PERSONAL SOUND BROWSER
A Collection of Tools to Search, Analyze and Collect Audio Files in a LAN and in the Internet

Sergio Cavaliere, Carmine Colucci
Università di Napoli "Federico II"
Dipartimento di Scienze Fisiche
Via Cinthia 80129 Napoly (Italy)

Keywords: Multimedia Databases, Audio Browsing.

Abstract: In this paper we present a toolbox aimed to search for audio files on the Internet, in a Local Area Network or in a single computer. Search is finalized both to analyze the collected files and to populate a multimedia archive for further use or analysis. The related tools to interface to a multimedia Data Base and analyze files is also provided. The toolbox is intended to be open in the sense that any user may customize it at will adding proprietary tools and methods. It is freely distributed and also open to contributions. The goal has been achieved building a Matlab Toolbox; this, as is well known, results in an open environment that anybody may customize at will. Research in the field of music and sound browsing analysis and classification is a large and open field in which a large amount of different solutions have been proposed in the literature. Deciding which sound parameters are suited to a kind of search or classification is still an open problem: we are therefore providing an open environment where anybody may customize at will tools and methods, an environment which, as a plus respect to other tools in the literature, starts from the very first stage of the process, searching and browsing directly from the Internet. Our work goes in this direction and proposes an open environment made of open tools for the purpose. The language used allows also, as a further benefit, the advantage of straightforward prototyping of new tools. Interested researchers are kindly invited to email the authors for the distribution of the toolbox.

1 INTRODUCTION

An open and very interesting field of research is the search for tools and methods to implement automatic identification, indexing and segmentation mechanisms suitable for musical audio and sounds (Tzanetakis & Cook, 2000, Zhang & Kuo, 2001, Wold , 1996, Foote, 1999, Peeters , 2002). This is a new concept in network and database navigation: the navigation in audio environment, signals and messages. This innovative type of navigation will also allow, in automated mode and at the several levels of the acoustical message content, to explore the large mass of sound and musical information that may be found in the Internet, by means of powerful audio search, interpretation and classification tools.

The indexing of these multimedia objects poses completely new problems, consisting in the identification and the analysis of the audio signal, by means of large set of parameters which may describe the audio content and allow both classification and search of sonic material for different purposes, including just listening, or using audio data for music composition or collect audio material for other purposes.

The starting idea is to provide an instrument which easily allows browsing in the Internet or in a personal computer or in Local Area Network, just as directory explorers allow listing directories, searching for particular files based on names, length, format or even content, and displaying file information and features. The same simple paradigm of file browsers should allow searching for audio file based on file name, format, length, but, most important, on content, sound parameters and others.

This is the idea of our *Personal Sound Browser*. After the search, the files just found may be added to a chosen Multimedia Data Base, in the form of links to their position in the Personal Computer or LAN or their Internet URL.

Cavaliere S. and Colucci C. (2006).
PERSONAL SOUND BROWSER - A Collection of Tools to Search, Analyze and Collect Audio Files in a LAN and in the Internet.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 335-338
DOI: 10.5220/0001571403350338

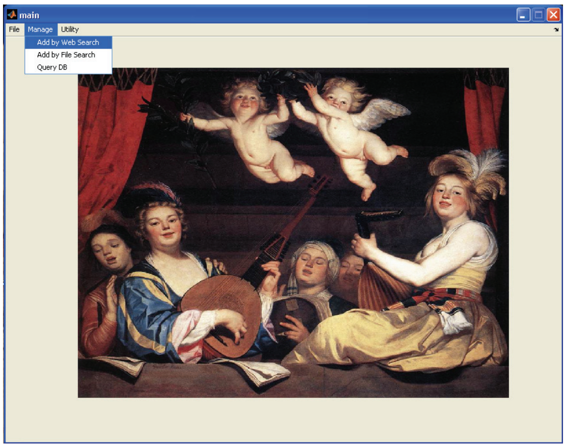

Figure 1: The top level interface of the application. Menù : add by WEB search - add by file search- Query DB.

2 THE PERSONAL SOUND BROWSER TOOLBOX

The toolbox allows searching in the Internet, in a Local Area Network (LAN) or in any storage device of a Personal Computer .

The toolbox provides the user with an interface to a Data Base *Access* (but any different DB is allowed), using which he may populate an existing archive; finally the archive may be visited, annotating sound files and a search in it may be performed both by example, that is providing a source file, or specifying a range of parameters for similarity comparison.

The whole toolbox is organized as a collection of files and routines which at the top level allow the following operations (see figure 1):

- Add by file Search
- Add by Web Search
- Query DB

2.1 Add by File Search and Web Search

The search, as already stated, may be performed in the computer (any storage device in it), just providing the starting directory and the number of levels to be analyzed in the directory tree. In the actual realization parameters for search operation and listing are grouped along the following organization:

file features: name-extension

sound features: number. of bits, number of channels, sampling rate

options for the download: download with preview-max dimension

In the same GUI, whose functionality is straightforward, the user may select the destination archive to populate, if for some files the user decides to store them in the archive, with predefined tables and structure.

A second interface GUI is accessed choosing the *adding by web search* menu entry.

This GUI is quite similar to the *Add by File search* interface and is used in order to state the modality of search; the search is then performed just reading the source page, looking for links to the chosen kind of files and adding these URLs to a list; then, in the same page, links to other pages are searched for, going down in the resulting tree by using a recursive depth first search (DFS) algorithm.

The search is performed using regular grammars and stops as far as a chosen number of levels is traversed or a predetermined number of pages is visited. The search may be delayed in time, pre-programming it to be performed at definite time, and at the end of the search a shut-down may be programmed. In this case, at the end of the search, the results, that is an *html* file containing the addresses of the files found in the search, is stored as a log file, for further processing.

Here also functions are grouped by category as shown in the following list:

- *Source for the search*
- *search options* include: name, extension, number of pages, number of levels for depth search, optional visualization of web pages during the search
- *sound features* such as: number of bits - number of channels - sampling rate
- *Download options* e.g maximum length, etc
- *General options*: delay the search, save the search, shutdown PC when search has finished.

While most of the items are readily understood from the name, we will point out some relevant features; first of all search in the Internet may be started from a specific URL; browsing at this address will mean that we are just looking for sounds, but also we are analyzing the content of the specific URL: in fact we may visualize the starting page and also the linked pages that will be visited during the search, if we decide to practice this kind of navigation; this will thus benefit also of information on the context of the sounds.

A second possible source for the search may be an *html* file saved by the user on the local disk, as a result of a search performed by any search engine. Our browser, in this case, stands on top of professional and efficient commercial search engines whose work we are enabled to refine just entering the name of the performed and stored search; from this our browser will start looping on the signalled URLs for further analysis of the content.

Finally a third case may be that of a search already performed by us by means of the Personal Sound Browser, whose results have been stored in our PC for this further analysis, as a log file.

Search is performed by means of depth first traversing algorithm, allowing exhaustive search of pages in the tree, up to a programmed number of levels.

The second stage in both GUIs, *the add by file search* and *add by web search,* is that of showing the results of the search, if any, and the analysis of individual files. The user may then ask for downloading which is performed by means of a call to the WGET free executable, from the GNU project:
(http://www.gnu.org/software/wget/wget.html).

For format different than *wave* the use of executables from the LAME project (http://lame.sourceforge.net/) allows audio format conversions.

The screen GUI has the following sections:

- *list of files*: where you may choose the file to download and analyze
- *signal plot*
  - spectrum: linear scale / db scale / mel scale for frequency axis
  - specgram linear frequency scale / mel scale
  - start time - final time of signal to display - play
  - view web page: option to view the source web page from which the file was downloaded.
- *Data Base operations*
  - play-list: this is a token to be stored in the Data Base, subjectively chosen by the user: it may be a stile, or a user defined collection or other
  - insert selected sound file
  - insert all sound files found

2.2 Browse the Data Base

The common interface to the DB, is provided by this functionality, which allows selecting files in the DB in order to fulfil some chosen features defined as their mean value and dispersion. The parameters have been computed on request, one by one, so that the user, displaying the features along time may carry on a thorough analysis of the signal.

Most of the features, are stored in form of time varying features, on the basis of a chosen time window, and their statistical distribution, including mean and standard deviation, for later comparison (Burred & Lerch, 2004).

The parameters chosen so far include, as suggested by the current literature:

- ZCR zero crossing rate
- RMS root mean square (see Figure 2)
- PITCH
- CENTROID weighted mean frequency
- ROLLOFF (related to energy distribution)
- FLUX (related to energy distribution)
- MFCC Mel frequency cepstral coefficients

These parameters may be easily extended at will by any user and are actually a starting point to be improved along lines easily found in the large literature in the field (Wold 1996, Scheirer 1997, Rossignol 1998, Brown 1999, Lu 2001, Zhang 2001, Zölzer 2002, Tzanetakis 2002, Burred Lerch 2004) .

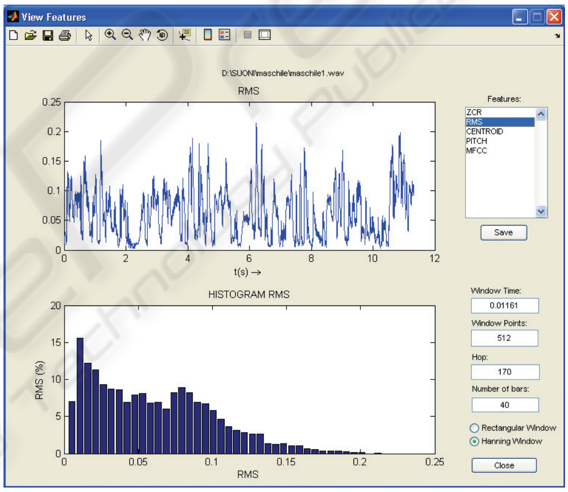

Figure 2: The screenshot for a time varying parameter: RMS value and histogram.

Search in the database is actually performed by means of a similarity criterion: the parameters of a chosen file are displayed in the form of their mean and standard deviation; the user may modify these parameters by hand or apply a multiplicative coefficient to the standard deviation; the search is then performed just looking for files whose mean value of the parameters fall in the programmed range; increasing the multiplicative coefficient of the standard deviation broadens the range of files collected, while further parameter refinement allows reducing the selected class at will, up to a desired class of sounds.

3 DEVELOPMENTS

Many improvement are programmed in our project, in all phases of its operation.

An improvement is that of using eventual XML content information for the search (Bellini Nesi 2001, Haus Longari 2002) and that of using text information from the URL by means of techniques from Natural Language Processing, to be added to the content information obtained by the signal: the context of the sound file, description, annotations and similar may in fact add useful information on it.

Other features to be used as means for classification and search will be added, from the large number identified by the literature (Peeters, Rodet 2002); an example is the kind of thumbnails recently introduced by one of the authors (Evangelista & Cavaliere 2005).

A second modality of search will also be implemented, based on histogram similarity using the Kullback-Leibler divergence or other measure. In this case the user will provide an example file or an entire class of files for the search; files are then searched for, which provide best fit to the statistical distribution of the parameters in the example file.

We are also working to an improvement of the program, consisting in a parallel version of it; parallelism will be achieved by a master computer which will divide the burden of annotation in chunks and will send tasks to slave computers (these mostly are in the LAN, but also might reside in any position in the network); these slaves, as soon as the user in them decides to open to parallel processing, will signal its presence in the net and will be waiting for the completion of the task. The maarester in fact will receive the address of the slaves which are ready and will send to it a specific task. The granularity of these tasks is easily identified in the analysis of the different sound files: the master just sends the address of the files in the Internet: the slave will download the sound file and, in turn, send back the computed sound parameters to be stored in the archive for further search.

The practice of our project has collected its first encouraging results, showing that it has configured a complete set of tools, which, installed in a Local Area Network, in a studio or also classroom or Research Laboratory, allows easily the efficient paradigm of a parallel archive with distributed storage and also distributed processing.

Also we realized that in spite of the use of high level interpreted languages the efficiency of the program is quite satisfying, while easiness of prototyping lets experiment easily new solutions: on the other end a compiled version of the Sound Browser speeds up both search and classification.

REFERENCES

Bellini, P. Nesi, P., 2001 WEDELMUSIC format: an XML music notation format for emerging applications Proceedings of the *First International Conference on Web Delivering of Music*.

Burred JJ, A Lerch 2004 Hierarchical Automatic Audio Signal Classification *Journal of the Audio Engineering Society*. Vol. 52, No. 7/8.

Evangelista G., Cavaliere S. 2005. Event Synchronous Wavelet transform approach to the extraction of Musical Thumbnails*, Proc. of the DAFX05 International Conference on Digital Audio Effects Madrid, Spain.*

Foote, J. 1999. An overview of audio information retrieval. *ACM Multimedia Systems*, 7:2–10.

Haus G, Longari M, 2002 Towards a Symbolic/Time-Based Music language based on XML Proc. First International IEEE Conference on Musical Applications Using XML (MAX2002), New York.

Lu L., Hao J., and HongJiang Z., 2001. A robust audio classification and segmentation method*. In Proc. ACM Multimedia, Ottawa, Canada*.

Pachet F, La Burthe A, Zils A, Aucouturier JJ - Popular music access: The Sony music browser *Journal of the American Society for Information Science and Technology*, Volume 55, Issue 12 , Pages 1037 – 1044.

Panagiotakis C, Tziritas G, 2005. A Speech/Music Discriminator Based on RMS and Zero-Crossings - IEEE Transactions on Multimedia.

Peeters G., Rodet X., 2002. Automatically selecting signal descriptors for sound classification. *In Proceedings of ICMC 2002*, Goteborg, Sweden.

Rossignol S., Rodet X., 1998. et al. Features extraction and temporal segmentation of acoustic signals. *In Proc. Int. Computer Music Conf. ICMC*, pages 199–202. ICMA.

Scheirer E., Slaney M., 1997. Construction and evaluation of a robust multifeature speech/music discriminator. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing ICASSP*, pages 1331–1334. IEEE.

Tzanetakis, G. Cook, P., 2000, MARSYAS: a framework for audio analysis . *Organised Sound, CambridgeUnivPress* 4(3), pages 169-177.

Tzanetakis G. and Cook P., 2002. Musical Genre Classification of Audio Signals *IEEE Transactions on Speech and Audio Processing,* VOL. 10, NO. 5, JULY p. 293.

Vinet H, Herrera P, Pachet F. , 2002. The Cuidado Project: New Applications Based on Audio and Music Content Description *Proc. ICMC.*

Wold E., Blum T., Keislar D., and Wheaton J., 1996. Content-based classification, search and retrieval of audio. *IEEE Multimedia*, 3(2).

Zhang T. and Kuo J., 2001. Audio Content Analysis for online Audiovisual Data Segmentation and Classification *IEEE Transactions on Speech and Audio Processing* (4):441–457, May.

Zölzer U. (ed.). 2002. *DAFX - Digital Audio Effects*. John Wiley & Sons.