

Science and Technology of Music and Sound: The IRCAM Roadmap

Hugues Vinet, STMS, IRCAM-CNRS

Hugues.Vinet@ircam.fr

Abstract

This article presents the roadmap for the research and development activities at IRCAM for the period from 2007 to 2010. Published in the context of the achievement of the Sound and Music Computing Roadmap by the S2S2 Consortium, it aims to deliver an extensive and synthetic vision of the identified directions for the research at IRCAM, through noticeably the introduction of the unifying notion of representation levels of musical information. The addressed subjects include sound analysis/synthesis, physical models, sound spatialisation, computer-aided composition, and interdisciplinary transversal themes concerning the links between the signal and symbolic worlds and more generally between different levels of music representations, the renewal of concepts underlying the management of time and interaction, and the issue of the elicitation and mediation of the musical knowledge.

1 Introduction

IRCAM¹ was founded in 1977 by composer and conductor Pierre Boulez under the aegis of the Centre Pompidou and the French Ministry of Culture and Communication. IRCAM is dedicated to promoting the relationship among music, science, and technology. Its main goal is to contribute to the renewal of contemporary music creation via advances in science and technology. In order to address the diversity of conceptual, methodological and technological issues raised in music, the array of specialities represented covers a wide range of scientific areas, including digital signal processing, computer science, acoustics, psychoacoustics, cognitive psychology, and musicology. The general scope of this research is designated—in the academic world—by the acronym STMS (*Science and Technology of Music and Sound*) the name of the lab jointly organized by IRCAM and CNRS, France's main research organization. Every year, approximately 20 composers are commissioned by IRCAM to create new works either using the latest results of the research carried out or through collaborations with the research teams in experimental projects. IRCAM is also committed to the dissemination of the works, technologies, and knowledge produced through various channels, including the production of concerts and live performances, training and education programs, and publications. Its software environments are distributed through the IRCAM Forum², a user group that represents 1,500 professionals worldwide. Educational activities include hosting the ATIAM³ Master 2 program, organized in collaboration with University Paris 6, and the Cursus Training Program in composition and computer-music intended for artists that has been recently extended to a 2-year course as part of the ECMCT⁴ Leonardo project involving a network of European cultural institutions.

¹ Institut de Recherche et Coordination Acoustique/ Musique – <http://www.ircam.fr>

² <http://forumnet.ircam.fr>

³Acoustics, Signal Processing and Computer Science Applied to Music - <http://recherche.ircam.fr/equipes/repmus/webatiam/>

⁴ European Course for Musical Composition and Technologies - <http://www.ecmct.eu.org/>

Since its establishment, IRCAM has remained the largest institution of its kind worldwide, bringing together currently more than 150 full-time collaborators including 90 researchers and engineers. It has been successively directed by Pierre Boulez (until 1991), Laurent Bayle (1992—2001), Bernard Stiegler (2002—2005), and Frank Madlener (2006—), who is also IRCAM's artistic director.

The evolution of the last decade has been marked by an acceleration of the range of activities carried out in order to meet increasing societal expectations and to address a broader audience. On the artistic side, a research group on performing arts was created, with the aim of exploring the modes of joint creations that associate music and other forms of performing arts including dance, drama, opera, and interactive multimedia. New applications for music education in junior and senior high schools were developed in collaboration with the French Ministry and Education in the framework of the Musique Lab 1 and Musique Lab 2 projects [Puig 2005]. This project is part of a more global line of action that focuses on cultural dissemination using new media, concentrating on musical works and related techniques, which has been identified as a key factor in their transmission to younger generations. Another recent and important concern, shared with other contemporary performing arts production centers, is the preservation of digital interactive works.

Beyond issues relating to contemporary music, the scientific and technical background developed in this framework has been also applied to many other fields of activity including sound and music production tools, electronic music distribution, telecommunications, games, and automobile manufacturing. The volume of the R&D activities has increased threefold in the last 12 years, thanks to the growing interest encountered in numerous application areas and to the subsequently increasing support of research funding agencies at European, national and regional levels in addition to that of industrial groups via collaborative R&D projects.

Due to the interdisciplinary nature of IRCAM's research topics, their renewal requires an intermittent organization of reporting and brainstorming sessions with all involved parties. The research orientations presented here are the result of a collective elaboration process that was an exceptional event in the 30-year history of IRCAM due to the level of collective participation it represented. The process was conducted at IRCAM between 2003 and 2006 and resulted in the main R&D objectives for the period from 2007 to 2010 as presented here. This roadmap was produced for IRCAM's research management and was not initially intended for publication. However, the publication of the SMC Roadmap⁵ in the framework of the S2S2 project, opened a debate that concerns the international musical research community, including IRCAM that did not have the opportunity to participate in these actions until now. This article, through the presentation of the IRCAM STMS roadmap, proposes an overview of the main research priorities for the coming years. Nine years after the last published synthesis on IRCAM research [Vinet 1999], it aims at providing an extensive and synthetic viewpoint on all the identified directions. Therefore, after a description of the roadmap elaboration process (Part 2), it introduces the notion of levels of representation of musical information (Part 3) that serves as the key conceptual background of this synthesis. The following parts present the four main transversal thematic orientations, following the structural

⁵ <http://www.soundandmusiccomputing.org/roadmap>

organization of the original roadmap, with small changes aimed at clarification. Part 4 (Core Science and Technology Areas) describes the lines of evolution of some of the main invested areas already: sound synthesis and processing, spatialisation, and computer-aided composition. The following sections correspond to interdisciplinary themes, concerning the links between various representations levels, including the signal and the symbol worlds (Part 5), the renewal of concepts for time and interaction (Part 6), and the issues of musical knowledge elicitation and mediation (Part 7).

2 Elaboration Process of the IRCAM Roadmap

This roadmap was produced as the result of a process that comprised several stages during two main periods. The first period (2003—mid-2005) involved more than 200 people, including most of the IRCAM collaborators, as well as several dozen composers from several different countries and representing a broad range of aesthetic sensibilities. The goal was to carry out an in-depth reflection on the main orientations of the musical research for the coming years at IRCAM, at the junction of the state of the art research and technology and the expectations and intuitions of a number of recognized artists. The first stage involved all of the R&D teams that were invited to produce a report formalizing the status of their main achievements and their proposals for new directions of research, in a form adapted to and understandable by the concerned artistic community, following brainstorming sessions. I then compiled these different reports, leading to the emergence of 20 different projects, divided into five main themes: *Composing Sound*, *Electronic Instruments*, *Interaction Concepts*, *Musical Space*, and *New Music Mediation Forms* [Vinet 2004]. A parallel can be made between this stage of achievement and the one resulting from the SMC Roadmap. Both were produced by a group of researchers of the same order of magnitude, but IRCAM's approach at this stage was more targeted to the issues of contemporary music. The main final difference has been that the IRCAM roadmap was further elaborated through a feedback and collaborative process involving representatives of the addressed targets, and was then used as a strategic agenda for an operational implementation of a currently ongoing research program.

The next stage consisted of presenting the summary report produced to a group of guest composers. Its proposals were subject to further exchanges in plenary sessions, each subject being handled by specialized subgroups that included members of each category of participant (researchers, developers, musical assistants, and composers). Several themes emerged from these exchanges. This led to the creation of project groups with a life span of several years and that reported their activity biannually to the collective on the following subjects, which are ordered from the most advanced to the most prospective ones:

- Score Following
- Sound Synthesis by Physical Modelling
- Sound spatialisation
- Gesture
- Voice
- Rhythm
- Computer-Aided Orchestration

Two other transversal subjects, integrating a number of issues raised in the other themes, were identified and led to further coordination actions between researchers:

- Signal/ Symbol
- Time-Related and Interactive Writing

The issues and objectives represented by these terms will be made clear through the development of the following sections. The whole process of plenary meetings and workgroups was coordinated by Andrew Gerzso and myself under IRCAM's Directors at that time, Bernard Stiegler and André Santelli.

The second period of production of the IRCAM roadmap, from 2005 to 2006, carried out under the direction of Frank Madlener, aimed to define the research program for the STMS lab for the four-year period from 2007 to 2010. The scope of the discussions was not focused only on concepts and applications for contemporary music, but concerned all dimensions of the research carried out at the institute and the potential applications for a broad range of activities. From the previously identified orientations and through further elaborations, each team was invited to produce its own four-year research agenda. The report I drew up identified four main transversal themes to structure all proposals that are, with few changes, the ones presented in this roadmap in sections 4 through 7. This scientific project was assessed positively by the IRCAM and CNRS scientific evaluation committees and its implementation is ongoing as of the date of publication of this article.

This roadmap production process can be seen as an experience of interdisciplinary research management. More specifically, it was intended as a means of counterbalancing a trend toward ever more specialized and disconnected activities by setting up the conditions of a collective action, gathering all the concerned skills and interests, and focusing on the production of a shared, structured view of common objectives.

3 The Representation Levels of Musical Information

This section is dedicated to the introduction of preliminary notions necessary for the presentation of the roadmap orientations. The goal is to characterize the different types of representations used for describing musical phenomena in technical systems. This section proposes a global view that integrates all kinds of representations in a single, unified scheme called *Representation Levels of Musical Information* for reasons explained hereinafter. The term "representation" refers here to the way musical information is represented internally in applications, i.e. essentially in data structures. It excludes by convention another aspect of the notion of representation, i.e. the way these internal data are mediated to the user through man-machine interfaces. These notions were first introduced in [Vinet 2003] and the following description is to be considered as an update and refinement of the initial concept and terminologies.

Most representations found in musical applications belong to one of the five following types:

- *Audio signals*
- *Symbolic representations*, describing discrete musical events in symbolic form, such as in score formats or MIDI notes

- *Control signals* in the form of slow-varying signals—intermediate between audio signals and symbolic descriptions—that describe the variation of musical parameters and their modulations. Such forms generally appear as input parameters for sound synthesizers and processors, as well as the output of gesture capture devices
- *Physical representations* that account for the physical state of a set of points within a spatial zone; the related data are obtained through measurements or are computed as outputs of physical models for sound synthesis or sound spatialisation
- *Semantic representations*, that describe aspects of musical materials using textual structures, beginning with any type of elementary high-level descriptors, in the terminology of music information retrieval and MPEG7 up to more structured data such as taxonomies, semantic networks, and ontologies

The adjective semantic was not initially chosen and remains relatively inappropriate in the case of music that is an abstract mode of expression and communication - except in the case of a few exceptions at the border of *musique concrète* and organized environmental sounds. It is still retained here as it is broadly used by other authors accounting for the use of language for describing music.

These different types of representation can be ordered, from bottom-up, by increasing levels of abstraction (Figure 1). These levels are related to the quantity of information that they convey per time unit, i.e. their data rate range provided optimal codings. The low-level representations are explicit and self-contained whereas higher-level ones rely on external, implicit knowledge.

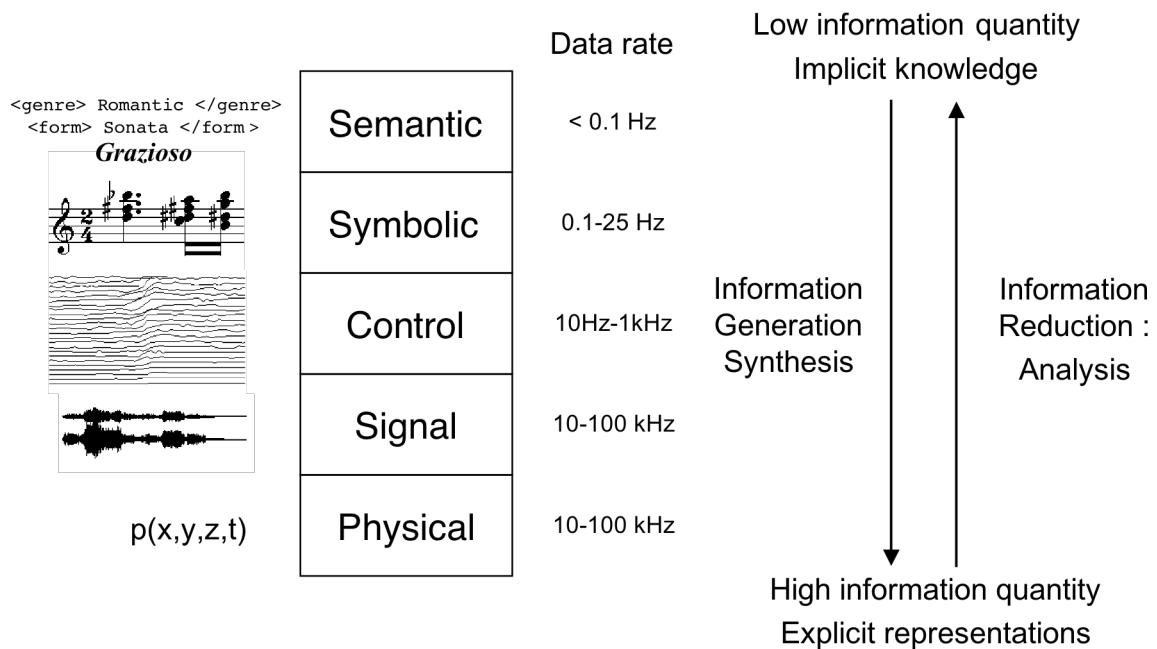


Figure 1: The Representation Levels of Musical Information

The conversions of representations across different levels are explained by the generic processes of analysis and synthesis. Analysis corresponds to the extraction and synthesis to the generation of information. For instance, on the synthesis side,

the information generated by an audio synthesis model from control signals is contained in the characteristics of the model, such as an instrument model, that produces certain classes of sounds in function of its input parameters. The production of control signals (gestures) from the score, i.e. from both symbolic (notes) and semantic data (nuances and performance notes) corresponds to the information brought by the performer in terms of interpretation choices and personal expression, through the activation of various kinds of relatively non-elicited knowledge dimensions related to his/her internalized musical background, motor abilities and personal sensitivity. The production of physical data from signal data is obtained through the combination of the acoustic radiations of the given sound sources with the room effect. Conversely, the analysis process corresponds, for conversions from physical to signal levels to a spatial reduction (sound capture); from signal to control levels to a parameter extraction over time (e.g. energy or fundamental frequency curve); from control to symbolic to a double digitization of the feature in terms of time segmentation and scale of values, after having applied a psychophysical conversion law between the measured values and the corresponding perceptual feature. Finally, the analysis carried out from lower-levels (signal to symbolic) to the semantic level corresponds to the integration of multiple pieces of information at a larger temporal scale in order to derive global characteristics.

One may object that there is no fundamental difference between the symbolic and semantic levels, that both relate to symbolic structures at different scales. A first answer is that musical objects considered at relatively different time scales (such as the note and movement levels) correspond *a priori* to different mental representations that shall be reflected in related technical representations. Another element of reflection concerns the kinds of features that are considered: the features recognized as part of the symbolic level are, through their use, the ones assimilated in the underlying music theories, such as pitch and loudness, and that are generally related to computable, numerical data. Contemporary music extends these theories by relying on additional and complimentary features, such as timbre or spatial features, that are moved from the semantic to the symbolic level as soon as they become an integral part of the musical vocabulary.

The concept of Representations Levels provides a useful grid for analysis of all kinds of musical applications from the viewpoint of their internal representations of musical phenomena. It also exhibits a large class of current research issues, concerning the conversion between levels through analysis/synthesis processes, or the building of representations appropriate to each level through a formalization process involving the elicitation of the related musical knowledge. These subjects constitute two of the main transversal orientations of the roadmap presented hereinafter (§5 and 7).

4 Core Science and Technology Areas

The research subjects presented in this section are the ones in which the greatest efforts have already been invested at IRCAM in the last years. These subjects have provided the foundations of current musical applications, with the exception of real-time interaction technologies that are presented separately. This section presents the evolution of these research themes through the continuation, enhancement, and extension of existing directions.

4.1 Audio Signal Synthesis and Processing

Even if initially designed in the framework of contemporary music creation (for extending orchestral sounds) and telecommunications (for voice processing), the applications for audio synthesis and processing techniques have spread beyond these areas and are widely used today in many fields such as audiovisual and cinema production, games, and vehicle simulators. Most related research at IRCAM relies on the conceptual scheme of analysis/synthesis models.

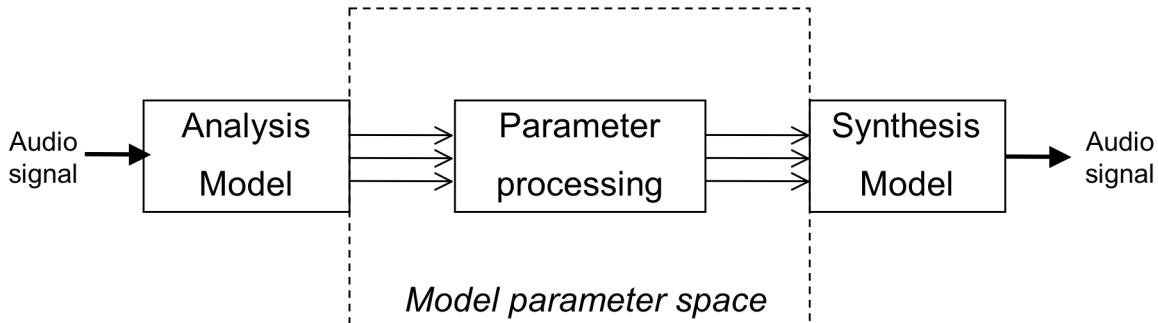


Figure 2: Processing architecture using an Analysis/ Synthesis model

A synthesis model has a set of input parameters and produces a given range of sounds according to the parameters' variations. The computation of these parameters from existing sounds is carried out by the analysis model, when this inverse function exists. Processing sounds using analysis/synthesis models then becomes a processing operation within the parameter space (Figure 2). Typical sound transformations obtained with models such as sinusoidal [Rodet 1998, Roebel 2006, 2007], pitch synchronous [Peeters et al. 1999] and phase vocoder [Dolson et al. 1999] include time-varying filtering, transposition, time-stretching, and hybridization between two sounds. This research has led to the development of AudioSculpt, a powerful time-frequency graphical editor based on the SuperVP phase vocoder [Bogaards et al. 2004]. One goal of current research is dedicated to obtaining high-quality processing functions, that preserve the naturalness of sounds, by integrating advanced features such as the automatic detection of transients for time-stretching [Roebel 2003], or an enhanced spectral envelope computation for transposing source-filter sounds [Roebel et al. 2005, Rodet et al. 2007, Villavicencio et al. 2007]. This research has been applied in particular to high-quality voice transformations such as making a spoken voice sing or changing the identity and voice quality of a speaker in film post-production. Major breakthroughs have been also identified in the analysis of polyphonic signals (analysis of multiple fundamental frequencies, source separation). Other important related issues concerning the automatic indexing of recorded music and the management of control signals are further described in section 5.

4.2 Synthesis by Physical Modelling

A significant effort has been devoted at IRCAM to physical models for audio synthesis, due in particular to the interest of composers in the richness and natural character of the sound materials produced. The main integration environment for all kinds of objects and interactions models is the Modalys software program [Adrien 1991, Iovino et al. 1997, Ellis et al. 2005], that has been recently completely reengineered, optimized and now runs as a standalone application, as well as in Max/MSP for real-time synthesis or in OpenMusic for enabling the synthesis control

via algorithmic musical processes. Modalys models any kind of one- or two-dimensional vibrating object using the same modal formalism (plates, membranes, tubes, strings). It is possible to build complex virtual instruments by specifying the elementary objects they are made of and the various kinds of interactions that relate one to each other, including their modes of excitation (strike, bow, reed, etc.). Some recent work carried out added the possibility of simulating 3D objects modeled by finite-element meshes [Bensoam 2003a, 2003b]. Recent and current research is dedicated to integrating state-of-the-art models of interactions and conceiving new ones, including flute and trumpet embouchures [Vergez et al. 2000], single and double reeds, piano hammer strikes and new models of bow, to extend the formalism of the models to weakly non-linear systems and to model the radiation of the virtual instrument in order to position the listening point anywhere in the space. Most work carried out on Modalys is still unpublished, but it has already been used in many musical works including for real-time performances. Another promising research direction concerns the efficient modelling of the non-linear propagation phenomena in tubes, including thermoviscous losses and flared tubes, using formalisms such as Volterra series and diffusive representations [Hélie et al. 2004, 2006, 2007a, 2007b]. An important research issue with complex physical models, both in the context of studio work and real-time performance is the one of their control: this subject is handled in section 5.

4.3 Sound Spatialisation

The term of sound spatialisation covers for sound production techniques that aim to provide spatial impressions to listeners. The issues considered here focus on reproduction techniques that are based on electroacoustic transducers, and do not touch on the problem of the acoustic room design. Sound spatialisation has become an important dimension of all application fields including audio synthesis and processing, such as music production, telecommunications, simulation and virtual reality, games and multimodal human-computer interfaces. The implementation of high-quality spatial rendering features is generally recognized as an important factor of the overall quality of multimedia systems, and can contribute in some situations to enhancing the global system performance (such as source spatial segregation in multimodal human-computer interfaces and simulators). In contemporary music, the control of the spatialisation of sound sources tends to become an autonomous musical dimension, crossing all aesthetic approaches.

The research topics concerned in sound spatialisation include electro-acoustic transducer systems for sound capture and reproduction, the representation and coding of sound scenes, and the study of spatial perception and cognition. The usual coding formats, including stereo and 5.1, rely on the juxtaposition of different audio reproduction channels and are therefore easy to operate, however they present the same inherent limitations: the optimal quality of reproduction is restricted to a limited spatial zone (the sweetspot), the coding is specific to a given reproduction setup, and the mixing of complex scenes, comprising many different sources, relies essentially on the empirical expertise of sound engineers, with few objective tools for assisting them. This latter remark also applies to the spatial diffusion of electro-acoustic music in concert that generally consists of routing the various input audio channels to a set of loudspeakers by applying different gains and filters to each matrix point. A more objective approach consists in describing the sound scenes to simulate in the form of sound sources and a listener positioned in a given virtual room. Geometrical models compute the impulse responses of the room transfer function from each source to the

listener position, and convolve them with the sources' signals. Their main drawbacks are computation cost and the necessity of updating the transfer function for each new position of the source when it is in movement. The approach followed at IRCAM in the Spat environment combines the simulation of the sources positions, directions and the room effect through the specification of perceptual parameters, that define the various subjective features of the target space quality independently of the chosen reproduction setup [Jot 1999]. The spatial simulation is then efficiently computed according to the used reproduction system, including binaural (headphones), transaural (stereo and quadraphonic), multi-loudspeaker intensity panning, and Wave Field synthesis systems (WFS). The Spat software program was designed for the Max/MSP environment; it is now available as portable libraries for integration into virtual and hardware mixing consoles. Both geometrical and perceptual approaches are standardized in MPEG4 [ISO/IEC 2000].

The identified priorities for current and future research include the experimentation of WFS systems in live performance situations, perceptual studies on spatial perception and cognition, and high-level concepts and tools for music composition and virtual reality applications.

Wave Field Synthesis Systems (WFS) aim at synthesizing the acoustic waves emitted by the virtual sources to simulate using multiple secondary sources surrounding the audience [Berkhout 1988]. One point of interest of this technique is that, unlike traditional loudspeaker systems, the positions of the simulated sources are perceptually consistent for all listener positions within an extended area. WFS systems also enable the simulation of plane waves, i.e. coming from the same incidence for all listener positions. After the successful implementation of small-scale systems for interactive installations, notably at Centre Pompidou⁶, the current efforts center on the experimentation of large-scale systems for real-time musical performances that will be installed in IRCAM's Espace de Projection. Current research issues include the automatic system calibration from specific measurements, the study and compensation of the room effect, the experimentation of new spatial effects such as sources with variable directivity, and the implementation of experimental studies aiming at characterizing the actual perception of such effects [Corteel 2006, 2007].

The synthesis of spatial sound scenes that preserve the consistency of sources positions across the different listener's positions was also the object of experimentation in the framework of the LISTEN project that aimed at synthesizing individual interactive sound scenes using binaural systems controlled by the real-time capture of the listeners' position and head orientation⁷. The new situations introduced by these systems, in which listeners can interact with a spatial sound scene by moving their body in the same space, modify the process of internalization of the spatial scene and still need to be better understood from the perspective of intermodal spatial cognition studies. This concern is being addressed in the CROSSMOD European project⁸ that considers in particular the influence of various modalities in order to optimize the multimodal rendering process in the context of virtual reality [Sarlat et al. 2006].

Another identified objective, related to these new possibilities of spatial sound simulation and interactivity, concerns the design of new concepts and tools for sound

⁶ Dada exhibition (2005), 'Danse avec loup' interactive installation (2006), Beckett exhibition (2007)

⁷ <http://listen.imk.fraunhofer.de>

⁸ <http://www-sop.inria.fr/reves/CrossmodPublic/>

engineers and composers. There is an interest for example in extending the usual paradigm of a sound scene made of punctual sound sources and following a given trajectory, to extended sources, with a variable directivity. Moreover, few concepts and tools are already in existence that specify a spatial interaction involving a listener in movement. More generally, the increasing use of sound spatialisation in interactive virtual or augmented reality systems involving 3D gesture capture and image synthesis also requires the definition of interaction metalanguages, compatible with the concepts, languages and tools used for each media, for specifying the interaction at a global level, using a common geometrical reference [Donikian et al. 2007]. A longer-term objective might be to design 4D multimodal editors for designing interactive spatial multimedia scenes that are variable over time.

4.4 Computer-Aided Composition

Whereas most music languages developed in the computer music community were designed primarily for audio synthesis and processing, there has been a strong expectation and interest by composers at IRCAM in the last 20 years in developing computer tools specifically devoted to the representation, manipulation, and algorithmic generation of symbolic musical structures that would assist them in elaborating materials for instrumental musical writing. This work has been implemented through two successive generations of software environments, PatchWork and OpenMusic, based on LISP/CLOS language frameworks [Assayag et al. 1999]. In order to be suitable for various aesthetic approaches, and to favour an experimental and intuitive approach through an interactive trial-and-error process, OpenMusic relies on a graphical language based on generic musical data structures, a large set of functions libraries elaborated by various composers for their own works, and graphical editors for various kinds of musical structures. OpenMusic has been used by numerous composers and is considered by many not only to be a computational tool for generating complex musical materials, but also as a reflexive and dynamic support for their musical thought [Agon et al. 2006]. The heuristic of functional programming it implements requires an analytical approach in order to identify the successive computational stages necessary for producing the expected materials. This programming mode has been completed with constraint-based programming that enables the specification of a problem through the expected properties of the result, especially in cases where a computation model cannot be easily formalized [Rueda et al. 2001, Truchet et al. 2004]. Constraint-based programming was applied to musically complex problems such as the paving of a time grid with a polyphony of different periodic sequences.

An important aspect of computer-aided composition tools is their connection to the signal world; both for the analysis of recorded audio materials into symbolic structures, and for sound synthesis from algorithmically generated symbolic structures. These functions, used in conjunction with sound analysis/synthesis models such as additive, phase vocoder, or physical models, are widely used by composers when elaborating orchestral materials to be performed with recorded sounds, or when producing synthetic sounds associated with instrumental material from the same musical structures.

The perspectives and current research issues in computer-aided composition are presented in the subsequent sections, including further developments of the connections between the signal and symbolic worlds, links between computer-aided composition and real-time environments for advanced interaction situations, and the

use of these environments for the analysis and formalization of the musical knowledge.

5 Signal/ Symbol: Cross-Level Approaches

The Signal/Symbol expression accounts for a broad class of current research themes that concern, in various types of applications, the relationships and simultaneous management of musical contents at different representation levels. It not only concerns relationships between the signal and symbolic levels through the control level, but also includes by extension, notably through recent advances in Music Information Retrieval, the semantic level.

5.1 Applications of Music Indexing

The development of the Music Information Retrieval (MIR) research community since 1999 replies to the irreversible development of music distribution via network-based services, through the structuring of a critical mass of participants, research themes, and methodologies. Although issues related to the mass distribution of popular music may seem far at a first glance from the concerns of contemporary music, IRCAM has been involved from the beginning in this subject that not only represents a new application field of its interdisciplinary research themes, but also a historical reconfiguration of the global technical system of music. This situation provides new opportunities for the dissemination of contemporary works using multimedia technologies and also for the emergence of new interactive and network-based forms. This latter aspect is further developed in sections 6.2.4 and 7.2.1. Another major interest in the technologies developed in the context of MIR is their potential applications far beyond commercial music distribution; music indexing and database technologies already represent the core foundations of new applications for sound synthesis and musical composition, such as concatenative synthesis and computer-aided orchestration. The term *Music Indexing* refers here to the general issue of the automatic analysis of signal and symbolic materials in order to infer related properties at higher levels. It encompasses the entire chain of computations from the extraction of low-level descriptors to the inference of high-level categories using machine-learning methods.

The first projects, to our knowledge, that experimented with search functions in sound databases from automatically computed content descriptors, were the MuscleFish sample browser [Keislar et al. 1999] and the Studio Online project, implemented at IRCAM between 1996 and 1998 [Ballet et al. 1999, Wöhrmann et al. 1999]. Both concerned the management of sound sample databases and were partially inspired from perceptual studies on timbre spaces [Grey 1977, McAdams 1999], that proposed a topology of elementary sounds in which the relative distances between them could potentially be used as a measure of perceptual dissimilarity. IRCAM was involved in the coordination of the first large-scale music indexing R&D projects carried out at a French national level with the ECRINS project [Mullon et al. 2002] and at a European level with the CUIDADO project (2001-2003) that systematically investigated all possible applications of the automatic extraction of musical contents from signal and symbolic materials [Vinet et al. 2002] and proposed inputs to the MPEG7 standard [ISO/IEC 2002, Peeters et al. 2000]. More recently, the SemanticHIFI European project⁹ was dedicated to designing future Hi-Fi systems

⁹ <http://shf.ircam.fr>

through the experimentation of advanced feature-based interfaces for the manipulation of musical recordings [Vinet 2005, 2006]. Numerous advances resulted from these studies, concerning the automatic extraction of various features, systems for the automatic classification of sounds, and the automatic extraction of the temporal structure of songs [Peeters 2005, 2006a, 2006b, 2007a, 2007b].

The community has invested considerable efforts in these issues. However, few of the existing results have yet been integrated in large-scale music distribution applications. This process, that may evolve rapidly, requires the active involvement of the right owners—including music Majors—in order, first, to test and assess the indexing methods at relevant scales in terms of number of items, but also to implement user feedback processes involving a sufficient number of subjects. The proposed features may not be limited to single-to-many distribution schemes but may include, in the spirit of Web 2.0 and the legalized evolution of peer-to-peer systems, the emergence of communities through personal production and sharing tools.

An important advance, whose general concept was introduced in the MPEG7 standard [ISO/IEC 2002], has been the idea of standardizing descriptions, in our case, of musical contents. Even if the actual impact of the existing standard, in terms of relevant and comprehensive descriptor structures, remains to be evaluated, at least this general idea of descriptors that accounts for specific musical features and can be automatically computed from the audio signals, provides a conceptual framework for potentially extending the vocabulary available to musicians and sound producers for specifying their materials in complement to the traditional symbolic categories. In order to elaborate relevant high-level descriptors for the targeted applications, a major current issue in MIR and, more generally, in music technology research, concerns the engineering of musical knowledge and in particular the methodologies for its elicitation and formalization. The current approach at IRCAM concerning this issue involving interdisciplinary coordination is presented in section 7.1. Our experience in different MIR-related projects shows that real-life music distribution applications involving multiple users require contextual (in particular relative to the subset corpus of items of interest for a given user), dynamic (through the evolution of the corpus over time), local scope (and potentially heterogeneous) high-level descriptions whose management addresses a number of interesting but delicate breakthroughs in the field of knowledge engineering.

Beyond the scope of electronic music distribution, new applications relying on the content-based management of multiple recorded sounds are currently being experimented with, as are new methods for sound synthesis and music composition. A first example is concatenative synthesis of musical sounds, that adapts methods for high-quality voice synthesis [Schwarz 2006, 2007]. A significant advantage of the related class of methods over modelling, provided that the original materials can be segmented into relevant units and that the expected manipulation is focused on concatenations, is that they retain all the complexity and related quality of sounds, whereas models necessarily imply a possibly audible simplification.

A second interesting case is the computer-aided orchestration project recently started at IRCAM, that provides a good example of knowledge elicitation and formalization into the development of an operational tool [Tardieu et al. 2007]. This knowledge of orchestration by composers has been mainly empirical and the existing written references are acknowledged to be quiet incomplete, whereas the underlying principles can be related to identified research topics concerning the acoustic rules of sounds combinations, the properties of perception, in terms of segregation/fusion of simultaneous sources and of similarity between sound aggregates, and the

formalization of instrumentation rules (i.e. the limits and constraints of the playability of existing instruments). The implementation of such a system also requires the handling of the issue of combinatorial search, that would necessitate for an exhaustive search to evaluate a number of combinations equal to the number of different samples (typically several tens of thousands) power the number of simultaneous sounds. The approach chosen consists of searching for the best candidates of combinations of sound samples in terms of perceptual similarity from the database that match a given target defined either by a sound file or by its symbolic description. Like with any kind of computer-aided tool, the goal is not to model the composer's knowledge in order to replace it with an automatic tool—which would be anyway technically unattainable—but to assist him/her with an interactive tool that models aspects of this knowledge and may provide new heuristics for discovering unusual sound combinations.

5.2 Audio Analysis/ Synthesis: Toward Higher-Level Control

The Analysis/Synthesis architecture of Figure 2 enables sound processing through the processing of the model parameters. However, the difficulty with this design is that these parameters may be difficult and too complex for the user to understand to be able to directly and independently manipulate it. A more relevant approach consists in providing the user with processing parameters at meaningful representation levels (i.e. the symbolic and the semantic levels) and implementing automatically the necessary analysis and synthesis functions for the conversion between the signal and higher-level representations. Therefore, a number of applications already exist that implement audio processing through a user-specification at the symbolic level per the scheme in Figure 3.

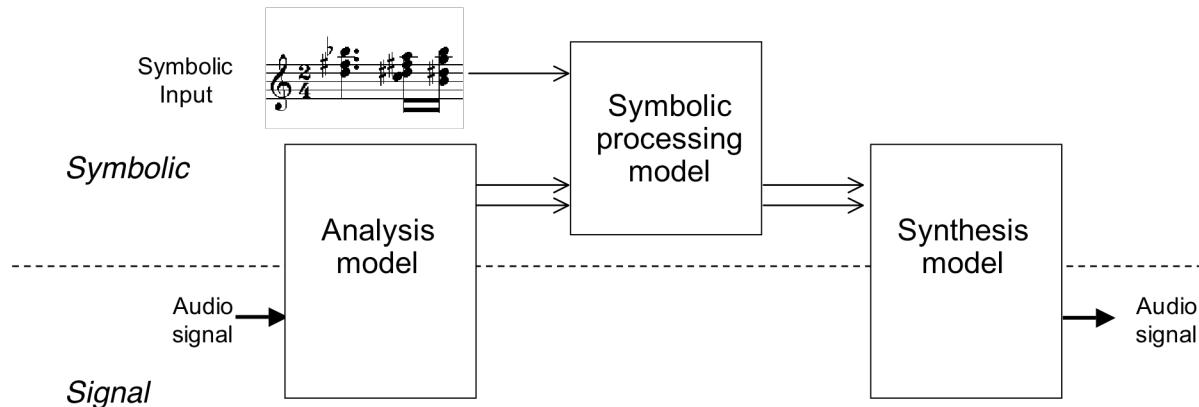


Figure 3 – Sound Analysis/ Synthesis Through Symbolic Specification

The GarageBand application by Apple¹⁰ makes it possible to integrate an existing music loop seamlessly into a musical sequence whose context is characterized by a tempo and a pitch scale: the loop, through its signal and symbolic data, is automatically stretched to the correct cycle duration and its individual notes transposed to match the pitch scale. Contrary to modular and programmable tools for composers, sound processing through symbolic specifications has been carried out for years by composers using a combination of the AudioSculpt and OpenMusic applications via a common symbolic data format called chord sequences, stored in SDIF format—the format for all audio analyses that is supported by all concerned

¹⁰ www.apple.com/ilife/garageband

IRCAM applications [Schwarz et al. 2000]. On the analysis side, an automatic segmentation of the signal is performed, creating segments with stable spectral properties. The spectral content of each segment is then reduced by detection of the spectral peaks into a discrete set of frequency values. This analysis performed in AudioSculpt provides chord sequences that can then be imported in OpenMusic that processes these data in the symbolic domain. The synthesis process back to the signal level can then be carried out using any kind of synthesizer with a symbolic control.

The refinement of possibilities of sound processing through the symbolic domain is dependent on the evolution of automatic audio analysis, in particular in the case of polyphonic sounds. When the symbolic data are available, efficient methods can be applied to align each note onset to the corresponding time in the audio file [Rodet et al. 2004, Kaprykowsky et al. 2006], which solves the issue of temporal segmentation and provides the pitch information from the score data [Dannenberg 2007]. Otherwise, the issue concerns the automatic transcription of polyphonic recordings, promising results already exist for a limited number of instruments in this domain [Yeh et al. 2005].

The next step in the high-level control of audio processing consists of enabling the specification of the sound result using perceptual features at the semantic level that may be entered either as absolute or relative target values. For instance, if a psychophysical bi-univocal relation exists between the perception of brightness and a feature like Spectral Centroid [Peeters et al. 2000]), the specification may be formulated as a target brightness value, or as a given percentage of increase in brightness. The issue is that an infinite number of possible solutions for this specification are possible since the feature represents a global characteristic. Actually, the same kinds of issues already exist with habitual processing functions. For instance, there are an unlimited number of possible choices for transposing a sound of a given interval, depending on the way its spectral envelope is affected. Once more, as explained in Figure 1, in the case of generation of information, a synthesis model of the feature effect—including several control parameters with outputs within the control level—must be designed for that purpose (Figure 4).

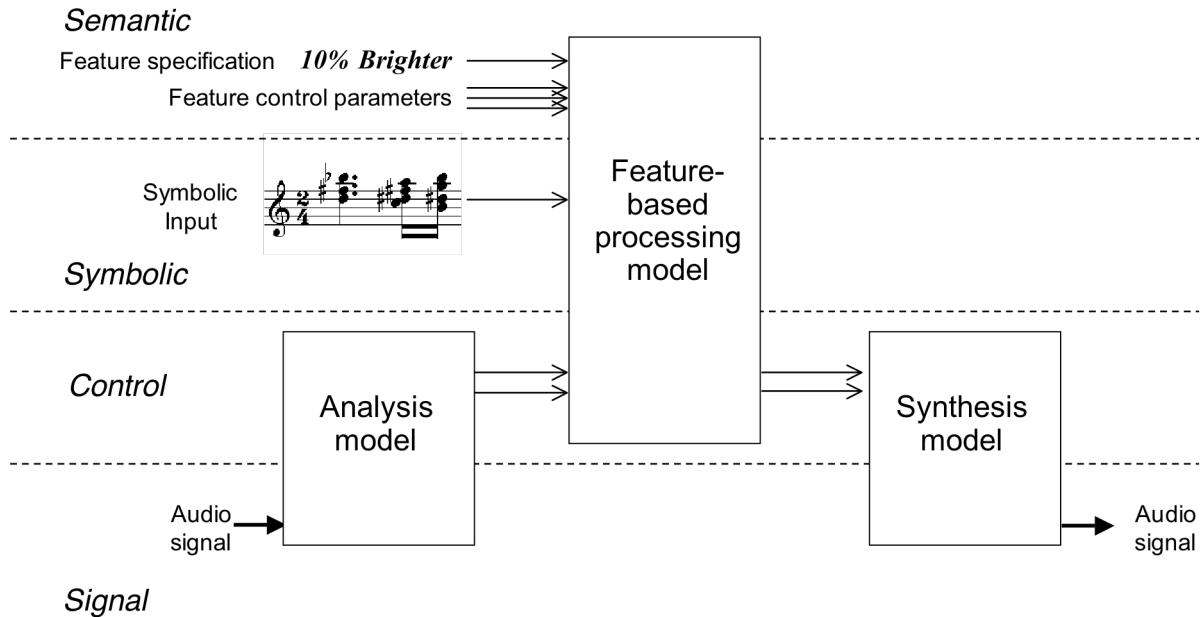


Figure 4 – Sound Analysis/ Synthesis Through Symbolic and Feature-Based Specification

One may notice that the connection between the processing model and the analysis and synthesis models needs to be carried out at the control level, i.e. using the full range of the analysis/synthesis model parametric space, and not only its simplification at the symbolic level like in Figure 3. The general issue of modelling the control information is the main subject of the next section.

5.3 Modelling the Control as the Main Place for Expressiveness

One difference between acoustic and electronic instruments, in terms of control, is that the latter present an energetic decoupling between the control gestures and the resulting sounds [Cadoz 2002]. This creates a level of freedom that opens up the possibility of various mapping strategies between the gestural data and the model input parameters [Wanderley et al. 2000], which corresponds to a MIMO (multiple input - multiple output) transfer function between multidimensional signals belonging to the control level. A first approach can consist in defining static transfer functions, i.e. that are not a function of time. However, in a number of cases a dynamic control may be necessary, including the case of physical models of instruments with sustained sounds, i.e. with an energy transfer over time. Considering the instrument as a dynamical system, methods from systems control such as inverse methods can be applied. This approach is currently being investigated in simple non-linear cases. Other methods implement machine learning for characterizing the variation of the input parameters from a set of recorded phrases including an array of playing and articulation modes [D'Haes 2004]. This simultaneous, time-varying control of several physical parameters is actually what a wind or bowed string instrument performer does in practice: once synthesis models approach the reality of the instruments dynamic behavior, their control requires the same kind of expertise as that of a performer.

However, the modelling of control signals is not necessary only in the case of complex instrument models, for producing realistic sounds playing target phrases in terms of note sequences, i.e. from a symbolic specification. The control signals also

capture the expressive aspects of the performance, including the precise timing of events on a continuous time axis (with a temporal resolution conformant to the control-related bandwidth), the continuous evolution of the various musical parameters, and the modulations of these parameters over time: the control data contain all aspects of the musical expressiveness whereas the related symbolic data are independent of any performance, if strictly considered as discrete values defined on a rhythmic time grid at a fixed tempo. This applies not only to sound synthesis, but also to analysis/synthesis sound processing architectures. In particular, coming back to Figure 4, the control data output from the analysis model includes the characteristics of the performance—if any exist—related to the input sound. Different strategies can then be considered for the management of these performance-related data in order to preserve them across changes at the symbolic level through demodulation and re-modulation functions, or to modify them according to processing functions that may make musical sense. Even if many studies already exist on performance analysis, there probably remains quite a lot of work to do in terms of eliciting rules of the performer’s action on sound in order to draw up usable performance models that would provide hints for controlling continuous variables such as the instantaneous tempo, the amplitude and phase of vibratos and tremolos, the amplitude curves for crescendi and decrescendi, or for articulations between successive notes according to the playing mode. The focus at IRCAM is centered on augmented instruments, i.e. acoustic instruments augmented with sensors, whose control capitalizes on the richness of the experiences of generations of performers [Bevilacqua et al. 2006, Rasamimanana et al. 2006]. The first stage consists of analyzing and characterizing the gestural signals produced by sensors in a given musical situation through segmentation and classification before identifying rules of mapping them to sound synthesis, processing, or interactive situations. Applications exist not only for music production, but also for technology-enhanced learning of music performance [Bevilacqua et al. 2007].

A parallel can be drawn with the expressive control of voice synthesis and processing. The symbolic level corresponds to phonemes, and phone related data can be automatically extracted from the signal level. The expressiveness is contained in the fine timing of the phonemes and the variation of a number of different parameters at the control level such as the fundamental frequency, the spectral envelope coefficients, the voiced/unvoiced frequency threshold. The goal of this line of research in voice—like with musical signals—is to provide various kinds of high-quality synthesis and transformation functions, that maintain the natural quality of the voice: changing the type or speaker (e.g. from a man to a woman or a child), changing the identity (to an existing voice), modifying various qualities (breeze, roughness, intonation, etc.) or the expressive content specified by “emotional” descriptors at a semantic level [Beller 2007].

6 New Concepts for Time and Interaction

The main software environments for music production and real-time interaction have relied on the same concepts of time and interaction for the last two decades, i.e. since the first design of Max [Puckette 1996]: linear time, event-based interaction. The question of the possible extension and renewal of these concepts in order to better answer to new artistic and production needs arises. Naturally, the temporal inscription of musical forms can be defined through any kind of abstract structure, and several examples of such formalisms can be found in contemporary works. However, the question addressed here concerns more specifically the way time and

interaction are explicitly managed in the design of software applications for music production, and by extension for the production of all kinds of temporal or interactive forms. A number of ongoing experiments, feedback from artists' experiences, and new artistic projects featuring, for instance, interactive systems associating dance and music, illustrate the limits of the existing tools and indicate lines of future research investigation. The purpose of this section is to propose a synthesis of these different approaches.

6.1 Representations of Temporal Structures

Most representations in media production software rely on a linear notion of time, that is expressed in graphical interfaces as a spatial axis, and in numerical representations as a real value. This linear representation, that applies the Cartesian concepts of space to time, is convenient for conceptualizing, operating, and objectively accounting for the inscription of temporal processes in usual technical media. However, its justification from the viewpoint of human cognition is not straightforward. The representation of temporal structures, as formalized in the western music theory, in the form of hierarchies of cycles related by integral subdivisions and multiples, seems to better account for some aspects of the human embodiment of musical performance. Rhythmic patterns can be fully represented digitally through a structure of integers that specify the occurrences, durations of the events, and the subdivision values at each cycle level, with possible operations such as addition and subtraction between rhythmic structures [Agon et al. 2002]. This rhythmic representation and the linear one, that would define each musical event through its start and end times as real values, contain mutually exclusive information. The question of the compatibility between these structures arises noticeably for the design of graphical representations in man-machine interfaces. Partial solutions exist such as the proportional notation, that uses a linear axis and replaces the note duration values with a segment displaying the note time interval span. In response to requests by composers for having notation editors combining linear and rhythmic representations, new interface designs are being developed [Agon et al. 2007].

Coming back to the issue of time cognition, exchanges with composers have made the lack of cognitive studies on rhythm obvious—in particular studies that would characterize the respective roles and relationships of musical dimensions contributing to a global rhythmic percept, including pitch (repeated phrases), intensity (accents), timbre (timbre melodies), and spatialisation (“spatial phrases”). Moreover, while rhythmical structures capture the symbolic aspects of performance they do not account for the continuous aspects that fall within the scope of the control level. The former paragraph stressed the importance of an improvement in performance rule knowledge, in particular for the control of the expressiveness of processed or synthesized sounds. In the broader perspective of human cognition of musical time, there is also a need for a better understanding of how temporal shapes at the motive or phrase scales are relevant components of how we structure time, or even contribute to a kind of internalization, at a motor level, if this makes any sense, of musical forms. Unlike music theories already formalized concerning the symbolic level, this aspect of musical knowledge, that musicians learn generally empirically through oral transmission and through their own practice, needs to be elicited and its understanding may lead to new formalisms of temporal structures at the control level [Leman et al. 2006].

6.2 Beyond Max-Like, Event-Based Interaction

The interaction paradigm in Max is event-based, i.e. computational processes are launched upon the reception and processing of messages that may be issued by various sources. This asynchronous information transfer mode from distributed sources already makes many situations of interaction possible but it presents limitations in terms of the management of temporal processes. A temporal process can be launched through a succession of messages or through the execution of a pre-defined cue list of events associated to defined time occurrences, but in both cases the possible control of the processes is limited. This event-based interaction mode is also inadapted for the management of parallel evolving processes that may exchange data or that are mutually synchronized. The following examples illustrate different concrete cases of approaches that aim to overcome these limits.

6.2.1 From Score to Multimodal Performance Following

Score following techniques are dedicated to the real-time detection of defined musical events during a performance on the basis of the knowledge of the score played by the performer and of the real-time analysis of performance data including the audio signal or the MIDI data of the instrument. The existing implementations provide accurate note/chord/trill detections within rapid phrases and have been used in live performance for various interactive pieces [Orio et al. 2003, Schwarz et al. 2004] but they still present the limit of conceiving interaction as a detection of note events. Recent developments enable the use of a learned performance as a reference instead of the score, so that it fits better with a given performer [Schwarz et al. 2005, Cont et al. 2005, 2006]. Moreover, getting rid of the score reference makes it possible to detect musical shapes other than note events. Recent works also enable not only the detection of events but also the following of gestures characterized by a vector of control signals issued from sensors or image analysis. From a recorded and learned reference gesture (or set of gestures) the system identifies in real-time which one is the most likely gesture to be produced and aligns the actual performance time to the related time to that of the found reference [Bevilacqua et al. 2007]. This technique has many potential applications for virtual instruments and interactive dance but also for feature-based interaction: the processed control vector can include any kind of control signals, including the real-time analysis of a music descriptor. This creates the potential of interactive systems based on new kinds of vocabularies of multimodal temporal shapes combining music descriptors beyond the traditional pitch/intensity categories, signals from gesture capture, and descriptors extracted from video analysis.

6.2.2 Generative Systems based on Symbol Sequences

The former systems are based on Hidden Markov Models (HMM), i.e. the modelling of temporal shapes as a succession of states with statistical rules of transition. However, this temporal digitization is hidden to the user. On the contrary, in sequence-based systems, the modelling relies on the segmentation of the temporal data flux at a bigger grain into a sequence of symbols. Transition laws between symbols can be modelled using various formalisms, including incremental parsing and generating (IPG), prediction suffix trees (PST), or Oracle Factors [Dubnov et al. 2003, 2007]. The modelling of transitions can then be used in generative systems for synthesizing new sequences with the same laws as the ones analyzed and therefore reproducing some aspects of the performer's 'style'. These techniques are applied to the realization of interactive generative systems for musical improvisation, such as

the OMax system that associates the Max/MSP and OpenMusic environments [Assayag et al. 2006]. These systems overcome the event-based, short-term interaction in Max by managing a structured, long-term memory of the performance through a parallel computation process running in the OpenMusic environment that communicates with Max in input for acquisition of the performance data and in output for the rendering of the computed sequences. It should also be noted that they are limited neither to the detection of symbolic events such as MIDI notes, nor to musical sequences. Just as in the previously described temporal shape following approach, these systems can potentially operate on any kind of multimodal data vector, provided that automatic temporal segmentation strategies can be implemented, and that a library of symbols can be computed from the segmented data. An interesting audiovisual experiment with the OMax system has been to analyze and generate video sequences of recorded music where only the sound track was used for the sequence segmentation and modelling.

6.2.3 From Absolute References to Temporal Constraints

In many applications managing temporal media, like sequencers, the timing of the events is defined relative to an absolute reference. This is consistent when this timing is known in advance (e.g. in the case of pre-recorded media). However, this may not be true in a number of cases, including interactive situations where events occur at an unpredictable time, or with sequencers of processes, whose duration depends on parameters not yet instantiated. In these cases, the user may want to ensure a certain level of global consistency by specifying the interaction situation through temporal constraints between events, such as Allen laws [Allen 1981]. Providing this possibility is the objective of several ongoing projects, based on various formal models of temporal constraints and resolution methods in real-time, including Petri Nets and Constraint Concurrent Programming [Allombert et al. 2006, Donikian et al. 2007]. Applications are not only envisioned for interactive music, but also for live performance in general (e.g. dance, lighting, VJ), virtual reality, games, and other situations involving the production of temporal media.

6.2.4 Toward New Interactive Musical Forms

The development of the game industry has shown that new interactive forms can emerge and develop at a broad scale, provided that a set of technical and marketing conditions are met, in particular for the standardization of user interfaces. In the music domain, the dissemination of interactive forms has been restricted to limited audiences interested in live media interaction and interactive installations until recently. The development of musical forms on a larger scale of diffusion has been limited for decades to stereo recordings via physical distribution supports in analog, digital, or compressed forms. The current reconfiguration of the music distribution system employing digital networks makes it possible, for the first time, to overcome a number of limitations of the exchange formats toward rich data sets that could include multi-channel audio data, symbolic data, and metadata and whose transmission could be scaled down to match the capacities of client terminals. The availability of rich musical data on client terminals could yield new interactive interfaces for browsing within the musical material as demonstrated in the SemanticHIFI project [Vinet 2006] or for collaborative distributed performance or production. These interfaces may feature a form of fusion of the functions of listening devices and electronic instruments that have traditionally been separated in the consumer electronic market. One can then consider the possible emergence of new interactive

musical forms that will be specifically composed for these interfaces, in particular through the implementation of interactive scenarios that manage the interaction at several temporal scales, including the high-level evolution of the piece and the immediate, instrumental user interaction. These new possibilities may represent a radical conceptual shift in composing time. IRCAM, as an institution dedicated to contemporary creation, must anticipate these potential evolutions by soliciting artistic experiments in that field.

7 Elicitation and Mediation of Musical Knowledge

The preceding sections have exhibited convergent needs of formalizing aspects of the musical knowledge involved in composition and performance. This formalization process may imply, in function of the available sources of information and of the pursued goals, various methodological frameworks supported by different scientific communities. A discussion on some of the main approaches can be found in the first part of this section. The second part presents several lines of action implementing different forms of mediation of the knowledge produced that is part of IRCAM's current priorities

7.1 Elicitation of Musical Knowledge

Existing music theories are far from extensively accounting for the musical knowledge involved in composition and performance. A scientific approach to the reality of music, that should be, at least etymologically the role assigned to musicology, implies to elicit aspects of this knowledge in given contexts and formalise them objectively. This issue can be addressed from different scientific perspectives. The approach taken at IRCAM, as an interdisciplinary interface lab, is to develop a clear view of the possibilities and limits of the methodological frameworks associated to each concerned scientific field and pragmatically combine them according to what each of them can bring in particular. The following sections propose some elements for reflection on this open subject of epistemological nature that concerns the different scientific communities involved in musical research.

7.1.1 Cognitive Psychology

Experimental psychology is the dominant methodological paradigm in cognitive sciences and the basis of research in psychoacoustics. A noticeable trend is that beyond perception studies, in the framework of the objective of the elicitation of performers' musical knowledge and also of studies in intermodal spatial cognition there is a growing interest in studying subjects involved in action/perception tasks, especially through their manipulation of interactive artifacts.

However, one may wonder about the limits of the complexity of situations that can be studied within experimental constraints: when using complex stimuli, like excerpts of contemporary music, the study's subject must focus on specific features in order to be manageable experimentally, with a risk of building an oversimplified model of the actual cognitive processes involved in the task.

Leaving other aspects related to the social situation of public performance that cannot be taken into account in laboratory experiments aside, a second important criterion to take into account is one of statistical validity: there are many features of human cognition, especially in the artistic field, that relate to individual singularities and that cannot *a priori* be handled by statistical methods. For these specific

purposes other methodological approaches, such as the ones presented in the next paragraph, may be better adapted.

7.1.2 Directions for a Contemporary Musicology

Musicology has traditionally been a science of texts—scores, pieces of writing by composers. New approaches, in particular at IRCAM, aim at extending its research topics to the analysis of musical practices, including composition, but also performance and listening. Methods developed in the field of cognitive anthropology, that aim to understand human activities through the study of their temporal process, are being applied to the analysis of the process of the musical composition of specific works [Theureau 2006a, Theureau et al. 2006b, Donin et al. 2005a, Donin 2006a, Donin et al. 2007]. The protocols followed include interviews with composers and/or the presence of researchers during the composition process in the artists' studio. The first experience, a complete genetic analysis of two works by composer Philippe Leroux, *Voi(rex)*, which premiered in 2003 and *Apocalypsis*, which premiered in 2006, has already been carried out and is documented through hypermedia supports [Donin 2006b]. Even if their scope is limited to singular works and subjects due to the time investment they require, these empirical methods make it possible to have an in-depth understanding of the works' interiors and of the underlying motivations and models used by the composer. Moreover, from an engineering perspective, the elicitation of the compositional process, including its technical environment, may provide software designers with useful information on the ways their applications are effectively employed.

This last remark leads to a second identified direction of the evolution of contemporary musicology toward a better integration of the advances of technology in its subjects and methods. First, a significant part of contemporary works written in the last decades cannot be studied in depth without a detailed understanding of the underlying technologies. Second, in terms of tools, recent results in automatic music indexing, operating from the signal and/or symbolic levels, can be applied to the context of musical analysis: they may help to highlight relationships between musical materials that would be too complex to obtain through manual operations. For instance, an application in performance analysis has compared several performances of a work through the analysis of precise timings of the notes obtained via recently developed methods for automatic score/signal alignment [Rodet et al. 2004, Donin 2005a]. Third, hypermedia technologies represent a powerful medium for the representation and mediation of musical analyses that can be used for communicating research results to the scientific community or as a dissemination support intended for a broader target audience, as developed in a latter section.

7.1.3 Computer-Aided Composition and Analysis

Computer-aided composition environments provide composers with computational tools and representations for formalizing their musical thoughts. Considered from an elicitation standpoint, the formalisms produced in this context represent the ideal support for understanding the conceptual bases of the production of musical materials (e.g. scores, sound files) present in works. These materials appear as instances of results of the computation processes implemented with a specific set of input parameters and other instances can be computed following the same model and be compared to the one chosen. In other cases, these environments can be used as computer-aided analysis tools for elaborating computational models of the works that will approach the actual materials. This is the case of works by composers

like Iannis Xenakis that were produced through a formal process documented by the composer and from which a model can be built using the formalisms of current computer environments [Agon et al. 2004]. Other approaches in the field of automatic indexing from the symbolic level are not limited to works relying on a formal model, but aim to provide a systematic analysis of the regularities found in the musical material [Lartillot 2005].

The field of computational musicology has developed in the last years, in particular through the federation of various international participants that are involved in mathematical formalisms for music through several inaugural events, including an international symposium on Set Theory held at IRCAM in 2003 [Andreatta et al. 2007], the foundation of the *Society for Mathematics and Computation in Music*, the launching of the *Journal of Mathematics and Music* in 2006, and the organization of the first edition of the Mathematics and Computation in Music international conference in May 2007 (MCM).

7.1.4 Inference Techniques and Knowledge Engineering

The use of inference techniques, including statistical analysis and machine learning techniques, has been generalized to many research approaches in the field of engineering such as music indexing and audio signal analysis/synthesis. One may understand this evolution as the necessity of complementing modelling approaches with information extracted from massive quantities of real-world data in order to account for phenomena of increasing complexity. In the specific case of supervised learning techniques, such as Bayesian networks, the goal is to compute relationships between automatically extracted data (bottom-up) and predefined information (top-down). One interesting aspect is that this predefined information can be specified not only on the basis of objective data, such as instrument families, but also of arbitrary user-defined data. An application of inference techniques applied to user-defined data was experimented in the framework of the SemanticHIFI project [Pachet et al. 2005]. The users could not only label their songs using the editorial metadata provided by external sources; they could also define their own classifications by providing prototypic examples for each class. The system then learned the characteristics of each class from the automatic analysis of the signals of the provided songs. This example illustrates the interest of integrating inference techniques in applications for enabling users to specify their own knowledge organization. In the context of computer-aided composition, these techniques may extend features of computer applications as reflexive supports of the user's musical thought (§4.4).

7.2 Mediation of Musical Knowledge

This section presents various forms of mediation of musical knowledge centred on different objectives and targets: dissemination of contemporary works through hypermedia supports and software applications for music education, dissemination and preservation of digital interactive works to enable their distant and future reproducibility, and sound design as an application of sound production in different application areas beyond the realm of artistic creation.

7.2.1 Cultural Dissemination and Musical Education

The dissemination of existing knowledge on contemporary works is an important aspect of a more global policy of dissemination of contemporary culture in society. Concerts and to a larger extent, scores, are accessible only to a minority and the only

diffusion supports available to a broader audience are audio recordings and books. Digital technologies offer appropriate media for documenting works and also for supporting new approaches in musical education.

A current objective at IRCAM, related to the research in contemporary musicology, is the design of hypermedia interfaces dedicated to the presentation of analyses of works. These interfaces rely on interactive scenarios integrating various synchronized media including audio recordings, the score display with animated annotations, other animated graphical structures, texts with hypermedia links to musical excerpts and other textual resources. This design has been assessed for the documentation of several works and its use is being generalized in the form of an authoring tool with generic document templates to produce publication supports for different media such as the Web and DVD-ROMs. This generic tool can then be used for producing series of hypermedia documentation on different corpuses of works using the same interface templates.

In the field of musical education, software applications are being designed in collaboration with teachers as a support for music courses in junior and senior high schools in the framework of the Musique Lab project [Puig 2005]. These software applications propose simple interfaces featuring various manipulations on the music material. Unlike traditional methods relying on music theory and listening, these applications support a hands-on approach to musical education, letting students progressively understand musical notions through their own practice and manipulation of musical materials.

7.2.2 Dissemination and Preservation of Digital Interactive Works

A number of contemporary musical works, especially those involving interactive situations, do not rely only on simple material supports like scores and audio files but on complex sets of data and programs executed during the performance. The reproduction of their performance requires, in addition to technical documentation describing the operations performed, the availability of a number of data in a structured form. This is the function of the MUSTICA database at IRCAM. This database is used as the production repository for all works produced and provides concert producers with remote access to these elements through a web-based interface¹¹. However, the programs used may no longer run on future operating systems so the preservation of these works is not only an issue of structured organization and long-term data storage but it also concerns the necessary processes of the adaptation of these data to the developments of technical environments. These issues are being dealt with in as a part of the CASPAR European project dedicated to the preservation of the European digital heritage through the implementation of the OAIS preservation standard and its validation in several application testbeds, including interactive performing arts¹². The main directions of investigation for the preservation of musical works concern the development of tools for facilitating the porting of Max/MSP patches and providing an abstract characterization of the operations involved in the performance of a piece independently of a given implementation [Bonardi et al. 2007, Guercio et al. 2007].

¹¹ <http://mustica.ircam.fr/>

¹² <http://www.casparpreserves.eu/>

7.2.3 Sound Design

Sound design is another unique case of the application of aspects of the knowledge developed in the context of musical research in other areas of activity in society. It concerns the controlled production of sounds that are intended to convey information in a given context such as alarms, automobile horns, vocal announcements, multimodal human-computer interfaces, or vocal servers. The design process includes phases of specification, of production of the sounds, and of their evaluation in terms of performances and/or preferences through laboratory experiments. The choices made during the production process must rely, insofar as possible, on objective criteria such as the characterization of the different types of sounds, of their subjective effects according to their sound characteristics, or of ‘instrumental’ models of the typical sound variations of objects of the environment in response to a user action. These issues are themes of research that aim to develop dedicated tools and methodologies for sound design, in particular in the framework of the CLOSED European project¹³.

8 Conclusion

The main orientations of the roadmap for IRCAM's research in the field of science and technology for music and sound have been presented. The first direction concerns extensions of several complementary techniques involved in music composition and sound production such as audio signal synthesis and processing, synthesis by physical modelling, sound spatialisation, and computer-aided composition. Special attention has been paid to the renewal of technical concepts for time and interaction that is essential in order to meet new artistic expectations, not only in the fields of music and performing arts, but also for temporal and interactive media in general. The concept of Representation Levels of Musical Information that has been introduced provides a unifying analysis grid for musical applications. It highlights a global need for the transversal themes addressed in the other two major orientations in this roadmap. The first theme, Signal/Symbol, concerns issues of conversion between different representation levels, that are encountered in several current research projects related especially to music indexing and to the high-level and expressive control of audio synthesis and processing. A particular issue that emerges from different convergent approaches to gesture analysis, parameter identification of synthesis models, performance modelling, expressive control of sound synthesis, and representation and cognition of multimodal temporal shapes is the understanding and the modelling of the control level. The second theme concerns the interdisciplinary issue of the elicitation of musical knowledge in its various aspects and of the means of mediation of this knowledge to various targets.

Contemporary music creation is not only a place for the inception of the culture of tomorrow. For decades, it has been a privileged framework for the development of interdisciplinary scientific and technological advances. Artistic creation is a demanding process that systematically tends to overcome or divert the limits—both conceptual and material—of the existing techniques, while developing a heightened sensitivity to the quality of the material produced. In this context, small groups of researchers, engineers, and artists have a great potential for innovation that may be initially targeted to a given artistic production, but that afterwards may be generalized and applied to other fields of activity. Many of today's advanced audio-processing

¹³ <http://closed.ircam.fr>

functions found in sound and music production studios were actually designed years earlier by engineers working in close collaboration with composers. Beyond professional applications, this transfer from artistic experimentation now has the potential to be extended under different forms to broader societal needs. Applications for musical education are being developed on the basis of functionalities of software applications initially developed for composers. New hypermedia forms are being designed as supports for the dissemination of musical works. The sounds of objects present in our daily environment are being specified, composed, and assessed using tools and methodologies that are the results of musical research. The current reconfiguration of the music distribution system toward network-based services opens up opportunities for new intelligent listening devices that may provide end-users with the possibility of interactive manipulation of musical materials and may favour the emergence of other new interactive forms. Musical research, for its both cultural and technological impact, is an inescapable component of tomorrow's cultural industries.

References

- Adrien, J.M. (1991). The Missing Link: Modal Synthesis. *Representations of Musical Signals*, G. De Poli, A. Picalli & C. Roads Eds, MIT Press.
- Agon, C., Haddad, K., & Assayag, G. (2002). Representation and rendering of rhythmic structures. Proceedings from WedelMusic 2002: *the International Conference on Web Delivering of Music*, IEEE Press, Florence.
- Agon, C., Andreatta, M., Assayag, G., & Schaub, S. (2004). Formal Aspects of Iannis Xenakis' "Symbolic Music": A Computer-Aided Exploration of Compositional Processes. *Journal of New Music Research*, 33(2), 145-159.
- Agon, C., Assayag, G., & Bresson, J. Eds (2006). *The OM Composer's Book*, 1, Collection Musique/Sciences, Delatour/IRCAM – France.
- Agon, C., & Bresson, J. (2007). Mixing time representations in a programmable score editor. Proceedings from SMC'07: *Sound and Music Computing Conference*, Lefkada, Greece, 2007.
- Allen, J. (1981). An interval based representation of temporal knowledge. Proceedings from the *seventh International Joint Conference on Artificial Intelligence*, 221–226.
- Alloumert, A., Assayag, G., Desainte-Catherine, M., & Rueda, C. (2006). Concurrent constraints models for interactive scores. Proceedings from SMC'06: *Sound and Music Computing Conference*, Marseille.
- Andreatta, M., Bardez, J.M., & Rahn, J. Eds (2007). *Around Set Theory. A French/American Musicological Meeting*. Delatour-IRCAM.
- Assayag, G., Rueda, C., Laurson, M., Agon, C., & Delerue, O. (1999). Computer-Assisted Composition at IRCAM : From PatchWork to OpenMusic. *Computer Music Journal*, 1999, 23(3), 59-72.
- Assayag, G., Bloch, G., & Chemillier, M. (2006). OMax-Ofon. Proceedings from SMC'06: *Sound and Music Computing Conference*, Marseille.
- Ballet G., R. Borghesi, P. Hoffmann, & F. Lévy (1999). Studio Online 3.0: An Internet "killer application" for remote access to IRCAM sounds and processing tools. Proceedings from JIM'99: *Journées d'Informatique Musicale*.
- Beller G. (2007). Content based transformation of expressivity in speech. Proceedings from *International Congress of Phonetic Sciences*, Saarbrucken, Germany.
- Bensoam, J. (2003a). *Représentation intégrale appliquée à la synthèse sonore par modélisation physique*. PhD Thesis, Université du Maine, Le Mans, Académie de Nantes.
- Bensoam, J., Caussé, R., Vergez, C., Misdariis, N., & Ellis, N. (2003b). Sound synthesis for three-dimensional object: dynamic contact between two arbitrary elastic bodies. Proceedings from SAMC'03: *Stockholm Music Acoustics Conference*, 1,369-372.
- Berkhout A. J. (1988). A Holographic Approach to Acoustic Control. *Journal of the Audio Engineering Society*, 36(12).
- Bevilacqua, F., Rasamimanana, N., Fléty, E., Lemouton, S., & Baschet, F. (2006). The augmented violin project: research, composition and performance report. Proceedings from NIME'06: *International Conference on New Interfaces for Musical Expression*, Paris, 402-406.

- Bevilacqua, F., Guédy, F., Fléty, E., Leroy, N., & Schnell, N. (2007). Wireless sensor interface and gesture-follower for music pedagogy. Proceedings from NIME'07: *International Conference on New Interfaces for Musical Expression*.
- Bogaards, N., Roebel, A., & Rodet, X. (2004). Sound analysis and processing with AudioSculpt 2. Proceedings from ICMC'04: *International Computer Music Conference*, Miami.
- Bonardi, A., & Barthélémy, J. (2007). Le patch comme document numérique : support de création et de constitution de connaissances pour les arts de la performance. Proceedings from CIDE 10: *10ème Colloque International sur le Document Electronique*, Nancy, 163-174.
- Bresson, J., Stroppa, M., & Agon, C. (2007). Generation and representation of data and events for the control of sound synthesis. Proceedings from SMC'07: *Sound and Music Computing Conference*, Lefkada, Greece.
- Cadoz, C. (2002). Continuum énergétique du geste au son, simulation multisensorielle interactive d'objets physiques. *Interfaces homme-machine et création musicale*, H. Vinet & F. Delalande Eds, Hermès Science, Paris, 165-181.
- Cont, A., Schwarz, D., & Schnell, N. (2005). Training Ircam's score follower. Proceedings from ICASSP'05: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Philadelphia.
- Cont, A. (2006). Realtime audio to score alignment for polyphonic music instruments using sparse non-negative constraints and hierarchical HMMs. Proceedings from ICASSP'06: *IEEE International Conference in Acoustics and Speech Signal Processing*, Toulouse.
- Corteel, E. (2006). Equalization in Extended Area Using Multichannel Inversion and Wave Field Synthesis. *Journal of the audio Engineering Society*, 54(12).
- Corteel, E. (2007). Synthesis of Directional Sources Using Wave Field Synthesis, Possibilities and Limitations. *EURASIP Journal on Advances in Signal Processing, Special issue on Spatial Sound and Virtual Acoustics*, 2007, article ID 90509.
- D'Haes, W. (2004). *Automatic Estimation of Control Parameters for Musical Synthesis*. PhD Thesis, Université d'Anvers, Belgium.
- Dannenberg, R. (2007). An intelligent multitrack editor. Proceedings from ICMC'07: *International Computer Music Conference*, Copenhagen.
- Dolson, M. & Laroche, J. (1999). Improved Phase Vocoder Time-Scale Modification of Audio. *IEEE Transactions on Speech and Audio Processing*, 7(3).
- Donikian, S., Delerue, O., Clenet, G., & Bianco, T. (2007). Towards a complete open source software environment for the creation of interactive poly-artistic works. Proceedings from ICMC'07: *International Computer Music Conference*, Copenhagen.
- Donin, N. (2004). Towards Organised Listening: Some Aspects of the "Signed Listening" Project, Ircam. *Organised Sound*, 9(1), 99-108.
- Donin, N. (2005a). Problèmes d'analyse de l'interprétation. Un essai de comparaison assistée par ordinateur d'enregistrements du premier prélude du Clavier bien tempéré. *Musurgia, Analyse et Pratique Musicales*, 12(4), 19-43.
- Donin, N. (2006a). Pour une "écoute informée" de la musique contemporaine : quelques travaux récents. *Circuit, Musiques contemporaines*, 16(3), Presse de l'Université de Montréal, 51-64.
- Donin, N., & Theureau, J. (2005b). Music composition in the wild: from the horizon of creative cognition to the time & situation of inquiry. Proceedings from EACE'05: *Annual Conference of the European Association of Cognitive Ergonomics*, Chania, 57-64.
- Donin, N., Goldszmidt, S., & Theureau, J. (2006b). De Voi(rex) à Apocalypsis, fragments d'une genèse. Exploration multimédia du travail de composition de Philippe Leroux, *L'inouï*, 2, IRCAM, Paris.
- Donin, N., & Theureau, J. (2007). Theoretical and Methodological Issues Related to Long Term Creative Cognition : the Case of Musical Composition. *Cognition Technology & Work*, 9(4).
- Dubnov, S., Assayag, G., Lartillot, O., & Bejerano, G. (2003). Using Machine-Learning Methods for Musical Style Modelling. *IEEE Computer*, 10(38), 73-80.
- Dubnov, S., Cont, A., & Assayag, G. (2007). Audio Oracle: A new algorithm for fast learning of audio structures. Proceedings from ICMC'07: *International Computer Music Conference*, Copenhagen.
- Ellis, N., Bensoam, J., & Caussé, R. (2005). Modalys demonstration. Proceedings from ICMC'05: *International Computer Music Conference*, Barcelona, 101-102.
- Grey, J.M. (1977). Multidimensional Perceptual Scaling of Musical Timbres. *Journal of the Acoustical Society of America*, 61, 1270-1277.
- Guercio, M., Barthélémy, J., & Bonardi, A. (2007). Authenticity issue in performing arts using live electronics. Proceedings from SMC'07: *Sound and Music Computing Conference*, Lefkada, 226-229.

- Hélie, T., & Hasler, M. (2004). Volterra Series for Solving Weakly Nonlinear Partial Differential Equations: Application to the Burgers Equation with Visco-Thermal Losses. *International Journal of Control*, 77(12), 1071-1082.
- Hélie, T., & Matignon, D. (2006). Diffusive Representations for Analyzing and Simulating Flared Acoustic Pipes with Visco-thermal Losses. *Mathematical Models and Methods in Applied Sciences*, 16(4), 503-536.
- Hélie, T., Matignon, D., & Mignot, R. (2007a). Criterion Design for Optimizing Low-Cost Approximations of Infinite-Dimensional Systems: Towards Efficient Real-Time Simulation. *International Journal of Tomography and Statistics*, 7, 13-18.
- Hélie, T., & Laroche, B. (2007b). On the Convergence of Volterra Series of Finite Dimensional Quadratic MIMO systems. *International Journal of Control*, in press.
- Iovino, F., Caussé, R. & Dudas, R. (1997). Recent work around Modalys and modal synthesis. Proceedings from ICMC'97: *International Computer Music Conference*.
- ISO/IEC 14496-1:2000. (2000) MPEG-4 Systems standard, 2nd Edition.
- ISO/IEC 15938-4:2002 (2002), Multimedia content description interface — Part 4: Audio.
- Jot, J.M. (1999). Real-Time Spatial Processing of Sounds for Music, Multimedia and Interactive Human-Computer Interfaces, *Multimedia Systems*, 7(1), 55-69.
- Kaprykowsky, H., & Rodet, X. (2006). Globally optimal short-time dynamic time warping application to score to audio alignment. Proceedings from ICASSP'06: *IEEE International Conference on Acoustics, Speech and Signal Processing*, Toulouse.
- Keislar, D., T. Blum, T., J. Wheaton, & E. Wold (1999). A content-ware sound browser. Proceedings from ICMC'99: *International Computer Music Conference*, Beijing.
- Lartillot, O. (2005). Multi-Dimensional Motivic Pattern Extraction Founded on Adaptive Redundancy Filtering. *Journal of New Music Research*, 34(4), 361-374.
- Leman, M., & Camurri, A (2006). Understanding Musical Expressiveness Using Multimedia Platforms. *Musicae Scientiae*, Special issue 2005-2006, ESCOM, 209-229.
- McAdams, S. (1999). Perspectives on the Contribution of Timbre to Musical Structure. *Computer Music Journal*, 23(2), 96-113.
- Mullon, P., Geslin, Y., & Jacob, M. (2002). Ecrins: an audio-content description environment for sound samples. Proceedings from ICMC'02: *International Computer Music Conference*, Göteborg.
- Orio, N., Lemouton, S., Schwarz, D., & Schnell, N. (2003). Score following: state of the art and new developments. Proceedings from NIME'03: *International Conference on New Interfaces for Musical Expression*, Montreal.
- Pachet, F., La Burthe, A., Aucouturier, J.J., & Beurive, A (2005). Editorial Metadata in Electronic Music Distribution Systems: Between Universalism and Isolationism. *Journal of New Music Research*, 34(2), 173-184.
- Puckette, M. (1986). The Patcher. Proceedings from ICMC'86: *International Computer Music Conference*, 420-425.
- Peeters, G. (2005). Rhythm classification using spectral rhythm patterns. Proceedings from ISMIR'05: *International Conference on Music Information Retrieval*, London.
- Peeters, G. (2006a). Template-Based Estimation of Time-Varying Tempo. *EURASIP Journal on Advances in Signal Processing*, Special Issue on Music Information Retrieval Based on Signal Processing.
- Peeters, G. (2006b). Chroma-based estimation of musical key from audio-signal analysis. Proceedings from ISMIR'06: *International Conference on Music Information Retrieval*, Victoria.
- Peeters, G. (2007a). A generic system for audio indexing: application to speech/ music segmentation and music genre. Proceedings from DAFX'07: *International Conference on Digital Effects*, Bordeaux.
- Peeters, G. (2007b). Sequence representation of music structure using higher-order similarity matrix and maximum-likelihood approach. Proceedings from ISMIR'07: *International Conference on Music Information Retrieval*, Vienna.
- Peeters, G., & Rodet, X. (1999). SINOLA: A new analysis/synthesis method using spectrum peak shape distortion, phase and reassigned spectrum. Proceedings from ICMC'99: *International Computer Music Conference*, Beijing.
- Peeters, G., McAdams, S., & Herrera, P. (2000). Instrument sound description in the context of MPEG-7. Proceedings from ICMC'00: *International Computer Music Conference*, Berlin, 166-169.
- Puig, V. (2005). Musique Lab 2 : A three level approach for music education at school. Proceedings from ICMC'05: *International Computer Music Conference*, Barcelona, 419-422.

- Rasamimanana, N., Fléty, E., & Bevilacqua, F. (2006). Gesture Analysis of Violin Bow Strokes. *Lecture Notes in Computer Science*, 3881, Sylvie Gibet, Nicolas Courty, Jean-François Kamp. Eds, Springer Berlin / Heidelberg, 145-155.
- Rodet, X., (1998). Sinusoidal + Residual Models for Musical Sound Signals Analysis/ Synthesis. *Applied Signal Processing*, 4(3), 131-141
- Rodet, X., Escribe, J., & Durigon, S. (2004). Improving score to audio alignment: percussion alignment and precise onset estimation. Proceedings from ICMC'04: *International Computer Music Conference*.
- Rodet, X., & Schwarz, D. (2007). Spectral Envelopes and Additive+Residual Analysis-Synthesis. *Analysis, Synthesis, and Perception of Musical Sounds*, J. W. Beauchamp Ed., Springer, New York, 175-227.
- Roebel, A. (2003). A new approach to transient processing in the phase vocoder. Proceedings from DAFX'03: *International Conference on Digital Audio Effects*, London, 344-349.
- Roebel, A. (2006). Adaptive Additive Modelling with Continuous Parameter Trajectories, *IEEE Transactions on Speech and Audio Processing*, 14(4), 1440-1453.
- Roebel, A. (2007). Frequency slope estimation and its application for non-stationary sinusoidal parameter estimation. Proceedings from DAFX'07: *International Conference on Digital Audio Effects*.
- Roebel, A., & Rodet, X. (2005). Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation. Proceedings from DAFX'05: *International Conference on Digital Audio Effects*, Madrid, 30-35.
- Rueda, C., Alvarez, G., Quesada, L., Tamura, G., Valencia, F., Diaz, J.F., & Assayag, G. (2001). Integrating Constraints and Concurrent Objects in Musical Applications: A Calculus and its Visual Language. *Constraints Journal*, 6(1), 2001. Kluwer Academic Publishers.
- Schwarz, D. (2006). Concatenative Sound Synthesis: The Early Years. *Journal of New Music Research*, 35(1), 3-22.
- Schwarz, D. (2007). Corpus-Based Concatenative Synthesis. *IEEE Signal Processing*, 24(2), 92-104.
- Schwarz, D. & Wright, M. (2000). Extensions and applications of the SDIF sound description interchange format. Proceedings from ICMC'00: *International Computer Music Conference*.
- Schwarz, D., Orio, N., & Schnell, N. (2004). Robust polyphonic Midi score following with hidden Markov models. Proceedings from ICMC'04: *International Computer Music Conference*, Miami.
- Schwarz, D., Cont, A., & Schnell, N. (2005). From Boulez to ballads: training Ircam's score follower. Proceedings from ICMC'05: *International Computer Music Conference*, Barcelona.
- Sarlat, L., Warusfel, O., & Viaud-Delmon, I. (2006). Ventriloquism Aftereffects Occur in the Rear Hemisphere. *Neuroscience Letter*, 404(3), 324-329.
- Tardieu, D., Carpentier, G., & Rodet, X. (2007). Computer-aided orchestration based on instrument models and genetic exploration. Proceedings from ICMC'07: *International Computer Music Conference*, Copenhagen.
- Theureau, J. (2006a). *Le cours d'action: Méthode développée*. Ed Jean Christol. OCTARES, Toulouse.
- Theureau, J., & Donin, N. (2006b). Comprendre une activité de composition musicale : les relations entre sujet, activité créatrice, environnement et conscience préréflexive. *Sujets, activités, environnements*, Barbier, J.-M. & Durand, M. Eds, Presses Universitaires de France, Paris, 221-251.
- Truchet, C., & Codognet, P. (2004). Musical Constraint Satisfaction Problems Solved with Adaptive Search. *Soft Computing*, 8(9), 633-640.
- Vergez, C., & Rodet, X. (2000). Dynamical Systems and Physical Models of Trumpet-like Instruments. Analytical study and Asymptotical Properties. *Acustica*, 86, 147-162.
- Villavicencio, F., Roebel, A., & Rodet, X. (2007). All-pole spectral envelope modelling with order selection for harmonic signals. Proceedings from ICASSP'07: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 49-52.
- Vinet, H. (1999). Recent Research and Development at Ircam. *Computer Music Journal*, 23(3), MIT Press, 9-17.
- Vinet, H. (2003). The Representation Levels of Music Information. *Lecture Notes in Computer Science*, 2771, Springer Verlag.
- Vinet, H. (2004). Synthèse du document Bilan-Perspective 2003. *Rapport d'activité IRCAM 2003*, 167-174, From <http://www.ircam.fr/uploads/media/RA2003.pdf>.
- Vinet, H. (2005). The SemanticHIFI project. Proceedings from ICMC'05: *International Computer Music Conference*, Barcelona.

- Vinet, H., Herrera, P. & Pachet, F. (2002). The CUIDADO project. Proceedings from ISMIR'02: *International Conference on Music Information Retrieval*, Ircam, Paris.
- Vinet, H. et al. (2006). *The SemanticHIFI Project – Final Public Report*. IRCAM, From: http://shf.ircam.fr/uploads/media/Final_Public_Report.pdf.
- Wanderley, M., & Battier M. Eds (2000). *Trends in Gestural Control of Music*. IRCAM-Centre Pompidou, Paris.
- Wöhrmann, R., & Ballet, G. (1999). Design and Architecture of Distributed Sound Processing and Database Systems for Web-Based Computer Music Applications. *Computer Music Journal*, 23(3), MIT Press, 73-84.
- Yeh, C., Röbel, A., & Rodet, X., (2005). Multiple fundamental frequency estimation of polyphonic music signals. Proceedings from ICASSP'05: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 3, Philadelphia, 225-228.