

Tools for content-based retrieval and transformation of audio using MPEG-7: the SPOffline and the MDTools

Òscar Celma¹, Emilia Gómez¹, Jordi Janer¹, Fabien Gouyon¹, Perfecto Herrera¹ and David García¹

¹*Music Technology Group, Institut Universitari de l'Audiovisual, Universitat Pompeu Fabra. Ocata 1, 08003 Barcelona, Spain. <http://www.iaa.upf.es/mtg>*

Correspondence should be addressed to Òscar Celma (oscar.celma@iaa.upf.es)

ABSTRACT

In this paper we present a set of applications for content-based retrieval and transformations of audio recordings. They illustrate diverse aspects of a common framework for music content description and structuring implemented using the MPEG-7 standard. MPEG-7 descriptions can be generated either manually or automatically, and are stored in a XML database. Retrieval services are implemented in the database. A set of musical transformations are defined directly at the level of musically meaningful MPEG-7 descriptors and are automatically mapped onto low-level audio signal transformations.

1. INTRODUCTION

The goal of this paper is to present a set of applications for content-based retrieval and transformation of audio based on the MPEG-7 Standard for content description [1]: the *Sound Palette Offline* application [2] and the *MDTools* application [3].

They are all based on a common framework for music content processing which is represented in figure 1. First, the audio signal is analyzed and annotated, in order to generate a set of features and descriptors. These features are structured into an XML file, according to MPEG-7 Description Scheme. The description is stored in a database that implements a set of search and retrieval functionalities. Finally, some content transformations are provided. High level transformations facilities are offered, mapping them to low-level signal transformations implemented by a content-based transformation engine.

As an example of this framework, the *Sound Palette* application has been thought as an editing, processing and mixing tool for sound designers and professional musicians. The difference between the *Sound*

Palette and other similar tools is that it provides a way to interact with the content of the audio, by offering audio sample management, editing and transformation features based on sound content description. Two versions of the *Sound Palette* have been developed in parallel: an “online” [4] and an “offline” version (*SPOffline*). In this paper, we concentrate on the second one, which was integrated as a final user-friendly application by the German company *Creamware Datentechnik GmbH*, commercial partner in the CUIDADO project.

Another application, namely *MDTools*, has been designed to help content providers in the creation and production phases for metadata annotation of multimedia content.

This paper is organised as follows: section 2 describes the above mentioned applications. Then, next sections explain the modules included in the block diagram (see figure 1) and link it with the applications developed.

2. APPLICATIONS

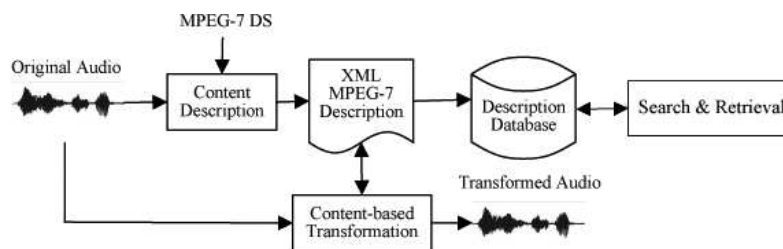


Fig. 1: Block diagram for music content processing.

2.1. SPOffline

The *SPOffline* application is restricted to the management of samples on the user's PC and includes sound edition, transformation functionalities and metadata description of the content. This tool generates content description of sound files, so the following functionalities are provided: content-based description, content-based navigation and retrieval –among the database of sounds already described– and, finally, content-based transformations.

The *SPOffline* distinguishes between three types of sounds: monophonic melodic phrases, drum loops, and unitary samples (e.g a single note). Each type of sound has associated a particular description according to its characteristics. The application mainly consists of the following modules (that we will explain in more detail in the following sections):

- **Sample Manager:** tool for generating content descriptions of audio recordings and inserting them into the database (explained in section 3).
- **Sample Editor:** is a tool for content-based transformation, synthesis of sounds and it also allows edition of descriptions (explained in section 4). This tool is integrated in the main application's window.
- **Sample Browser:** is a tool for navigating through the database using content-based queries based on MPEG-7 descriptors (explained in section 5.1). The queries help us to find and retrieve audio recordings.

Figure 2 shows a screenshot of the *SPOffline* application.

2.2. MDTools: Multimedia Description Tools

An important role in the framework depicted in figure 1 is the overall description of the content itself, that is the metadata information associated within the audio files.

MDTools application is an authoring tool to describe a multimedia production. The application is based on MPEG-7 Multimedia Description Scheme (MDS) and it allows to describe the creation and production process, information about the media, segmentation of the content and to create classification schemes.

The main components of the application are:

- **Timeline and segment edition:** to edit temporal segmentation of the multimedia content. It creates a hierarchical structure of the content.
- **Basic segment information:** it includes time information (start time, and duration), text annotations, segment identifier, segment type (audio, video, audiovisual, images, etc.) and StructuralUnit –a term to describe structural aspect of the segment–.
- **Classification Schemes:** it allows to create taxonomies and term definitions. A user can create his/her own terms and definitions and link it to the MPEG-7 description of a multimedia document.
- **Media Information:** Media files can be linked to a multimedia segmentation. Furthermore, MPEG-7 media profiles can be created to describe different coding schemes and file types (mp3, ogg, wav, aiff, etc. for audio or avi, mpg, etc. for video) which are used to adapt to client bandwidth. Moreover, segments that contain audio can be listened within the application.

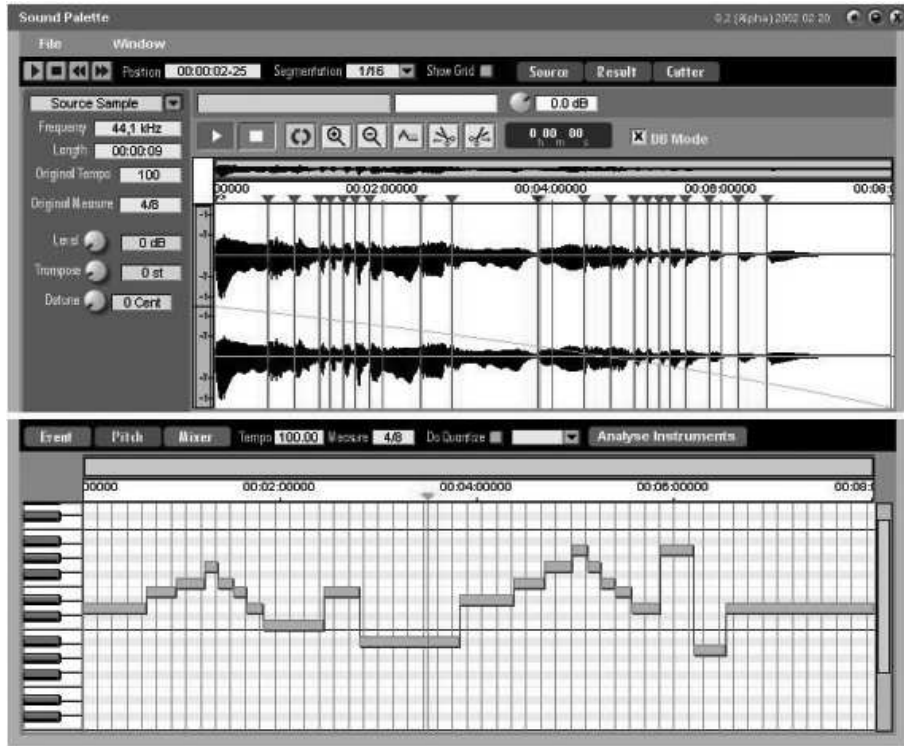


Fig. 2: Melodic content editing and transformation section of the *Sound Palette Offline*.

- **Creation and Production Information:** it allows to describe information such as creators (author, singers, etc.), title, abstract, as well as the production staff of a multimedia production –using terms that refers to a classification scheme–. Related material of the segment can be linked with an URI (Universal Resource Identifier).

Figure 3 shows a screenshot of the *MDTools* application.

3. CONTENT-BASED DESCRIPTION

The **Sample Manager** tool included in *SPOffline* generates content description of sound files. For a new audio file, the user selects the sound type (Phrase, Loop or Sample), and fills some additional information such as Author, Date, etc. Then, the analysis algorithm automatically describes the audio file and creates a file following MPEG-7 descriptors. Once the description is generated and stored

into the database, it may be loaded onto the *Sample Editor*, where the information is shown to the user. The main descriptors categories are related to melody (shown in figure 2), rhythm and instrument. More details about the implemented features and the description scheme for each type of audio material is explained in [5]. The user can also refine these descriptions, which are automatically updated in the database. Finally, the content descriptions can be exported as XML files for other uses.

4. CONTENT-BASED TRANSFORMATION

The *SPOffline* application integrates some transformation functionalities into its **Sample Editor** module. This is a tool for content-based transformation and synthesis of sounds that processes an original recording using content descriptors. Using the implemented algorithms, we can either stretch/compress a note's length, or change its pitch. On the other hand, we can use the content description to generate a complete new synthesized sound

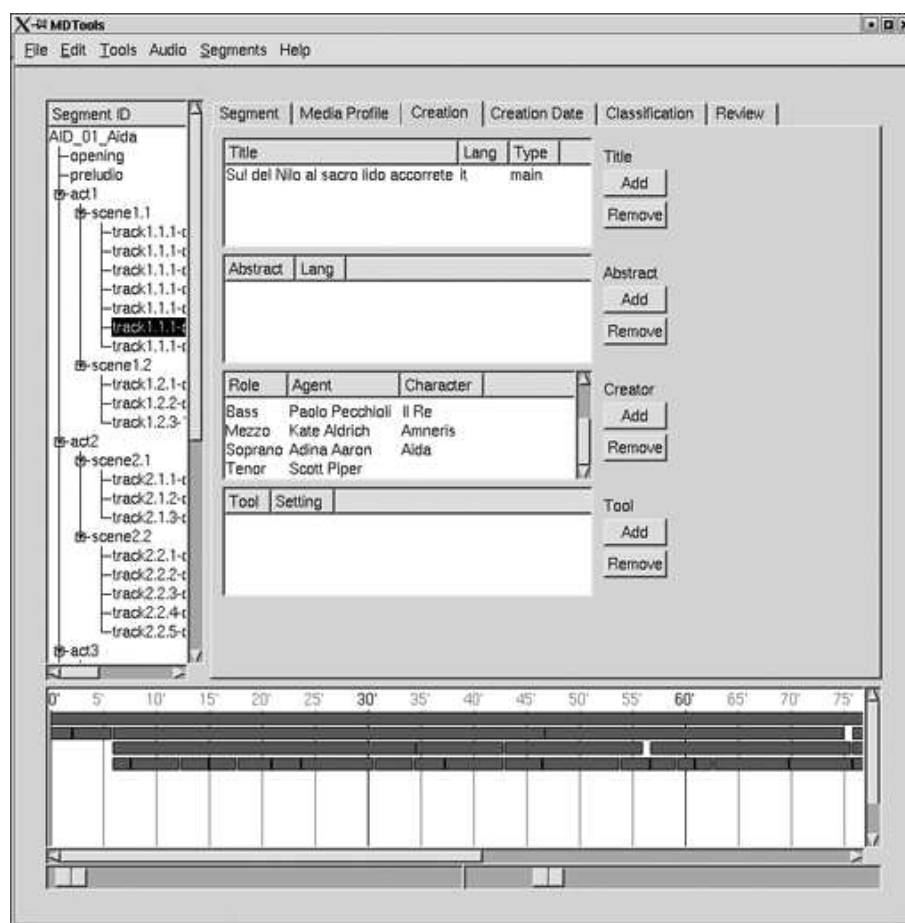


Fig. 3: MDTools application screenshot.

by using samples retrieved from queries in the *Sample Browser* (see section 5.1). The *Sample Editor* includes a *MIDI-like* multitrack, where the user can insert new events in several tracks and link them with other sounds. These sounds must have their content description already in the database. In this case, a mixer assistant is responsible for adjusting the levels of the different tracks and mix them together to a stereo output.

5. SEARCH AND RETRIEVAL

Once the MPEG-7 descriptions are stored, a search and retrieval module queries the database to get multimedia segments. Thus, in the *SPOffline* application a user can access and retrieve segments,

stored in a local database, to edit and manipulate it. On the other hand, multimedia productions described with *MDTools* are stored on a server database and then, using a web browser, a user can search and retrieve audiovisual files.

5.1. Content-based retrieval

As we have mentioned, retrieval facilities are implemented within the *SPOffline* into the **Sample Browser**. This is the main tool for interfacing with the content descriptions stored in the database. The queries are constructed by specifying one or more of the twenty-one criteria options [5][6]:

- Metadata: Author, Date, Description Name, Audio Path, Sound type

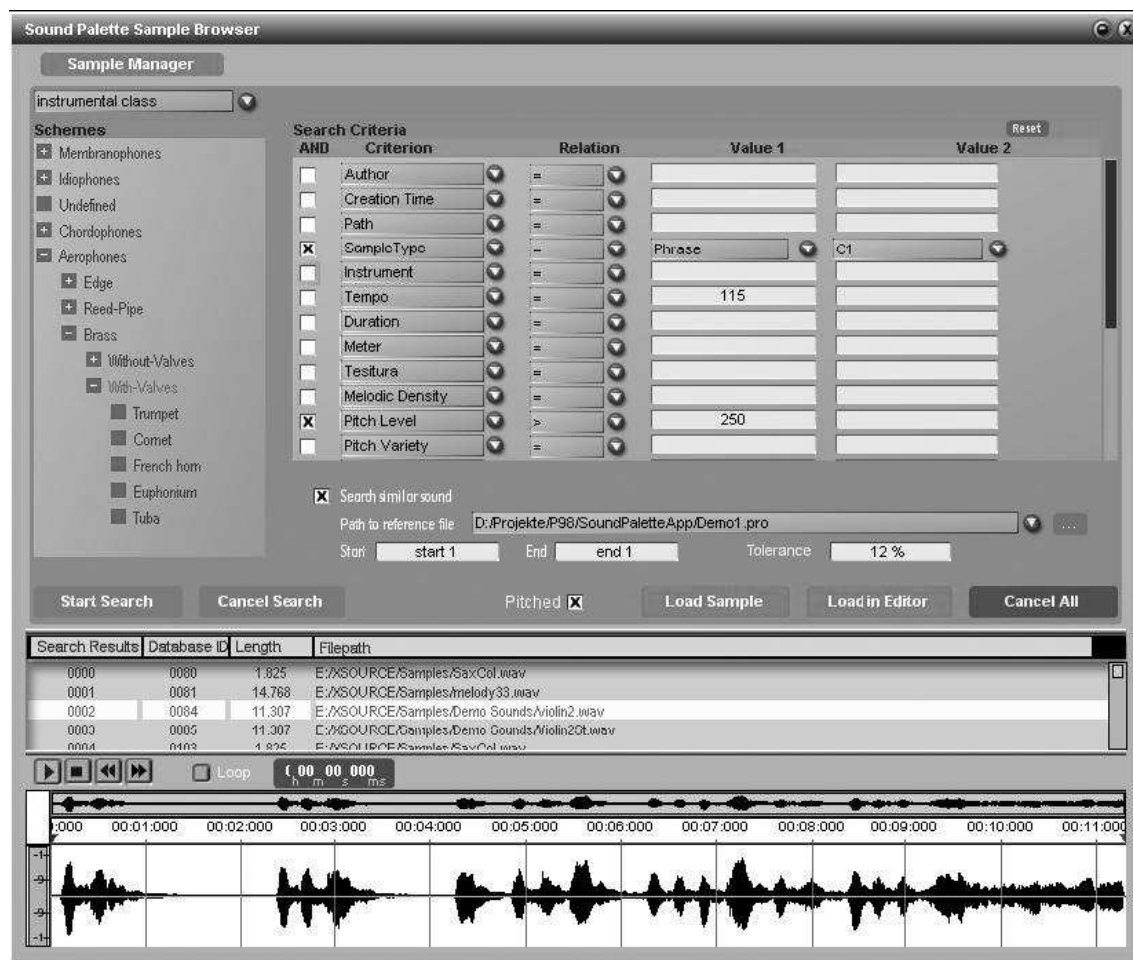


Fig. 4: Browsing and retrieval section of the *Sound Palette Offline* application.

- Basic: Duration, Instrument
- Melody related: Tessitura, Melodic Density, Pitch Level, Pitch Variety, Melodic Contour, Duration Variety, Direction, Dissonance
- Rhythm related: Tempo, Meter, Tick, Swing factor
- Morphological related: Envelope, Mass, Mass Profile, Pitch Profile

Given a query, the system retrieves a list with all audio descriptions in the database that match the desired criteria. By browsing the list, we can listen

to the retrieved audios and eventually load them in the *Sample Editor* for further operations. Moreover, a special hierarchical taxonomy for the *Instrument* criterion was created, as shown on the left of figure 4. In the current state of the application, *instrument* information is not generated automatically, and thus it should be inserted manually a posteriori. A screenshot is shown in figure 4.

5.2. Multimedia and metadata retrieval: MPEG-7 web browser

In the context of multimedia productions (generated with *MDTools*), an MPEG-7-based retrieval system has been implemented to search and retrieve audio-visual and metadata information regarding to the

productions [3].

For simple pull applications (i.e, the client request data from the server) the *Client-Server* computing model is suitable. Hence, a web-based application queries the database –using MPEG-7 descriptors as input– and the browser displays the multimedia segments with all its associated metadata information.

6. IMPLEMENTATION

Both applications have been developed using CLAM [7]. CLAM (C++ Library for Audio and Music) is a free software [8] framework developed at Music Technology Group¹ in order to share the development efforts among projects.

CLAM framework provides music processing algorithms, a way of interconnecting them, and a set of utilities such as exporting XML/MPEG-7 files, dealing with audio devices and audio files.

The *Sample Manager* and the *Sample Editor* use CLAM to implement and connect the processing algorithms. In fact, processing algorithms developed within *SPoffline* are going to be integrated into CLAM framework in the near future.

On the other hand, *MDTools* has been implemented using two building blocks. A GUI interface based on *TrollTech Qt* [9] library and the CLAM framework utilities –basic structures and XML/MPEG-7 support–.

7. ACKNOWLEDGMENTS

This work has been partially founded by the CUIDADO project² and the Opendrama project³.

8. REFERENCES

- [1] Manjunath, B. S., Salembier, P., Sikora, T. “Introduction to MPEG 7: Multimedia Content Description Language”. Ed. Wiley, 2002.
 - [2] Vinet, H., Herrera, P. and Pachet, F. 2002. “The CUIDADO Project”. Proceedings of IS-MIR 2002 - 3rd International Conference on Music Information Retrieval, Paris, France, 2002 October 13–17.
 - [3] Celma, Ò. and Mieza, E. “An opera information system based on MPEG-7”. Proceedings of the AES 25th International Conference, London, United Kingdom, 2004 June 17–19.
 - [4] Mullon, P. “Ecrins: an audio-content description environment for sound samples”. Proceedings of ICMC-2002. Göteborg, Sweden, 2002 September 17–21.
 - [5] Gómez, E., Gouyon, F., Herrera, P. and Amatriain, X. 2003. “Using and enhancing the current MPEG-7 standard for a music content processing tool”. Proceedings of Audio Engineering Society, 114th Convention Amsterdam, The Netherlands, 2003 March 22–25.
 - [6] Ricard, J. Herrera, P. 2003. “Using morphological description for generic sound retrieval”. Proceedings of Fourth International Conference on Music Information Retrieval Baltimore, USA, 2003, October 26–30.
 - [7] CLAM Framework. “C++ Library for Audio and Music”. Music Technology Group, Universitat Pompeu Fabra. <http://www.iua.upf.es/mtg/clam>
 - [8] Free Software Foundation. <http://www.fsf.org>
 - [9] QT-Trolltech. “A multiplatform, C++ application development framework”. <http://www.trolltech.com/qt>
-
- ¹<http://www.iua.upf.es/mtg>
²<http://www.cuidado.mu>
³<http://www.opendrama.com>