Browsing inside a Music Track, the Experimentation Case Study

Guillaume Boutard[1], Samuel Goldszmidt[1], Geoffroy Peeters[1]

[1] Ircam – CNRS/STMS,
Place Igor Stravinsky 1 - 75004 Paris - France
boutard@ircam.fr, goldszmidt@ircam.fr, peeters@ircam.fr

Abstract. During the European project Semantic HIFI, a new HIFI system with innovative features was implemented. Among these features is an innovative GUI for browsing inside the music structure. This particular feature was tested during a set of user sessions. The very positive feedback did not conceal several lacks pointed out by the users. For example, users pointed out difficulties to understand the GUI and requested new advanced features. The corresponding specification updates will lead toward a powerful semantic tool for structural browsing including both advanced intra and inter audio structure browsing, as well as P2P meta-data exchange possibilities.

Keywords: user sessions, intra and inter-document browsing, music structure discovery, content-description, GUI, meta-data, P2P sharing.

1 Introduction

Semantic-HIFI is a European I.S.T. project dealing with the development of a new generation of HIFI system. The Semantic-HIFI system offers innovative indexing, browsing, rendering, performing, authoring and sharing features.

These features have been implemented in a hardware HIFI system (see Fig. 2) and tested during a set of user sessions. The goal of these sessions was to validate and further elaborate each of the proposed features. These sessions took place in Paris at "La Villette – Cité des Sciences et de l'Industrie" in early July 2006. First sessions were dedicated to basic features (such as browsing by editorial meta-data in a music collection), last sessions to advanced features such as browsing inside a music track. In this paper, we review the results of the later.

About twenty people subscribed to the sessions through either Ircam or "La Villette – Cité des Sciences et de l'Industrie". Eight of them participated to the session on browsing inside a track. People had different backgrounds being either sound professionals (three of them dealing with Hifi Sytem, multimedia or music business) or standard HIFI systems end users (five of them). Each session lasted two hours.

2 Browsing inside a music track: concept and graphical interface

2.1 Representing music track by structure

Navigation into a music collection (browsing in a music database) has been the subject of many researches since a decade. Browsing by artist, title, music genre or years are now common features. The possibility to extract automatically the content of a music track has recently opened new possibilities. Query by humming, search by tempo, rhythm, key or by orchestration are new ways for searching music. All these features, which are included in the HIFI system, however only concern what is called "intra-document" navigation. Finally, the goal of the user is always to listen to a music track. In most systems, this is always achieved using the old CD player paradigm, i.e. with a start/ stop/ blind-forward/ blind-backward/ pause buttons and displaying the time elapsed (remaining) since (to) the beginning (end) of the track.

A new listening paradigm, which allows the user to interact with the music content, has been proposed in [1] [2] [3]. The old CD player interface is replaced by a map, which represents the temporal structure of the music track. The user is then able to browse in this representation by clicking on it (pointing with its finger in the case of the touch-screen of the HIFI system). The structure represents similar musical events (occurring at various times in a track) by similar visual elements. A chorus will be represented by a specific rectangle and each occurrence of the chorus in the track will be represented by the same rectangle.

Providing time indexes inside a music track is not a new idea. Indeed the first CDs production already contained such information and the first CD players were able to access these indexes (see [4] for a description of Red Book standard). The novelty comes from the possibility to automatically extract these indexes using signal processing algorithms and the possibility to visualize the corresponding structure in a HIFI system.

In our HIFI system, the automatic extraction of the structure is done using a signal processing modules developed at Ircam [1] [2] [3]. This module represents a track as a succession of state. A *state* is defined as a set of contiguous times, which contains similar acoustical information. Examples of this are the musical background of a verse segment or of a chorus segment, which is usually constant during the segment. The algorithm used by the module starts by representing a music track by a succession of feature vectors over time. Each vector represents the local timbre characteristics of the music track. Time-constrained clustering algorithms are then used to provide the state representation from the succession of feature vectors over time (see [3] for a full description).

The output of the module is stored in an XML file, which represents the various segments by their location in the track (starting and ending time) and labels (represented by numbers such as 1,2,3) describing the states they belong to.

This description provides a semantic of the structure of a music track however it does not provide a semantic of each segment. Indeed, it is still difficult to assign a semantic to each part (i.e. telling which number among the above 1,2,3 is the chorus and which one is the verse) without making strong assumptions on the music genre (such assumptions are in fact not justified for many tracks). The extracted structure is then read by an advanced media player [5], which allows the user to interact with the song structure during its listening (see Fig. 1)

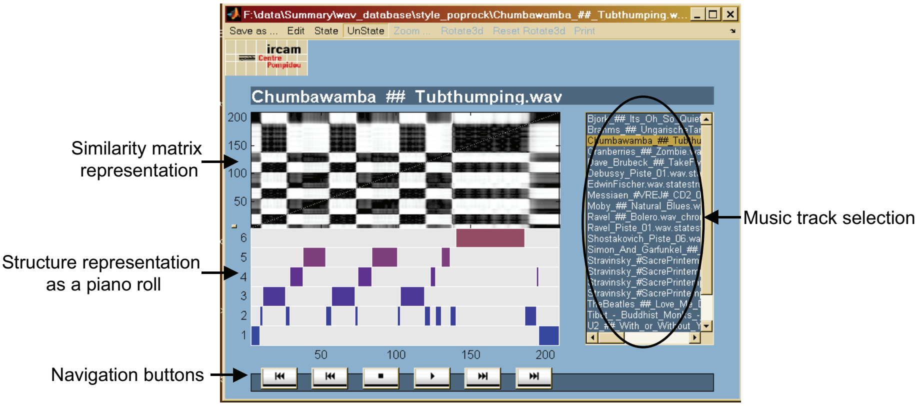


Fig. 1. Prototype of advanced media player (as proposed in [5]) displaying the track similarity matrix (top part of the display) and the structure representation as a piano-roll (lower part of the display) for track Chumbawamba "Tubthumping" [EMI].

A specific version of this player has been developed for the Semantic-HIFI system. This version has been developed using Adobe© Macromedia Flash technology. The Semantic HIFI system relies on a Linux Debian DEMUDI kernel, graphical interfaces of the system use Flash 7 version for Linux.

2.2 Interface for browsing inside a music track

Interface Requirements. This tool was designed not only to be compliant with the HIFI system (see Fig. 2) touch-screen but also with the remote control of the system, which is a PDA (see Fig. 3). It then had to be kept as simple as possible but still being intuitive and efficient. The possible interaction between the GUI and the audio server system would be handled via a bi-directional OSC [6] XML-socket, exchanging time-based meta-data for the current track.

The GUI is meant to propose a simple and intuitive display. This was achieved by focusing on the structure representation part of Fig. 1 (provided in the XML file) instead of the similarity matrix, balancing the loss of mathematical information with the ergonomics and the comprehensibility. Each time a user wants to display the structure of a given track, the corresponding XML file is loaded and the graphical interface is dynamically created.

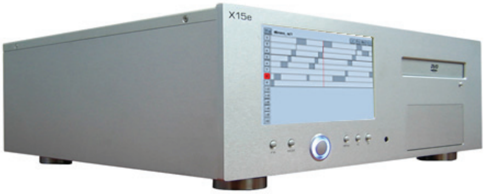


Fig. 2. The Semantic HIFI

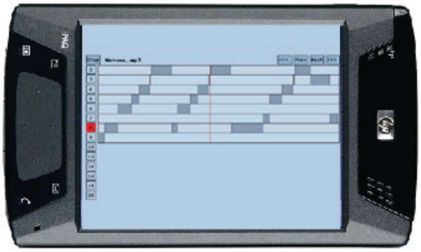
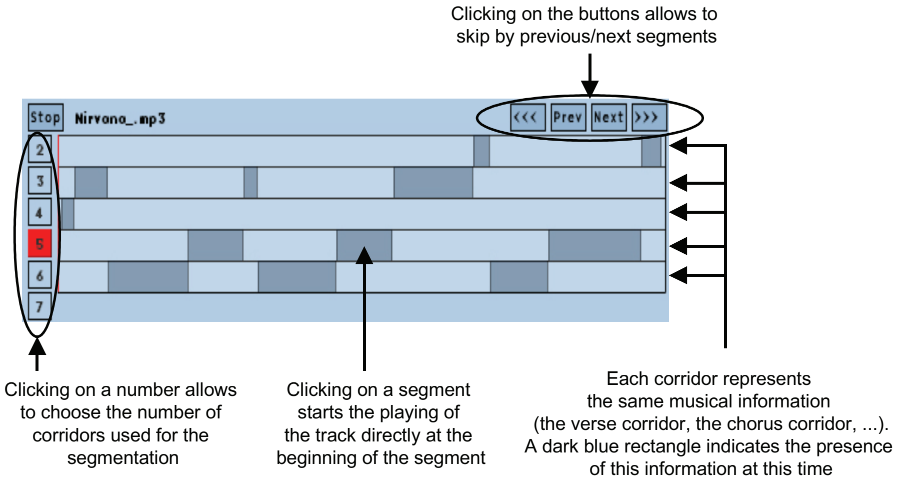


Fig. 3. The Semantic HIFI PDA Remote Control

Interface proposed. The visualization tool displays the track structure as an audio/midi sequencer representation (see Fig. 4). Time is represented in the horizontal axis.

Horizontal corridors represent the various states of the track (one corridor for state 1, one corridor for state 2 and so on). The number of corridors then corresponds to the number of states used by the extraction module. Inside a corridor, a dark blue area indicates the presence of this state during this period of time. Therefore, two areas in the same corridor indicate two occurrences of the same state at various times. During playing, a red cursor (vertical navigation bar) scrolls through time (like in a music sequencer).



Clicking on the buttons allows to skip by previous/next segments

Clicking on a number allows to choose the number of corridors used for the segmentation

Clicking on a segment starts the playing of the track directly at the beginning of the segment

Each corridor represents the same musical information (the verse corridor, the chorus corridor, ...). A dark blue rectangle indicates the presence of this information at this time

Fig. 4. HIFI system interface for browsing inside a music track. The track “Smells like teen spirit” by Nirvana is represented using five states.

Interaction with the interface. Using this display, the user can perform a set of actions.

- Click directly on a dark blue segment. In this case, the action performed is to start playing the sound file at the beginning time code of the selected segment. However, when reaching the end of the selected segments, the sound file will go on playing.

- Click anywhere outside the dark blue segments. In this case, the action performed is to start playing from the click position.

Four buttons are positioned on the top right corner of the display: “<<<”, “Previous”, “Next” and “>>>”.

- The buttons “Prev” and “Next” allow the user to skip back or forward to neighboring segments whatever state they represent (jumping from verse to chorus to bridge and so on)
- The buttons “<<<” and “>>>” allows the user to skip back or forward by previous or next segment representing the same state (jumping from the first verse to the second verse to third verse and so on).

The “Stop” button, in the top left of the display, stops the sound file.

The left part of the figure represents vertically the total number of states used for the segmentation hence the number of corridors. “2” will represent the song using only two corridors, “4” will use only four corridors. The larger this number is the finer the time decomposition will be. For example, in Fig. 4, the user has chosen a visualization using 5 states.

Button locations. The first elements one sees in a GUI are the ones located in the top left area. In our interface, these elements are the ones related to the level of details (number of states) of the segmentation. The user is therefore encouraged to compare the various levels of segmentation. This favors the comprehensibility of the temporal aspect of the horizontal axis: the grouping of similar segments into larger segments (various parts of the chorus collected into a single segment).

3 Validation of the interface by user Sessions

3.1 Protocol

A five steps protocol was set up:

1. Introduction: the host introduces the whole system and the feature to be tested.
2. Presentation: the host explains the tasks the user will have to fulfill.
3. Experiments: the user goes through the tasks he's been asked to fulfill.
4. Discussion: the user reports his experience.
5. Report: the user fills two forms. The first one describes her/his musical background (and other personal meta-data), the second describes her/his feeling about the feature tested.

Eight users tested the browsing inside a track feature. Only one music track was tested [Nirvana “Smells like teen spirit”] [7]. During the whole process, a video camera was shooting what seemed to be the most important moments such as approvals, criticisms or remarks made by any user in order to produce a highlighting movie of the test sessions.

3.2 User Feedback and Specification updates

All users found the feature interesting and useful. It was thought very innovating. However, some ergonomics and comprehensibility lacks were pointed out and suggestions were proposed in order to improve the feature. These are summarized into Table 1. Following, is a discussion on each comment and proposal.

Table 1. User proposals as revealed by the form

Feature Ergonomics	
Weak point specified by users	**Users proposal**
Structural parts are not labeled.	The users could assign a label to each block of the structure
	Exchange structures by the mean of the sharing system (with their names) (made by the author himself or by end-users).
Structural parts are not distinguishable enough.	Assign a color or a spectral representation to each block.
	Add a timeline with tempo annotation.

Labeling. As expected, the first comment pointed out by the user concerns the lack of semantic information about the various segments provided by the algorithm. “Which one is the chorus, the verse, the bridge…?”. After explaining that the extraction module could not provide this information (see above), the users have expressed the need to tag this semantic information her/himself. As proposed by users, labeling could lead to two possibilities. The first one is to assign manually a label to a given corridor (assign the name “chorus” or “verse” to a corridor). The second one is to manually highlight one segment in a song (in this case the highlighted segment is not necessarily repeated) and assign to it a label. This would allow users to browse inside several tracks simultaneously and allows linking different segments from different songs or from various performances. One could then play and compare for instance the same chorus from different performances. This would lead to a semantic cartography of a whole music database based on intra-document structures.

Sharing. As suggested by one user it makes sense to exchange such labels between users. This means using a meta-data exchange mechanism such as P2P[1]. We propose here a framework for such structure annotation exchange network.

At the present time, different users run on their own HIFI system the same content-extraction algorithms (among which the extraction of the structure) often on the same songs. In order to speed up the access to this structural information, it was thought that the sharing of meta-data could handle the structure extracted. The meta-data related to the semantic of the structure could then also be shared through different users. When one looks for these meta-data, one must be able to rely on the fact that this structural information has been extracted from the same song and with the same extractor as its own. The first certification can be achieved using an audio fingerprint

[1] P2P is already the base of the Semantic HIFI system meta-data exchange mechanism.

mechanism. For the second, it is foreseen to include the extractor name and version into shared meta-data. This implies the implementation of several features:

1. An authoring tool for structural meta-data specifications. This tool could first be foreseen as a labeling tool and then enhanced for further meta-data associations such as historical, cultural meta-data ... The most simple way to handle this would be to add an edition mode to the current GUI.
2. A sharing meta-data model enhancement that would add these structural meta-data to the previously designed project model.
3. A method for retrieving these meta-data from the sharing system based on audio fingerprinting.

Comprehensibility. Another comment concerns the comprehensibility of the provided representation of the segmentation. Several proposals were made in order to improve ergonomics, such as adding a timeline with tempi (or just beat marking). This would be however difficult to achieve since this would imply drawing 480 beat markers for a 4 minutes track at 120 bpm. Another suggestion was to add a visual dimension to the segments such as different visual icons for various states.

Authoritative segmentation and authoritative semantic description was also suggested: the possibility to get ‘listening guides’ of music track structure made by a qualified person such as the composer, a musicologist … This would reinforce the sharing option previously discussed. It would also be seen by users as a starting point for personal labeling.

Editing. Some editing features were requested by users. The first feature is the ability to re-order vertically the corridors. The second is to be able to shift horizontally the proposed boundaries of the segments (when the user does not agree with the extraction module) to add new segments or delete them[2]. This would lead finally to a completely user-defined inner structure. Several requests focused on being able to compare various segmentations together: display simultaneously segmentation with various numbers of states. This would require somewhat of a visual transparency mechanism. Another request was the possibility to compare two segments from different songs. This would imply including a way to link segments in our system. This would create the possibilities described in the 'labeling' section, and new possibilities of ‘segments playlist’ mechanism inside one or several track structures, and to some internal remix features according to the segmentation.

5 Conclusion

In this paper we have reported on user sessions of the browsing inside a music track interface of the Semantic-HIFI system. These user sessions revealed high expectations for a feature like this. User suggestions revealed that this tool could be used for many different applications: music summary, intra-document segment comparison; performance studies, inter-document segment comparison, segment

[2] Considering that some users did not agree on the proposed segmentation, we asked them explicitly if, despite they do not agree with the proposed segmentation, they found it useful. The answer was unanimously yes.

playlist generation, segment annotation tool. Considering the scale of this user test it should be noted that the present results are to be taken with precaution. It appears that P2P network would be the appropriate tool to allow developing these features. It seems obvious that additional use of textual meta-data P2P sharing and new browsing features -as previously described- will be leading to some powerful features for semantic inner browsing.

6 Acknowledgments

Part of this work has been released under the framework of the Semantic-HIFI project [8]; http://shf.ircam.fr. Many thanks to all Semantic HIFI partners.

References

1. Peeters, G., Laburthe, A., Rodet, X.: "Toward Automatic Music Audio Summary Generation from Signal Analysis," in Proc. of ISMIR, Paris, France, 2002.
2. Peeters, G.: "Patent AH/EMA-FR 03 07667: Audio summary," 2003.
3. Peeters, G.: "Deriving Musical Structures from Signal Analysis for Music Audio Summary Generation: Sequence and State Approach," in *CMMR 2003 (LNCS 2771)*, *Lecture Notes in Computer Science*, U. K. Wiil, Ed.: Springer-Verlag Berlin Heidelberg 2004, 2004, pp. 142-165.
4. Compact-Disc-Red-Book-Standard, http://www.licensing.philips.com/information/cd/audio/.
5. Peeters, G.: "Indexation et accès au contenu musical," in *Les Nouveaux Dossiers de l'Audiovisuel*, vol. 3, 2005. http://www.ina.fr/produits/publications/nouveaux_da/3/annexes/peeters.light.fr.html
6. Open-Sound-Control, http://www.opensoundcontrol.org/.
7. Nirvana: "Smells Like Teen Spirit", in *Nevermind album*: Geffen, 1991.
8. Vinet, H.: "The Semantic Hifi Project", in Proc. of ICMC, Barcelona, Spain, 2005.