will be able to select from the built in phrases (or able to add more phrases) to annotate music files. Automatic association of the objective content will be provided through the built in rules that we have created for the ontology. The design plan of the proposed ontology was not to constrain users with those rules only, more rules may be created to incorporate users' choices and specific interest (e.g. cultural) – addition of more rules will be considered as a future direction to improve the ***mpeg-7Music*** ontology.

While modeling music annotation ontology the following simple words have been selected as music consumers' phrases to support for subjective content. These words have been chosen from a previous research [50] and online music module [51]. Fig. 4 shows below a snapshot of those phrases.

The phrases compiled by the survey conducted at MIT Media Lab [50] describing timbral features and melodic features presented in [51] (as shown in Fig. 4) are unstructured and are very subjective – because a single phrase may be used to derive multiple meaning by different music users. Besides, such phrases are quite ambiguous in meaning as may be understood by music consumers to describe a particular music item and associate these phrases with different musical dimensions. As these phrases have not been organized into any hierarchy and structure these are not suitable to be used for categorization or annotation of musical objects. As a result it requires us organizing these phrases into a conceptual structure for further utilization of these compiled phrases in facilitating the task of semantic annotation of music.

3.3. Proposed Music Ontology and inference rules

The Simplest way to represent semantics and concepts is the controlled vocabulary that actually is a list of terms, enumerated explicitly. The more structured ones are taxonomy, thesauri and ontology respectively [44]. The term ontology has been applied in many different ways [23], but the core meaning within computer science is a model for describing the world that consists of a set of types, where there is only one type, one property as well as context, and very few relationships. In practice thesauri are not considered ontologies because their descriptive power is far too weak, precisely because of limited vocabulary.

There are two views in the contemporary literature on what makes a controlled vocabulary to qualify as ontology [26]. One view emphasizes on representing the controlled vocabulary using ontology encoding languages e.g. OWL and the other view conceives ontology as a set of concepts, properties, instances and inference rules. The ***mpeg-7Music*** annotation ontology qualifies to be lightweight concept ontology to enable music producers to annotate any music segment from both the view mentioned above as it is represented using the dominant semantic web ontology representation language i.e. OWL.

The idea of Semantic Web as envisioned by Tim Berners Lee [7] provides a knowledge infrastructure to explicitly represent conceptualizations of domain knowledge in the form of ontology in different levels [17]. At present the most important

Table 1
mpeg-7Music ontology class concepts and (individual) instances.

No.	Concepts	Sub-classes	Instances	
1	ResourceFormat	mp3, wav		
2	MusicSegments			
3	MusicalConcepts	MelodicExpressions	Shape	archShaped, falling, rising
			Motion	conjunct, disjunct, leap
		TimbralExpressions	Sharpness	acoustic, brassy, crunchy, hard, metallic, resonant, ringing
			Brightness	gentle, airy, cold, clean
4	TimeLine	MediaTime, PhysicalTime		

Table 2
mpeg-7Music ontology properties.

No.	Property	Property Type	Detail
1	characterizedBy	Symmetric	links ResourceFormat with MusicalSegments
2	relatesTo	Transitive	links ResourceFormat with MusicalConcepts
3	denotedBy	Functional	links MusicalSegments with TimeLine and its subclasses
4	attached	Object	links MusicSegment with TimeLine
5	timbralDescriptor	Data type	Denotes different MPEG-7 timbral descriptors e.g. Spectral Centroid, LogAttackTime and so on
6	melodicDescriptor	Data type	Denotes audio fundamental frequency type and contour type

ontology languages are Extensible Mark-up Language (**XML**), XML Schema (**XMLS**), Resource Description Framework (**RDF**), RDF-Schema (**RDFS**) and Web Ontology Language/Ontology Web Language (**OWL**). Semantic web supports representation of ontology in standard languages namely the Resource Description Framework (RDF), RDF-Schema (RDFS) and Ontology Web Language (OWL). The richness of the semantics represented in each paradigm increases from XML to RDF and RDF to OWL [16]. So, there are two basic issues that need to be addressed while encoding any ontology. Firstly, the expressiveness of the underlying semantics presented by the ontology; secondly, the inference capability that can be achieved for the semantic representation in finite computing time.

The W3C's Web Ontology Working Group has defined OWL as three different sub languages – OWL Full, OWL DL and OWL Lite. The choice among adopting the three sublanguages of OWL depends on the extent to which users require the expressive constructs and inference capability. Generally, the richer express-ability is the more inefficient the reasoning support as well as computability and hence trade-off is required between express-ability and reasoning support when we choose an ontology representation language. We used OWL-DL to encode the ***mpeg-7Music*** ontology.

The ***mpeg-7Music*** ontology extends ABC ontology's AudioSegment concept by adding MusicalSegment concept and relationships representing the semantics of MPEG-7 Audio features. It also formed inheritance relationship with Music Ontology [46] by importing the TimeLine concept and creates MediaTime subclass to Timeline that actually linked to MPEG-7 basicTimePoint and basicTimeDuration data types. Moreover, it uniquely added MusicalConcepts class to cater for Timbral and Melodic Expressions as understood by ordinary listeners and these expressions had got defined association with MPEG-7 scalar and vector data types by appropriate data and object type properties. Tables 1 and 2 show a summary of the class concepts, class instances and properties created by the mpeg-7Music Ontology.

The MPEG-7 Audio Encoder tool [14] was used to extract the low level audio features for each of the music file. To create the grouping of consumers' tag as subclasses under MelodicExpressions and TimbralExpressions classes, this research relied on MPEG-7 data content quality insights as used in [51] and [40] respectively. For example, motion of a melody may be either termed as conjunct or disjunct or leap. Depending on whether a melody rises and falls slowly/quickly it is generally referred to as *conjunct or disjunct respectively.* If it is a mix of both then it is called *leap*.

Having defined the detailed ontology structure the rest of the problem was how it could be enabled to associate MPEG-7 descriptors mapping with MusicSegment class during annotation by music producers.

In order to associate aforementioned timbral descriptors with MusicSegment class few basic rules were defined to translate low-level acoustic terms gathered from MPEG-7 description such as 'spectral centroid' (SC) to more familiar commonly-used intermediate-level terms like 'long' and 'short'. Works carried out in [18] and [50] researched such links between low-level features and commonsense terms. They found, for example, that humans consider a piece of music characterizing dull or bright based on perceptual dimension linking to Spectral centroid (SC) – this fact was used to construct the following association rule to link the timbralDescriptor i.e. SC with MusicalSegments. The timbre description tool combines several low level descriptors for harmonic (*HarmonicSpectralCentroid*(*HSC*), *HarmonicSpectralDeviation*(*HSD*), *HarmonicSpectralSpread*(*HSS*), *HarmonicSpectralVariation*(*HSV*) *& LogAttackTime*(*LAT*)) and percussive (*SpectralCentroid*(*SC*), *TemporalCentroid*(*TC*) *& LogAttackTime*(*LAT*)) sounds.

Table 3
Timbral Rules defined in ***mpeg-7Music*** Ontology.

Timbral Rules
Rule 1 – BrightnessFeature Association Rule: ***mpeg-7Music***:MusicSegment(?*x*) ∧ ***mpeg-7Music***:TimbralDescriptor(SC) → ***mpeg-7Music***:characterizedBy(?*x*, Brightness)
Rule 2 – SharpnessFeature Association Rule: ***mpeg-7Music***:MusicSegment(?*x*) ∧ ***mpeg-7Music***:TimbralDescriptor(HSC) ∧ ***mpeg-7Music***:TimbralDescriptor(HSS) → ***mpeg-7Music***:characterizedBy(?*x*, Sharpness)

Table 4
Melodic Rules defined in ***mpeg-7Music*** Ontology.

Melodic Rules
Rule 3 – MotionFeature Association Rule: ***mpeg-7Music***:MusicSegment(?*x*) ∧ ***mpeg-7Music***:MelodicDescriptor(AFFT) → ***mpeg-7Music***:characterizedBy(?*x*, Motion)
Rule 4 – ShapeFeature Association Rule: ***mpeg-7Music***:MusicSegment(?*x*) ∧ ***mpeg-7Music***:MelodicDescriptor(contourValue) → ***mpeg-7Music***:characterizedBy(?*x*, Shape)

The above rules (in Table 3 and Table 4) associate different MPEG-7 data descriptors (those were modeled as data type properties in ***mpeg-7Music*** *ontology*) with MusicSegments instances using the characterizedBy object property (which is actually a symmetric property).

The TimbralExpression class in the ***mpeg-7Music*** ontology contains two broad categories of music consumers' phrases named as Sharpness and Brightness (as shown in Table 1) and each of these sub-categories contains the instances of individual phrases borrowed from [50] that only provide a set of key phrases but ***mpeg-7Music*** was added with category hierarchy for these phrases and the timbral tags that were used for the canonical rules (Rule 1 and Rule 2 in Table 3) according to research efforts of [2] and [39]. As mentioned above spectral descriptors creates a relation to verbal attributes used to describe timbre perception of complex sounds e.g. smooth-rough or light-dark by correlating verbal attributes with one or more perceptual dimensions – "dry" correlated with the log of the attack time dimension, "round" correlated with the spectral timbral descriptors, "brilliant/bright" correlated with spectral centroid while "metallic" was correlated with three perceptual dimensions such as spectral centroid (SC), harmonic spectral centroid (HSC) and harmonic spectral spread (HSS).

The music producers can use this rule to mark the music segment to have bright timbre (using instance from TimbralExpression class) then they can easily do that using Rule 1 and 2 (mentioned above); these rules were created to form the abstraction layer for associating spectral centroid data type value of the underlying audio.

Similarly melodic rules have been defined using motion and shape features that associates melodic descriptors from mpeg-7 audio data types such as fundamental frequency and contour value respectively (Rule 3 and Rule 4 in Table 4). Rule 3 associates Motion concept with the audio fundamental frequency (AFFT) descriptor and Rule 4 defines Shape concept using *MelodyCountour DS*. The MelodyCountour DS uses five step contours (generally termed as shape of the melody) representing the interval difference between adjacent notes ranging between -2 and $+2$ [15]. Melody contour or shape of a melody may be categorized as "rising", "falling" or "arch-shaped" melody [51]. Besides, the melodic and rhythmic information may be transcribed into melody contour calculated from MPEG-7 descriptor *AudioFundamentalFrequency* (AFFT) using five contour values as specified in MPEG-7 Audio standard [59]. MelodicDescriptors take values from melodyContourType which is related to AudioFundamentalFrequency. The AudioFundamentalFrequency descriptor holds the frequency information [59]. Melodic expressions may be derived [51] verbally from shape of melody described by melody contour descriptor values, e.g. rising melody (for ascending notes denoted by contour value 1 & 2), arch shaped melody (in which melody rises and falls and again slowly) etc. Arch shaped melodies are easy to understand by ordinary users and may be described by contour value -1, 0 and -1.

SWRL[3] (Semantic Web Rule Language) notations were used to specify rules from the proposed ontology classes, properties and individuals expressed in OWL. The novelty of ***mpeg-7Music*** ontology is the inclusion of numerous built-in relations that creates the connection between low level audio data types with intermediate level terms to facilitate user annotation. For ease of implementation we used Protégé 3.4.1 to encode the mpeg-7Music Ontology that comes with SWRL Tab covering OWL 1.0 features only. A detail OWL representation of this ontology is available at http://carus.aces.shu.ac.uk/coreontology/mpeg-7/mpeg-7Music.xml.

[3] http://www.daml.org/2003/11/swrl/.

4. Evaluating the *mpeg-7Music* annotation ontology

The ***mpeg-7Music*** ontology was evaluated from three perspectives – functional, structural and usability profiles [22]. Ontologies are semiotic objects that may be structurally evaluated by looking at it as an information object by checking the formal semantics of the ontology topologically and as well as consistency of the semantics represented by the logical properties. Functional dimension focuses on the intended conceptualization specified by the ontology and specifically looks at it as a language as well as its components [49]. From the usability viewpoint, it is important to look at the ontology profile (annotations) that typically addresses the communication context i.e. pragmatics of an ontology. Due to the complexity of content and structure that characterizes music objects, ontology driven representation of semantics concerning music audio prompted us to evaluate the proposed music annotation ontology from three different perspectives. In essence we used the following three dimensions for evaluating the proposed ontology:

1. Standard requirements for multimedia ontology: *Functional dimension* – The design of the ***mpeg-7Music*** annotation ontology was guided by a set of requirements as specified in the general requirements for Common Multimedia Ontology Framework as proposed in [48]. To be interoperable and to provide prospect for standardization it must conform to the requirements for a common multimedia ontology framework [19]. In constructing the ontology with those precise requirements in mind it was easy for us to trace back to the requirements for a multimedia ontology to be satisfied that the ***mpeg-7Music*** ontology conforms.
2. Validating the ontology using prominent ontology validator: *Structural dimension* – Protégé (version 3.4.1) was used to encode the ***mpeg-7Music*** ontology; it also provides open source APIs[4] to use existing reasoners to integrate into semantic web applications. Besides, reasoners, Protege provides Pellet for consistency checking of the ontology. Pellet is also a free ware tool to check OWL mark-up for problems beyond simple syntax errors; that will examine the OWL content of the proposed for a variety of potential errors and reports them along with the location of the errors in the files. When using the direct Pellet 1.5.1 reasoner distributed with the Protégé 3.4, the ***mpeg-7Music*** ontology appeared consistent and it classified the inferred concepts without any error.
3. Demonstrating ease of use of the ontology via a custom built Annotation tool: *Usability Dimension* – Surveying the existing annotation tools available for multimedia annotation a customized tool was developed to evaluate the applicability of the proposed annotation ontology. It provides the music producers with the ability to annotate by connecting music audio's acoustic features with ontological semantics; these auto-generated annotations form the knowledgebase for search algorithms to generate recommendations/suggestions for music producers to proceed with the task of annotation. The design of the simple semantic annotation tool was focused on three main requirements to be fulfilled by such a tool. First, the tool should provide a platform to support annotation of heterogeneous formats of digital music; secondly, the system should provide a standard way to store the created annotations and thirdly it should create an ontology based annotation process. Then the ***mpeg-7Music*** ontology was evaluated to prove its potential to be used in simple annotation platform to empower music producers to annotate digital music.

5. Conclusion

The novelty of ***mpeg-7Music*** annotation ontology lies in the fact that, to the best of our knowledge there is no music annotation task ontology that creates a unique opportunity for music producers to annotate music with its audio properties as represented by the MPEG-7 encoding. Existing multimedia ontologies only deal with image and video annotation and do not consider the unique requirements to be addressed for music annotation. Besides none of the existing research considers designing ontology for supporting the annotation task of digital music coming in different content formats and so a customized semantic annotation tool was designed to demonstrate the applicability of the ***mpeg-7Music*** ontology. Significantly, a notable aspect of this ontology is that it is designed according to two standards – first the semantic web ontology standard (OWL 1.0) and second the multimedia description standard (MPEG-7 Audio).

The current state of music tagging could be improved significantly if a structured metadata scheme is used for tagging by music producers so ***mpeg-7Music*** was structured as an ontology to provide meaningful metadata. Music information retrieval systems implementing semantic search techniques also require structured metadata to provide satisfactory search results against textual query; furthermore, existing MPEG-7 compliant multimedia ontologies are not appropriate for use by music producers' annotation. Our proposed ***mpeg-7Music*** ontology is novel because it can be used in the music search engines to fulfill the requirements for structured metadata equally important additionally it is MPEG-7 compliant.

Finally we claim novelty in saying that the ***mpeg-7Music*** ontology creates a bridge with music consumers' tags and MPEG-7 acoustic metadata and extends upper level multimedia ontology i.e. the ABC ontology and the Music Ontology (Raimond et al., 2007). Moreover, ***mpeg-7Music*** ontology was designed by creating a mapping with MPEG-7 Audio data types and was encoded in OWL1.0 syntax and thus it creates a standard interoperable representation.

---
[4] http://protege.stanford.edu/plugins/owl/api/ReasonerAPIExamples.html.

References

[1] S. Abels, L. Haak, A. Hahn, Identifying Ontology Integration Methods and their applicability in the context of product classification and knowledge integration tasks, Technical Report: Department of Business Information Systems University of Oldenburg, Report No. WI-OL-TR-01-2005, Date of Publication: October 2005.
[2] G. Agostini, M. Longari, E. Pollastri, Musical instrument timbres classification with spectral features, EURASIP J. Appl. Signal Process. 1 (2003) 5–14.
[3] L. Ahn, L. Dabbish, Labeling images with a computer game, in: SIGCHI, 2004.
[4] S. Ali, P. Aarabi, A novel interface for audio search, in: IEEE International Conference on Multimedia and Expo, 2006.
[5] S. Baumann, A. Klüter, M. Norlien, Using natural language input and audio analysis for a human-oriented MIR system, in: Proceedings of the 2nd International Conference on Web Delivering of Music, Darmstadt, Germany, 2002.
[6] R. Berkan, Semantic search: an antidote for poor relevancy, ReadWriteWeb.com archive, published on May 29, 2007, available online at http://www.readwriteweb.com/archives/semantic_search_antidote_for_poor_relevancy.php [accessed January 25, 2009].
[7] T. Berners-Lee, Semantic web stack, 2008, [electronic print] available at http://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html [accessed June 18, 2009].
[8] Binyu Zang, Yinsheng Li, Wei Xie, Zhuangjian Chen, Chen-Fang Tsai, Christopher Laing, An ontological engineering approach for automating inspection and quarantine at airports, in: Web and Mobile Information Systems 2006, J. Comput. System Sci. 74 (March 2006) 196–210.
[9] M.A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, M. Slaney, Content-based music information retrieval: current directions and future challenges, Proc. IEEE. 96 (4) (2008) 668–696.
[10] O. Celma, P. Cano, P. Herrera, Search sounds: An audio crawler focused on weblogs, in: Proceedings of the international conference on music information retrieval (ISMIR), 2006.
[11] Y. Chang, S. Amanda, Multimedia Chinese web search engines: a survey, in: Fourth International Conference on Information Technology, ITNG '07, April 2–4, 2007 pp. 481–486.
[12] Claudio Gutierrez, Carlos A. Hurtado, Alberto O. Mendelzon, Jorge Perez, Foundations of semantic web databases, in: Database Theory, J. Comput. System Sci. 77 (3) (May 2011) 520–541.
[13] N. Corthaut, S. Govaerts, K. Verbert, E. Duval, Connecting the dots: music metadata generation, schemas and applications, ISMIR 2008 – Session 2c – Knowledge Representation, Tags, Metadata, 2008.
[14] H. Crysandt, Hierarchical sound classification using Mpeg-7, in: IEEE 7th Workshop on Multimedia Signal Processing, October 30/November 2, 2005, pp. 1–4.
[15] N. Day, Search and Browsing, Introduction to MPEG-7: Multimedia Content Description Interface, John Willey and Sons, Ltd., 2002 (Chapter 20).
[16] Diego Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Riccardo Rosati, View-based query answering in description logics: Semantics and complexity, J. Comput. System Sci. 78 (1) (2012) 26–46.
[17] L. Ding, P. Kolari, Z. Ding, S. Avancha, Using ontologies in the semantic web: a survey, in ontologies: a handbook of principles, in: Concepts and Applications in Information Systems, Springer, 2007, pp. 79–113.
[18] S. Donnadieu, Mental Representation of the Timbre of Complex Sounds, Analysis, Synthesis, and Perception of Musical Sounds, Springer, New York, 2007.
[19] H. Eleftherohorinou, V. Zervaki, A. Gounaris, V. Papastathis, Y. Kompatsiaris, P. Hobson, Towards a Common Multimedia. Ontology Framework, Analysis of the Contributions to Call for a Common multimedia Ontology Framework Requirement, April 2006, available online: http://www.acemedia.org/aceMedia/files/multimedia_ontology/cfr/MM-Ontologies-Reqs-v1.3.pdf (last accessed: 01.04.2007).
[20] M. Fernandez, V. Lopez, M. Sabou, V. Uren, D. Vallet, E. Motta, P. Castells, Semantic search meets the web, in: 2008 IEEE International Conference on Semantic Computing, August 4–7, 2008, pp. 253–260.
[21] C. Gang, T. Wang, P. Herrera, A novel music retrieval system with relevance feedback, in: 3rd International Conference on Innovative Computing Information and Control, 2008, ICICIC '08, June 18–20, 2008, pp. 158–158.
[22] A. Gangemi, C. Catenacci, M. Ciaramita, J. Lehmann, Modelling ontology evaluation and validation, ESWC 2006 (2006) 140–154.
[23] L. Garshol, Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all, J. Inf. Sci. 30 (4) (2004) 378–391.
[24] A. Ghias, J. Logan, D. Chamberlin, B.C. Smith, Query by humming: musical information retrieval in an audio database, in: Proceedings of the 3rd ACM International Conference on Multimedia, San Francisco, CA, USA, 1995.
[25] O. Goussevskaia, K. Michael, L. Michael, W. Roger, From web to map: exploring the world of music, in: International Conference on Web Intelligence and Intelligent Agent Technology, 2008 IEEE/WIC/ACM vol. 1, December 9–12, 2008, pp. 242–248.
[26] J. Graybeal, P. Alexander, What is an Ontology? in: The MMI Guides: Navigating the World of Marine Metadata, 2009, http://marinemetadata.org/guides/vocabs/ont/definition.
[27] D. Hai, F.K. Hussain, E. Chang, A survey in semantic search technologies, in: 2nd IEEE International Conference on Digital Ecosystems and Technologies, 2008, DEST 2008, February 26–29, 2008 pp. 403–408.
[28] L. Hollink, Semantic annotation for retrieval of visual resources, PhD thesis, VU University Amsterdam, Netherland, 2006.
[29] J. Hunter, Enhancing the semantic interoperability of multimedia through a core ontology, IEEE Trans. Circuits Syst. Video Tech. 13 (1) (January 2003) 49–58.
[30] Jacek Sroka, Jan Hidders, Paolo Missier, Carole Goble, A formal semantics for the Taverna 2 workflow model, in: The 2nd International Workshop on Workflow Management and Application in Grid Environments & The 3rd International Workshop on Workflow Management and Applications in Grid Environments, September 2010, J. Comput. System Sci. 76 (6) (2009) 490–508, Special Issue: Scientific Workflow.
[31] Kanzaki, The web KANZAKI, 2003 (last updated October 31, 2007), available online at http://www.kanzaki.com/ns/music.rdf [accessed on April 20, 2009].
[32] B. Kostek, Computing with words concept applied to musical information retrieval, Electron. Notes Theor. Comput. Sci. 82 (4) (March 2003) 141–152.
[33] Last.fm, Last.fm music recommendation service, 2008 (last updated November 21, 2008), available online at http://www.last.fm/about [accessed on April 20, 2009].
[34] T. Li, M. Ogihara, Toward intelligent music information retrieval, IEEE Trans. Multimedia 8 (3) (June 2006) 564–574.
[35] A.T. Lindsay, I. Burnett, S. Quackenbush, M. Jackson, Introduction to MPEG-7: Multimedia Content Description Interface, John Willey and Sons, Ltd., 2002 (Chapter 17).
[36] N.C. Maddage, H. Li, M.S. Kankanhalli, Music structure based vector space retrieval, in: Proceedings of 29th Annual International ACM SIGIR Conference, Seattle, USA, 2006.
[37] J.M. Martinez, R. Koenen, F. Pereira, MPEG-7: the generic multimedia content description standard, part 1, IEEE Multimedia 9 (2) (April–June 2002) 78–87.
[38] P. Midwinter, Is Google a Semantic Search Engine?, published on March 26, 2007, available online at http://www.readwriteweb.com/archives/is_google_a_semantic_search_engine.php [accessed September 10, 2008].
[39] N. Misdariis, A. Minard, P. Susini, G. Lemaitre, S. McAdams, E. Parizet, Environmental sound perception: metadescription and modeling based on independent primary studies, EURASIP J. Audio Speech Music Process. 2010 (2010), Article ID 362013, 26 p.

[40] D. Mitrovic, Z. Matthias, E. Horst, Analysis of the data quality of audio descriptions of environmental sounds, Research Article from: J. Digital Inf. Manag. (April 1, 2007).
[41] D. ONeill, ID3, 2009, available online at http://www.id3.org/ [accessed 30 October 2010].
[42] T.B. Passin, Explorer's Guide to the Semantic Web, Greenwich Manning Publications, 2004 (Chapter 4).
[43] J.C. Perez, Google rolls out semantic search capabilities, PCWorld Business Center, Tuesday, March 24, 2009, available online at http://www.pcworld.com/businesscenter/article/161869/google_rolls_out_semantic_search_capabilities.html [accessed 5 May 2009].
[44] W. Pidcock, What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model?, January 15, 2003, available online at http://www.metamodel.com/article.php?story=20030115211223271 [accessed on December 18, 2007].
[45] Y. Raimond, S. Abdallah, M. Sandler, The Music Ontology, Austrian Computer Society, 2007.
[46] L. Reeve, H. Han, Web technologies and applications (WTA): Survey of semantic annotation platforms, in: Proceedings of the 2005 ACM Symposium on Applied Computing SAC '05, March 2005.
[47] M.M. Ruxanda, A. Nanopoulos, C.S. Jensen, Y. Manolopoulos, Ranking music data by relevance and importance, in: 2008 IEEE International Conference on Multimedia and Expo, June 23/April 26, 2008, pp. 549–552.
[48] C. Saathoff, S. Schenk, S. Staab, Requirements for a multimedia ontology framework, in: 3rd European Semantic Web Conference, Montenegro, June 11–14, 2006.
[49] Santiago Figueira, Daniel Gorin, Rafael Grimson, On the formal semantics of IF-like logics, in: Workshop on Logic, Language, Information and Computation, J. Comput. System Sci. 76 (5) (August 2010) 333–346.
[50] M. Sarkar, B. Vercoe, Y. Yang, Words that describe timbre: a study of auditory perception through language, in: Language and Music as Cognitive Systems Conference (LMCS-2007), Cambridge, UK, May 11–13, 2007.
[51] C. Schmidt-Jones, Melody – connexions module, 2010, available online at http://cnx.org/content/m11647/latest/ [accessed April 30, 2010].
[52] A. Storr, Music and the Mind, HarperCollins Publishers, London, 1997.
[53] I.S.H. Suyoto, A.L. Uitdenbogerd, F. Scholer, Searching musical audio using symbolic queries. Audio, speech, and language processing, EEE Trans. 16 (2) (February 2008) 372–381.
[54] Chrisa Tsinaraki, Panagiotis Polydoros, Stavros Christodoulakis, Interoperability support between MPEG-7/21 and OWL in DS-MIRF, IEEE Trans. Knowledge Data Engin. 19 (2) (February 2007) 219–232.
[55] D. Turnbull, L. Barrington, G. Lanckriet, Five approaches to collecting tags for music, in: ISMIR 2008, 2008.
[56] H. Vinet, The representation levels of music information, in: Computer Music Modelling and Retrieval, in: Lecture Notes in Comput. Sci., vol. 2771, Springer, 2004, pp. 203–216.
[57] B. Whitman, Learning the meaning of music, PhD thesis, MIT, April 14, 2004.
[58] G. Zou, B. Zhang, Y. Gan, J. Zhang, An ontology-based methodology for semantic expansion search, in: Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008, FSKD '08, vol. 5, October 18–20, 2008, pp. 453–457.
[59] Jan-Mark Batke, Gunnar Eisenberg, Philipp Weishaupt, Thomas Sikora, Query by humming system using MPEG-7 descriptors, in: 116th Convention of Audio Engineering Society, Berlin, Germany, May 2004.