AN MPEG-7 DATABASE SYSTEM AND APPLICATION FOR CONTENT-BASED MANAGEMENT AND RETRIEVAL OF MUSIC

Otto Wust
Music Technology Group
Universitat Pompeu Fabra

Òscar Celma
Music Technology Group
Universitat Pompeu Fabra

ABSTRACT

Computer users are gaining access to and are starting to accumulate moderately large collections of multimedia files, in particular of audio content, and therefore demand new applications and systems capable of effectively retrieving and manipulating these multimedia objects. Content-based retrieval of multimedia files is typically based on searching within a feature space, defined as a collection of parameters that have been extracted from the content and which describe it in a relevant way for that particular retrieval application. The MPEG-7 standard offers tools to model these metadata in an interoperable and extensible way, and can therefore be considered as a framework for building content-based audio retrieval systems.

This paper highlights the most relevant aspects considered during the design and implementation of a DBMS-driven MPEG-7 layer on top of which a content-based music retrieval system has been built. A particular focus is set on the data modeling and database architechture issues.

1. INTRODUCTION

Large amounts of digital audio are nowadays widely available. This can range from musical pieces or songs for download over the internet by a casual computer user, to highly specialized sound libraries delivered on DVD targeted for the professional musician.

As the number of available audio media increases, so does the need for a user to locate audio clips in an efficient way. The field of information retrieval deals with the modeling, indexing and accessing of information mainly within digital libraries. It has been extensively studied for many years [1], mostly focusing on retrieving textual information using text-based methods.

Database management systems (DBMS) have been widely used to implement efficient text based information retrieval systems, solving many of the problems encountered in that field. However, in the area of multimedia, and in particular in audio and music information retrieval, there is still a lot of ongoing research and open questions concerning implementation and architectural aspects such as scalability of the systems, as well as functional issues related to the usability.

This paper addresses the possibility of using available DBMS in combination with MPEG-7 standard [2] as the key building block for music retrieval applications that can satisfy the needs of a user. First, the MPEG-7 standard is briefly reviewed with special focus on aspects that must be taken into consideration when designing and implementing it within a database management system. Then, a particular MPEG-7 database implementation is described. This was developed throughout the CUIDADO project [1], and is targeted for content-based processing of music MPEG-7 descriptions. Some of the relevant features are illustrated in the section that follows along a content-based audio editor, manipulation, and authoring application, which is buit on top of the described database system. Finally conclusions are drawn.

2. MPEG-7 AS A DATA MODEL

MPEG-7 has been promoted as a standard with the objective to provide a common interface for audiovisual content description in multimedia environments. This should allow that different MPEG-7 systems or modules can easily interoperate.

The standard is based on the notion of Descriptors (D) and Description Schemes (DS). The former represent a model for specific high or low level features that can be annotated for a given media object. The latter just represent a grouping of a series of Descriptors or further Description Schemes in a particular functional area.

The definition of the MPEG-7 standard relies on further standards of the MPEG family and heavily on the XML language and XML-Schema which are used in its representation and its definition. MPEG-7 itself is provided in the form of an extensible XML-Schema defining an object oriented type hierarchy which delivers a set of predefined descriptors grouped into its functional description schemes.

As a matter of example, we consider the Agent DS defined in the standard, which allows to represent data for persons, groups or organizations. The following example

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profi t or commercial advantage and that copies bear this notice and the full citation on the fi rst page.
© 2004 Universitat Pompeu Fabra.

[1] http://www.cuidado.mu

in XML shows an agent descriptor which defines a particular person:

```
<Agent xsi:type="PersonType">
  <Name xml:lang="en">
    <GivenName xml:lang="en">
      John
    </GivenName>
    <FamilyName xml:lang="en">
      Smith
    </FamilyName>
  </Name>
</Agent>
```

We recall the properties of the XML language as an adequate means for interchanging structured self-described information, and the XML-Schema language as a powerful tool (which overcomes many limitations of the Document Type Definitions) for describing the validation rules of a given XML document.

Some of the main difficulties in supporting MPEG-7 at database level are a consequence of the object-oriented definition, the actual size and complexity of the current MPEG-7 schema, and at a further level, due to the flexibility in the structure of the XML documents the schema admits as valid. This can lead to functional difficulties when considering an MPEG-7 database implementation, since different MPEG-7 applications could use and/or generate semantically equivalent descriptions in structurally and syntactically different yet MPEG-7 valid documents. Although it should be considered an application issue, it has an impact on the generality of database implementation, as it can lead to application specific fine-tuning requirements.

Even if we restrict ourselves to the case of a single precisely defined application, a further issue in an MPEG-7 database implementation is that the standard does not define how the searching or indexing has to be made on the described data. It also does not make any assumptions about the internal storage format.

There are other factors which are not inherently related to the problem of modeling for multimedia, but which have to be taken into account when designing a real system for music information retrieval. The most relevant consideration in this area is probably the fact that many descriptors will contain textual labels, which may need to be presented in several user languages and more important searched in a consistent user language independent way. MPEG-7 provides the structure, but does not specify any rules on how a search over multi-language data has to be done.

3. DATABASE IMPLEMENTATION

The database implementation described here was developed within the CUIDADO project for its use by *Sound Palette*, a music authoring and post production application which targets exhaustive content based retrieval and processing functionalities, as described within the next section and in [3]. For the actual database implementation, an Oracle 9i Release2 object-relational DBMS was chosen since it incorporates the XDB component which treats XML in an efficient and native way. On top of it an MPEG-7 specific layer was built. The system involves:

- **A repository of XML-Schemas** which are registered within the DB; in paricular, the MPEG-7 XML-Schema and the schema extensions specifically developed within the CUIDADO project. In the XML-Schema registration process a number of underlying data structures are inferred from the data definitions encoutered in the XML-Schema and maintained by the DBMS for storing the actual content in a structured manner. A mapping of MPEG-7 schema elements to the underlying internal data structures is therefore established within XDB.
- **A repository of descriptions**; the actual content generated and manipulated by the application's users is inserted into a combination of tables specifically defined for the *Sound Palette* application and the underlying object-relational data structures created from the schema definitions during the schema registration process.
- **A set of indexes** which serve for speeding up queries and for helping to maintain referential integrity within particular elements or attributes in the XML descriptions.

3.1. Access to content

Two interfaces exist to interact with the MPEG-7 database layer. A low-level SQL-based interface provides direct access to tables. In the *Sound Palette* database several content tables store the XML descriptions in as many rows as required, generally we have one single record per MPEG-7 document which is stored in one or another table depending on its functional context. These tables include a column of Oracle's XMLTYPE data type which is linked to the CUIDADO-extended MPEG-7 XML-Schema registered with the DBMS. The SQL-based interface —that also incorporates functions to support the XPath standard— provides access and location to the elements and attributes within the XML descriptions. This low-level interface has also a very low application dependency.

The following example shows the syntax of a query to obtain the names of terms in an MPEG-7 classification scheme for those terms with a termID attribute "Piano". XPath syntax is used within the statement for two different purposes: first in the filter section (WHERE clause) to restrict the number of candidate descriptions, then on the returned documents to extract particular nodes, in this case the Name elements.

```
SELECT
  extract( x.content,'/child::node()/Name',
  'xmnls="urn:mpeg:mpeg7:schema:2001"')
  .getStringVal()
```

```
FROM CUI_CLASSIFICATION_SCHEMES x
WHERE existsnode(x.content,
  '//Term[@termID="Piano"]') > 0
```

In this case, for simplicity, access is to a single table (CUI_CLASSIFICATION_SCHEMES) only, but a statement can equally join other tables that store either MPEG-7 XML or otherwise simple object relational content.

The second interface that was built is a high level API providing application-specific functions, in this case very specific to the functionality required by the CUIDADO project. It serves to hide the tables and the complexity of the SQL statements to the applications.

3.2. Storage

The storage is hidden by both of the interfaces. It is performed in underlying object relational data strucutres, which are automatically created by the DBMS during XML-Schema registration. At insertion time, the incoming description in XML format is parsed and stored distributed across all the underlying database objects which map the MPEG-7 data structures defined in the XML-Schemas registered.

This structured storage strategy results in a retrieval performance directly comparable to that of a relational database. Since data is stored within native data types, common B-Tree indexes can be used to efficiently access the target documents without the neccesity to parse each of them. As a drawback, we have to note the small overhead required for reconstructing the XML from the underlying data strucutres. This can however be neglected in most of the typical queries which imply a search over a large number of documents to return just a fraction of the total.

Furthermore, the structured storage automatically avoids the difficulties that are encountered when numerical or temporal data is stored as text within the textual XML format, and range queries can be performed naturally.

An additional feature is the compression effect obtained. XML tags contained in the description documents, can account for a high percentage of the overall document volume. Tags, however do not need to be stored as the internal object types inherently mantain the MPEG-7 structure of the descriptions.

A further indexing structure has been created to facilitate preservation of the referential integrity in descriptions across different description documents, and somehow extending the notion of ID and IDREF. For instance, the MPEG-7 layer guaranties that an element within a classification scheme cannot be deleted or modified if it is being used (referenced) in a description.

3.3. Search by similarity

Searches by similarity do usually not constitute a problem in terms of database model per se, since they can be reduced to the problem of evaluating a distance function on a set of relevant features. There are however other types of database problems related to these types of searches, typically related to the low scalability of the systems. In order to do an exhaustive search by similarity, the reference object must theoretically be compared to all the objects in the database, by evaluating the distance function, and frequently the cost of computation of that distance is high, making such an approach not feasible when the database is large.

A simple search by timbre similarity function was implemented according to [4] which scales well over a corpus of 100000 simulated descriptions, which were created artificially with random values in the features used by the function.

3.4. Extensions

Although this particular database implementation has been made using only a specific subset of the MPEG-7 descriptors, those for audio content management, we believe that it can be considered generic from the point of view of metadata management and retrieval based on metadata information, since at database level it solely relies on XML and XML-Schema functionality, and therefore this approach should be able to be applied to the field of content-based modeling and retrieval of video applications as well, with very little adaptation efforts.

A further feature of the presented MPEG-7 database layer is that it supports extensibility, in conformance to the extensible design of the MPEG-7 XML-Schema. In the case of the (*Sound Palette*) application, a set of new descriptors for melodic and rhythm description not originally available in the standard, had to be introduced. The extension is made in form of an additional application specific XML-Schema document which imports the standard MPEG-7 XML-Schema document and creates a series of specialized descriptors such as the Scale, the Meter, the Key, the MelodyContour in the form of extensions of types already supplied by the standard [5].

4. APPLICATION EXAMPLE

The *Sound Palette* is an application for content based processing and authoring of music and is compatible with MPEG-7 standard descriptions of audio. It has been designed for users who own large libraries of sounds and loops and offers novel ways to interact and work with audio. Figure 1 shows a screenshot of the application's content based features. A drum loop has been analyzed and segmented. This process has generated a number of descriptors which have been inserted into the database. The waveform for the drum loop is displayed in the upper region (2 channels since it is a stereo file), and the temporal segments have been separated with vertical lines.

In this case, the segmentation has also been made for the different instruments that build up the drum loop. This is visible in the lower region, which shows the two dimensions of the segmentation. The x-axis is still the time base, while the y-axis carries the different instruments en-

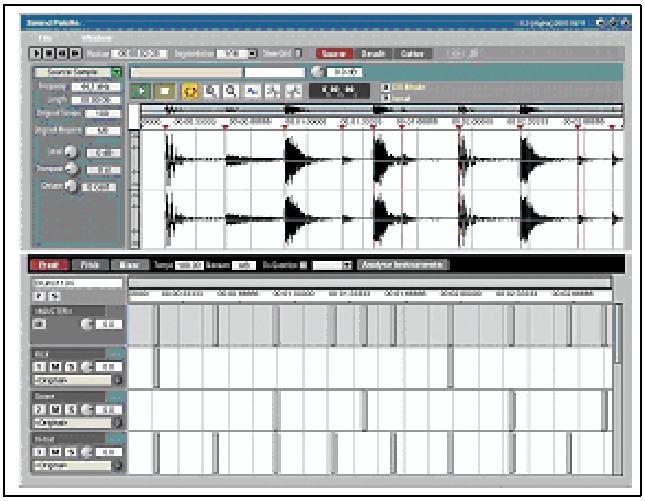

Figure 1. Screenshot of *SoundPalette*

countered as a result from the segmentation. The beats corresponding to four different percussion instruments are visible on the lower region.

This information, stored in MPEG-7 format, can be then used to locate appropriate substitution sounds in the audio database. In the retrieval, many different MPEG-7 descriptors can be considered, such as related to physical properties (sampling rate, file format), related to ownership or rights holders, or more interesting parameters directly related to the actual content, such asSoundPalette the timbre, the energy, the pitch, etc. In addition, the user has the choice to set tolerances for any of the numerical descriptors, in order to modify the retrieval accuracy of the system.

Furthermore, *Sound Palette* and the underlying database allow the user to organize the content collection with MPEG-7 classification schemes. These allow the user to browse through a hierarchy of categories, as opposed to just retrieving by issuing queries. In order to enhance the browsing experience, relations within descriptions are supported at database level.

All this metadata information is stored in the database presented above in MPEG-7 compliant format, and can be used in order to either set further filters when searching for a particular content, or to directly retrieve the media from the virtual containers that the terms within the classification scheme constitute.

5. CONCLUSIONS

Applications based on the MPEG-7 standard are emerging in the areas of multimedia archive, digital broadcasting, digital library, etc. MPEG-7 provides description mechanisms for multimedia content; however, applications are still immature and are not really explored in concrete fields. In this article we have highlighted the database aspects of a system for content based processing and authoring of music in which MPEG-7 has been successfully used.

The MPEG-7 standard has offered adequate tools to model features employed to describe musical content, all in a satisfactory way from the user requirements perspective. Furthermore, the XML based technologies that the MPEG-7 standard employs, have minimized the development effort by allowing the use of a widely available database management system capable of efficiently managing XML. The MPEG-7 database layer developed provides features not encountered in a general XML database and also hides to the application much of the complexity involved in persistently storing and manipulating XML descriptions.

Those descriptors required by the application but not defined by the MPEG-7 standard have been also integrated without major development effort, which leads to the conclusion that the proposed approach is generally viable.

6. ACKNOWLEDGEMENTS

Part of this work has been supported by the European Commission, under the CUIDADO project (IST-1999-20194).

7. REFERENCES

[1] Frakes, w., Baeza-Yates, R. *Information Retrieval: data structures and algorithms* Prentice Hall, New Jersey, 1992

[2] Manjunath, B. S., Salembier, P. and Sikora, T. ”Introduction to MPEG 7: Multimedia Content Description Language”. Ed. Wiley, 2002.

[3] Celma, O., et. al., ”Tools for Content-Based Retrieval and Transformation of Audio Using MPEG-7: The SPOffline and the MDTools” *Proceedings of 25th International AES Conference*. London, UK, 2004.

[4] Peeters, G. McAdams, S. Herrera, P. ”Instrument Description in the Context of MPEG-7” *Proceedings of International Computer Music Conference*. Berlin, Germany, 2000.

[5] Gómez, E. Klapuri, A. Meudic, B. ”Melody Description and Extraction in the Context of Music Content Processing” *Journal of New Music Research Vol.32.1*, 2003.