

	SemanticHIFI IST-507913 <i>Browsing, listening, interacting, sharing on future Hi-fi Systems</i>
Final Report Covering period: 01/12/03-30/11/06 Report version: 1 Report preparation date: 28/11/06 Writers: Hugues Vinet, all partners Control : IRCAM Classification: Public Contract start date: 01/12/03 Duration: 36 months Project co-ordinator: Hugues Vinet, IRCAM Partners: - Institut de Recherche et de Coordination Acoustique/Musique, (IRCAM), France - Ben-Gurion University, (BGU), Israel - Sony France, (SONY-CSL), France - Fundacio Universitat Pompeu Fabra, (UPF), Spain - Fraunhofer Gesellschaft Institute for Digital Media Technology/Institute for Integrated Circuits, (FhG), Germany - Native Instruments, (NI), Germany - Sony Deutschland, (SONY-EuTEC), Germany	
 	Project supported by the European Community under the "Information Society Technology" Programme (2003-2006)

TABLE OF CONTENTS

0	OBJECTIVES AND CONTENT OF THE DOCUMENT	4
1	PROJECT OVERVIEW	4
1.1	EXECUTIVE SUMMARY	4
1.2	CONTEXT AND OBJECTIVES	5
1.3	PROJECT ORGANIZATION	9
1.4	CONSORTIUM STRUCTURE AND PARTNERS' ROLES	10
2	PROJECT IMPLEMENTATION	12
2.1	WORKPLAN SYNTHESIS	12
2.2	ACTUAL WORKPLAN IMPLEMENTATION STAGES	12
2.3	USER FEEDBACK METHODOLOGIES	14
3.3.1	<i>Hi-fi System</i>	15
	Proposed improvements of the first version of Hi-fi System	15
	Important Issues for Future Work	16
3.3.2	<i>Sharing</i>	17
	General Description	17
	Features of Interest to be Developed	17
3.3.3	<i>Authoring Application</i>	18
	Proposed improvements of the current Authoring Application system	18
	Important issues for the future work	19
3	PROJECT RESULTS AND ACHIEVEMENTS	20
3.1	OBJECTIVES AND METHODOLOGY	20
3.2	TPOLOGY OF RESULTS	21
3.3	INDEXING MODULES (9)	22
	RESULT N° IRCAM-1: Music Structure extraction	22
	RESULT N° IRCAM-2: Music audio summary generator	23
	RESULT N° IRCAM-3: Tempo, beat, rhythm feature generator	23
	RESULT N° IRCAM-4: Musical key generator	24
	RESULT N° FhG-1: AudioID	24
	RESULT N° CSL-4: Timbre Similarity	25
	RESULT N° CSL-5: Playlist Generation	26
	RESULT N° BGU-1: Source separation	26
	RESULT N° NI-1: Beat & Tempo Detection	27
3.4	ADVANCED AUDIO PLAYERS (6)	27
	RESULT N° IRCAM-13: Interface for navigation within the track temporal structure	27
	RESULT N° UPF-5: Player for the Hi-fi System	27
	RESULT N° BGU-2: Browsing by Lyrics, BGU	28
	RESULT N° NI-2: Native Mix Playback	28
	RESULT N° IRCAM-10: Score Following	28
	RESULT N° NI-3: Seamless navigation in a track	29
3.5	AUDIO SYNTHESIS AND PROCESSING MODULES (8)	29
	RESULT N° UPF-1: Voice Transformation	29
	RESULT N° UPF-2: Voice Instrumentizer	30
	RESULT N° UPF-3: DJ Voice-Controller	30
	RESULT N° UPF-4: Rhythm transformations	31
	RESULT N° IRCAM-11: "X-Micks" Song Morphing Plugin	32
	RESULT N° CSL-10: SongSampler	33
	RESULT N° NI-4: Composition of Segments (Beatmasher)	34
	RESULT N° NI-5: Real-time processing (tempo-synced FX and filters)	34
3.6	USER INTERFACE PROTOTYPES (8)	34
	RESULT N° IRCAM-5: Cocktail party EarBrowser	34
	RESULT N° IRCAM-6: LScanner	35
	RESULT N° IRCAM-7: Visualization tool for the control of Spatialization	35
	RESULT N° IRCAM-8: "Cera Player" Realtime remixing tool	37
	RESULT N° IRCAM-9: "Webern Player" Audio/score synchro & annotation tool	37
	RESULT N° FhG-2: Query by Humming browsing module	38
	RESULT N° CSL-8: BabyBrowser	39
	RESULT N° CSL-9: EyeTune	39
3.7	SOFTWARE FRAMEWORKS (10)	40
	RESULT N° FhG-3: P2P Sharing System	40
	RESULT N° IRCAM-12: Shared metadata model	41

RESULT N° CSL-1: MCM, Multimedia Content Management	41
RESULT N° CSL-2: EDS, Extractor Discovery System.....	42
RESULT N° CSL-3: Music Browser	43
RESULT N° CSL-6: Automatic Generalization of Personal Categories.....	44
RESULT N° CSL-7: Picture Grabbing	46
RESULT N° NI-6: Browsing Module.....	46
RESULT N° NI-7: Multimedia rendering - Embedded Web-Browser and Socket communication.....	47
RESULT N° NI-8: VST-Host	47
3.8 TARGET APPLICATIONS (2)	47
RESULT N° NI-9: Authoring Application	47
RESULT N° EuTEC-1: Hi-fi System Prototype	49
4 DISSEMINATION ACTIONS AND MATERIALS	50
4.1 PUBLIC DISSEMINATION	50
<i>Press</i>	50
<i>Radio</i>	50
<i>TV</i>	51
<i>Public presentations</i>	51
4.2 PROFESSIONAL DISSEMINATION	51
<i>Fairs and exhibitions</i>	51
<i>Technical workshops</i>	52
4.3 SCIENTIFIC DISSEMINATION.....	53
<i>Conferences</i>	53
<i>Journals</i>	56
4.4 CONTRIBUTION TO STANDARDS	57
<i>Proposals</i>	57
<i>Workshop organization</i>	58
4.5 PATENTS.....	58
4.6 PHD DISSERTATIONS.....	58
5 CONCLUSION.....	59

0 Objectives and Content of the Document

This public report provides an overview of the work and main achievements realized as part of the SemanticHIFI project, an R&D project supported by the European Commission in the framework of the FP6 Information Science and Technology (IST) Programme, Networked Audiovisual Systems and Home Platforms (NAVSHIP) Action Line.

Part 1 provides an overview of the project including objectives, workplan organization, and the consortium structure. Part 2 follows with a presentation of the main stages of the project's implementation, including a synthetic report from the user feedback sessions performed on the first versions of the applications. The second half of the document is dedicated to the presentation of the project's main achievements in terms of technological results (Part 3), dissemination actions and materials (Part 4), and a general conclusion (Part 5).

A more complete and detailed description of these various achievements is provided in the project Public Reports specific to each project workpackage: WP2 (Indexing), WP3 (Browsing), WP4 (Rendering), WP5 (Performing), WP6 (Authoring), WP7 (Sharing), and WP8 (Hi-fi System).

1 Project Overview

1.1 Executive Summary

BROWSING, LISTENING, INTERACTING, PERFORMING, SHARING ON FUTURE HI-FI SYSTEMS

In the context of large-scale digital music distribution, the goal of this project is to develop a new generation of Hi-fi Systems, offering new functionalities for browsing, interacting, rendering, personalizing, and editing musical material. This next generation of hard-disk based Hi-fi Systems will drastically change the relationship that the general public has with music and multimedia content. Users will be able to interact with music up to the point of blurring the traditional limits between playing, performing, and remixing. These Hi-fi Systems will be listening stations as much as they will be open instruments.

Technically, with IPv6, every Hi-fi System will have an IP address and will systemically use metadata extraction and exploitation techniques allowing semantic or thematic browsing in large content catalogues on the Web and via file sharing systems. It will also make available manipulation features on the music material.

The main innovations include a more in-depth access to the audio content and structure, through state-of-the-art semantic musical metadata extraction and exploitation (partially from MPEG-7): temporal segmentation, polyphonic, melodic, and high-level descriptions such as genre. Users will be offered innovative possibilities for manipulation, edition, re-composition, organization, sharing, and interactivity with the audio material as well as true 3D audio rendering and real-time music remixing.

Using music from CDs, DVDs or online services, consumers will be able to freely exchange metadata information enriching their music collection and letting them to listen to remix sessions performed by other users.

Bringing together leading European music research and industrial partners, SemanticHIFI delivers next generation tools for Hi-fi Systems and modular PC applications, validated by musicians and experts alike.

The objective is to deliver three complementary and interoperable technical components:

- **A Hi-fi System** that provides, in the form of a dedicated hardware platform, a set of advanced features mainly targeted to various access modes and interfaces in the audio data and metadata materials:
 - title database management and browsing, with automatic metadata and playlist generation and a system of user-defined categories learned and generalized from track examples,
 - intra-track navigation using various criteria and interfaces: automatically computed high-level temporal structure, lyrics synchronized with audio, hypermedia interfaces, etc.
 - sharing of various kinds of metadata with other Hi-fi Systems and Authoring applications, including the possibility of executing mixfiles.
- the **Authoring Tool** software targets more advanced users and runs on a separate PC. This tool complements the Hi-fi System by providing authoring features for the audio data/metadata material, as well as more advanced interactive and editing functions:
 - indexing and metadata generation from audio files,
 - content-based browsing between tracks and within tracks using automatically-generated content descriptors,
 - metadata editing for access in HIFI and sharing systems,
 - generation of mixes and playlists from audio tracks including automatic synchronization between tracks and the possibility of sharing the resulting mixfiles and playlists on a peer-to-peer system,
 - real-time and non real-time audio file processing and personal authoring tools,
 - performing features : spectro-temporal loop editing and morphing, voice-controlled instruments and effects, voice transformation effect, rhythm transformation effect.
- the **Metadata sharing system** available online as a file sharing system connected to a centralized server which enables multiple users to share:
 - o their metadata on audio files (playlists, personal categories, various other user-produced track metadata, etc.)
 - o their authoring work on audio files through authoring tool scripts (mixfiles), that let end users reproduce the different processes performed on audio files, provided that these files are referenced on their Hi-fi System.

1.2 Context and Objectives

Within the last few years considerable technological changes have brought new social, cultural, and economic models into the cultural domain. Music was the first media to benefit from, and stimulate, the development of network protocols (streaming) and compression schemes (mp3, mpeg4, aac). These trends can be summarized as follows:

- content storage cost are decreasing

- value is brought by a tight and innovative combination of technology and content,
- metadata management technologies are at the heart of future information and entertainment systems as they allow access to more data – browsing over Internet and databases – and to better quality data – high level or semantic indexing,
- Intellectual Property Rights Protection (IPR) plays an important role within the distribution of electronic music, video, and other data,
- easier access to content and to content description allows users not only to retrieve and listen, but to modify and perform,
- as proved by the development of mp3 players on mobile telephone, automotive, and Hi-fi Systems, valid business models come from a suitable combination of freedom, instability, discovery, sharing (Internet), quality, and simplicity (dedicated equipment).

Many debates and competition address the issue of where this simplicity will end, the issue of convergence, be it on a PC, on mobile phones, on a TV set, on a DVD player, on a gaming console, or an audio Hi-fi System.

Led by leading European research organizations and industrial music companies, this project focuses on hard-disk based audio Hi-fi Systems with dedicated user interfaces. In addition, authoring and personal publication tools on a PC will be developed as compatible options for this system, validated with the help of musicians and expert users. The goal of the project is to go beyond existing Hi-Fi interfaces with limited control and access to audio material (track selection of a CD, Play, Stop, Record buttons, volume, treble/bass faders, etc.) and to design, develop, and assess innovative and relevant new functions for domestic Hi-fi Systems as a direct application of state-of-the art and targeted research in audio and music technology.

The main principles on which these new features rely include:

- More in-depth access to the audio content and structure, limited up until now to the digital audio signal, through adapted description structures (as standardized in MPEG-7) in terms of temporal segmentation, polyphony, melodic structures, and high-level descriptions (music genre, etc.),
- As a consequence of this kind of access there are increased possibilities of manipulation, edition, re-composition, organization, sharing, and interactivity with the audio material,
- A new step forward into high quality audio rendering through high-level control.

Related scientific and technological objectives can be expressed as follows:

Indexing

The project goes beyond traditional classification, indexing and management systems of audio files on local hard disks, on external storage supports (Audio CDs indexed or converted in files, CD-ROMs, DVDs), and Internet. The indexing modules target the automatic extraction of musical features from the signals of the audio recordings relevant to two main function classes:

- *inter-document management and browsing*: through global descriptors of the musical content of the titles : audio fingerprints, query by humming, analysis of timbre, BPM, rhythmic structure, key, etc,
- *intra-document browsing*: through the analysis and visualization of the internal musical structures of the music piece—high-level temporal structure, polyphony, score, and/or lyrics synchronized to the audio.

Browsing

Innovative browsing features are the visible consequence of the aforementioned indexing technologies with:

- Automatic classification of titles from user-defined categories (categories),
- Query by example (search by similarity with selected excerpts),
- Query by humming: searching for pieces of music by their melody, automatic extraction of melodies from audio material,
- Navigation through any combination of various descriptors of the musical content: artist description, keywords, automatically extracted musical attributes (e.g. timbre, tempo, rhythmic structure, key, etc.), and automatic playlist generation (using metadata management and constraints).

High-Level Listening Control Interfaces

The impressive rendering capabilities of personal home cinema are useless without user control capabilities. The project brings:

- High-quality, variable speed music playback (slow down and/or accelerate) without modification of pitch (time stretching),
- Real-time spatial audio scene synthesis and spatial rendering through a 2D interface that lets the user to move positions of the instruments (or polyphonic voices as independent audio tracks) and the listener,
- Navigation within the high-level temporal structure of the music piece: intro, verses, chorus, etc.,
- Real-time display of lyrics (when relevant) synchronized with the audio playback,
- Execution of small hypermedia interfaces (Flash) displaying relevant graphical information (such as musical analyses produced by musicologists) synchronized with the audio playback.

Interacting and Performing

Even with limited control devices (TV remote control, microphone, game joystick, DV camera, etc.) performing with the system is an important goal of the project bringing:

- Conductor-like control: tempo variations by time stretching in real time using home controllers,
- Voice-controlled instruments: bass, winds, percussions (beat-boxing),
- Voice processes,
- Automatic accompaniment system, synchronized to the user performance (interactive karaoke),
- Song sampler: keyboard playback and concatenative synthesis of song segments.

Authoring

In addition to the Hi-fi System an authoring environment is designed for use on a PC, but in some cases can lead to simple commands integrated in the Hi-fi System for:

- Tools for loop de-mixing and re-mixing,
- Composition of segments: real-time mixing,
- Recording of all filtering and processing parameters for non-real time authoring,
- Composition: sequencing, mixing, morphing using content information,
- Publication of playlists.

Peer-to-Peer Sharing

The project architecture is based on a peer-to-peer middleware that connects individual application nodes (Hi-fi systems, Authoring applications) and enables them to share various kinds of data through mechanisms of user registration and groups. Within the SemanticHIFI project peer-to-peer technology is used for sharing different kinds of non-copyrighted materials like user-produced metadata and categories, playlists, “DJ-mix-files” (generated by

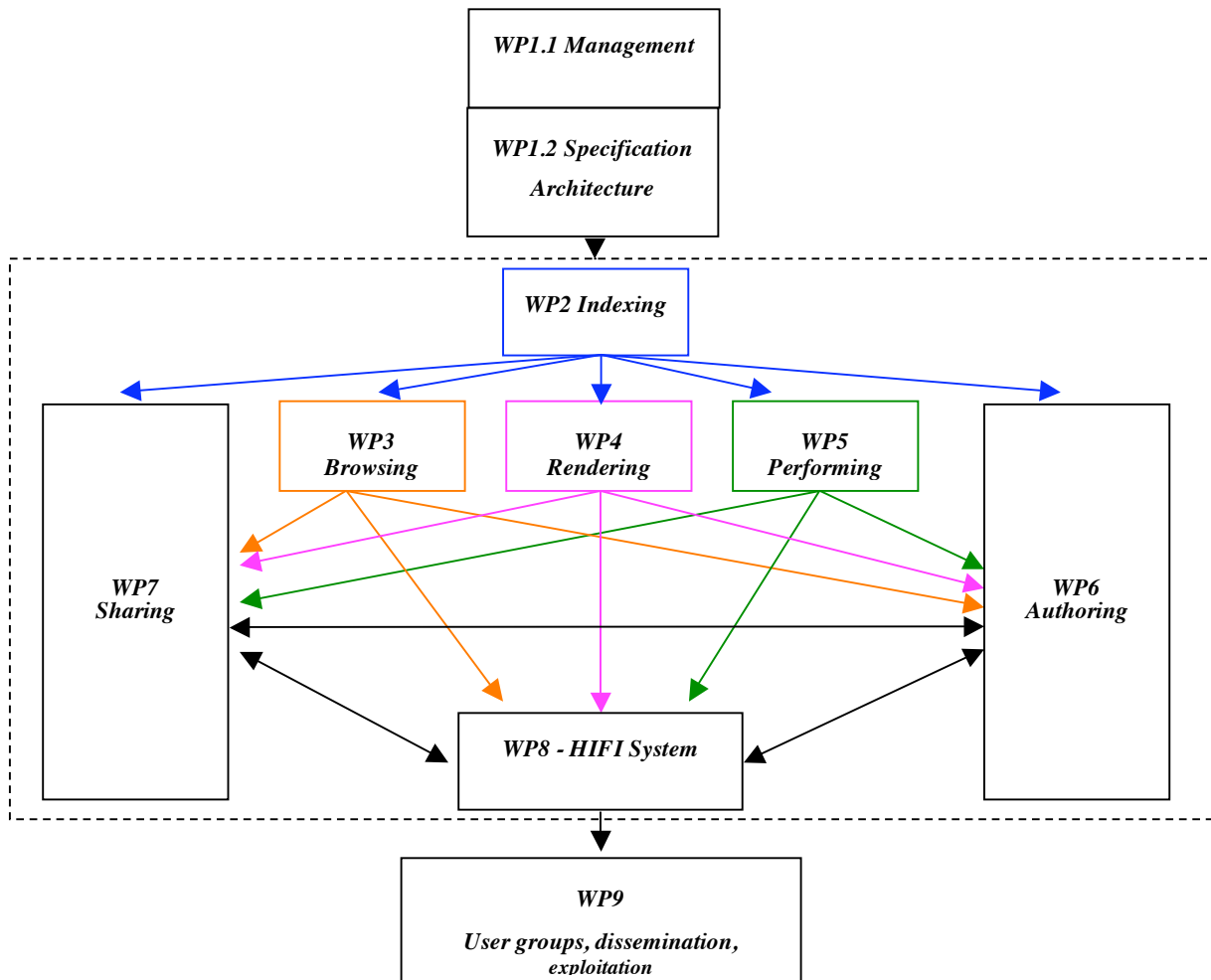
authoring applications), and messages between peers (chat functionality). The sharing of musical data (audio files) is not allowed by the P2P sharing system.

In order to meet the demand of the project to prevent the sharing of copyright protected data, the following mechanisms are part of the P2P sharing system:

- filtering of files to be shared (only a defined set of file types should be sharable at all),
- a secure W3C XML-Signature based license mechanism for storing information about the sharing process (e.g. peer name, group in which a file is allowed to be shared, a link to the shared file, the digital signature of the sharing peer),
- support for secure user groups (access to particular files can be ensured by group membership).

1.3 Project Organization

The project is organized according to the following synopsis.



In order to achieve its goals, the Semantic HIFI workplan has been organized into 9 main workpackages.

WP1 is dedicated to **management**, **specifications** involving users throughout the project, and **technical coordination**. It is divided into WP1.1 (technical and administrative management) and WP1.2 (specifications, coordination of the technical architecture). Leader : IRCAM.

WP2 (**Indexing**) aims at performing basic research on the core modules both at the content extraction level, classification and indexing methods adapted to personal use on top of which the various functional workpackages (WP3 to 7) are based. The objective is to conduct basic research on algorithms for signal extraction, automatic indexing, segmentation, audio summary and source separation. Leader: IRCAM

WP3 (Browsing) is dedicated to **track management** and **browsing** features. The objective is to design and implement an integrated browsing tool allowing all browsing scenarios in a home-based context where navigation time must be reduced and graphical interfaces must be simple. Leader: Sony CSL.

WP4 (Rendering) is dedicated to **audio** and **multimedia rendering**. The objective is to bring simple and intuitive control interfaces for in-track interactive navigation. Leader : IRCAM.

WP5 (**Performing**) addresses performance and real-time processing. Performing music while listening with simple “instruments”, home devices, and voice as an addition to the played music in an edutainment context (instrumental performance learning) or in a creative context. Real-time transformations of the music by using different input controls. Real-time interaction and performance with the music collection using Voice or “conducting instruments”: tempo variations by time stretching in real time using existing control devices (TV command, microphone, Game joystick). Leader: University Pompeu Fabra.

The next workpackages are dedicated to the target technical components:

WP6 (Authoring) aims to the development of a PC based **Authoring tool** for audio and metadata material editing and composition (mixing), in particular for the preparation of materials to be played, browsed, performed and shared on the Hi-fi System. Leader: Native Instruments.

WP7 (Sharing) is dedicated to the development of the **peer-to-peer sharing middleware** architecture. The objective is to develop functionality to share (peer-to-peer) user-generated metadata over the Internet, through a system enabling to prevent the exchange of copyrighted data and including automatically computed references to the used copyrighted materials. Leader: Fraunhofer IDMT.

WP8 (Hi-fi System) aims at developing the **Hi-fi System** application (personal intelligent listening system), user tests and documentation. Leader: Sony NCSE/ EuTEC

WP9 is dedicated to **dissemination** and **management of user feedback**. Research & Development, demonstration and management phases are running in parallel thanks to a user-feedback loop methodology. Leader: IRCAM.

1.4 Consortium Structure and Partners' Roles

The consortium gathers the following partners:

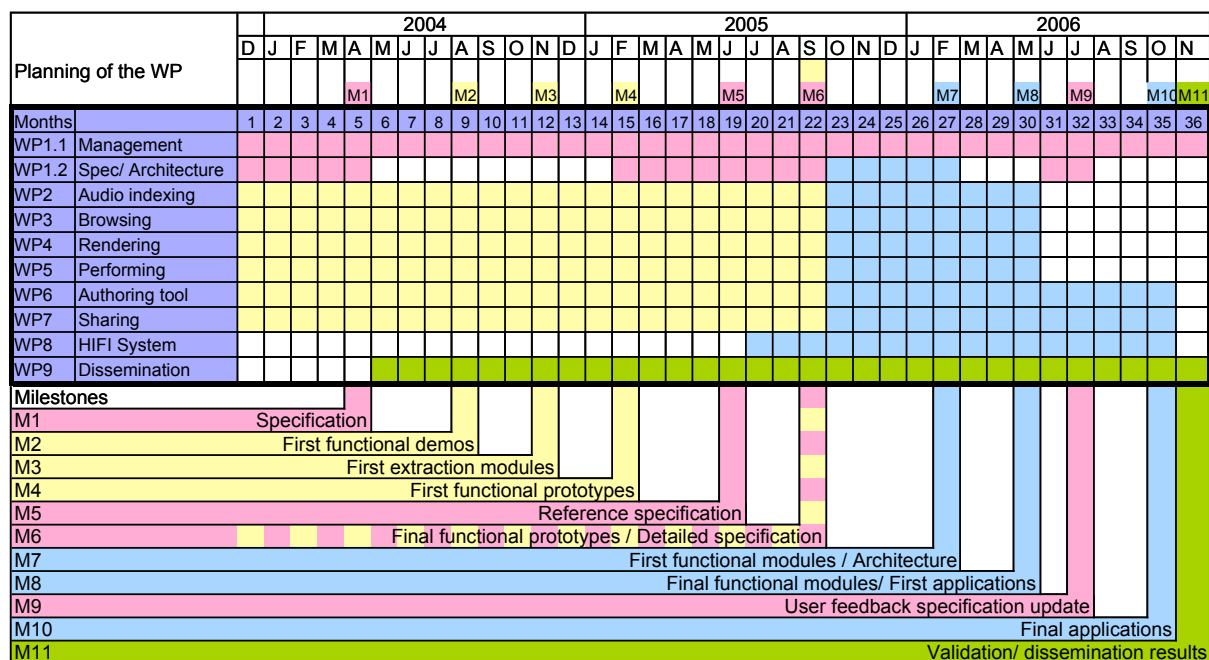
1. Institut de Recherche et Coordination Acoustique/ Musique (IRCAM , F)
2. Fraunhofer Gesellschaft, Institute for Digital Media Technology/Institute for Integrated Circuits (Fhg, D)
3. Sony France (Sony-CSL, F)
4. Fundacio Universitat Pompeu Fabra (UPF, E)
5. Ben Gurion University (BGU, IS)
6. Native Instruments (NI, D)
7. Sony NSC Europe (Sony-NCSE, B)
8. Sony Deutschland (Sony-EuTEC, D)

Participant	Country	Role	Function	Note
IRCAM	F	Coordinator Research Lab Music Production Music Archive	Coordination Research Partner WP1.1, 1.2, 2,4,9 Leader	Project management - Research in audio indexing and processing, sound spatialization, man-machine interfaces. Technical coordination. Management of user specification and feedback. Coordination of dissemination.
Fraunhofer- Gesellschaft	D	Research Lab	Research Partner WP7 Leader	R&D related to Audio ID and Query by humming. Coordination and development of Sharing.
Sony CSL	F	Research Lab	Research Partner WP3 leader	R&D dedicated to Browsing and Performing.
UPF	ES	Research Lab	Research Partner WP5 leader	R&D dedicated to Performing.
BGU	IS	Research Lab	Research Partner	R&D dedicated to source separation and score/audio and lyrics/audio synchronization.
NI	D	Music production software publisher	Application integrator and developer WP6 leader	Coordination and development of the Authoring application.
Sony NSC Europe	B	Engineering Consumer electronics manufacturer	Application integrator and developer WP8 leader (until September 2004)	Coordination and development of the Hifi- System.
Sony EuTEC	D	Engineering Consumer electronics manufacturer	Application integrator and developer WP8 leader (from February 2005)	Coordination and development of the Hifi- System.

2 Project Implementation

2.1 Workplan Synthesis

One of the main project challenges was to combine a high degree of innovation—beyond the existing state-of-the-art in several scientific and technology areas—with the demonstration of their validity through their integration in full-featured, unified, and consistent application prototypes compliant with targeted users' expectations. This objective had been foreseen in the design of the global project task schedule presented hereinafter.



The specification, R&D, and application development activities are displayed respectively in pink, yellow, and blue. After the initial specification phase (milestone M1), which consisted in specifying the project Technical Annex in particular concerning the R&D activities and the application target features, the project included two main phases. The first phase (up to milestone M6) was dedicated to the parallel implementation of various research activities (WP2 to 7, in yellow) and aimed at developing the target application features. The second phase started from the Reference et Detailed specifications (milestones M5&M6) and was dedicated to the applications development through the integration of the various features resulting from the various R&D tasks. This schedule included, through several intermediary milestones, preliminary demonstrations and prototypes deliveries, as well as the implication of user through requirement and feedback processes at various stages.

2.2 Actual Workplan Implementation Stages

On the basis of this global schedule, which was respected, despite unexpected events (such as the resignation of Sony NCSE from the project in August 2004 and its replacement with Sony EuTEC from February 2005), the following successive stages have been implemented:

1. Start of all R&D activities (WP2 to 7) and formalization, by each technology provider, of proposals of **R&D objectives** and **refined workplan**;
2. Identification of **use cases** (individual features) and **user scenarios** (sequences of use cases) related to each of the developed features; these use cases and scenarios, formalized in a uniform way for all targeted features, resulted from a bottom-up process involving all project technology providers;
3. First stage of **user requirement surveys** performed on populations of potential users, based on the use cases and on prototypes of the target functionalities; these use cases were performed by the application integrators through their registered user basis;
4. Release of the **Initial specification** (May 2004), specifying the **workplan for all WPs** and the **targeted application features** resulting from the initial proposals of stage 1 and the user requirements of stage 3;
5. Adaptation and **focusing of the ongoing R&D tasks** to the results of this initial specification. Intermediate prototype deliveries through Milestones M2, M3, M4.
6. Finalization of a **list of potential target functions**, from the actual/ potential results of all R&D tasks;
7. **Second stage of requirements** performed with potential users and concerned business units by the industrial partners on the basis of the **target function list**;
8. **Selection of functions** for each of the two applications as the result of stage 7;
9. Functional specifications of the target applications;
10. Design of the technical architecture necessary for implementing these functions through a modular system enabling all technology providers to contribute using well-defined API's. Formalization in two close stages through the **Reference** (June 2005) and **Detailed** (September 2005) **Specifications**;
11. **Finalization of prototypes demonstrating individual functions** ending the research phase (September 2005) and start of a **second phase in the R&D tasks** consisting in adapting, optimizing and implementing the various functions as modules compliant to the target applications architecture;
12. Setup of the technical architecture in the applications, start of integration of the functional modules, and **release of a first version of the prototype applications** (May 2006);
13. **User feedback experiments** based on these first versions of the applications (June 2006). These experiments are presented more in detail in the next section;
14. **Updated functional and technical specification** according to the results of the user feedbacks (August);
15. **New versions of the applications** fixing problems and integrating proposals arising from the user feedback experiments (September 2006);
16. **Identification of all project results** through a common template and **formalization**, by all owning partners, of **their exploitation plans** (as partly described in section 3 of this document);
17. **Final and documented versions of the applications**, validating the project concepts for integration into future products (October 2006).

¹ Application programming interfaces

2.3 User Feedback Methodologies

This section describes the main results of the large user feedback experiments that took place in summer 2006, to feed up the final specification update, and a synthesis of their main conclusions. Those experiments, including video reporting, methodology and questionnaire analysis, have been precisely described within the corresponding deliverable.

One full week was dedicated to the Hi-fi System experiments, from Tuesday June 20, 2006 to Sunday June 25, 2006, in cooperation with the Carrefour Numérique at the Cité des Sciences et de l'Industrie in Paris, taking advantage of the national Music festival. Monday June 19, 2006 was dedicated to the installation of our equipment (the Cité is closed on Mondays).

The Cité des Sciences et de l'Industrie (<http://www.cite-sciences.fr/>) is one a the largest European sites that welcomes the general public where they can discover science and technology.

The experiment sessions involved only French users and were held in French.

A conference was planned on the Tuesday evening to welcome the general public and to demonstrate the Hi-fi System prototype.

A schedule was drawn up and included six different experiments, each test was held twice, by two different sets of users.

Each of the experiments was designed to last two hours. To allow a logical progression throughout the experiment, they were organized by days or by sets of days as follows:

- First set of experiments: (Exp.1 + Exp. 2 = Day 1) + (Exp.3 + Exp. 4 = Day 2)
- Second set of experiment: Exp.5 + Exp. 6 = Day 1

Users were “hired” with the helop of the Cité. Visitors were told about of the testing project.

Around twenty users accepted the challenge and sent an e-mail to sign up for the tests. They were asked to contact the SemanticHiFi test session organizer for detailed explanations and to give us the names of their favorite song titles to work on.

Hi-fi System developers were also asked to be there to raise and/or solve technical questions, to tune and/or parameter their modules and to see what happened.

To welcome the group of users and the developers the SemanticHIFI test session organizer asked for the help of some students from an engineering school. These students also participated in the design of test questionnaires.

They were asked:

- to welcome the users, to explain the protocols, to supervise the filling in of the user questionnaires;
- to fill in the test-leader questionnaire, specially designed to get feedback from those so-called test-leaders;
- to call on the reporter to intervene every time a potentially interesting sequence was discovered.

The reporter, equipped with a professional camera and sound recording device, was asked:

- to film video of any potentially interesting sequences;
- to produce as many video sequences as discovered user feedback sequences (selected with the help of the SemanticHIFI test session organizer);

- to produce a movie presenting all video sequences;
- to produce an hyper-document presenting all video sequences;
- to produce a small movie presenting the test sessions in general.

We were successful in welcoming twelve different users and getting back ten useful user feedback questionnaires. Four test-leader questionnaires were also filled in successfully.

The reporter filmed more than five hours of video rushes, and put together several videos. A video providing an overview of these reactions is available online on the project Web site at: http://shf.IRCAM.fr/user_feedback.html.

Analyzing the different questionnaires, we first designed a categorization of interesting user feedback results that we presented as a table. This task was particularly difficult to carry out properly and as objectively as possible, due to a technical bias: How to classify/consider a bad evaluation of relevant to a given feature, if the technical that feature is not technically mature enough to properly present its intended function?

We then did the same thing with the video sequences. This allowed us to complete our table by cross-reference. Then, we asked the reporter to produce the most important movie, in French with English subtitles, as a more traditional/global way to illustrate the results.

The purpose of that section was to show the main results of those experiments in terms of enhancement of the project and re-specification toward a better fit to user needs and desires.

3.3.1 Hi-fi System

The user tests that were carried out with the Hi-fi System basically resulted in two main kinds of thoughts:

- The first is that many of the advanced music selection and navigation concepts are very interesting for users. In particular, the concepts of personal classification, the in-track navigation, the possibility of setting constraints for any metadata field, as well as the ability to navigate through content by using successive filters was appreciated by the test users. This met with our expectations regarding these features. Users even made suggestions for a better use of some of these advanced features (like visualization of song structure or user defined classification), which were not foreseen by the partners.

- The second result from the user tests is the fact that the GUI design of the system with so many features is a real challenge from a usability point of view, particularly when the main device for handling the most part of features is a handheld device such as a PDA, which imposes a number of constraints in terms of readability, the size of the screen, and in manipulation (particularly when it come to use the fingers in that manipulation instead of a pen). While some features were accepted quite well, many usability difficulties and suggestions were reported.

Proposed improvements of the first version of Hi-fi System

Usability Issues

A lot of usability related requests are based on the experience users have with a PC where a rich user interface may be implemented. It is generally agreed that most of the tools we know from the PC such as scrollbars, cascading menus, drag and drop, and double-clicking do not work efficiently on a PDA from a usability point of view. At the same while, considering the use of Game Boys (for example the latest Nintendo DS product), it should be noticed that these features are used generally in the main user interface as well as in games.

Usability is in general a difficult issue and especially challenging on small devices like PDA's. From the current experience in the SemanticHIFI project we identified a high demand for Hi-Fi related usability issues. The whole topic is extremely challenging and we could only address the issues related to usability in part within the project. This included the ergonomics of the navigation within the interface, the management of multiple selections, the reduction of click numbers for a given function, the graphical display of the various descriptors columns, the addition of a new search functionality for artists, tracks and genre, etc.

User Defined Genres

In the process of creating their own genres, users frequently face the difficulty of assigning songs to a category that they recognize as not correctly assigned by the system when they don't know exactly which category to assign to them to at the time. In the meantime, these songs seem to introduce instability into the algorithm. The user is faced with a problem s/he doesn't know how to solve, and there is a risk of resignation.

In order to try to solve this problem, it has been imagined to have a category "other" (or more precisely "none") which will not be taken into account by the system, and which will not introduce instability in the solving of the algorithm.

Important Issues for Future Work

From the test sessions we got some very interesting hints on future extensions that could not be realized within the course of the project due to limited time and resources, but may feed further developments.

Remixing Starting From the Song Structure

The remixing of a song structure is one of the features that is really attractive for users. Starting from the song structure, the user has the possibility of changing the order of items, removing, or replacing items with a very simple and intuitive user interface (almost completely developed). Such mixes could be shared in the same way as is already possible with playlists and mix files from an authoring system.

Enhancing the Song Structure

Other song structure related features people would like to see are a timeline with tempo notation, as well as different visualizations of the blocks of song segmentation such as different colors or a spectral representation of the segments.

Moreover, it could be interesting for users to have the ability to qualify the song structures (for example, by naming structure items in order to identify them, such as "refrain – couplet").

Sharing Enhanced Song Structures

It could be interesting to be able to exchange the previously defined qualified structures by using the sharing system, and/or to acquire these qualified structures from the producer (e.g. author) or a qualified person (a musicologist).

Creation of Sub-Categories in User-Defined Genres

Finally, the possibility of creating sub-categories within the user-defined genres might be an attractive feature for future products.

3.3.2 Sharing

General Description

The sharing system has not been totally tested, since its integration was not fully available at the time of the user tests. But the aims and goals of the sharing system have been explained to users and users have made suggestions about the use of that system.

The following suggestions have been made:

- Exchanges of enhanced song structures by the mean of the sharing system, containing qualification of segments (names of segments, like “Refrain”, Couplet”, “Main Theme”. Qualified people should decide on these qualifications (authors themselves, or musicologists, etc.). The implementation of this feature is strongly related to the availability of an authoring feature, letting users qualify the segments of that structure themselves.
- Sharing of time-consuming descriptions. Users suggested this feature in order to cope with the problem of availability of songs during the computation of descriptions (like Query by Humming, song structure). Nevertheless, this issue has already been addressed in the paragraph related to the SHF platform, and the decision has been taken to postpone the development.

It should be noticed that none of these enhancements are foreseen for implementation in the remaining time schedule of the project.

Features of Interest to be Developed

Sharing Editorial Metadata

According to the logic feedback of feature 5, it would be relevant to share track internal structure qualifying metadata on the P2P. The implementation of this feature needs the addition of the following elements in the system:

- A tool for authoring enhanced structures – this tool can be kept simple in the beginning and give the user the ability to give a name to each element of the structure.
- A new item in the shared metadata model, “SongStructure”
- A function for searching qualified structures on the sharing system. This tool can be very simple and can be launched automatically when an item is inserted in the system, or it can be triggered by a simple time in the User Interface.

The definition of the structure itself (as an XML file) should be enhanced in order to accept qualifiers.

Sharing of Time-Consuming Descriptors

In order to be able to exchange time-consuming descriptors, it could be sufficient to add this kind of descriptors to the MusicTrack data model. These descriptors could be exchanged in the same way as simple descriptors (e.g. key, bpm, and so on). There is no need for an additional tool, the current system will automatically download these descriptors.

Implementation of that feature will remain extremely simple.

3.3.3 Authoring Application

The user tests for the Authoring Application including the sharing functionalities were conducted from Monday June 19th 2006 to Wednesday 28th June 2006 at the Sound Design Studio at Native Instruments in Berlin.

Approach and Conduction

- 10 participants that represent real users in experience, age, etc.
- Users from three main target groups: beginners, advanced users, experts
- Participants were asked to perform actual use cases
- Observing and recording of what participants do and say by 2 observants
- Recorded questionnaire and discussion afterwards

Technical Setup

- 2 personal computers (client & server)
- Authoring Application and Sharing System installed
- Audiodata (incl. audio from participants)
- Metadata already prepared (playlists & musictrack data)
- Convenient and homelike studio environment

The user tests performed basically showed two things.

The first is that the central technological advances raised high interest by the users. In particular the test users appreciated the concepts of non-real time mix editing, advanced real-time effects, and sharing of meta-data. This is in accordance with our expectations resulting from earlier user requirement studies (D6.1) regarding these features.

The second result from the user tests is that the GUI design of a system with so many features and such a broad target audience from beginner to professional is a real challenge from a usability point of view. Nevertheless, a lot of the implemented features were understood and accepted very well. Still, quite a few usability issues and suggestions were reported; a lot of which we already were aware of due to our regular user feedback mechanisms (forums, beta-tests, etc.) Due to limited resources the tested prototype was not optimized in terms of usability in all areas, especially sharing. Non real-time mix editing was not available at all. For that reason, bugs reported and features requested were often already foreseen for implementation in the next software iteration to be released after the test session.

Proposed improvements of the current Authoring Application system

Usability

As mentioned the Authoring Application has potential for increased ease of use. This ranges from the used fonts, GUI elements and arrangements to flow of the user actions.

Rating mechanism for playlist and mix data

We propose to add one or more additional fields by which the users can search and sort search results of shared playlists and mix files. It would create an added value for the users to be able to search for 'the best' mix containing a certain file. There are several options to realize such

a rating, e.g., by collective user ratings, by download statistics, etc. Each of those methods would require significant additional resources and infrastructure.

Important issues for the future work

Users would like to remix a song based on the individual instruments. To realize this feature one of the existing formats for this kind of application would need to be adapted for distribution of audio material or a new distribution format would need to be established. Current state of the art does not seem to allow for such kind of feature based on standard stereo material..

3 Project Results and Achievements

3.1 Objectives and Methodology

The goal of this section is to characterize, identify, and formalize the main project results in their various forms. The term “results” refers here to elements of knowledge or technology produced as part of the project, and that are potentially subject to further exploitation in various contexts (industrial, educational, cultural, etc.). It does not include their dissemination in various forms, such as articles, conferences, workshops, fairs, patents, PhD theses, etc. This latter aspect is handled in section 4 of this document.

The implemented approach has been to start from the project’s structural organization that isolates each partner’s activity in each workpackage (WP) as part of a sub-workpackage (sub-WP). The result is that each sub-WP corresponds to a well-defined functional feature and provides an appropriate level of granularity for identifying the results and their owner(s). Then, each providing partner filled in a grid for each identified result, following a common template, in order to specify its various characteristics (features, implementation) and criteria leading to exploitation plans. An example of such completed grid is provided hereinafter.

RESULT N° IRCAM-1: Music Structure extraction & graphical representation	
A	DESCRIPTION OF THE RESULT
Features & obtained functionalities	Automatic music structure extraction allows graphical or buttons-based intra-document navigation, allows music remixing based on content, keyphrase detection
Innovative aspects	Algorithm: hierarchical approach, likelihood sequence based approach. It segments a music piece in homogeneous segments.
B	POTENTIAL MARKET
Competition	Audio segmentation & Music information retrieval community
Business applications	New functionalities for all systems allowing navigation inside a music database (hardware or software or web-based) Consumer electronics(Hifi systems), online music stores and P2P networks, degrading products
Key prospects	Not disclosed in current document
C	INTELLECTUAL PROPERTY ISSUES
Owners	Ircam
Dissemination status	French Patent AH/EMA-FR 03 07667
D	BUSINESS MODEL
How	C/C++ module for third party product integration
Who	Direct marketing by owner or OEM integration in third party
Licensing	
Additional requirements	programming skills for integration
Regulations	N/A apart from music copyright issues (indirect impact)
Standards	This algorithm is new but would gain being pushed in de facto standards like CDDDB metadata, or public standards like ID3 tags for MP3, or public standards like MPEG 7, MPEG 21, etc., MIDI & OSC

3.2 Typology of Results

Due to the numerous results obtained (43), another stage of structuring them appeared necessary in order to facilitate their understanding and potential exploitation. Therefore, a division of the results into a limited number of different types has been conceived. This typology is based on two main kinds of criteria, which characterize the way they can be used for exploitation: classes of functionalities, and implementation forms. Six different types of results have been identified according to these combined functionality/implementation criteria, from bottom-up :

- *Indexing modules* are software modules taking in input a sound file and producing in output a metadata structure accounting for a specific musical feature or description. These modules include no user interface and are targeted to be included in any software environment featuring automatic audio and/or music analysis functionalities.
- *Advanced audio players* propose novel ways of browsing within a music track while listening to it. They are generally made of a module taking in input the sound file (or stream) and performing the audio playback and a user interface, that controls the playback and/or displays relevant information synchronous to its execution.
- *Audio synthesis and processing modules* are software modules producing audio files (or streams), either from specific input controls (synthesis), or as the processing of an input audio file (or stream). Most of the ones produced as part of the project are developed in their final versions in the form of VST plug-ins (industry standard by Steinberg²) and include a graphical interface for the various controls. Most audio sequencer applications, including the project Authoring Application, are compliant with this standard so that the specific functionalities included in the plug-ins complete their basic features.
- *User interface prototypes* provide new modes of interaction with the musical material. They generally illustrate a new man-machine interface concept through an experimental implementation.
- *Software frameworks* provide a set of specific, consistent and sometimes complex functionalities, through the interoperability of several software components through a dedicated software architecture.
- *Target applications* integrate a number of functionalities in a consistent environment, both from the technical architecture and user experience viewpoints. The project target applications are the Hi-fi System and the Authoring Application.

² <http://www.steinberg.net>

The table below illustrates the number of results of each type produced by each project partner.

	Ircam	FhG	Sony CSL	BGU	UPF	NI	Sony EuTEC	TOTAL
Indexing Modules	4	1	2	1		1		9
Advanced Audio Players	2			1	1	2		6
Synthesis/ Processing Modules	1		1		4	2		8
User Interface Prototypes	5	1	2					8
Software Frameworks	1	1	5			3		10
Target Applications						1	1	2
TOTAL	13	3	10	2	5	9	1	43

The next sections provide a summarized description of the various results for each result type. The results are numbered first by the producing and owning partner identifier, then by a unique result number by partner.

3.3 Indexing Modules (9)

RESULT N° IRCAM-1: Music Structure extraction

Automatic music structure (segmentation) discovery aims at providing insights into the temporal organization of a music track by analyzing its acoustical content in terms of repetitions. It then represents a music track as either a set of states or as a set of sequences.

- 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.
- A *sequence* is defined as a set of successive times which is similar to another set of successive times but the times inside a set are not necessarily identical to each other. It is therefore a specific case of a state. Example of this are the various melodies repeated in a music track.

Once extracted, the structure can be used for:

- intra-document browsing
- automatic audio summary generation

Both functionalities target the Hi-fi System (player of the system and preview of the database content) and the Authoring tool (giving inside in the Traktor DJ files being mixed).

Intra-document browsing displays the automatically extracted structure in a piano-roll way in order to let the user browse inside a track. The user is able to access immediately (by clicking on a map) the various parts of the track, perform forward/backward by segments, perform forward/backward by verse or by chorus or by melodies (in the case of the sequence representation).

The feature extraction front-end has been extended to represent new kind of features. In the last version, it extracts timbre-related features (MFCC), tonality-related features (chroma/PCP) or both type of features. The extension to tonality has been done to allow the processing of non-timbre based music (classical, jazz music). The types of temporal modeling being considered have also been extended: mean, variance and dynamic features modeling. The modeling length can be modified in order to allow the processing of very long files.

RESULT N° IRCAM-2: Music audio summary generator

Audio summary uses the extracted structure in order to create a short audio extract (usually less than 30s) which collects the various contents or the key-phrase of a music track. It uses a beat-synchronous concatenation of the various parts of a music track according to the parts estimated during the structure process. The beat-synchronous algorithm allows keeping a musical feeling to the summary.

The module developed in SHF performs the three following tasks:

- 1) estimate the structure of a music track,
- 2) provide a visual map of a music track,
- 3) provide an audio summary of a music track.

In SHF, the audio summary generation has been extended to include the possibility to generate it using various methods:

- Output of the state approach; option: chose the number of states to be used in the summary
- Output of the sequence approach: option: chose the number of sequences to be used in the summary
- Only the main sequence of the piece of music

Various improvements have been done in order to increase the quality of the audio summary.

The main improvement consists in creating the audio summary using directly the input soundfile (previous version used the internal resampled/mono-converted audio signal): the audio summary can now be stereo and 44.1KHz.

The estimation of the beat markers (for beat-synchronous summary generation) is now performed directly inside the module (WP rhythm description).

Finally a new algorithm for constraining the length of the summary to a specific duration has been written.

RESULT N° IRCAM-3: Tempo, beat, rhythm feature generator

The module for rhythm description extracts high-level features related to rhythm characteristics directly from audio signal analysis.

It provides three groups of features:

1. Instantaneous features (time-varying tempo and beat marker locations). This features are to be used for later audio processing (Authoring tool and Ircam X-Micks processing);
2. Global features: global tempo, global meter, percussivity index, periodicity index. This features are to be used for browsing in a database (Hi-fi System);
3. A rhythmical pattern to be used for search by similar rhythm in a database (Hi-fi System).

It should be noted that another module developed by Native Instrument (tempo and phase extraction) targets specifically percussive based music (dance and electronic music). The present module targets the general class of music including non-percussive based music (jazz, classical, variety music). In this case important factors are the detection of onsets, potential quick variation of tempo and potential non binary meters.

RESULT N° IRCAM-4: Musical key generator

The module for musical key description extracts high level features related to musical key characteristics directly from audio signal analysis.

It provides a set of global features to be used for browsing or search by similarity in a database (Hi-fi System). The following features are extracted and store in an xml file:

- Global key-note (C,C#,D,...) and mode (Major, minor) of a music track
- Global tuning of a music track
- An harmonic pattern to be used for search by similar rhythm in a database (Hi-fi System).

These features target specifically music for which tonal information plays an important role (classical music).

Musical key can be estimated from the knowledge of the score of the music. However because a transcription is most of the time not available and because automatic transcription algorithms (multi-pitch detection) are still limited to small polyphony and costly (difficult to use for a real task of large database management), the module we have developed is based on a lighter technology: the Chroma or Pitch Class Profile approach [PeetersDAFX2006, PeetersISMIR2006]. While this technology does not allow the extraction of the various notes, it allows an efficient estimation of the global key-note, mode and tuning of the track. We have also started at the end of the project the estimation of the chord succession over time of a music track [Papadopoulos2006]. We add to these features the extraction of an harmonic pattern which can be used for search by similarity.

RESULT N° FhG-1: AudioID

AudioID is a system that performs an automated identification of audio signals. It recognizes real-world audio signals by comparing their fingerprint with a fingerprint database of sounds that have been previously extracted. This functionality can be considered as the algorithmic equivalent of human recognition of a song from the memory of the recognizing person. A fingerprint contains the “essence” of an audio item and its extraction algorithm has been standardized within the MPEG-7 standard. To identify music, a fingerprint must be extracted from the audio item and later compared with the fingerprint database. On a positive comparison result, the gathered metadata are replied and further processed. The following picture describes the necessary steps:

AudioID consists of two different parts. The first part has been finalized and standardized in 2001. The extraction of the fingerprint from the first version is done in the following manner:

1. Windowing of the audio signal into different parts, each with a window length of 30ms
2. Weightening of each part using a hamming window
3. Performing an FFT on each part
4. Spacing of the spectrum logarithmically into subparts
5. Calculate geometric mean/ arithmetic mean with each bin of each subpart (called Spectrum Flatness)
6. The resulting vector of each part has to be stored within a matrix

During the SemanticHIFI project, Fraunhofer IDMT established a second version of the fingerprint format, since the first version was not robust enough to classify extremely distorted audio signals as for e.g. GSM-coded audio signals. The second version has been standardized in an amendment of the MPEG-7 audio standard. The extraction algorithm is nearly the same as from the first version. The only difference is the processing of the logarithmically spaced subparts (5.). Within each subpart the bin values have to be squared and thereafter added.

The resulting matrixes of both versions are scalable. The raw values can be compressed by building the mean of a number of successive vectors (up to 32). If the fingerprints are scaled, its size decreases. Its duration for classification decreases too. The disadvantage is the decreasing classification performance.

During the identifying process, an audio example needs to be extracted. The classification process compares the resulting fingerprint with a database of fingerprints using a nearest neighbour search algorithm. The result list is thereafter sorted according to its distance to the query fingerprint. Based on the distance vector a likelihood value can be calculated to estimate a classification confidence.

RESULT N° CSL-4: Timbre Similarity

Finding songs similar to a given song *X* means looking for the nearest neighbours of *X* in the context of some similarity distance. It is the case for Query-by-Humming, where the similarity distance is between melodies. For the project, *timbre similarity* was also considered a very important feature. With timbre similarity, the user can ask for songs which "sound like" some song he/she likes, not in terms of melody but in terms of more global spectral information.

Timbre similarity requires two main components: an algorithm to compute the timbre of a song, and an algorithm to compute the distance between the timbre of any two songs. WP3 provides both components, as well as means to supervise the computation and storage of timbre and timbre distances as a whole.

At the application level timbre similarity is available in the "Similarity" panel of the Music Browser. In this panel, the user can select a song (the *seed*) and ask for similar songs in terms of timbre (other similarities than timbre can also be defined but we won't enter the details). In general the distances have already been precomputed and the query is just about displaying the songs in increasing distance from the seed, which is very fast. The user can then immediately listen to these similar songs. When the distances haven't been computed, which is the case for songs freshly added to the database, the "Similarity" panel offers to compute and store the missing information.

At the programming level, the computation of timbre is provided as a computable MCM field. It is very easy to handle. Computation of timbre distances is a little less straightforward because it requires in general to consider the distances between all pairs of songs. Anyway, MCM provides the notion of distance relation which encapsulates all the necessary methods. Most of the time these methods have been highly optimized.

RESULT N° CSL-5: Playlist Generation

An important concept when browsing music databases is the creation of a *playlist*, i.e. an ordered list of songs that can be listened to in a sequence. A simple approach to playlist generation is to store the results of individual queries over the database, i.e. build playlists with "rock" songs, or "songs which sound like x ", etc. Playlists can also be built automatically, in a background process, as the user listens to individual songs. For instance, each song played by the user is automatically stored in a playlist labeled by user and date, e.g. "John's music on Tuesday".

In WP3, playlists can also be constructed automatically based on various constraints. For instance, a playlist can be generated with 10 songs starting with a given song (say, The Beatles – Yesterday) and ending with another (say, Beethoven's 9th Symphony) and in which the timbres of the songs are as continuous as possible from one song to the next. Such a playlist could e.g. select songs such as a guitar concerto piece from the classical repertoire, which shares the acoustic guitar timbre of the Beatles' song and the orchestral textures of the Beethoven piece.

With whatever means it is generated a playlist can always be edited, copied, renamed, saved, etc. Several useful operations are provided, such as "shuffle" or "remove duplicates".

RESULT N° BGU-1: Source separation

This result refers to an original method by BGU, initiated as part of the project, aimed at the automatic separation of specific components in the audio signals.

The objective of demixing modules is to enrich existing audio content by trying to reconstitute the original separated sound sources. This audio content can then be remixed at the rendering stage and, to some extent, adapted to the listener's preferences using applications such as IRCAM's audio spatialization technology.

This advanced audio processing tool performs the separation of parts or instruments from within a multi-track audio recording. The main goal of this tool is to give the listener the ability to manipulate audio in ways not available previously and to enable artistic liberty normally available only in the studio.

Originally, virtual mixing was dependent upon alignment to the existing score of the recording available as a MIDI file. This enabled the direct harmonic separation of the instruments based on the information appearing in the score. A more advanced approach has been taken which renders the use of score alignment optional. Even so, the score alignment can be used for melody or instrument identification.

Virtual Mixing presents a high-level challenge, requiring the use of statistical sound processing techniques. These techniques employ a model-based pitch and voicing (or harmonic) tracking algorithm. The problem of multi-track recording separation has been approached from many directions since the turn of the century but has not yet been applied in a working user-based system such as Semantic Hi-Fi.

Three new approaches have been developed for the purpose of musical source separation and decomposition. The first is top-down analysis, in which analysis priority is given to higher notes. The second is temporal alignment of the analysis based on previous knowledge of the score. The third approach is harmonic sharing. A large number of notes in a composition share harmonics. This fact is taken into consideration. Priority is given to lower harmonics, i.e. the fundamental, and decreases towards the overtones of the note.

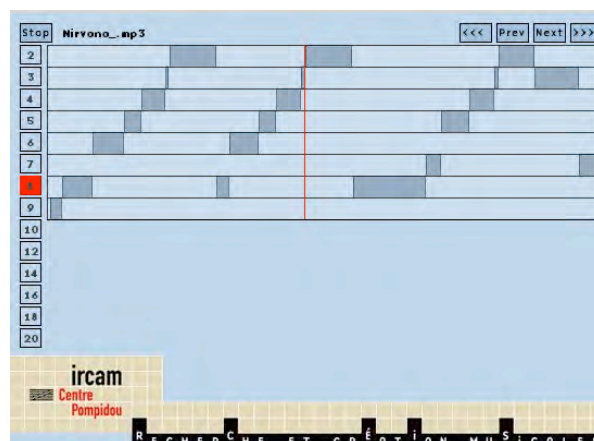
RESULT N° NI-1: Beat & Tempo Detection

Starting from an existing module used in the Traktor application, Native Instruments developed an improved transient detection and improved bpm calculation algorithm adapted to Dance music.

3.4 Advanced Audio Players (6)

RESULT N° IRCAM-13: Interface for navigation within the track temporal structure

This interface displays the temporal structure of a track as computed by the corresponding indexing module (RESULT N° IRCAM-1). The user can select the number of states that best fits the track structure. Clicking on a segment launches the playback at its starting time. This interface is integrated in the Hi-fi application prototype.



Graphical interface for the temporal navigation within the track

RESULT N° UPF-5: Player for the Hi-fi System

The SHF (Semantic HIFI) player component is responsible for reading sound files, decoding them and playing them back. Its functionality can be controlled over a UDP connection, making it possible to control it through custom controller components. The player's functionality also includes sound processing tasks, mainly via plug-ins, but also internally. Functionality like shuffle, repeat and playlist handling are built into the player.

The Player component is controllable by Open Sound Control messages (OSC).

OSC³ is a protocol specifically designed for the communication between synthesizers and computers and due to its flexible naming scheme and parameter types ideal for the envisioned task.

³ <http://opensoundcontrol.org/>

OSC communicates via messages, which consist of an address and arguments. The address field can be structured hierarchically in order to access several components with the same features. In the following specification the first word in italic and starting with a “/” will be the address of the command. The arguments will be either of type “float”, “boolean” or “string”, which will be indicated by the first letter of the argument “b” for boolean and “f” for float, strings will be encapsulated in angle brackets.

RESULT N° BGU-2: Browsing by Lyrics, BGU

The main idea of the browsing by lyrics (BBL) is to move back and forth inside the audio file by moving back and forth in the song lyrics. Instead of using time of song, we are using the song’s lyrics in order to skip Forward and Backward. During this period, we implement this idea to java code, also we provide a new idea to align the lyrics easily. A new GUI has been developed in order to combine all the new functions due to the conclusions from the last meeting in Paris.

Song lyrics are aligned by rows. This appears as the best resolution for optional users.

The basic commands of this tool are *play*, *pause*, *stop*, *skip forward/backward*, *FF*, *RW*. The *play* button is starting a clock that counts the cycles of 250 milliseconds. The clock has the time that the music playing so it can easily show or erase the lyrics in time. Pressing *play* while a row is shown on screen will skip back to the beginning of the row. The *pause* button pause the clock and the *stop* button stops the lyrics (and music) and return the clock to its starting position (zero). The *RW* and *FF* buttons skip line each cycle (the lines shows on the GUI) forward or backward until the pause or play button press, or until it comes to the end or start of the lyrics, in this case it leaves the lyrics bar clean. The *skip_RW* and *skip_FF* skip one row each press. It has two possibilities: one, in case that pause is already pressed each skip will show as the lyrics of the next/previous row. In case that the music (and lyrics) are playing (no pause has been taking place) each press will skip the line, the clock and the music to the new state.

RESULT N° NI-2: Native Mix Playback

This software module enable the playback of Mix files produced by the Authoring Application, i.e. to reproduce the mix, with variable parameters over time, from the original audio tracks. This module is integrated as part of the Hi-fi System.

RESULT N° IRCAM-10: Score Following

Score following is the key to an interaction with a written score/song based on the metaphor of a performer with an accompanist or band. While for many Pop and Rock music songs the usual karaoke setup of a performer singing along with a simple recording of the accompaniment is sufficient, especially for a repertoire of classical and jazz songs, ballads and traditionals, the synchronization of the accompaniment to the performer is indispensable.

The Score Following Player designed in the framework of this project adapts and improves a score following algorithm originally developed for contemporary music performance in order to create an easy to use and robust automatic accompaniment application accepting monophonic audio and MIDI input from the performer. The included audio and MIDI following modules use the same core algorithm based on Hidden Markov Models (HMM).

Score following is the real-time alignment of a known musical score to the audio signal produced by a musician playing this score in order to synchronise the electronic part of the music to the performer, leaving him with all possibilities of expressive performance. It now has a history of about twenty years as a research and musical topic.

For an introduction and state of the art on score following and details of the system developed by IRCAM's Real-Time Applications team, a review of past attempts in score following literature, focusing on the adaptability and learning aspects of the algorithms, specially of importance for our work is available in the WP5 public report. Robust score following is still an open problem in the computer music field.

During the SemanticHIFI project, the training algorithm for the HMM core algorithm of the score follower was improved to make the system usable by untrained singers. Training in this context is to adapt the parameters of the following algorithm to a certain instrument and to a certain style of performance. A novel automatic discriminative training algorithm was introduced, which, unlike classical methods for HMM-training, is not concerned with modelling the music signal but with correctly choosing the sequence of music events that, was performed.

The learning for the Score following Player prototype application was performed on a database of recordings of various male and female inexperienced singers. After the automatic offline learning, the system provides significantly improved alignment off the score following.

RESULT N° NI-3: Seamless navigation in a track

This player, included in the Authoring Application, enables to move within a track during playback by skipping a chosen number of beats or bars while keeping a clean audio result as if the corresponding time interval had been removed from the audio file.

3.5 Audio Synthesis and Processing Modules (8)

RESULT N° UPF-1: Voice Transformation

Voice transformation has been an area of exploration probably ever since Lester Polfus (best know as Les Paul) started playing with tape recorders to achieve echo and delay voice transformations. Voice effects in music production became really popular in the mid sixties and seventies basically due to the advances in electronics and also due to the experimentation spirit of those days. Actually, it is really rare (nearly impossible) to find a popular music production in which no effect has been applied. Furthermore, it is very usual to apply alien modifications to the voice to transform it up to the point the voice loses its human character (especially in more dance and electronic productions).

In this sub-workpackage we offer voice transformations which, using state of the art signal processing algorithms, manage to convert and modify the voice of the users in amazing, yet musically meaningful ways that preserve the natural qualities of the voice.

Voice transformation is achieved with the *ComboVox* plugin, which allows the user to sing along with the music played by the system, while having his or her voice corrected or modified in meaningful transformations, such as adding vibrato, transposing or correcting pitch or modifying timbre. With *ComboVox* users are able to transform the character of their voice in real-time. It allows transformations such as *gender change*, *robot effect*, *ogre effect*, and others. It is based on manipulations of pitch and timbre.

The underlying technique in the voice conversion process is called Voice Pulse Modeling. This approach tries to combine the strengths of classical time and frequency domain techniques into a single framework, providing both an independent control of each voice

pulse and flexible timbre and phase modification capabilities. A more complete and detailed description of the algorithm can be found in related publications [Bonada, DAFX 04].

The graphical user interface of the ComboVox plugin

RESULT N° UPF-2: Voice Instrumentizer

Voice instrumentation is a step further into voice transformation. In this case, the voice of the performer is not used anymore as a modified voice; instead the maximum information from the voice is extracted and mapped into different instrumental synthesis algorithms, so to allow singers to directly control different musical instruments such as bass, drums or synthesizers that will be meaningfully added to the music playback.

This feature explores the possibilities of the singing voice as a natural and simple, yet sophisticated musical instrument. It provides new ways of control for digital musical instruments, accomplishing at the same time an expression control improvement for musicians and a low-entry cost for novices.

A voice analysis module extracts information about the characteristics of the voice signal in real-time. Depending on these descriptors, the synthesis module takes one sample from a database and transforms it. The attained effect is that the voice drives an instrument sound, or what we call “instrumentizing” the voice.

In the current implementation two instrument sounds are integrated: bass guitar and flute. With the taken instruments, we cover different instrument families which have usually distinct musical purposes. Bass guitar sounds are mainly used as rhythm accompaniment, while the flute acts usually as soloist.

In the GUI, one slider allows transposition of one octave up/down. Another slider allows to change the envelope amplitude.



Graphical interface of the Instrumentizer plug-in

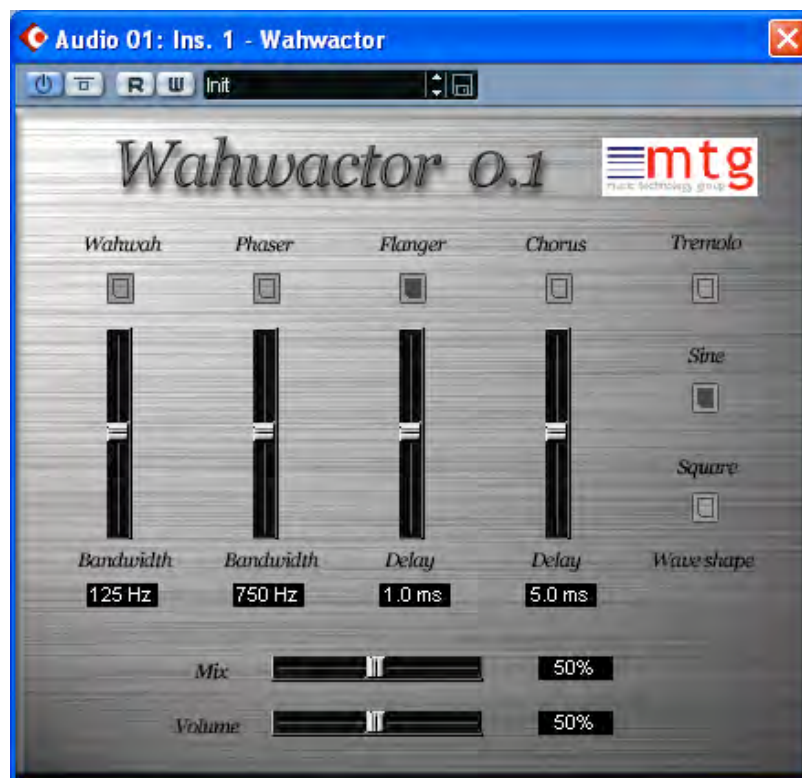
RESULT N° UPF-3: DJ Voice-Controller

The **Wahwactor** is a two-input and one-output system. Out of the two input tracks, one of the tracks may be considered a control rather than a proper input since it is the one in charge of driving the transformations to be applied to the audio signal. In the context of the Wahwactor,

the audio signal is typically a guitar signal and the control signal is a voice [wa-wa] utterance signal.

First, the voice signal is analyzed to pick up a meaningful descriptor that, after a simple conversion (shift, scale and smooth), is used as the centre frequency of the wah-wah filter, through which the guitar signal is sent to be mixed with the original.

To work in real-time, the Wahwactor uses a frame-by-frame algorithm described by the diagram illustrated in Figure 8. The voice signal is sampled at 44100 Hz and analyzed using a 2100 sample Hamming window and a 2048 point Fast Fourier Transform. The guitar signal is sampled at 44100 Hz and filtered using 2100 sample length buffers. The algorithm uses a 1050 sample hop size so that we have a 50% overlap in synthesis. This overlap is necessary to smooth the filter phase frame to frame variations.



Graphical interface of the Wahwactor plug-in

RESULT N° UPF-4: Rhythm transformations

This feature was initially intended to allow, “conducting or modifying the tempo and/or the dynamics of the music being played by the system using the HIFI remote command as a baton or tapping the beat”. Since during the evolution of the project, the consortium decided that no additional sensors would be implemented in the remote command, the feature has been finally implemented, allowing the user to tap the beat in some button or key.

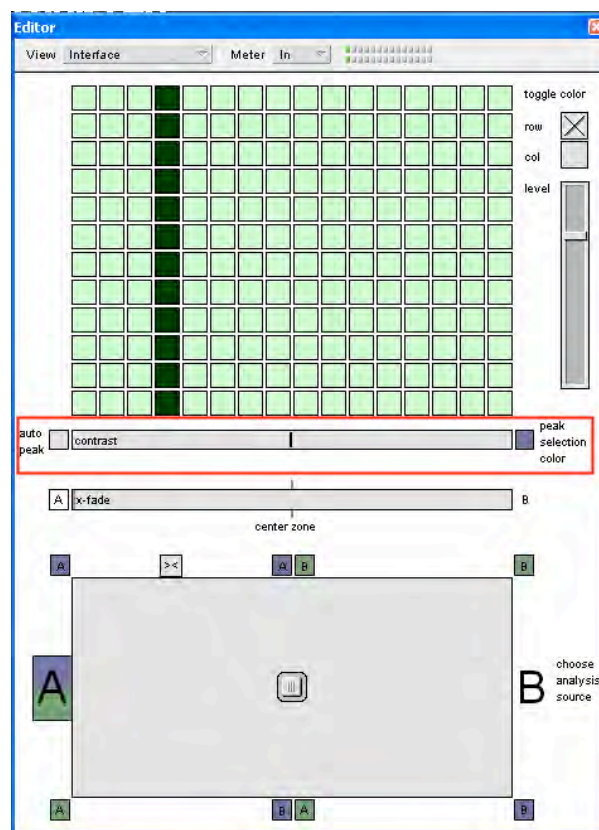
This feature involves three complementary techniques: (1) detection of the user periodic inputs, (2) on-the-fly detection of the background music beats, so that tempo changes can be synchronized with the user, and (3) high-quality time-stretching and pitch preserving techniques, which will allow the continuous modification of the music playback speed, without affecting its pitch nor its timbre. These modifications also include the possibility of changing the swing feel of the song (without affecting the average tempo), either by conducting or, simpler, by means of a dedicated “swing” slider.

A rhythm analysis module extracts information of tempo and beat location. Based on this rhythm information, we apply different transformations: "Tempo variation", and "Groove". This type of manipulation is generally referred as Content-based transformations. In addition, user interaction plays also an important role in this system. Tempo variations can be controlled either by tapping the rhythm with a MIDI interface or by using an external audio signal as tempo control. We foresee several use-cases, focusing on live performance situations.

RESULT N° IRCAM-11: “X-Micks” Song Morphing Plugin

The *X-Micks Song Morphing Module* allows for interpolating and mixing of two songs in the authoring application. Using the module the user can create transitions from one song to another and create hybrids of two songs other than by usual cross-fade.

The module has two stereo inputs for two beat-synchronised song inputs and provides one stereo output with the hybridised sound. Even if the control of module is slightly more complex than a cross-fader, a simple and intuitive user interface (see figure 1) has been developed for the module.



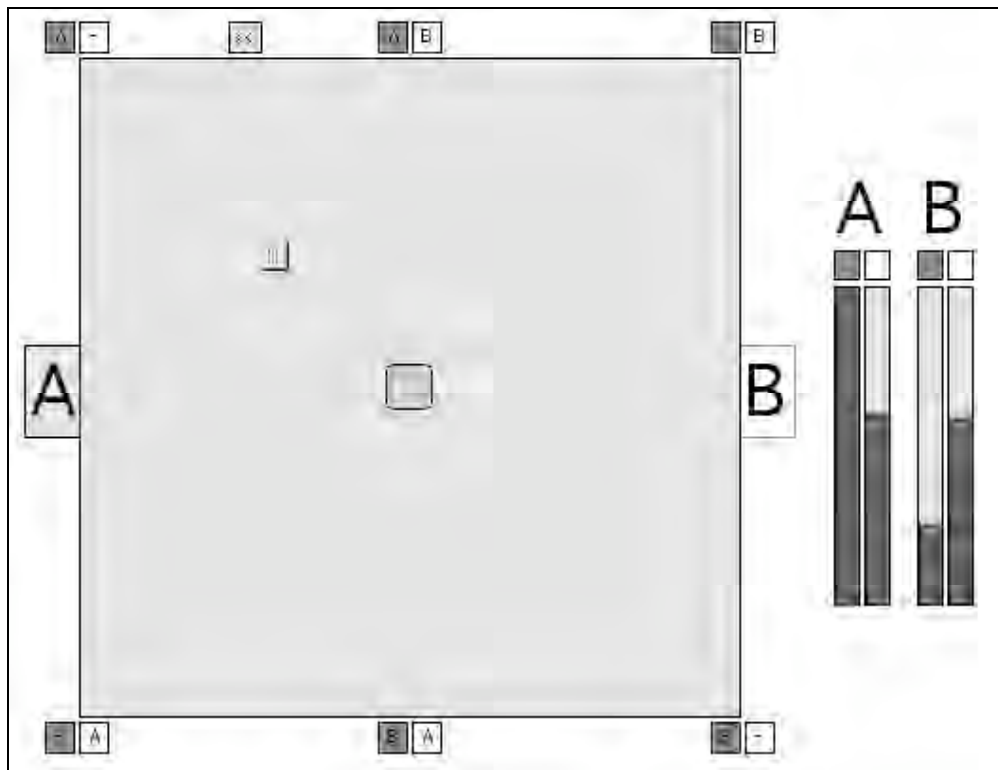
X-Micks VST plugin interface: Highlighted are the auto peak function and upper fader

The chosen approach is based on beat-synchronous band filtering: The *X-Micks* module decomposes the incoming sounds into 12 perceptive filter bands, which are controlled beat by beat synchronously to the incoming soundfiles. The user interacts with the module via a matrix display representing a beat pattern (typically 16 beats).

Each square of the matrix represents a certain frequency band at a certain beat in a rhythmic pattern of the beat-synchronised audio streams input to the module. Selecting and deselecting the squares of the matrix, the user can choose for each beat, in which way the audio output will

be composed mixing frequency bands from the audio inputs. The two audio inputs and the filters controlled by the matrix are beat-synchronized.

The actual mixture of the two input sounds represented by the selected and unselected matrix squares is determined by a second interface.



GUI prototype of the 2D mixing controller and experimental level indicators showing the current mixing of the two incoming audio tracks A and B to the selected (dark grey) and unselected (white) matrix regions. The large fields A and B left and right to the square of the controller are buttons to choose the automatic matrix selection by real-time analysis.

The matrix squares can be automatically selected by a real-time audio analysis. The automatic algorithm would select the strongest components of one of the incoming audio streams. The user can select the analysed input stream. The automatic selection of matrix regions can be seen as a very simplified spectrogram where time is quantized into discrete beats, frequency to 12 bands and the energy to two levels, on and off.

The interaction with the provided matrix representation is intuitive and creates a novel paradigm merging multiple widespread audio processing concepts and representations into one: the drum step sequencer, the spectrogram, the equalizer and the vocoder. The usual cross-fader is replaced by a two-dimensional controller.

RESULT N° CSL-10: SongSampler

The "SongSampler" is for musicians who want to experience novel ways of interacting with a music title. It automatically analyses the sound fragments present in any music title and maps these fragments to a synthesizer keyboard or any MIDI instrument. It produces an instrument that plays the same fragments as the original audio file. Playing with such an instrument is supposed to enhance the feeling of appropriation by letting listeners play their own music with the sounds of their favorite tunes.

The SongSampler is a portable extension panel for the Music Browser.

RESULT N° NI-4: Composition of Segments (Beatmasher)

The Beat Masher is a unique effect, developed as part of the Authoring Application, that isn't based on any classic effect type. It essentially samples a bar of music into a buffer which can then be transformed, and mashed. The buttons and knobs of this effect have to be explained thoroughly and in detail:

- Tap sets the tempo when the Beat Masher is not synced to the deck or the Master. If the effect is running in Sync, this button has no effect.
- Action starts the sampling until the buffer is full. Then, it repeats the recorded audio and warps it accordingly to the settings of the effect.
- Length defines the length of the Loop recorded in the buffer. The amount is always based on beats, and from left to right the values are: 1/32th (minimum value), 1/16th, 2/16th, 3/16th, 1/8th (centre position), 3/8th, 2/4th, 3/4th and one bar (maximum value).
- Sync synchronizes the tempo to the deck or to the Master if the Beat Masher is used as a master effect.
- Rotate changes the position of the Loop within the sampled bar. This function is most effective at short to minimum setting.
- Reverse plays the Loop backwards. If this is combined with a Gate Value set between 8 am and 10 am the effect is very obvious because the original signal is being punctured by short bursts of the reversed Loop.
- Gate works in two different modes. If you move it from the centre towards

the maximum value, it progressively mutes sections of the Loop until only one 16th of the Loop is audible at 100%. When moved from the centre towards minimum value, the original signal is being mixed into the loop, resulting in the most interesting effects. If Gate is in the centre position, it plays the Loop exactly as defined by the Length knob.

Note: You will not be able to get the optimum result out of the Beat Masher effect without triggering it by hotkeys or with the help of a MIDI controller, as this effect develops its character only when several parameters are being tweaked at the same time

RESULT N° NI-5: Real-time processing (tempo-synced FX and filters)

For the manipulation of audio-material played on a Deck of the Authoring Application, the extensive FX and Filters section provides a set of state-of-the-art effect and filter modules. Since these effects and filters can be synchronized with the tempo detected during the indexing steps the user can integrate these features comfortably in his live-performance or mix-recording. To increase the flexibility to control these effects or filters the user can modify its parameters directly at the mixer section or at separated control panel. The FX contributed by the partners (Groovator, Instrumentizer and Xmix) are also controled via these controls. The advanced user can open additional user interfaces dedicated to these advanced FX.

3.6 User Interface Prototypes (8)

RESULT N° IRCAM-5: Cocktail party EarBrowser

The “Cocktail Party Browser” provides an original means for browsing in a database of music pieces based on streams separation. Different titles are simultaneously presented to the user as audio streams organized in a sound scene paradigm. The browsing process is translated into

spatial navigation. The audio stream corresponding to the title under focus is presented in the frontal position, while secondary related streams are presented in the background. The challenge in this application lies in the creation of the desired “cocktail party effect” that allows putting the focus on a musical piece while remaining conscious of several other pieces played in the background.

The method used for this prototype consists in performing the spatialization simultaneously over three different music title. A number of “spatial positions” have been defined: front, front left, front right, back, back left and back right. When the user chooses either the front left or front right title to replace the front one, a smooth transition operates and the selected title comes to the front whereas the two other titles are soberly replaced by two new corresponding titles.

RESULT N° IRCAM-6: LScanner

This research relates to the domain of visualization and man – machine interaction in order to imagine meaningful ways of controlling the auditory sound scene. Indeed, as specified in use case WP4.3 (man/machine interface section) it is imperative that the sound scene (in the example of “on-the-fly mixing” application for instance) be controlled from a remote position matching preferably the listening position of the user.

The “L-Scanner” prototype is composed of a portable device that allows scanning the physical environment and representing a three dimensional model of the environment as well as the sound sources according to their spatial position in the manner of augmented reality systems.

The system displays a 3D model of the world that has been prepared beforehand. On top of this scene are represented the several sound scene whose position are transmitted by the ListenSpace environment. Finally a tracking system attached to the control interface allow determining the position of the listener and its orientation in order to adapt the graphical representation.



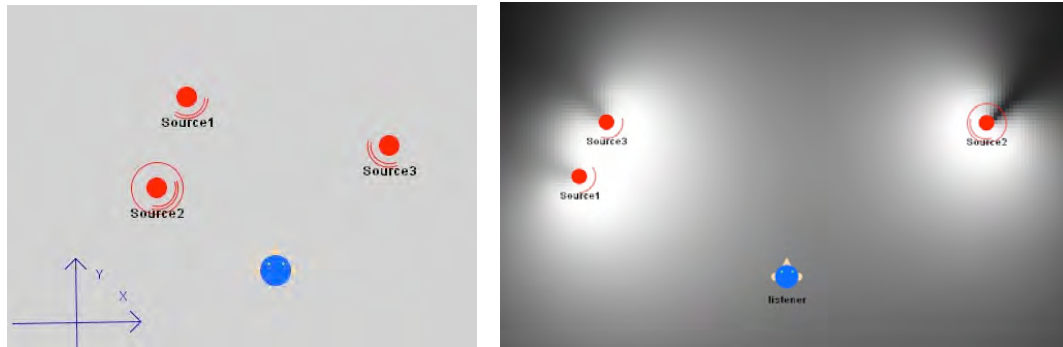
Utilization of the L-Scanner interface for controlling spatialization other a wave field synthesis restitution system

RESULT N° IRCAM-7: Visualization tool for the control of Spatialization

The visualization tool aims at providing useful information to the user while monitoring the spatialization of sources in a sound scene. The key idea is to exploit the background area of interactive graphical user interfaces to superimpose valuable information for the user. This information describes the dependency of a given perceptual criterion with regard to the

parameter currently controlled by the user and is represented in the form of a grey scale. Thus, the user can visualize and anticipate what would be the effect of a given action such as the movement of a sound source in the scene according to this criterion.

The system requires that a source is defined as “selected” and represents for each location of the sound scene what would the value of the criterion if the target source was moved to this location.



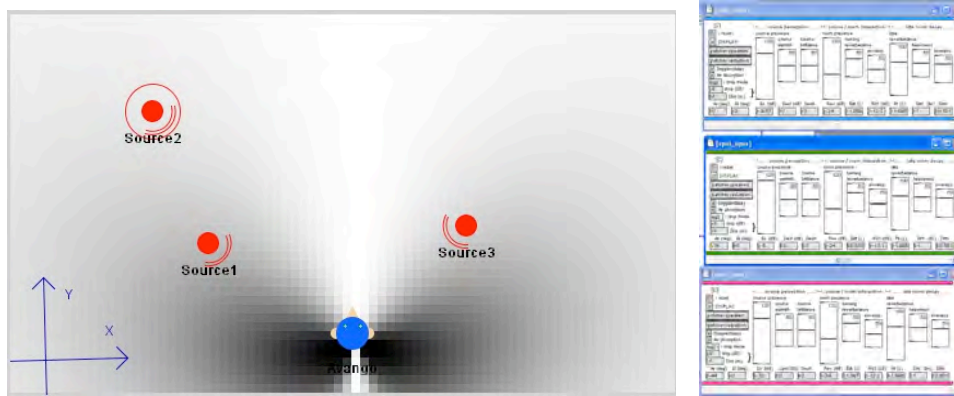
***2D representation & interaction space for a sound scene (left)
and superimposition of the radiation pattern of sound sources (right)***

The first figure gives an example of background representation: it consists in superimposing the radiation pattern of sound sources to the actual sound scene. This indicates to the user the positions of the listening reference point where the sources will be hearable and the positions where they can not be heard.

This system can be applied to a number of perceptual criteria. The representation of these criteria will guide the user in order to help reaching a consistent auditory result. Each of these criteria can be based on metadata either calculated in advance or estimated in real time. Example of such metadata is the loudness value of each sound source, its global attenuation (resulting from spatialization parameters) or general contextual parameters such as the rendering setup description and the rendering algorithm used.

Several criteria have been implemented and demonstrated. These criteria include for instance the “spatial uniformity” (depending on the rendering setup and the panning law used) that claims that “a good mix implies a homogeneous spatial distribution of energy at the ears of the listener” and uses the loudness value of each sound source calculated in real time as metadata information.

This prototype does not define a graphical user interface in itself but adds a method for drawing the background of an eagle-view like user interface representing a sound scene. The second figure gives an example of such representation for a sound scene composed of three sources. The criterion represented is the “spatial uniformity” in the context of a stereo (amplitude panning) rendering. The “Source2” is the target source: its location in a mid grey area indicates that the auditory result is slightly unbalanced and the system shows with a bright color the areas where this source could be moved in order to best satisfy the criterion.



Spatial uniformity criterion

RESULT N° IRCAM-8: “Cera Player” Realtime remixing tool

The application proposed consists in a “step sequencer” playing several sounds extracted (using filtering method and time segmentation) from a given original audio sample and reconstructed synchronously along the original sample. This example was integrated to the demonstrating system by transforming the main module into the form of a specific “player” module. A corresponding application of the Hi-fi System was created by associating to this player a number of spatialization modules and choosing MusicSpace to control the spatialization of each instrument stream.

The originality in this prototype lies in the fact that the music, when delivered to the listener, is not “pre-recorded” as in actual audio CDs. In the case of the Cera Player, part of the music is predefined and part of it comes as a set of samples that will be arranged and played in real-time according to the interaction with the listener. This concept of synthesizing the music ‘on the fly’ pushes further the idea of “active listening” already brought up in the MusicSpace project.

RESULT N° IRCAM-9: “Webern Player” Audio/score synchro & annotation tool

The hypermedia analysis of Webern Op.5 n°II is a flash prototype available integrated on the touch screen of the semantic Hi-fi or standalone on a PDA. It presents two score-based listening guides of a Webern piece based on the score written by Nicolas Donin, musicologist and researcher at Ircam.

The user can play the music track and follow it on the score, where some annotations guide its listening. Due to the small size of the touch screen on the Semantic Hi-fi system or the screen of a PDA, the analysis is entirely based on the usage of shapes and colors for the different annotations. The text is only present as a title for these two analyses.

The song and the score are linked by a cursor, but at any time, the user can:

- Navigate through the score with drag and drop,
- Reach other time code in the song with a click on the timeline.

While listening to the piece, the user can compare the two analyses, switching from one to another without the sound stops or the position of the score changes.

There are two types of annotation. Some that are always represented on the score and others that appear and disappear at a time code, not necessarily when the cursor reaches it. This allows the author to help the user anticipate some song structure, important phrases, breaks ...

The hypermedia analysis has been developed using Flash Macromedia Professional 8.

The first step was the multimedia synchronization of the score and the music track. After, annotations were added to the score for each of these two analysis.



Operation of the Webern op 5 n°II hypermedia analysis interface on the Hi-fi System PDA

RESULT N° FhG-2: Query by Humming browsing module

The Query by Humming module allows users to search songs in their database by singing the melody (or a part thereof) of the requested song. To achieve this goal, it comprises two main functionalities:

- Indexing the songs in the database: During this process, the melodies of the songs in the database are extracted and stored.
- Searching for a song: In this case, a query sung by the user is compared to the database of extracted melodies, and the songs containing the most similar melodies are returned together with a value indicating the similarity.

A working Query-by-Humming system existed prior to the project and served as a basis for new developments. Several innovations were achieved:

- The search algorithm used to compare melodies was improved by a complete redesign, which lead to raising the recall rate by 10%.
- The melody extraction process for sung queries was redesigned, so that runtime resource needs be compatible with the Hi-fi System while retaining quality of search results.
- The Extraction of the main melody from polyphonic audio (pieces stored in the database) for melody indexing is now automatic.
- These processes were encapsulated in libraries with interfaces that can be used within a content management system.

Query-by-Humming has been integrated in the Music Browser as an extension panel. It uses MCM to get a list of the songs in a database and do indexation. When searching for a song, the proper similar MCM song items are returned.

RESULT N° CSL-8: BabyBrowser

The "BabyBrowser", is a video browser especially dedicated for children. It offers limited use of descriptors but focuses on ease of navigation for users without the ability to read. Each song can be played and visualized with its corresponding music video clips and various interaction modes are designed for browsing these clips. By combining a text synthesizer and a simple textual input, the motivation to find the right video clip is used as an incentive for the child to learn to read and write.



A child using the BabyBrowser

The BabyBrowser is built on top of MCM and uses additional packages for voice synthesis. It was only tested on Windows.

RESULT N° CSL-9: EyeTune

The "EyeTune" application is a webcam-based gesture recognition music browser. It allows users to perform simple control like the volume, but also to select music using high-level metadata such as genre, country, artist, etc. The simplicity of the application (no need for any mouse or keyboard) makes it particularly adapted to the family context. It can be controlled by more than one person at the same time. Specific gestures were investigated to allow controls adapted to music access such as tempo tapping for query by tempo, or line drawing for creating personalized playlists with given properties, e.g. « increasing tempo ».



A user of the EyeTune application, selecting music by metadata through gesture recognition



The Eye Tune system in a home environment

The EyeTune interface and technology is a completely novel approach to music access. It builds up on the idea of the EyeToy Sony PlayStation2 technology, but incorporates a number of specific gestures and concepts critical to the context of music selection.

EyeTune is built on top of MCM and uses additional native Windows libraries for video input and gesture recognition. As such, it only works on Windows.

3.7 Software Frameworks (10)

RESULT N° FhG-3: P2P Sharing System

In order to meet the demands of the project, the P2P sharing system is provided with some – to our knowledge - unique features like certificate and central web service based secure user group management, the use of control mechanisms (licenses, filtering), the ability of metadata sharing, the use of semantic fingerprints for the search of audio file related metadata via a central matching server.

- *Secure user's group management* : In addition to the default membership services provided by the JXTA Framework, a new centrally managed and certificate based membership service is introduced in order to fulfil the requirement for secure user groups.
- *White-List Filter*: This component verifies if a file to be shared is allowed for sharing. Within the context of the Semantic Hifi project there are three valid types of files:
 - metadata files (XML, containing properties like specified in the SemanticHIFI metadata model)
 - play lists (XML, Native Instruments “Traktor” compliant)
 - mix-files (binary, Native Instruments “Traktor” compliant)

Depending on the type there are different tests which have to be passed successfully in order to be recognized as valid file.

- *Sharing license mechanisms*: The purposes of the license component are to complement the JXTA secure group concept by preventing “leaking out” content to other groups then

the specified ones and on the other hand to establish traceability of shared content by including the digital signature of the peer who shared the content. In order to fulfill these tasks, the following premises have to be met: only files with a valid license are allowed to be shared and downloaded and every user is provided with a digital certificate (PKCS#12). The W3C XML Signature standard is used for the generation of licenses. In case of infringement, e.g. an attacker has managed to bypass the filter component and has shared unauthorized files, the signature within the license can be used for obtaining his identity. Based on this information the user could be banned from the system. Licenses are shared within the “Base Group”, thus they are available for every user of the Semantic Hi-fi System.

- *Semantic fingerprints (AudioID identification)*: in order to link audio files to related metadata, the AudioID (Fraunhofer IDMT, WP2) system is used. However, the sharing system does not use AudioID fingerprints directly for sharing, instead a unique identifier is used (TUID). The sharing system provides means for fingerprint based search queries.

RESULT N° IRCAM-12: Shared metadata model

This result consists in the metadata model developed for the project peer-to-peer sharing features. The main entities of the meta-model are classes and properties. The properties can be text properties or numerical (float and int). This distinction might not be implemented in the sharing library though. Properties are defined as being optional or mandatory.

For each object, what can be shared is a meta-data file and an optional data file. The meta-data file should contain as much information as possible, allowing knowledge exchange between users. This document tries to collect a set of agreed properties, but a meta-data file is not limited to contain just the properties listed here. It can contain any property.

When an object is shared on the p2p network it has to be advertised by a set of properties. These properties will be available for searching by the user. Not all properties are advertised, since this would lead to a useless network traffic and RDV peer charge. For this reason beneath each property in the data model you will find a column Adv. that tells whether or not the property has to be advertised.

The shared properties of the music track are encoded in a simple XML file compliant with the RDF standard. This is an abstract format that can be used to generate also meta-data files for other classes. In this way, the shared metadata model can be extended to other objects or properties, and the objects or properties can be exchanged or advertised by the mean of the sharing system, provided that they comply to the advertisement of core properties.

RESULT N° CSL-1: MCM, Multimedia Content Management

The track management and browsing features developed as part of the project are based on a portable content management system called MCM (Multimedia Content Management) which development started right after the beginning of the project. The MCM API (Application Programming Interface) aims at simplifying the design of applications dealing with databases of multimedia objects, such as songs, video clips, etc. MCM is used in the project both for the development of WP3 applications and for the global specification of the metadata servers.

MCM is a set of Java classes offering the following data structures and functionalities:

- Multimedia items (e.g., songs, artists) existing synchronously both in database and in application memory.
- Fields (metadata) for each of these items (e.g., song's tempo, artist's name).

- Field values for each item are read/written to/from database, and can be cached in memory for applications requiring more CPU power, like playlist generation.
- Fields can be computable, i.e. their value is the output of an extractor, either computed online or offline in batch mode. The computation of field values can be scheduled in various ways, allowing an efficient management of possibly heavy computations (e.g. computation in background with low priority)
- Items can be linked to other items using fields (e.g., the value of the "artist" field of a "song" item represents the artist associated to the song).
- Items can be linked together through general purpose relations (e.g., timbre or cultural similarity linking songs together).
- Items, fields, relations can be dynamically added, updated, retrieved or deleted from the database at any time during the life of an application.
- The general structure of databases can be managed with schemas (e.g., the *song schema* defines the default structure of a database made of songs and artists).

MCM supports the important notion of *property* (or logical assertion) which can be instantiated on the database's items. Properties can be checked on specific items, and used to retrieve all matching items. Properties may be as simple as "rock songs", or arbitrarily complex, such as properties on the database structure ("countries which are referenced by at least one artist whose genre is rock"), or arbitrary matching algorithms ("artists whose name is close to Beetle"). Properties support logical operations (negation, conjunction, disjunction, etc.). Properties are automatically compiled into optimized SQL code.

RESULT N° CSL-2: EDS, Extractor Discovery System

Audio descriptors are traditionally designed by manually combining LLDs (Low-Level Descriptors) using ad-hoc algorithms or machine-learning algorithms. It requires a lot of experience and trial / error to achieve good results.

The EDS system (Extractor Discovery System) is a generic scheme which aims at simplifying as much as possible the design of high-level audio descriptors from audio signals. It is even able to *automatically* produce a fully-fledged audio extractor (an executable) from a database of labeled audio examples, including the automatic search of LLDs, combinations of LLDs, and machine-learning algorithms.

The key idea of EDS is to substitute the basic LLDs with arbitrary complex compositions of signal processing operators. EDS automatically composes operators to make features of signal processing functions that are optimal for a given descriptor extraction task. The search for specific features is based on genetic programming, a well-known technique for exploring search spaces of function compositions. Resulting features are then fed to a machine-learning model such as a GMM or SVM to produce a fully-fledged extractor program.

The global architecture of EDS consists in two parts: modeling the descriptor, and synthesizing the extractor. Both parts can be fully automatic.

The modeling of the descriptor is the main part of EDS. It consists in searching for a set of relevant features using a genetic search algorithm, and then to search for the optimal model for the descriptor which combines these features. The genetic programming engine automatically composes signal processing operators to build arbitrarily complex functions. Each built function is given a fitness value which represents how well the function performs to extract a given descriptor on a given learning database.

The evaluation of a function is very costly, as it involves complex signal processing on whole audio databases. Therefore, to limit the search, a set of heuristics are introduced to improve

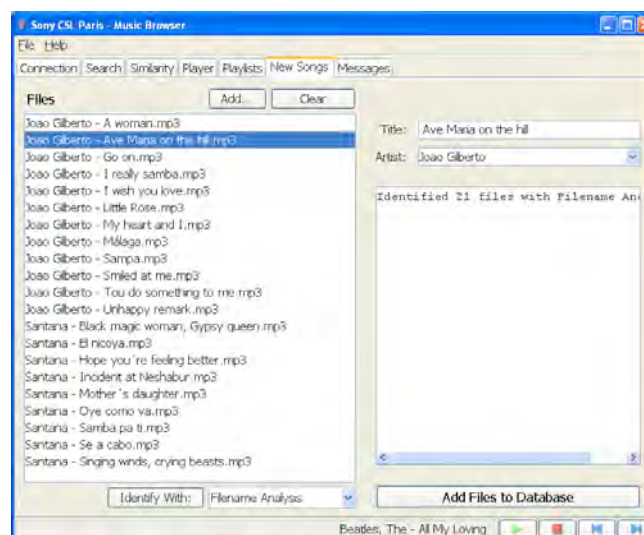
the relevance of the created functions, as well as rewriting rules to simplify functions before their evaluation. Once the system has found relevant features, it combines them into various machine learning models, and then optimizes the model parameters.

The synthesis part consists in generating an executable file to compute the best model on any audio signal. This part allows the validation of this model on arbitrary audio signals, to predict their value for the modeled descriptor.

RESULT N° CSL-3: Music Browser

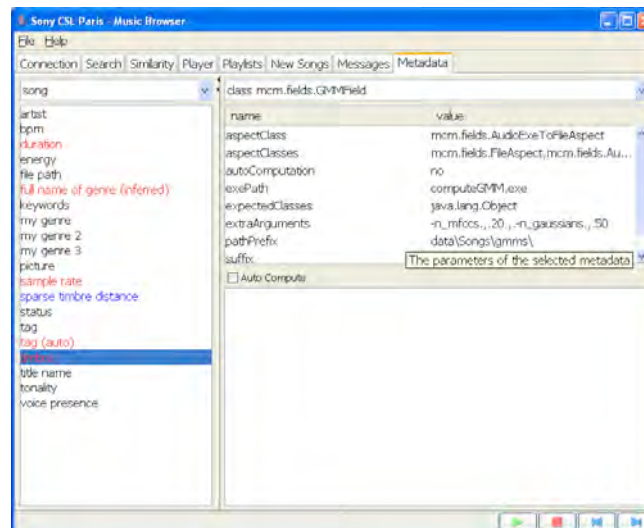
The main application developed in the context of WP3 is called the "Music Browser". It is a generic, non-specialized browsing tool including browsing functions and creation of general and personal metadata. It is built on top of MCM and emphasizes the use of plug-ins called "extension panels". All panels in the Music Browser are indeed extension panels with possible inter-communication. Most of the tools which were developed for WP3, as shown in this document, are or have been integrated as new extension panels.

With the Browser, the user can create a new MCM database from scratch and immediately import songs. The songs are automatically recognized using a variety of techniques: ID3 tags, filename analysis, fingerprinting, etc.



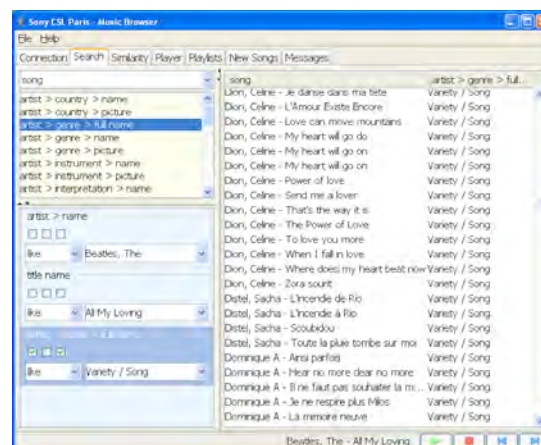
Adding new songs to a database

A panel allows the creation of new metadata/descriptors, but the application adds a lot of default descriptors to any new database, including computable fields like *song duration* and *timbre*.



The "Metadata" panel of the Music Browser

The user has access to all the available metadata/descriptors in a database. Descriptors can be used to query for, e.g., rock songs with low energy and acoustic guitar. Distributions of descriptor values on the whole sets of songs can be visualized, allowing to find most represented genres or to monitor the computation of a classifier.



The "Search" panel of the Music Browser

RESULT N° CSL-6: Automatic Generalization of Personal Categories

When the user is listening to a song, he might be tempted to associate the song to some personal category (e.g., "Evening Music"). Later on he might like to retrieve all songs from this particular category and organize them in a playlist, etc. Basic features of this browsing mode include the creation of new personal categories and the association of a song to a category. These features are already available in the Music Browser through metadata creation/edition and queries.

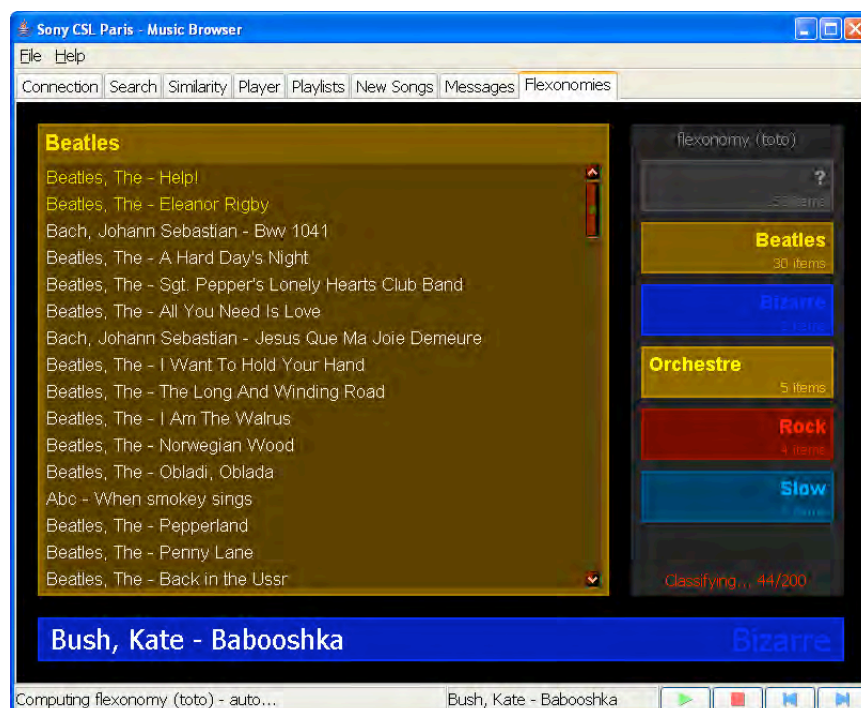
A much more advanced feature of this browsing mode is the *automatic generalization* of personal categories. It means the automatic placement of any song into one of the categories. With such feature it isn't necessary for the user to manually associate every song to a category: a specific algorithm would do it instead. For this feature to work the meaning of each personal category has to be "learned". If such a meaning could be formally defined by the user, it would be easy to check every song in the database against every category definition. But in general it is not a good idea to ask the user about such a definition. On the contrary, in a simple system a simple definition should be derived.

The "Flexonomies" panel of the Music Browser was developed to experiment such an automatic generalization process based on *audio analysis* of the songs. The timbre was first experimented, but using another audio extractor is also possible. (Indeed, what was used is not the timbre as defined for *timbre similarity* above, but a timbre feature from SONY CSL's EDS system which is also able to automatically learn the best signal processing feature to use for a specific problem. Here the problem could really be the definition of each personal category.)

So, in the panel a personal category **C** is defined by the timbre of the songs which were manually associated to **C** by the user. A song which was not associated to any category by the user would be automatically put into **C** if it were close to the songs making **C**, *and* not closer to any of the songs making another category **D**. So it depends on all the timbre distances with the songs making the existing categories.

This browsing mode is *fundamentally very interactive*: when the user does any change to his personal categories, the result of the previous automatic generalization is obsolete because the definition of at least one of the categories changed and the songs which do not appear in the definition of any category (i.e., were not manually *tagged* by the user) have to be checked against all categories once again.

An interesting feature of the "Flexonomies" panel is that it deals quite well with this kind of interaction by doing on-the-fly automatic generalization while the user is creating/editing his personal categories. Because timbre has to be computed only once, what changes is just the definition of categories and the distances between songs and categories. So the whole process can be optimized, and this is what the panel does. A few additional tradeoffs are also used to make the process even faster.

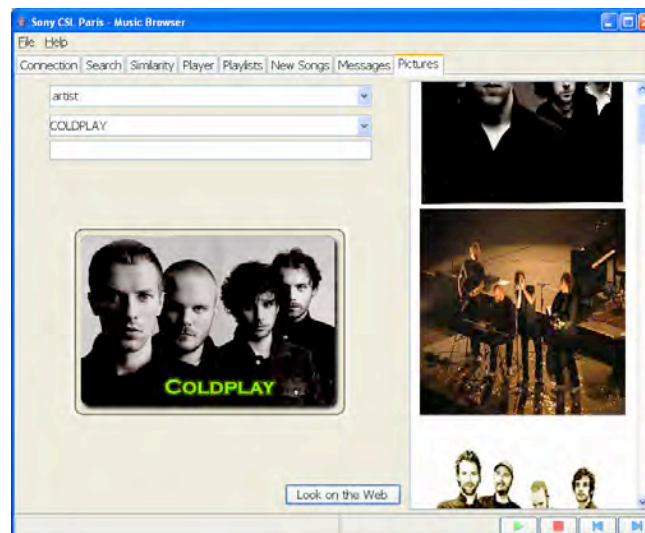


The "Flexonomies" panel of the Music Browser

RESULT N° CSL-7: Picture Grabbing

Browsing music databases could be just about listening. For the user experience however it is always more interesting to have pictures or videos. For instance, being able to display the photo of an artist or an album can improve a user interface a lot. Yet, it is not rational in a home Hi-fi System to have a picture for every artist or album in the world. It should rather depend on the songs which are actually on the system.

In the Music Browser a "Pictures" panel has been developed for this browsing mode. By default it displays pictures of artists, but it can be changed to any other type of items (e.g., country). The panel has a button which triggers a query to Google images for the current artist and parses the resulting HTML page so to find the URLs of the suggested pictures. The first 10 pictures are then displayed from the URLs (for the URLs which are still valid). If the user clicks on one of them, it becomes the default picture associated to the artist and is stored on local disk so that the corresponding URL is not queried anymore (saving network bandwidth).



The "Pictures" panel of the Music Browser

RESULT N° NI-6: Browsing Module

A number of interface browsing features were implemented by Native Instrument for DJ/Electronic music in the context of WP6 (Authoring):

- New search feature "Search by category": Type a phrase, then select where to search (album, artists, label etc.)
- "Show all" buttons in most list fields allows finding all tracks sharing the same artist, label, BPM etc.
- Undo functionality returning to the view before the last search command.
- Categories in the list window can be sorted by drag & drop and hidden via context menu.
- Sorting by icon column allows swapping of playlists.
- Task bar with 10 targets can be used for sorting tracks into playlists or as always visible shortcuts to any target in the browser tree.

- Time stamped collection backups for increased data security.
- Convenient edit pane to edit song metadata.

RESULT N° NI-7: Multimedia rendering - Embedded Web-Browser and Socket communication

A Web-Browser has been developed and embedded, which enables the Authoring Application to display web-content in a custom window. This component is used to integrate the Sharing Component. For the communication between the Authoring Application and the Sharing Component both applications contain socket-objects.

RESULT N° NI-8: VST-Host

The new VST-Host of the Authoring Application provides the fundamental functionalities and mechanisms to integrate external modules which manipulate audio-material in real-time. These modules, so-called “VST-Plug-Ins” were provided by the WP-partners IRCAM and UPF. Just as the internal effects of the Authoring Application these modules can be selected comfortably from a list of available modules. To modify parameters of the VST-Plug Ins the user has access to the GUI of the respective Plug-In-GUI. Furthermore a mechanism to synchronize Authoring Application and VST-Plug Ins has been implemented utilizing the standard mechanisms provided in the VST API. To integrate the “Instrumentizer”-Plug-In a microphone input has been implemented.

3.8 Target Applications (2)

RESULT N° NI-9: Authoring Application

The Authoring Application, developed from Native Instruments’ Traktor product, is aimed at musical content creation, potentially recorded as so called Native Mix. In the following all features developed for advanced real-time mixing in the frame of are given. These features are integrated in one final application and have been developed completely new in the frame of SHF. The following section provides an overview of the main developed features. An extensive description is given in the project WP6 Public Report.

Advanced EQ and Master section

The advanced EQ and Master section is the central mechanism to control the audio-output of the Authoring Application. All audio-streams of the Decks used by the user and the external audio inputs, eg, mic inputs, can be routed into this section and manipulated by its extensive set of parameters. The EQ and Master section consists of a perfectly emulated Allen&Heath Xone:92™ 4-channel club mixer, including its adjustable cross-fader curve, cross-fader assignable filters and high-end EQs. Since the Authoring Application provides 4 Decks to playback tracks, a central control panel was developed which allows the user to assign all Decks in a free definable combination to the Crossfader. The EQ and Master section itself is also free customizable. Each of the 4 controls from the EQ-Control to the Master Control can be turned on and off, which enables the user to design his environment intuitively. Extending state-of-the-art applications the EQ section is exchangeable. The user can choose from 4 different types of EQs which are outstanding emulations of real existing DJ-EQs like the Ecler Nuo4™. Also different cross fader curves are available.

4 fully-featured playback decks including an optional 2 deck mode

With the Decks the Authoring Application provides 4 completely new designed virtual turntables that behave like any real DJ turntable. In order to playback audio-files the user can now comfortably load songs into a deck (e.g. by drag & drop). The graphical representation of the track is generated, and displayed in 2 different ways, which give more control of the audio file to the user as state-of-the-art applications do. The Waveform Display gives a detailed overview of the audio material in a range the user can define himself. In the Stripe Display the user gets an overview of the whole track. By the usage of this display the user can navigate intuitively in the track and set Cue Points and Loop Points (see below). All 4 Decks can be used simultaneously. Each of these Decks consists also of control-elements, which enable the user to synchronize the decks and to modify the playback (speed, loops) and the pitch (key) of a track. Decks are the central mechanism to playback audio-files in the Authoring Application.

Advanced remixing functionality (loop processing)

The Authoring Application provides new real-time loop editing features which extend state-of-the-art looping features. One way to loop in a track is the Loop Points feature. The user sets a point in a track that defines the start of a section to be looped by pushing a single button on the Deck. Since the loop-length is definable in relation to the beat-grid, the end-point is set automatically and the loop runs seamlessly. Once the Loop is set the entire loop interval can be moved, even while the track is playing. Beyond that the loop-length can be modified afterwards directly at the respective deck.

Composition of segments: real-time mixing: see RESULT N° NI-4.

Seamless navigation in a track: see RESULT N° NI-3.

Real-time processing (tempo-synced FX and filters): see RESULT N° NI-5.

Recording all mix parameters in a non-destructive mix-file

A performing session with the Authoring Application can be recorded as a so called Native Mix saving all DJ moves as control data. For the first time a DJ mix can be recorded as a non-destructive, editable control file. The original tracks remain untouched by the mix. It allows to save hours of performance in a file size suitable for convenient sharing and distribution over the internet. Additionally, it allows convenient dubbing of a mix and to manipulate the mix in non-real time using a graphical user interface.

During playback of the mix the individual tracks together with the control data is rendered to give the original mix. The actions performed by the artists during the recording can be displayed while the pre-recorded mix is played back. A mix can be shared on the sharing system and can be played back if the retrieving user possesses the identical tracks as the creator – without infringing intellectual property rights. Mix playback module

The pre-recorded mixes can be loaded and played back with this module if the tracks used to create the mix are also present. The tracks are linked in the mix using a special Audio ID created solely for this special purpose. The mix playback module is running on the Hi-fi System and in the Authoring Application. OSC commands are used as API to control the mix playback.

Non-real time Mix-Editing Feature

This module provides functionalities to visualize and edit recorded mixes in non-real time extensively. Control values as well as time and position information can be edited, inserted or

deleted in a intuitive way combining the feeling of a live DJ performance and familiar stuido production tools to a completely new way to interact with audio material.

Publication of Playlists

The publication of playlists as html files has been implemented. The user can customize the fields he would like to see in the playlists (track number, duration, artist, title, ...) and their order. XSLT is used to transform the XML playlists of NI to html code. This enables advanced users to customize the printed playlist in any form he wishes and facilitates the publication on the internet and via P2P networks.

Sharing

An intuitive user interface for the execution of the search request, presentation of search results and downloading of files has been designed and implemented. Additional features like the Search Result Preview let the user get information about the content of a Playlists or Mixfile in advance.

RESULT N°EuTEC-1: Hi-fi System Prototype

A fully operational prototype of a Hi-fi System has been developed that realises a selected set of features that has been developed in the course of the SHF project. The Hi-fi System can be operated by a PDA style advanced remote controller in order to support sophisticated GUI operations in order to facilitate the new operating method provided by the various audio related technologies. The Hi-fi System is equipped with a touch screen in order to provide advanced metadata access such as displaying of artist pictures as well as the navigation on the song structure.



The remote controller provides a convenient way of accessing all the methods for advanced selection of content, such as individual song categories for automatic song classification. Furthermore it enables enhanced metadata access such a synchronized lyrics display.



An extensive and detailed description of the achievements related to the Hi-fi System is provided in the project WP8 Public Report.

4 Dissemination actions and materials

4.1 Public Dissemination

Press

WPI

Vinet, H. « L'Ircam se met à l'écoute de la seconde révolution de la musique », *Le Monde*, 16th October 2002

Stiegler, B. « Créer une lutherie électronique », *La Recherche*, October 2004

Vinet, H. Peeters G., and al. « L'Ircam recherche la chaîne hi-fi du futur », *01net*, 21st October 2004

Vinet, H. « La hi-fi du futur, véritable instrument de musique », *Le Figaro*, 16th February 2005

http://www.lefigaro.fr/sciences/20060401.FIG000000737_la_hi-fi_du_futur_veritable_instrument_de_musique.html?095854

Vinet, H. « Technologies musicales en perspective –SemanticHIFI », *Les nouveaux dossiers de l'audiovisuel*, INA, February-March 2005

Peeters, G. « Indexation et accès au contenu musical », *Les Nouveaux Dossiers de l'Audiovisuel* INA, February-March 2005

Vinet, H., Peeters, G., « Autour de la semaine de du son » *Recording*, March 2005

Vinet, H., *Science et Vie Junior*, May 2005

Vinet, H. « Le projet SemanticHIFI », *Champs culturels*, 2005

Vinet, H. « Pour la chaîne Hi-fi du futur, l'Ircam mise sur l'informatique », *Les Echos*, 12th October 2005

Vinet, H., Peeters G. et al. « Je t'MP3 moi non plus », « Le futur fait ses gammes », *Télérama*, 2nd November 2005

Vinet, H. « Imaginer la Hi-fi du futur », *Univers Hebdo*, 20th November 2005

Vinet, H. « Active listening gives meaning to digital music », *IST Results*, 21st November 2005

<http://istresults.cordis.lu/index.cfm/section/news/tpl/article/BrowsingType/Features/ID/79378>

Vinet, H. « La chaîne hi-fi qui obéira au doigt et à l'œil », *Le Monde*, 4-5th December 2005

Vinet, H., « La chaîne hi-fi du futur », *Le Figaro - Figaroscope*, 2nd November 2006.

Vinet, H., « SemanticHIFI », *Culture et Recherche*, issue #110, November 2006.

Radio

WPI

Vinet, H., Puig, V., Peeters, G., Radio France International, 21st Oct 2004

Vinet, H. « L'invité des sciences », *France Info*, 6th January 2005,

http://www.radiofrance.fr/chaines/france-info/chroniques/itwsciences/index.php?m=3&chro_diff_id=95000111

Vinet, H., Puig V., Peeters G. *Le magazine du multimedia*, Radio-France/RFI. Paris, France, 2005.

Vinet, H. & Pachet, F., 11 oct 05, « Science frictions » broadcast, *France Culture*, 29th October 2005.

Vinet, H. « Le projet SemanticHIFI », *01 Business broadcast; BFM*, <http://www.radiobfm.com/archive/player.php?id=2317&date=200605112255&duree=65>, May 11, 11:00 pm, 2006.

Vinet, H., *HSF Ilmenau: Radio Interview with M. Saupe*, May 15th, 2006

TV

Peeters, G. « SemanticHIFI, chaîne fidèle », video, *Télérama.fr*, 2005.

Peeters, G. « La chaîne HIFI du futur », *France 3 19/20 Edition Ile-de-France*, 2006.

Public presentations

WP1

Vinet, H., Peeters, G. & Delerue, O., « Le projet SemanticHIFI », *Resonances Ircam, Public Open Doors Conference*, October 2004

Vinet, H. « Le projet SemanticHIFI », *La Semaine du son*, Ircam, Paris, France, 2005

Vinet, H., Peeters, G. « Le projet européen SemanticHIFI », *Ircam Open Doors Conferences Resonances 2006*, June 11, 2006.

Vinet, H. , « Indexation automatisée des documents musicaux, Exemples d'applications », Round table on « Comprendre et transmettre la musique avec le multimédia » Bibliothèque Publique d'Information, October 23rd, 2006

WP2

Peeters, G. « Naviguer dans sa discothèque », *La Semaine du son*, Paris, France, 2005.

Peeters, G. « Description automatique et classification des sons instrumentaux », *Journées de la SFA (Description automatique et perception de la musique)*, ENST, Paris, 2005.

Peeters, G. « Caractérisation du timbre des sons musicaux », *journées Europeaniano*, 2006.

Peeters, G. « Systèmes d'identification et de navigation intra-documentaires », *Journée Valorisation du patrimoine de la création musicale*, 2006.

WP3

Sailer, C. « Query by Humming and interactive browsing on future HiFi-Systems », *presentation and panel discussion*, maerzmusik 2006, Berlin, March 21st, 2006.

WP4

Warusfel O., Delerue O. Workshop on Music Education: invited talk entitled « Ecoute de la musique spatialisée », *Resonances*, Ircam, June 16th 2006.

WP5

Loscos, A. « The Semantic HiFi Project. Performance prototypes », *School of Computing*, National University of Singapore, May 15, 2006.

Janer, J. « Voice as a musical controller », McGill University, *Music Technology Area*, Montréal, Canada, October 15, 2005.

Janer, J. « Leisure and Voice Control », Presentation at the *2nd Summer School in Sound and Music Computing*, Barcelona; July 23, 2006.

4.2 Professional Dissemination

Fairs and exhibitions

WP3

Fraunhofer IDMT, National Audio Broadcast Conference and Exhibition, Las Vegas, USA, April 2004.

WP5

UPF-MTG, European AES Pro Audio Expo & Convention, Barcelona, May 28-31, 2005.

WP7

Fraunhofer IDMT, MIDEM, Cannes, January 23rd-27th, 2005.

WP9

Ircam, **Sony-CSL**, *IST2003*, 2-4th September 2004

Ircam, *NetAtHome* 2003, 30th November-1st December 2003

Ircam, *KT Web Workshop*, 2nd-3rd December 2003

FhG, *Cebit 2004*, 18th-24th March 2004

Ircam, *Frankfurt Musik Messe*, 31st March-3rd April 2004

Ircam, *Villette Numérique*, 10th April 2004

UPF, *IST2004*, 4th-6th October 2004

Ircam, **Native Instruments**, **Fraunhofer IDMT**, *Musik Messe Frankfurt*, March-April 2006.

Ircam, Audio Engineering Society, Paris, May 2006

Technical workshopsWP1

Puig, V., NAVSHP AVISPA Concertation meeting, Cannes, 29th October 2003

Rousseaux, F., NAVSHP AVISPA Concertation meeting, Brussels, 11st March 2004

Vinet, H., “Représentations numériques et extraction des informations musicales”, Séminaire Apprentissage du Laboratoire d’informatique de l’Université Paris 6 (LIP6), 13rd May 2004.

Vinet, H., NAVSHP AVISPA Concertation meeting, Nice, December 2nd 2004

Vinet, H. « The SemanticHIFI European Project », Net At Home 2004, Nice, December 3rd, 2004

Vinet, H. Panel on Connected Appliances, Net At Home 2004, Nice, December 3rd, 2004

Vinet, H., « The SemanticHIFI Project », Workshop « Making music smarter : semantic audio research project in Europe », 118th Convention of the Audio Engineering Society, Barcelona, May 28th-31st, 2005

Puig, V., NAVSHP AVISPA Concertation meeting, Toulouse, 27th-28th October 2005

Vinet, H., “The SemanticHIFI Project”, Ircam-AES Technical Tour, AES Paris, May 23rd, 2006

Vinet, H., “Indexation des documents musicaux et applications - Le projet européen SemanticHIFI”, Association des professionnels de l’information et de la documentation (ADBS), March 9th, 2006; 300 persons.

Vinet, H., “Le projet SemanticHIFI”, Workshop on Music Education (top National Education Executives on Music Education), Resonances, Ircam, June 16th 2006 (80 persons)..

Lescurieux, O., NAVSHP Concertation meeting, Brussels, 24th-25th October 2006

WP3

Beurivé, A. « Personal Music Categories in the Music Browser », *Cité des Sciences, La Villette, Paris*, 21st and 23th June 2006.

Pachet, F. « Creation and Automatic Generalization of Personal Music Categories in the Music Browser », *Laboratorio di Informatica, Facoltà di Scienze della Formazione, University of Bologna, Italy*, May 24-26, 2006.

Pachet, F. « Music Information Retrieval », *University of Bologna, Italy*, 20-25th May 2005.

Riconno, P., Beurivé, A., Aucouturier, J.-J. and Defrenne, S. « Musical Games With the Music Browser », *Cité des Sciences, La Villette, Paris*, 20th Sept. – 3rd Oct. 2004.

Pachet, F., Beurivé, A., Aucouturier, J.-J. « Musical Games With the Music Browser », *Cité des Sciences, La Villette, Paris*, 15-18th June 2004.

Pachet, F. « Music Information Retrieval », *University of Bologna, Italy*, 26-28th May 2004.

WP4

Delerue O. « Control Interfaces for Sound Spatialisation », *AES Technical Tour* organized by IRCAM, May 23rd 2006.

WP7

Hasselbach, J. « Semantic Hifi: XMLSignature-basierte P2P-Sharing-Lizenzen für das JXTA Framework », *4th XML-Signature Workshop (XML-Uni)*, University of Koblenz-Landau, Germany, May 11-12, 2006, URL: <http://www.tu-ilmenau.de/fakmn/Veranstaltungen.2980.0.html>

4.3 Scientific Dissemination

Conferences

WP1

Vinet, H. « Description des contenus musicaux et applications », *invited talk*, International Conference on Digital Document, *La Rochelle, France*, 22nd-25th June 2004

Vinet, H. « The SemanticHIFI Project », International Computer Music Conference, Barcelona, September 2005.

Vinet, H. « Applications musicales du traitement de signal : synthèse et prospective », *GRETSI2005*, Louvain la neuve, Belgium, 9th September 2005.

Vinet, H. « The SemanticHIFI Project : Content-based Management and Manipulation of Musical Recordings », *EWIMT2005*, London, 2005.

Rousseaux, F., Bonardi, A., Poncelet R. « Online Music Distribution: a Case Study », *International Workshop for Technology, Economy, Social and Legal Aspects of Virtual Goods*, Ilmenau (Allemagne), 27-29 mai 2004.

Rousseaux, F., Bonardi, A., Zeppenfeld, C. « Une mise en scène inspirée de la fouille interactive de données », *Conférence Internationale sur le Document Electronique*, La Rochelle, pp. 205-112, 2004.

Rousseaux, F., Bonardi, A., Akdag, A. « Music-ripping: vers une conciliation de l'art et de la culture à travers la Toile », *SETIT'04*, Sousse, pp. 325-335, 2004.

Rousseaux, F., Bonardi, A. « New Tools and Practices for Electronic Music Distribution », *ISICT'03*, Dublin, 2003.

Rousseaux, F., Bouaziz, T. « Why Not to Imitate Collectors Using their Desire, rather than Logicians Using their Knowledge, to Address our Content Browsing Issues? », *WSPI'06*, 2006

Blanchard, F., Herbin, M., Rousseaux, F. « Compendium de données multidimensionnelles par une image couleur », *Atelier "Visualisation des connaissances" des journées Extraction et Gestion des Connaissances EGC 2005*, Paris, 19-21 janvier 2005

Rousseaux, F., Bonardi, A. « A Style of Theater Production directly Inspired by Interactive Data Mining », *Workshop Style and Meaning in Language, Art, Music, and Design*, *Conférence Internationale AAAI*, Washington, 2004

Rousseaux, F., Bonardi, A. « Reconcile Art and Culture on the Web : lessen the Importance of Instanciation so Creation Can Better Fiction », *First Workshop on Philosophy and Informatics*, Cologne, pp. 97-105, 31 mars – 1er avril 2004, également <http://CEUR-WS.org/Vol-112/>

WP2

- Peeters, G.** « Time Variable Tempo Detection and Beat Marking », *Proc. of ICMC*, Barcelona, Spain, 2005.
- Peeters, G.** « Rhythm Classification using spectral rhythm patterns », *Proc. of ISMIR*, London, UK, 2005.
- Peeters, G.** « MIREX 2005: Tempo detection and beat marking for perceptual tempo induction », *Proc. of ISMIR*, London, UK, 2005
- Peeters, G.** « Music Pitch Representation by Periodicity measures based on Combined Temporal and Spectral Representations », *Proc. of IEEE ICASSP*, Toulouse, France, 2006.
- Peeters, G.** « Musical key estimation of audio signal based on HMM modeling of chroma vectors », *Proc. of DAFX*, McGill, Montreal, Canada, 2006.
- Peeters, G.** « Chroma-based estimation of musical key from audio-signal analysis », *Proc. of ISMIR*, Victoria, Canada, 2006.
- Boutard, G., Goldszmidt S., and al.** « Browsing inside a Music Track, the Experimentation Case Study », *Workshop on Learning the Semantics of Audio Signals (LSAS)*, Athens, Greece, 2006.
- Ben-Shalom, A., Dubnov, S.** « Optimal Filtering of an Instrument », *Proceedings of International Computer Music Conference*, Miami, November 2004.
- Hadar O., Bykhovsky D., Goldwasser G., and Fisher E.** « A musical source separation system with lyrics alignment » in *10th WSEAS CSCC Multiconference*, Vouliagmeni Beach, Athens, Greece, July 10-15, 2006.
- Hadar O.**, Invited talk, « Semantic Hi-Fi consortium: Source separation and browsing by lyrics », *NATO ASI on Coding and Analysis of Multiple Access*, 26 August - 5 September 2006, Budapest, Hungary
- O. Hadar**, « Semantic HIFI Consortium*: Source Separation and Browsing by Lyrics on Future Hifi Systems » in *ECEC*, NJIT, USA March 29, 2006.

WP3

- Roy, P., Aucouturier, J.-J., Pachet, F. and Beurivé, A.** « Exploiting the Tradeoffs Between Precision and Cpu-time to Speed Up Nearest Neighbor Search », *Proceedings of the International Conference on Music Information Retrieval*, London, September 2005.
- Aucouturier, J.-J. and Pachet, F.** « Ringomatic: A Real-Time Interactive Drummer Using Constraint-Satisfaction and Drum Sound Descriptors », *Proceedings of the International Conference on Music Information Retrieval*, London, September 2005.
- Aucouturier, J.-J. and Pachet, F.** « Tools and Architecture for the Evaluation of Similarity Measures: Case Study of Timbre Similarity », *Proceedings of International Conference on Music Information Retrieval*, Barcelona, 2004.
- Aucouturier, J.-J., Pachet, F. and Hanappe, P.** « From Sound Sampling To Song Sampling », *Proceedings of the International Conference on Music Information Retrieval*, Barcelona, 2004.
- Pachet, F.** « The HiFi of the Future: Toward new modes of Music-ing » In Perrot, X., editor, *Proceedings of ICHIM 04*, 2004.
- Aucouturier, J.-J. and Pachet, F.** « Music Similarity Measures: What's the Use? », In Ircam, editor, *Proceedings of the 3rd International Symposium on Music Information Retrieval*, October 2002.
- Aucouturier, J.J. and Pachet, F.** « Scaling up Music Playlist Generation », *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, Lausanne, August 2002.
- Sailer, C.** « Using string alignment in a query-by-humming system for real world applications », 150th ASA Fall Meeting, Minneapolis, October 21st, 2005.
- Sailer, C., Rosenbauer, K.** « A bottom-up approach to chord detection », *Proceedings of ICMC 2006*, New Orleans, 2006

Sailer, C., MIREX 2006: «Two Note Based Approaches to Query by Singing/Humming », *Contribution to the Query by Singing/Humming task*, ISMIR 2006, Victoria, B.C., Canada

Sailer, C., Dressler, K. MIREX 2006: « Finding Cover Songs by Melodic Similarity », *Contribution to the Audio Cover Song task*, ISMIR 2006, Victoria B.C., Canada

Dressler, K. « Sinusoidal extraction using an efficient implementation of a multi-resolution FFT », in *Proc. of the Int. Conf. on Digital Audio Effects (DAFx-06)*, Montreal, Quebec, Canada, Sept. 18–20, 2006, pp. 247–252.

Dressler, K. MIREX 2005 « Extraction of the Melody Pitch Contour from Polyphonic Audio », ISMIR 2005, London, UK

WP4

Delerue, O., Warusfel O. « Mixage Mobile », *IHM06 (Interaction Homme – Machine)*, conference in Montreal, April 18th – 21st 2006.

Delerue, O. « Visualization of perceptual parameters in interactive user interfaces: Application to the control of sound spatialization », *AES 120th Audio Engineering Society Convention*, Paris, May 20th – 23rd 2006.

WP5

Alonso, M. Geiger, G. Jordà, S. « An Internet Browser Plug-in for Real-time Audio Synthesis », *Proceedings of International Computer Music Conference 2004*, Miami, September 2004.

Alonso, M. Geiger, G. Jordà, S. « An Internet Browser Plug-in for Real-time Audio Synthesis », *Proceedings of 4th International Conference on Web Delivering of Music*, Barcelona, October 2004.

Bonada, J. « Voice Solo to Unison Choir Transformation », *Proceedings of 118th Audio Engineering Society Convention*, Barcelona, May 2005.

Fabig, L., Janer, J. « Transforming Singing Voice Expression - The Sweetness Effect », *Proceedings of 7th International Conference on Digital Audio Effects*, 2004.

Hazan, A. « BillaBoop: real-time voice-driven drum generator », *Proceedings of 118th Audio Engineering Society Convention (AES 118)*, Barcelona, Spain, 2005.

Hazan, A. « Performing expressive rhythms with BillaBoop voice-driven drum generator », *Proceedings of 8th International Conference on Digital Audio Effects (DAFX 05)*, Madrid, Spain, 2005.

Hazan, A. « Towards automatic transcription of expressive oral percussive performances », *Proceedings of International Conference on Intelligent User Interfaces (IUI 2005)*, San Diego, 2005.

Janer, J. Bonada, J. Blaauw, M. « Performance-driven control for sample-based singing voice synthesis », *Proceedings of 9th Intl. Conference on Digital Audio Effects*, Montréal, September 2006.

Janer, J. Bonada, J. Jordà, S. « Groovator - an implementation of real-time rhythm transformations », *Proceedings of 121th Audio Engineering Society Convention*, San Francisco, September 2006.

Janer, J. « Voice-controlled plucked bass guitar through two synthesis techniques », *New Interfaces for Musical Expression Proceedings*, Vancouver, Canada, May 2005.

Janer, J. « Feature Extraction for Voice-driven Synthesis », *Proceedings of the AES 118th Convention*, Barcelona, May 2005.

Janer, J., Loscos, A. « Morphing techniques for enhanced scat singing », *Proceedings of 8th International Conference on Digital Audio Effects*, 2005.

Jordà, S. « Digital Instruments and Players: Part II – Diversity, Freedom and Control », *Proceedings of International Computer Music Conference 2004*, Miami, September 2004.

Jordà, S. « Multi-user Instruments: Models, Examples and Promises », *Proceedings of 2005 International Conference on New Interfaces for Musical Expression*, Vancouver, May 2005.

Jordà, S., Alonso, M. « Mary had a little scoreTable* or the reacTable* goes melodic », *Proceedings of the 2006 Conference on New interfaces For Musical Expression*, Paris, June 2006.

Jordà, S., Kaltenbrunner, M., Geiger, G., Bencina, R. « The reacTable* », *Proceedings of International Computer Music Conference*, Barcelona, September 2005.

Loscós, A. and Bonada, J. « Emulating Rough And Growl Voice In Spectral Domain », *Proceedings of 7th International Conference on Digital Audio Effects*, 2004.

Loscós, A., Aussenac, T. « The Wahwactor: a voice-controlled pedal », *Proceedings NIME 05, International Conference on New Interfaces for Musical Expression*, Vancouver, Canada, May 26-28 2005.

Loscós, A., Celma, O. « Larynxophone: Using Voice As A Wind Controller », *Proceedings of International Computer Music Conference*, 2005.

Cont Arshia, Schwarz Diemo, Schnell Norbert, « Training Ircam's Score Follower ». IEEE International Conference on Acoustics, Speech, and Signal Processing. Philadelphia : March 2005.

Schwarz Diemo, Cont Arshia, Schnell Norbert, « From Boulez to Ballads: Training Ircam's Score Follower ». International Computer Music Conference (ICMC). Barcelona : September 2005.

Cont Arshia, « Realtime Audio to Score Alignment for Polyphonic Music Instruments Using Sparse Non-negative constraints and Hierarchical HMMs ». IEEE International Conference in Acoustics and Speech Signal Processing (ICASSP). Toulouse : Mai 2006

Cont Arshia, « Realtime Multiple Pitch Observation using Sparse Non-negative Constraints ». International Symposium on Music Information Retrieval (ISMIR). Victoria : Octobre 2006.

Diemo Schwarz, Arshia Cont. « Score Following at Ircam ». ISMIR, Vancouver, Canada: October 2006.

Diemo Schwarz, Arshia Cont. « MIREX 2006: Score Following Evaluation ». Music Information Retrieval Evaluation Exchange Contest Organisation, ISMIR, Vancouver, Canada: October 2006.

WP6

Norbert Schnell, Diemo Schwarz, Remy Müller. « X-Micks – Interactive Real-Time Content Based Audio Processing », Digital Audio Effects (DAFx), Montreal, Canada: September 2006.

WP7

Hasselbach, J. « P2P sharing within the Semantic Hifi European Project », Axmedis 2006 – *Second International Conference of Cross Media Content for Multi-Channel Distribution*, Leeds, UK, December 12-15, 2006, URL: <http://www.axmedis.org/axmedis2006/>

Nützel, J., Kaufmann, M. « Sharing Systems for Future HiFi Systems », *Wedelmusic 2004 - 4th International Conference on Web Delivering of Music*, Barcelona, Spain September 13-14, 2004, URL: <http://www.juergen-nuetzel.de/papers.html>

Journals

WPI

Vinet, H. « Explorer la musique à l'âge numérique », Cahiers de médiologie N°18, “Révolutions industrielles de la musique”, Nicolas Donin & Bernard Stiegler Editors, Paris, 2004, Fayard. 2005.

Vinet, H. « Les nouveaux musiquants », L'Inouï, Ed. Nicolas Donin, Bernard Stiegler. (Léo Scheer, Paris), 2005.

Vinet, H. “Le projet SemanticHIFI : Manipulation par le contenu d'enregistrements musicaux », Circuit, vol.16 N°3, Presses de l'Université de Montréal, 2006.

Vinet, H. “Applications musicales du traitement de signal : Synthèse et prospective”, *Traitement du signal*, Vol.23 N°5-6, GRETSI, December 2006.

Rousseaux, F., Bouaziz, T. “Collecting or Classifying?”, *Lecture Notes in Computer Science*, Volume 3782/2005, Springer-Verlag, LNAI 3782, Professional Knowledge Management WM2005, Althoff & al. (Eds), Décembre 2005.

Rousseaux, F., Bonardi, A. “Music-ripping : des pratiques qui provoquent la musicologie”, *MUSICAE SCIENTIAE*, Janvier 2004.

Rousseaux, F. “Par delà les Connaissances inventées par les informaticiens : les Collections ?”, *Intellectica*, 2006

Rousseaux, F., Bouaziz, T. “Comment, sans nuire aux pratiques courantes des informaticiens, dépasser l'artifice des Connaissances ?”, *TSI-Hermès*, Numéro spécial « documents numériques », accepté, à paraître en 2006

Rousseaux, F., Bonardi, A. “Similarité en intension vs en extension : à la croisée de l'informatique et du théâtre”, *Revue d'intelligence artificielle*, 2005, volume 19, n° 1, pp. 281-288

WP2

Peeters, G. « Deriving Musical Structures from Signal Analysis for Music Audio Summary Generation: "Sequence" and "State" Approach », *CMMR 2003 (LNCS 2771)*, U. K. Wiil, Springer-Verlag Berlin Heidelberg 2004: 142-165, 2004.

Peeters, G. « Template-based estimation of time-varying tempo », *Eurasip, J. on Applied Signal Processing*, (Special Issue on Music Information Retrieval Based on Signal Processing), 2006.

Hadar O., Bykhovsky D., Goldwasser G., and Fisher E. « A musical source separation system with lyrics alignment », *WSEAS Transactions on Systems*, Issue 10, Vol. 5, pp. 2464-2467, October 2006.

WP3

Pachet, F., Aucouturier, J.-J., Laburthe, A., Zils, A. and Beurivé, A. « The Cuidado Music Browser: an end-to-end Electronic Music Distribution System », *Multimedia Tools and Applications, Special Issue on the CBMI03 Conference*, 2006.

Pachet, F. « Knowledge Management and Musical Metadata », In *Encyclopedia of Knowledge Management*, Idea Group, 2005.

Pachet, F., Laburthe, A., Aucouturier, J.-J. and Beurivé, A. « Editorial Metadata in Electronic Music Distribution Systems: Between Universalism and Isolationism », *Journal of New Music Research*, 34(2):173-184, June 2005.

Aucouturier J.-J., Pachet, F. and Sandler, M. « The Way It Sounds: Timbre Models For Analysis and Retrieval of Polyphonic Music Signals », *IEEE Transactions of Multimedia*, December 2005.

Pachet, F. « Les nouveaux enjeux de la réification », *L'Objet*, 10(4), 2004.

Pachet, F., Laburthe, A., Zils, A., Aucouturier, J.-J. « Popular Music Access: The Sony Music Browser », *Journal of the American Society for Information Science*, 55(12):1037 - 1044, 2004.

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.

4.4 Contribution to standards

Proposals

WP2

Fraunhofer IDMT: Proposal of EnhancedAudioSignatureDS for MPEG-7 (input document m13586 at the 77th MPEG-meeting in Klagenfurt), concerned with conformance and testing of audio identification.

Ircam, successful standardization of some descriptors for Music tonality: These descriptors and the related extraction method on MIDI data allows to automatically detect the Musical scale of a given music (minor or major). This is of particular interest in Cinema production and home listening context where major and minor often correspond to joyful and sad. July 25-27, 2005 and October 17-20, 2005

Workshop organization

WP2

MPEG-7 Workshop session on June 17, 2006, 14h30 – 17h45

Coordinator : Peeters, G., Ircam

Presentations :

- **Peeters, G., Ircam**, « What is MPEG7?, How to get into MPEG-7?, How to read MPEG7?, Patent issues »
- **Herre J., FHG**, « Scalable series and its performance for audio identification at various scale, aliasing problem of ss »
- **Jacob M., Ircam**, « MDS and databases »
- **Casey M., City University**, « LLD, SpectralProjection, the LLAMAS project, Low-Level Audio Metadata, projet GNU, StructureMusic.com, Video scene recognition using audio content »
- **Gomez E., UPF**, « CUIDADO SoundPalette Offline, MDS Tools, OpenDrama MPEG-7 site »

4.5 Patents

WP2

Peeters, G. (2004), « Patent Audio summary » (FR04/01493, 2004/06/16), Europe 04767355.3, Japan 516296/2006.

4.6 PhD Dissertations

WP3

Aucouturier, J.-J. « Dix expériences sur la modélisation du timbre polyphonique », (main sections in english), University Paris 6, June 2006.

WP5

Jordà, S. « Digital Lutherie: Crafting musical computers for new musics' performance and improvisation », UPF, Barcelona, March, 2005.

Janer, J. « Voice Control for Digital Musical Instruments », UPF Ph.D. Dissertation, 2006. (Thesis defense requested).

Loscos, A. « Singing Voice Transformations on Spectral Domains », UPF Ph.D. Dissertation, 2006. (Thesis defense requested).

5 Conclusion

The project gathered some of the most active labs and companies involved in music and audio technologies worldwide. Despite difficulties and challenges encountered, its implementation was conformant to the scheduled work plan and produced many research advances and innovations beyond the existing state-of-the-art, in several fields: automatic audio analysis for music information management, retrieval, browsing, and sharing; man-machine interfaces in various forms dedicated to interaction with musical materials (e.g. intra-track browsing, real-time performance, content-based processing and rearranging of existing recorded materials). Moreover, the project demonstrated, through the development of two target applications (Hi-fi System and Authoring Application), the possibility of integrating most of the obtained research results in consistent technical and user-friendly environments, through iterative design and development cycles involving potential users. Indeed, one of the main project challenges was to combine the implementation of innovative research-driven activities in various knowledge and technology areas with the market-driven finalization of full-featured application prototypes. The related choices have been implemented in order to maximize the number of integrated research results while remaining conformant to the identified market expectations in terms of functional features and technical constraints.

Another important concern has been to facilitate the integration and further exploitation of all project results in multiple forms through modular implementations corresponding to several complexity layers (functional modules, frameworks, applications). Moreover, the development of two different applications, one relying on an existing commercial product and enhancing it (Authoring Application/ Traktor), the other one developed from scratch following a new application concept (Hi-fi System), represent two complementary cases of potential exploitation of the project results.

It is worth mentioning that the main lines followed, aimed at promoting new ways of “active listening”, personal management, real-time performing, and related metadata sharing of recorded musical materials through content-based manipulation features are quite unique for a project of this size, and tend to overcome “usual” models of electronic music distribution that rely on client-server architectures and fixed metadata models. The originality and potential of these concepts have met great interest in the general public as shown by the numerous related press and media reports. As a matter of fact, the barriers between traditional technical musical functions (composition software, instruments, audio production tools, album-based physical delivery of recordings, listening devices) tend to vanish in the digital world continuum; this facilitates the emergence of new practices in which perception (listening) and action (personal organization, selection, re-production, production, performance, sharing, and distribution) are interrelated and feed each other, as in many spheres of human activity. The project has demonstrated the technical feasibility of such advanced manipulation interfaces. It paves the way for future music delivery and access modes, when electronic distribution will have replaced the current model based on physical supports of audio recordings, freeing the digital coding of musical contents from the constraints of these supports and enabling extended digital musical representations and richer user experiences.