THE SEMANTICHIFI PROJECT : CONTENT-BASED MANAGEMENT AND MANIPULATION OF MUSICAL RECORDINGS

H. Vinet

*IRCAM-STMS; France
1, place Igor Stravinsky, F-75004 PARIS, FRANCE
vinet@ircam.fr
Fax : +33 1 44 78 15 40
http://www.ircam.fr/recherche.html?L=1

Keywords: Hi-fi Systems, Music Information Retrieval, Music Cognition, Digital Signal Processing, Man-Machine Interfaces.

Abstract

The SemanticHIFI project aims at designing and prototyping tomorrow's Hi-fi systems, which will provide music lovers with innovative functions of management and manipulation of musical contents. The limitations of current equipments are mainly related to those of the music distribution media (album-based audio recordings in stereo format), with poor control features and interfaces (album/ track selection, play, stop, volume, etc.). Enabling the manipulation of richer media and related metadata (either distributed with the audio recordings or computed by the user using dedicated indexing tools) opens a wide range of new functionalities: personal indexing and classification of music titles, content-based browsing in personal catalogues, browsing within titles with automatic segmentation and de-mixing tools, 3D audio rendering and assisted mixing features, etc. Moreover, the manipulation of interactive music contents will be made accessible to music consumers, through dedicated performing, and authoring tools. They will then have the possibility of publishing and sharing their personal work with others, through a dedicated peer-to-peer sharing middleware specifically designed for preserving the rights of the used digital media. All these features are the result of the various R&D tasks and experiments performed as part of the project and represent the state-of-the art in various research fields : digital audio signal processing, music information retrieval, man-machine interfaces and peer-to-peer networks. Moreover, the project includes an integration phase, which aims at producing full-featured applications prototypes, designed to fit identified market needs, through technical choices compatible with these markets. This article proposes an overview of the project, by presenting its background and objectives, the main scientific issues and breakthroughs it addresses in relation to the description and extraction of musical information, the main applications features it aims at developing and choices made for their integration into application prototypes compliant with market needs.

1 Project background and main objectives

The development of music technologies over the last decades, in particular in the field of digital audio signal processing, has produced increasingly sophisticated applications for music production, mainly dedicated to professional users, but had a limited impact on end-user applications. More recently, the deployment of the Internet network, combined with the generalization of audio compression techniques promoted in particular as MPEG standards, has opened new ways for the distribution of recorded music, but also new interfaces for accessing, manipulating and listening to music. The development of peer-to-peer exchange protocols, enabling anybody to share copyrighted contents, has been mainly presented as a threat against the traditional music distribution industry, but less as a sociologically interesting phenomenon to be taken into account in new business models. Among the various changes introduced by these technologies, one can mention the replacement of the album as the basic distribution item with the title, and the increasing complexity of management due to the multiplication of the potential number of items[1], which requires new ways of managing musical contents.

Let alone mobile phone-based music distribution, many existing models rely on the fact that the reference access and management device is the computer, even though sound files can be afterwards transferred to personal listening devices. As a successful example, Apple's model combines a computer application, online distribution of audio files and editorial metadata (iTunes/ CDDB), and personal listening devices (iPod). The combination of computers and personal listening devices tends to essentially promote individual situations of listening, as opposed to the traditional Hi-fi system in the living room. One can then wonder about the evolution of the notion of Hi-fi systems. Shall they be definitely replaced with computers as general-purpose digital media manipulation devices? Before existing developments of home networks and media servers come to maturity as products, the main answer

[1] Standard computer hard disks currently enable to store several tens of thousands of titles in compressed form.

in current commercial offer is oriented to Home Cinemas, i.e., as far as sound is concerned, 5.1 reproduction systems, less dedicated to pure music listening than to video sound tracks. However, the generalization of the DVD format as the main production support for recorded video, including 6 DTS PCM-quality audio tracks, or of even higher quality formats such as DVD-Audio or SACD, is to be considered as a significant evolution factor of sound production methods, which had been dedicated for decades to stereo formats.

As a matter of fact, even if new network-based distribution models, facilitated by the availability of good quality compression formats, have started to develop, they do not provide any radical change on the kinds of used musical representations. These representations still rely on stereo recordings, whereas the current state-of-the-art in music technology enables a rich set of representation types along several complementary abstraction levels: the physical, signal, symbolic and knowledge levels [17]. This limitation on supported music representations has direct consequences on possibilities of manipulation of music materials offered to music lovers, which have been limited up to now, even on the most recent Internet distribution systems, to very basic controls : play, stop, track selection, volume, equalization, balance, etc.

The objective of the SemanticHIFI project is to design and implement a new concept of Hi-fi systems, which overcome all aforementioned limitations through the following orientations :

- to implement state-of-the-art research in audio signal processing and analysis in order to propose high-level content-based manipulation of music materials, targeted to non-expert users,
- to consider two complementary ways of obtaining these rich musical contents, either through specific providers as a result of a production process, or through personal indexing tools made available to the system users,
- to go beyond most existing applications of Music Information Retrieval research, generally focused on data models at the server side, by targeting client devices, with man-machine interfaces specifically designed for non-expert users. This is the case of the former CUIDADO project [16], whose resulting technologies are being adapted and further developed in the context of SemanticHIFI,
- on the basis of these rich music representation formats and easy-to-use client devices, to propose advanced, interactive functions of manipulation of musical contents, including inter- and intra- document browsing, 3D audio rendering and assisted mixing, real-time performing, authoring and peer-to peer sharing of the user-produced material. These various targeted functions are presented more in detail in the following sections.

The project is supported by the European Commission (IST Programme) and gathers circa 30 researchers and engineers from five research labs among the most advanced worldwide in music technologies[2] and two commercial companies which are international leaders in their fields[3]. Its main goal is to implement, within a 36 months period ending in November 2006, specific R&D tasks aimed at the development of all these functions in the form of three full-featured, interoperable application prototypes : the Hi-fi system itself, targeted to non-expert users, an Authoring application, dedicated to more advanced users and a sharing middleware and server.

The implementation of all targeted features is the result of a number of tradeoffs related to technical and marketing criteria, but also and mainly to the current state-of-the-art in the concerned scientific fields. Before presenting the various features, the main scientific issues and breakthroughs addressed in the project are introduced in the next section.

2 Description and extraction of musical contents

Despite of the project title, which is to be considered as a metaphorical acronym, one cannot speak of semantics in the case of music. Even if music can relate to specific aspects of human life (religion, war, dance, or other collective rituals and events), or can express some feelings (one of the main functions of music in cinema and TV), music remains an independent way of expression with no specific meaning by itself. Continuing this comparison with language at a syntactical level, it appears that the musical content is highly structured, according to musical theories defining elements of vocabulary such as pitch and intensity scales, time metrics based on periodic pulses, etc. All this information is contained in music scores, from which various kinds of digital representations exist (MIDI[4], MusicXML[5]) and could be ideally considered for setting up musical descriptions for different applications. However, this approach practically meets the following obstacles:

- *Production vs. reception:* the main purpose of musical theories is to define structures for producing music, but not for describing relevant features from a reception viewpoint. As a matter of fact, there is an unlimited number of possible, subjective viewpoints on a given musical piece,
- *Temporal scope* : in particular, the microstructures used for writing music (at the note level) are not necessarily relevant, whereas structural elements at a higher level of description, such as melodic patterns, cannot be isolated in an unique way in the general case,
- *Musical skills*: if the targeted applications are intended to non-musician users, the musical material present in the score is generally too complex for them, and the description structures must account for musical knowledge at a level accessible to these users;
- *Incompleteness of the score information*: the score does not contain important aspects of the musical expression present in recordings, such as sound characteristics (particularly important in electronic music), performance aspects, etc. This

[2] IRCAM (Project coordinator, F), Sony Computer Science Laboratory (Sony-CSL,F), Fraunhofer IDMT (FhG, G), University Pompeu Fabra (UPF, E), Ben Gurion University (BGU, Is).

[3] Native Instruments(G) and Sony European Technology Center(G)

[4] www.midi.org

[5] www.recordare.com/xml.html

information is present in the audio signal and can be, at least theoretically, extracted from it,

- *Evolution of musical theories* Even though the musical works listened to nowadays rely as a vast majority on a vocabulary issued from stabilized musical theories from the 19th century (in particular concerning their melodic, harmonic and rhythmic aspects); these theories have dramatically evolved since the beginning of the 20th century, up to a point of unlimited complexity in contemporary works, which often define their own structural systems. Even if such sophisticated works are quantitatively negligible in the mainstream of commercial music, they qualitatively represent important aspects of the European cultural background and their dissemination requires specific efforts. This issue is particularly important for an institution such as IRCAM as the SemanticHIFI project coordinator, and as a center dedicated to the production and dissemination of contemporary music, and is addressed specifically in the project.

All aforementioned points evidence the necessity of elaborating in the field of music cognition in order to formalize appropriate knowledge structures that will be used as metadata for manipulating the musical contents. Practically, and according to the current state-of-the-art in music cognition, the knowledge elements used in the project rely on the following complementary approaches :

- *editorial information* (title, composer and performers' names, etc.);
- subjective, *user-defined categories* such as genres (classical, folk, jazz, etc.), possibly organized through hierarchical structures;
- *"unary" numerical descriptors*, automatically extracted from the audio signal, which globally account for specific properties of a given piece : orchestral "timbre", tempo, intensity, voice presence, etc. These descriptors are used for inter-document browsing, in particular through the computation of *similarity distances* between sounds, which combine various criteria with appropriate weights and enable to implement, from a reference sound file, a *search by example* according to this similarity distance;
- *audio fingerprints*, which characterize the contents of an audio file in a compact digital signature and are used in the project for identifying the presence of a given title in a Hi-fi node on a peer-to-peer network;
- *structural information* on the music piece *at a macroscopic level*, that will be used for intra-document navigation and personalized spatialization : temporal structure, polyphonic decomposition into various voices or instruments, etc.;
- for simple pieces such as songs, *analysis of the main melody*, in order to enable "query by humming" search features;
- for songs, *use of the lyrics*, synchronized with the audio signal, both for navigating into the audio content by lyrics and for enabling the display of lyrics during playback;
- *simplified score representation*, automatically synchronized with the audio recordings, in order to implement functions such as the comparison of different performances;
- *musicological analyses*, which go more in depth into the elicitation of the inherent structures of a piece, and are produced *in the form of interactive multimedia interfaces*, that can be played back in the Hi-fi system.

Together with music cognition as a top-down process, another major scientific breakthrough in such project is related to the state-of-the art in digital audio signal analysis, which defines the limits of possibilities of extraction of relevant information from the audio signals as a bottom-up process. These automatically extracted data are generally referred to as "low-level descriptors", and are then mapped, through statistical analysis and machine learning techniques, to the high-level knowledge descriptions, so that these descriptions can be automatically inferred from the signal information.

3 Inter-document browsing

The personal management of music files in the Hi-fi system is handled through evolutions of Sony-CSL's Music Browser application framework, in particular through the development of the MCM data management framework [11]. No hypothesis is made on the way the sound files came to the user's system hard disk : this feature is supposed to be part of any complementary music distribution system (online services, CD or DVD ripping, etc.). Once sound files have been loaded to his (her) system, the user can import editorial data from several online providers, which deliver information such as title and artists names and related musical genres. An original feature of the Music Browser is also that high-level descriptors, such as tempo, intensity, orchestral "timbre" [2], can be computed automatically from the audio signals using personal indexing tools. The project also includes dedicated research for extending these automatically extracted high-level descriptors to features such as complex tempo analysis and marking (for variable tempo pieces such as in classical music)[14] and for automatic analysis of the key signature. The user can also define his (her) own high-level classification categories using arbitrary textual labels, and let the system learn the computation of the low-level descriptors associated to the classes he (she) has defined from a few sound examples, using the EDS system[18]. Once sound files are labeled, various content-based navigation features are available, in particular using user-customable similarity distances between sound files. A special adaptation of FhG's query by humming algorithm [7], which delivers lists of sound files containing a melody sung by the user, is also integrated. As an original way of fast browsing between audio files, the system also includes the automatic generation of music summaries[6], which enable the fast listening of the main variations of the musical contents. The available functions, as a result of the CUIDADO project, also include the possibility of automatic generation of title playlists, specified through global constraints based on musical properties : increasing

[6] As a result of an automatic analysis of the temporal structure of the piece, presented in the next section.

tempo, genre continuity, global percentage of occurrence for each genre or artist, etc.

4 Intra-document browsing and spatial rendering

Inter-document browsing is made possible through the use of "unary" descriptors, which provide a global value, associated to a relevant musical parameter, related to the title content. The project also aims at overcoming traditional playback interfaces through the use of descriptions of the internal musical structures of the pieces, which enable innovative intra-document navigation features. In other words, the objective is to refine the Hi-fi system features as a listening instrument. Three kinds of such features are developed. The two first ones correspond to analyses along the two main dimensions of the musical scores : time (horizontal dimension) and polyphonic and spatial superimposition (vertical dimension). The last approach corresponds to more elaborated analyses in the form of interactive hypermedia documents related to the piece.

4.1 Navigation through the temporal structure

Several complementary approaches are taken for obtaining information on the temporal structure of a recorded musical piece. One of them automatically analyses the musical content from the signal as a succession of states, within which the musical content is relatively stable [13]. This enables to exhibit the main parts of the piece such as the introduction, the verses, the chorus, etc., and to produce a graphical representation of the state sequence, which provides an overview of its temporal structure and lets the user start the playback at any beginning of state segment by clicking on its representation. An example of such temporal decomposition is shown in Figure 1. The system also generates "musical summaries", which are sound files obtained by the concatenation of a sound fragment related to the beginning of each different state and which provide, in ten or twenty seconds, the main variations of the whole piece. This concatenation takes care of the beat synchronicity of all used fragments, so that the pulse is preserved in the resulting audio summary. Another approach consists in using, when available, an existing representation of the musical structure, such as a MIDI file or lyrics, and automatically aligning each symbolic event to the sound file. This enables to navigate within the piece by specifying the corresponding symbols, such as lyrics, this feature being developed by BGU. Another experimented application consists in comparing various performances of the same piece through the alignment of each recording to a reference MIDI file.

4.2 Navigation through the polyphony and spatialization

Even if instruments or instrument groups are often recorded separately in multitrack format, this information is lost in the final distribution format such as stereo through the mixing stage. Since network-based diffusion or even the DVD format enable to distribute more channels, there is an interest in preserving this polyphonic information up to the user and offering him (her) interfaces for browsing within it, or to propose a pre-mix containing the main voices of polyphony. This is made possible through the use of a spatialization interface which combines Sony-CSL's MusicSpace interface [5,12] with IRCAM's Spat® 3D audio rendering engine [9]. Spat enables to simulate the 3D audio spatial effect produced by sound sources (as input signals) at a given position and resonating in a room whose acoustical quality is specified through a set of perceptual parameters. In the developed prototype, icons of the various available instruments are represented in a 2D space, which also contains the avatar of the user as the listening point. The user can modify the position of each instrument, as well as his (her) own virtual position, and the system provides in real-time a 3D audio rendering of the sound at his (her) position, with programmable constraints on the position variations in order to preserve the spatial image. The MusicSpace interface used for this prototype is shown in Figure 2, with the visualization of the constraints set on the positions of the various instrument positions. This spatial renderer is compatible with various reproduction systems, such as binaural (headphones), stereo and transaural, and multi-loudspeaker systems such as standard 5.1. When the decomposition of the signal into various sound tracks is not available, the system includes a personal analysis algorithm, which enables, in specific conditions, to separate a lead instrument and its accompaniment and re-spatialize them with different levels [3].

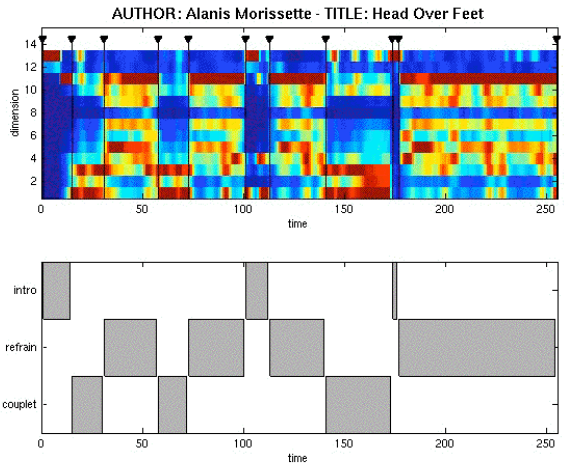

Figure 1: Example of automatic extraction of the temporal structure of a piece by signal analysis.
Credit: Geoffroy Peeters. ©IRCAM

4.3 Interactive multimedia interfaces

The system also includes a player of hypermedia documents, which combine 2D graphical interfaces (e.g. at Macromedia Flash format) with sound files or real-time synthesis. This enables the production and delivery of interactive analyses of musical works based on graphical representations synchronized with the audio, such as in IRCAM's "Signed listening" project [6], which overcome the limitations of

automatic analysis algorithms, and enable to various analytical expert viewpoints on a given piece.
Delivering such browsing interfaces to the music lovers also opens a space for composers to produce new musical forms as interactive pieces, in which the various elements of materials can be discovered separately, and which break the traditional limits of temporal linearity or fixed polyphony. Composers working with IRCAM have already expressed their interest in composing for these new interactive media.

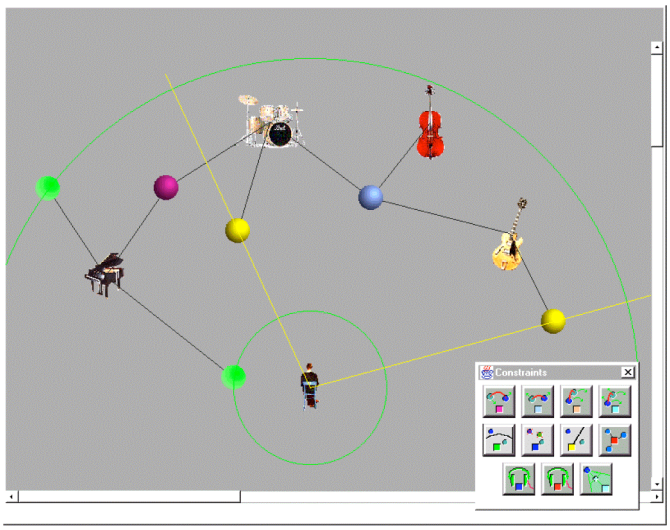

Figure 2: The MusicSpace interface
Credit: Olivier Delerue, François Pachet. ©Sony-CSL

5 Performing

The targeted performing functions represent a step further in terms of interactivity, since the user control is extended to other input modes, such as voice, or gesture. Whereas, in the intra-document browsing functions, the Hi-fi system is considered as a listening instrument, the objective here is to extend it to a simple musical instrument, which makes such performance accessible to non-musicians through a number of appropriate interaction metaphors [8]. A number of them are developed by the group of UPF participating in the project : "conductor" (real-time tempo modification through gesture analysis of beat tapping), beat-boxing and voice-controlled instruments (e.g. trumpet) and effects (e.g. wah-wah pedal), "advanced karaoke", enabling modifications of the voice quality [4] and producing a choir effect from a single voice. Applications of IRCAM's score following technique [15,10] are also considered for producing interactive pieces in which an automatic accompaniment is produced synchronously in real-time with the user's vocal or instrumental performance. Following another approach, the "Song Sampler" by Sony-CSL consists in automatically indexing music recordings from the database and mapping their various parts to the keys of a MIDI keyboard [1].

6 Authoring

The Authoring application is a PC-based application, derived from Native Instrument's Traktor software for the production of DJ performances. Aimed at advanced users, it features offline editing and real-time content-based manipulation of recorded audio materials, including inter- and intra-document browsing, automatic playlist generation, beat-synchronous transitions between pieces, etc. One of its functions is the preparation of musical materials to be played in the Hi-fi system, such as playlists. As an important feature, user operations are recorded in specific "Mix files", which enable the user to re-play them, but also to share his (her) authoring work with other users through the peer-to-peer system without having to distribute the protected materials.

7 Peer-to-peer sharing and DRM

The application components (instances of Hi-fi systems and Authoring applications) are interfaced to each other and to the Internet world through a peer-to-peer sharing middleware, developed by FhG and the 4FriendsOnly company in collaboration with IRCAM. This middleware is based on Sun's JXTA[7] free software and includes client modules linked to each application instance and centralized servers for common functionalities. The system includes the notion of users and user groups, and personal contents can be made accessible to specific user groups, with mechanisms of advertisement, discovery and transfer of the published data. The original approach developed by the project is to promote the use of peer-to-peer networks for direct data exchanges between users, while preserving the access rights of protected items. This is made possible by allowing users to share only materials, including metadata, that they have produced themselves. Each shared material contains two kinds of data : a distribution license, which grants specific user groups access rights to this material, and a unique identifier (fingerprint) of all used protected materials. A user will be able to access to and use metadata produced by another user only if he (she) has the correct access rights, and if the referenced protected materials, identified through their fingerprint are already present in his (her) own system. Any kind of user-produced data can be shared this way, including personal indexing and classification metadata, spatialization, performance, and authoring data. In particular, since automatic indexing algorithms may require a lengthy analysis process, there should be an interest for users to be able to get computation results produced by others.

8 Integration

Once produced and experimentally validated as results of R&D tasks, the features described in the former sections are to be integrated, as part of a currently ongoing second stage of the project, into the development of two application prototypes : the Hi-fi system, developed by Sony EuTEC, and the Authoring Application, developed by Native Instruments.

[7] www.jxta.org

8.1 Methodology

The design and development of the applications follows a process including the following stages:

1. Identification and formalization of use cases (individual features) and user scenarios (sequences of use cases) related to each of the developed features;
2. First stage of user requirement surveys performed on populations of potential users, based on the use cases and on prototypes of the target functionalities;
3. Adaptation of the ongoing R&D tasks to the results of these user feebacks;
4. Finalization of a list of potential target functions, as the result of all R&D tasks;
5. Second stage of requirements performed with potential users and concerned business units by the industrial partners on the basis of the target function list;
6. Selection of functions for each of the two applications as the result of stage 5;
7. Functional specifications of the target applications;
8. Design of the technical architecture necessary for implementing these functions through a modular system enabling all technology providers to contribute using well-defined API[8]s;
9. Development of a first version of the prototype applications;
10. User feedback experiments based on these first versions of the applications;
11. Updated functional and technical specification according to the results of the user feedbacks;
12. Final versions of the prototype applications, validating the project concepts for integration into future products.

At the time of the editing of this article, steps 1 to 5 have been performed and steps 6 to 9 are ongoing. So it is out of the scope of current article to describe in detail the final choices operated for integration. Section 8.2 provides a general presentation of the systems components, and, in order to give a concrete idea of the targeted features, section 8.3 proposes examples of simplified user scenarios that may be operated on the developed systems.

8.2 Targeted systems

The Hi-fi system is designed as an independent hardware device, which includes a high-capacity hard disk, an Internet connection, a colour graphical display, and a processor powerful enough for real-time audio processing. Its user interface uses its own display, as well as a PDA as a remote controller. The Authoring application is a PC-based application, and relies on Native Instrument's Traktor[9] product for DJ authoring and performance. The Hi-fi system is intended to end-users and includes a set of features adapted to this target: personal title indexing and management, inter- and intra-document browsing, playlist generation, sharing of user-produced metadata, etc. The Authoring application is intended to more advanced users and includes inter- and intra-document browsing, authoring, performing and sharing features. All Hi-fi and Authoring application instances are interoperable through the use of the peer-to-peer middleware.

8.3 Examples of user scenarios

Sound files import
The user imports sound files to the Hi-fi system database by ripping titles from a CD/DVD or downloads them from any commercial Web portal.
Personal labelling and classification
The user submits the imported titles to an online metadata server of his (her) choice (CDDB, Musicbrainz, etc.) and imports available editorial information (title name, artists, year, genre, etc.). He (she) can edit and complete this information, and add personal labelling categories to some titles. He (she) can then launch the generalization system that will learn the audio characteristics of his (her) personal categories and compute them for selected titles. He (she) can also launch the computation of low-level descriptors (duration, orchestral timbre, intensity, tempo, key, rhythm description) as well as audio summaries related to a selection of sound files.
Inter-document browsing
The user can specify search criteria using any combination of metadata, including editorial metadata, personal categories, and low-level descriptors. He (she) can also browse in his (her) collection by starting from a given title and search for similar ones according to any kinds of low-level descriptors: the system answers by providing an ordered list of titles according to the selected descriptors. The playback of the obtained titles can be launched from the original sound files or from their audio summaries.
Content-based playlist generation
The user can specify the computation of a playlist according to global criteria, such as: total duration, increasing tempo, percentages of occurrences of selected labels, continuity from one title to the next one according to any similarity criteria (key/tonality, rhythm structure, BPM, timbre, etc.).
Basic playback
The user selects a sound file and starts its playback. Various kinds of related information are displayed on the Hi-fi monitor (title, artists, cover photograph, time, etc.) and, when relevant, the lyrics are displayed synchronously.
Intra-document browsing
The user can listen to a title in an accelerated or slowed down way using a time-stretching function. He (she) can also display the temporal structure of a given title as a succession of various states (intro, verses, instrument solo etc.). He (she) can then select any state and listen to the corresponding sound. He (she) can also launch a small interactive application among various ones (techno loop transcription, polyphonic spatialization, hypermedia analysis, etc.).
Authoring
Using the authoring PC (or Mac) application, the user can create mixes using various features such as automatic tempo

[8] Application programming interfaces
9 www.traktor-dj.com/traktor210home.info

synchronization, morphing, and any kind of VST plug-in. In particular, he (she) can control audio synthesis and processing functions through his (her) voice using various performance plug-ins: wah-wah effect, trumpet or bass synthesis, beat-boxing, etc.

Sharing

The user can publish on the peer-to peer network all kinds of metadata (personal labels, low-level descriptors), playlists and mixes he (she) has produced in either of the applications. He (she) can grant access rights to selections of these materials to specific user groups. Another user in the peer-to-peer network belonging to one of these user groups can then access to these metadata and load them into his (her) system, which informs him (her) of the related sound files already present and the ones missing. When all sound files are present, he (she) can then run the playlists and mixes on his (her) system, and add the selected metadata to his (her) title database.

9 Conclusion

An overview of the SemanticHIFI project has been presented. The original concept of Hi-fi system it develops goes well beyond the functionalities of traditional systems, through the use and combination of a rich set of music representation formats and advanced music management and manipulation features based on these formats (inter- and intra-document browsing, performing, authoring, peer-to-peer sharing). These features allow more interaction with the musical material and break the traditional limits between composition, performance and passive listening, by giving non-musician end-users access to these functionalities through appropriate musical knowledge modeling and manipulation interfaces. The delivery of such music manipulation interfaces designed for end-users paves the way of possible extensions of music production formats towards richer contents and more interactive forms. Beyond the experimental validation of these functionalities in a research context, the project integrates them in fully operational, interoperable, application prototypes specifically designed to meet identified market needs, thus giving its innovative results the best chances of introduction and promotion to these markets.

Acknowledgements

This project is supported by the Information Society Technologies (IST) programme of the European Commission, in the Networked AudioVisual Systems and Home Platforms (NAVSHP) Action Line.

References

[1] J.J. Aucouturier, F. Pachet, P. Hanappe. “From Sound Sampling To Song Sampling”, Proceedings of the International Conference on Music Information Retrieval (ISMIR), USA, (2004).

[2] J.J. Aucouturier, F. Pachet. “Improving Timbre Similarity: How high is the sky?” Journal of Negative Results in Speech and Audio Sciences, 1(1), (2004).

[3] A. Ben-Shalom, D. Dubnov. “Optimal Filtering of an Instrument Sound in a Mixed Recording Given Approximate Pitch Prior” Proceedings of the International Computer Music Conference, Miami, USA, (2004).

[4] J. Bonada. “High Quality Voice Transformation Based on Modeling Radiated Voice Pulses in Frequency Domain”, Proceedings of the International Conference on Digital Audio Effects, Naples, (2004).

[5] O. Delerue. “Spatialisation du son et programmation par contraintes : le système MusicSpace”, PhD Dissertation, Université Paris VI / Sony / IRCAM, (2004).

[6] N. Donin. “Towards organised listening: some aspects of the ‘Signed Listening’ project, IRCAM”, Organised Sound 9(1), (2004).

[7] T. Heinz, A. Brückmann. “Using a Physiological Ear Model for Automatic Melody Transcription and Sound Source Recognition”, Proceedings of the 114th Convention of the Audio Engineering Society, Amsterdam, (2003).

[8] S. Jorda. “Instruments and Players: Some thoughts on digital lutherie”, Journal of New Music Research, 33(3), (2005).

[9] J.M. Jot. “Efficient Models for Distance and Reverberation Rendering in Computer Music and Virtual Audio Reality", Proceedings of the International Computer Music Conference. San Francisco, USA, (1997).

[10] N. Orio, S. LeMouton, D. Schwarz, N. Schnell. “Score Following : State of the Art and New Developments”, Proceedings of the International Conference on New Interfaces for Musical Expression (NIME03), (2003).

[11] F. Pachet, J.J. Aucouturier, A. La Burthe, A. Beurive. “The Cuidado Music Browser : an end-to-end Electronic Music Distribution System”, Multimedia Tools and Applications, Special Issue on the CBMI03 Conference, (2004).

[12] F. Pachet, O. Delerue. “MidiSpace: a temporal constraint-based music spatializer”, Proceedings of the 6th ACM International Conference on Multimedia Bristol, England, (1998).

[13] G. Peeters. “Deriving Musical Structures from Signal Analysis for Music Audio Summary Generation : “Sequence” and “State” approach”, Lecture Notes in Computer Science, Springer Verlag, Volume 2771, (2004).

[14] G. Peeters. “Time Variable Tempo Detection”, Proceedings of the International Computer Music Conference, Barcelona, (2005).

[15] D. Schwarz, A. Cont., N. Schnell. “From Boulez to Ballads, Training IRCAM’s Score Follower”, Proceedings of the International Computer Music Conference, Barcelona, (2005).

[16] H. Vinet, P. Herrera, F. Pachet. “The CUIDADO Project”, Proceedings of the International Conference on Music Information Retrieval (ISMIR), Paris, (2002).

[17] H. Vinet. “The Representation Levels of Music Information”, Lecture Notes in Computer Science, Springer Verlag, Volume 2771, (2004).

[18] A. Zils, F. Pachet. “Automatic Extraction of Music Descriptors from Acoustic Signals using EDS”, Proceedings of the 116th Convention of the Audio Engineering Society, Berlin, Germany, (2004).