THE SEMANTIC HIFI PROJECT

Hugues Vinet
IRCAM-CNRS STMS
1, place Igor Stravinsky
75004 PARIS – FRANCE
hugues.vinet@ircam.fr

ABSTRACT

The SemanticHIFI European project aims at designing and prototyping tomorrow's Hi-fi systems, which will provide music lovers with innovative functions of access 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 through 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. Last but not least, the users 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.

1. OBJECTIVES

The deployment of the Internet network over the last decade, combined with the generalization of audio compression techniques promoted in particular as MPEG standards, has radically changed the way recorded music can be distributed, accessed, and listened to. 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. It seems that the main trends as technical answers go through more protection, both in the distribution networks themselves, and in the use of closed media coding formats which limit the access to given conditions (number of copies, limited subscription period, proprietary listening devices, etc.). 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 tend to essentially promote individual situations of listening, as opposed to the traditional Hi-fi system in the living room. What about the evolution of the notion of Hi-fi systems, then? Shall they be definitely replaced with computers? Before current developments of home networks and media servers come to maturity as products, the main answer 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.

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

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

through personal indexing tools made available to the system users,

- to go beyond most existing applications of Music Information Retrieval research, generally focussed 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[1] and two commercial companies[2]. 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 *Hifi-system* itself, targeted to non-expert users, an *Authoring* application, dedicated to more advanced users and a *sharing* middleware and server. The Hi-fi system will include a high-capacity hard disk, an Internet connexion, a high-quality spatial audio reproduction system, and a processor powerful enough for real-time audio processing. Its user interface may use the TV screen as a display, as well as one or several simple input devices such as a remote control, a PDA, a microphone and a camera for the analysis of the user's gestures. The Authoring application will be PC-based, and will rely on Native Instrument's Traktor[3] product.

2. INTER-DOCUMENT BROWSING

The personal management of music files in the Hi-fi system is handled through evolutions of the Music Browser by Sony-CSL, 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 has defined from a few sound examples, using the EDS system[18]. Once sound files are labelled, 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[4], 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 : increasing tempo, genre continuity, global percentage of occurrence for each genre or artist, etc.

3. 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 musical dimensions : time and polyphony/space. The last one corresponds to more elaborated analyses in the form of interactive hypermedia documents related to the piece.

3.1. 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 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. The system also generates "musical

[1] IRCAM (Project coordinator), Sony-CSL, Fraunhofer IDMT (FhG), University Pompeu Fabra (UPF), Ben Gurion University (BGU).

[2] Native Instruments and Sony Technology Center in Stuttgart.

[3] www.traktor-dj.com/traktor210home.info

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

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 is made beat-synchronous, 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 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.

3.2. Navigation through the polyphony and spatial rendering

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]. 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. 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. Once more, 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].

3.3. Interactive media

The system also includes a player of hypermedia files, 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].

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, such as Jonathan Harvey, have already expressed their interest in composing for these new interactive media.

4. 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 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].

5. 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.

6. 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 free software[1] and includes client modules linked to each application instance and centralized servers for common

[1] www.jxta.org

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 access rights to specific user groups, 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.

7. 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 (multichannel audio, symbolic representations, low- and high-level music descriptors) and dedicated database and sharing middleware. 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. At the time of this publication, this vision reflects the various experiments, made in a research context during the first half of the project, which at least fulfill criteria of technical feasibility, but it will need at a later stage to be faced and adapted to marketing opportunities and related technical constraints, possibly through the application of the various developed technologies in different products.

8. REFERENCES

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

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

[3] Ben-Shalom, A. and Dubnov, S. "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] Bonada, J. "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] Delerue, O. "Spatialisation du son et programmation par contraintes : le système MusicSpace". Thèse de Doctorat, Université Paris VI / Sony / IRCAM, 2004.

[6] Donin, N. "Towards organised listening: some aspects of the 'Signed Listening' project, IRCAM" *Organised Sound 9(1).*

[7] Heinz, T. and Brückmann, A. "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] Jorda, S. "Instruments and Players: Some thoughts on digital lutherie", *Journal of New Music Research,* 33(3).

[9] Jot, J.M. "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] Orio, N., LeMouton, S., Schwarz, D. and Schnell, N. "Score Following : State of the Art and New Developments" *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME03),* 2003.

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

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

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

[14] Peeters, G. "Time Variable Tempo Detection", *Proceedings of the International Computer Music Conference*; Barcelona, 2005.

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

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

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

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