9th Symposium on Neural Network Applications in Electrical Engineering (NEUREL), Belgrade, Sept. 25–27, 2008

ENRICHED ACCESS TO DIGITAL AUDIOVISUAL CONTENT

Ivan Damnjanovic, Chris Landone, Josh Reiss and Ebroul Izquierdo

Queen Mary, University of London
Mile End Road, E1 4NS, London
United Kingdom
{ivan.damnjanovic, chris.landone, josh.reiss, ebroul.izquierdo}@elec.qmul.ac.uk

ABSTRACT

This paper presents access engine to digital audio and related content developed under IST FP6 project EASAIER. The main driving force for the project was the lack of qualitative solutions for access to digital sound archives. An innovative remote access system which extends beyond standard content management and retrieval systems, addresses a range of issues identified, such as inconsistent formats of archived materials with related media often in separate collections and related metadata given in non-standard specialist format, incomplete or even erroneous. The system focuses on sound archives, libraries, museums, broadcast archives, and music schools, but the tools may be used by anyone interested in accessing archived material; amateur or professional, regardless of the material involved. The system functionalities; enhanced cross media retrieval, multi-media synchronisation, audio and video processing, analysis and visualisation tools, enable the user to experiment with the materials in exciting new ways.

Index Terms— Sound Archives, Multimedia Retrieval, Music Ontology, marking, looping, time-scaling.

1. INTRODUCTION

A number of key features that are required in order to enrich sound and music archives were indentified by thorough user needs studies [1] and extensive research by the JISC [2-4]. These findings stress the need for web-based access, integration of other media, time-stretching functionality, and alignment of scores with audio, among many others. A recent Scottish study into training needs analysis in e-learning [5] reported that audio is still an under-used technology.

We focused on key areas still lacking a deep, systematic, and focused approach: multi and cross- media retrieval, interactivity tools, integration of speech and music processing methods, and systemic archive analysis. In order to cope with these kinds of problems, innovative audio processing, data mining, and visualization techniques, have been developed and integrated into the system. These will be deployed in several sound archives in order to demonstrate a qualitative jump in usability, effectiveness and accessibility.

The major challenges identified during architecture design are:

- Speech and music technologies integration to enable user to have common access point for archives and web resources
- Related materials (image, video, text) access regardless of its source (the archive itself or the web).
- A *common set of metadata* and a mapping for various existing archive ontologies.
- A common timeline to enable easier access to the documents and its segments.
- Integrating both similarity and metadata retrieval
- Synchronisation of media components for enriched access and visualisation

2. SYSTEM ARCHITECTURE

The system architecture is typical client/server architecture. It is design to support the possibility to express relations between every data assets. Data and metadata are treated at the same level, linking one to each other.

The front-end consists of two types of clients: web client and advanced user application. The web client application (Figure 1) allows the user to browse and query the archive according to the retrieval system functionalities. Its functionalities are restricted due to the web application framework and subsequent limitations in speed, memory and processing power. Simple retrieval, playback and visualization are provided but expensive processing like real-time audio filters are not supported. The advanced user application, beside functionalities of web client, allows enriched access, visualization of an audio file and its related metadata and media. The user is able to interact with audio resources through separating and identifying sources, processing and modifying the audio, and aligning various sources at playback.

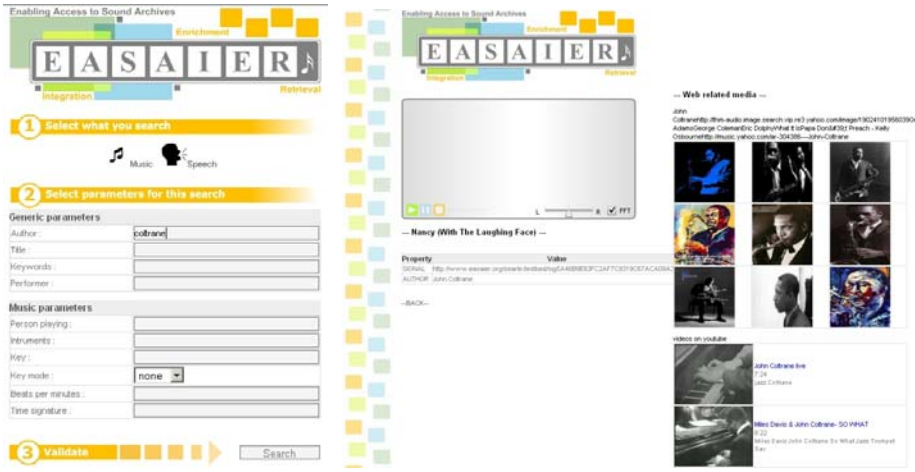

Figure 1 Query Interface and web related media access page (web client)

On the back-end side, there are components to support advanced retrieval functionalities:

- The database storage composed of RDF storage for storing objects identifiers, metadata and related media in a semantic form and media storage for binary data (audio, video, images…)
- The SPARQL end-point to query the RDF storage and return results to the front-end applications (by the way of a web server for the web client).
- The database administration application which is in charge of database administration (add, modify or remove data from the database, manage user and rights). The administration tool is assisted by a feature extraction module to extract automatically metadata from a multimedia file.

Available metadata is increased by the extraction of various low, mid and high level features which enables greater search functionality, such as: the data that describes encoding of actual content of the asset, metadata embedded in the asset, low-level features (MFCC …), and human readable content descriptions (key, tempo, etc). The automatic extraction of metadata is carried out by a stand-alone service that communicates with the binary storage, RDF storage and Archiver components of the system.

To address the types of metadata from low-level features to higher level editorial metadata, the retrieval engine is build around the Music Ontology [6] together with its extension the Speech Ontology. In addition, for purpose of integrating existing archives the system provides automatic mapping to a metadata representation standards, such as Dublin Core [7]. However, to fulfill custom metadata representation mapping needs to be done manually. Such a case was the HOTBED [8] archive that we mapped to the Music Ontology, so the system and tools can be used for querying and processing HOTBED material.

The Music Ontology is built on the Timeline ontology [9] and the Event ontology [10]. The Timeline ontology defines a TimeLine concept, which represents a coherent backbone for addressing temporal information. An instance of this class covers the physical time line: the one on which we can address a date. Another instance may back an audio signal, and can be used to address a particular sample. An Event concept defined in the Event ontology, having a number of factors (such as a musical instrument, for example), agents (such as a particular performer), products (such as the physical sound that a performance produces) and a location in space and time (according to the Timeline ontology). This definition is broad enough to include performances, and compositions, but also structural segmentation, chord extraction, onset detection, etc. In addition, the Friend-of-a-friend (FOAF [11]) ontology and the Functional Requirements for Bibliographic Records ontology (FRBR [12]) are incorporated in the Music Ontology. First enables modeling of musical groups, artists, labels, and other music-related agents. Latter is used for its concepts of Work, Manifestation and Item.

3. ENHANCED CONTENT RETRIEVAL

The system integrates multiple retrieval engines, allowing for searching of content and metadata using multiple techniques and modalities**.** Music retrieval engine consists of the generation of compact representations for both the query and the collection and the search for similarities between these representations. Most music retrieval systems use low-level features which allow fingerprinting of audio files, but are limited to only exact match retrieval. By using appropriate higher level features, ranked lists of audio files are obtained related to the query through melodic and harmonic similarity.

The music retrieval incorporates both audio similarity search and search on features. The audio similarity metric used in Soundbite [13] (audio similarity engine used in the system), is undisclosed, whereas the similarity metric for searching on metadata is part of the open system architecture.

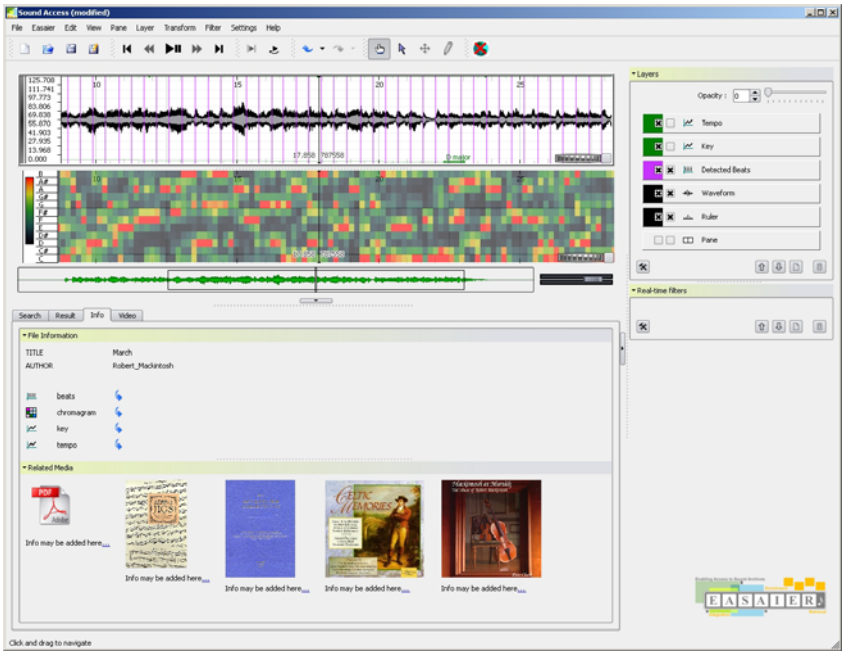

Figure 3: Client Interface audio presentation and related media

For example, Soundbite can calculate normalized similarity value between two assets to be 0.7. The user also assigns the key of a query to be C major, hence its confidence is 1. On the other hand, the track in the collection can have automatically estimated key to be C major with a confidence of 0.7. Clearly, these two assets will be more similar then the query and another acoustically similar asset (with same value of 0.7) that is in C major with confidence of 0.6. Weightings w_1, w_2,… w_N are assigned to each feature as well as acoustic similarity calculated by Soundbite w_{AS}, giving a fully flexible similarity measure that takes into account both the importance of each feature and the confidence in the estimation of that feature, for both query and retrieved track. The speech retrieval features complement the music retrieval features in order to support the interrogation of archives with mixed content. In this case, the speech and music parts of the sound materials are managed separately, using the adequate algorithms. The speech/non-speech/music segmentation ensures the separate preparation, sound object identification and indexing. In addition cross-media retrieval allows the user to search media in various formats (audio recordings, video recordings, notated scores, images etc…) and find related material across different media (Figure 3).

Retrieved asset is presented to the user in intuitive and configurable interface enabling enriched access to the content. Tools to allow time aligned textual markup are provided to the user. Sections of audio can be selected manually or automatically and looped seamlessly for learning or analysis purposes. Advanced audio signal processing tools allow the user to specify how they listen to and interact with the media content. Figure 3 shows a screen shot of the client interface with audio content in both the time and frequency domain.

A source separation tool allows a user to listen to individual instruments within the piece of music while a noise reduction tool can be used to eliminate unwanted artefacts. The real-time equalizer tool is comprised of a simple 5 band (fixed) equalization tool for simple audio personalisation, but also features an expert mode which allows users to draw a freehand curve. Time-stretching allows a user to slow down (or speed up) recordings, without modifying the pitch in real-time. This enables a music student or musicologist for example, to easily learn or analyze a piece of music. The ability to speed up the audio content also gives the user the ability to browse long segments of audio rapidly. The same technology also allows for pitch-shifting of the audio without affecting the time scale. A key innovation also allows the video stream to be synchronised with the audio during time and pitch scaling. It is also possible to zoom the video content, allowing closer inspection of an instrumentalist's particular technique. This suite of enriched access tools is presented in real-time with all functionalities accessible simultaneously.

4. CONCLUSIONS

We presented a state-of-the-art access system for sound archives, incorporating multiple, integrated retrieval systems, and enriched access tools which allow manipulation of the resources. The user requirements for

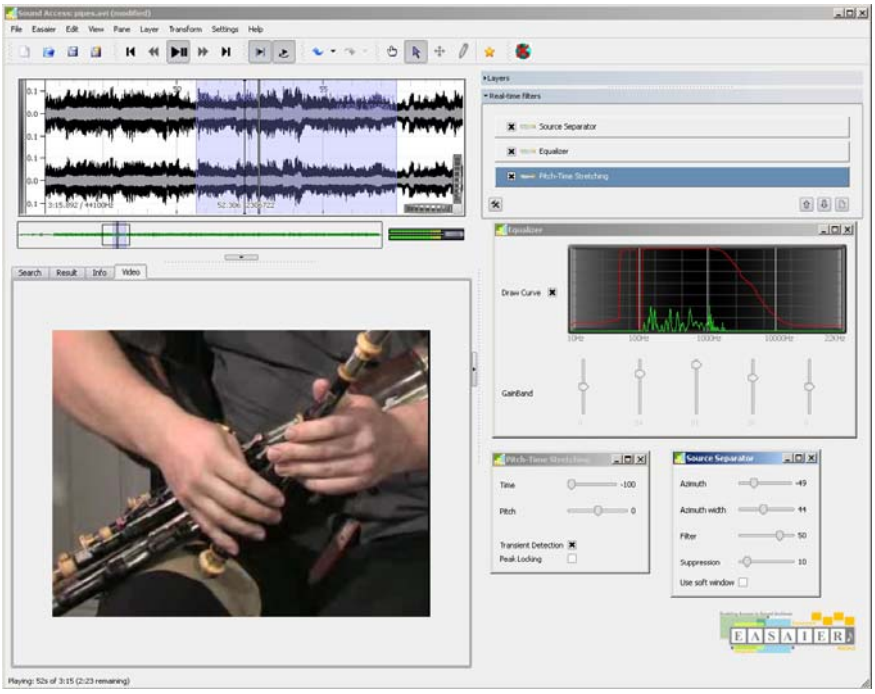

Figure 4: Client Interface with video time scaling

such a system are carefully studied and considered during design process. System architecture and integrated technologies, such as Music ontology, integrated audio similarity and metadata retrieval, addressing mentioned challenges are described. In the next stage, are aim is to deployment and user evaluation of the system on a large scale, benefiting sound archives in cultural heritage institutions which until now have been unable to provide advanced access systems to their users.

ACKNOWLEDGMENTS

This work has been partially supported by the European Community the COST Action 292 and FP6 project EASAIER (IST-033902): www.easaier.org.

REFERENCES

[1] S. Barrett, C. Duffy, and K. Marshalsay, "HOTBED (Handing On Tradition By Electronic Dissemination)," Royal Scottish Academy of Music and Drama, Glasgow, Report March 2004.www.hotbed.ac.uk

[2] M. Asensio, "JISC User Requirement Study for a Moving Pictures and Sound Portal," The Joint Information Systems Committee, Final Report November 2003. www.jisc.ac.uk/index.cfm?name=project_study_picsounds

[3] "British Library/JISC Online Audio Usability Evaluation Workshop," Joint Information Systems Committee (JISC), London, UK 11 October 2004. www.jisc.ac.uk/index.cfm?name=workshop_html

[4] S. Dempster, "Report on the British Library and Joint Information Systems Committee Usability Evaluation Workshop, 20th October 2004," JISC Moving Pictures and Sound Working Group, London, UK 20 October 2004

[5] "Higher Education Training Needs Analysis (HETNA)," Scottish Higher Education Funding Council (SHEFC), Sheffield, UK November 2004. www.shefc.ac.uk/about_us/departments/learning_teaching/hetna/hetna.html

[6] Y. Raimond, S. Abdallah, Mark Sandler and Frederick Giasson, "The Music Ontology", Proceedings of the International Conference on Music Information Retrieval, 2007

[7] S. Weibel, J. Kunze, C. Lagoze, and M. Wolf, RFC 2413 - Dublin Core Metadata for Resource Discovery, 1998

[8] S. Barrett, C. Duffy, and K. Marshalsay, "HOTBED (Handing On Tradition By Electronic Dissemination)," Royal Scottish Academy of Music and Drama, Glasgow, Report March 2004. www.hotbed.ac.uk

[9] Y. Raimond and S. A. Abdallah, "The Timeline Ontology", 2006

[10] Y. Raimond and S. A. Abdallah, "The Event Ontology", 2006

[11] Dan Brickley and Libby Miller, FOAF Vocubulary Specification, 2005

[12] I. Davis and R. Newman, Expression of Core FRBR Concepts in RDF, 2005

[13] M. Levy, M. Sandler and C. Sutton, "Soundbite", http://www.isophonics.net/SoundBite