Filling Gaps between Current Musicological Practice and Computer Technology at IRCAM

Michael Fingerhut and Nicolas Donin
Ircam – Centre Pompidou

Knowledge build-up is a process which involves complex interactions between intellectual pursuits and the tools used to examine reality. While the interdependence between research and its instruments is more readily apparent in such fields as (say) neurophysics or microbiology, it is usually obscured in musicology, where the nature of the knowledge which is produced is rarely explicitly correlated to the devices which allow for its emergence.

Computers provide new means to put in relation, organize, process, ascribe meaning to, and reuse a wide variety of musical information, that which lends itself to digitization (from traces of the compositional process such as sketches, notes, etc., to computer "patches", musical scores, books and other forms of publications about the work; recordings of live events and on information about them, etc.), in massive in-depth and broad scopes, and thus cannot but have a major impact on contemporary musicology. Their use addresses a multiplicity of related domains (acoustical, perceptual, musical, technological, historical, social, legal…), and levels of interpretation (physical, symbolic, semantic, cognitive…).

At the crossroad of the musical creative process, production and performance on the one hand, and research and development in the related sciences and technologies, IRCAM holds a particular place which allows for the examination of these interdependences in conjunction with the development of specific tools. In this paper, we will attempt at presenting some examples of the musicologist's ideal instrumentation emerging from this reflection, as well as some of the concepts and tools which are already in use or in the course of realization.

■

Computers have been used in classroom situation at least since the late 50s with the innovative Plato project at the University of Illinois at Urbana-Champaign [Plato 1993, 1994]. Since then, a plethora of educational CD-ROM and software packages have appeared, including in music-teaching situations[1]. However, it is hardly the case that computers have been used for active interaction with the material which is used in music theory or history classes except for very particular situations (such as for a specific piece of music, say), nor have they provided yet such an environment for the musicologist.

In a 2000 paper (see [Bonardi 2000]), Alain Bonardi listed technical requirements from such an environment which, in his view, would be of help to musicologists:

1. Rich view of the work including: score, sonogram, diagrams…; sound recording(s); symbolic representations…
2. Views of related works.
3. Annotation of a view.
4. Form recognition and identification.

At the basis of his analysis there is the need to access (find and extract) and use information which is related to musical works (as embedded in the works themselves), to the processes in which they are involved (creation and performance) and to their context (historical, social, economical…). While this kind of information can usually be found in libraries on physical media, the development of digital technologies (computer speed, storage volume, networking – and their implication on programming and systems), together with the phenomenal increase in the availability of contents in digital form locally and remotely have affected these practices. Figure 1 maps the principal types of information pertaining to this domain and their

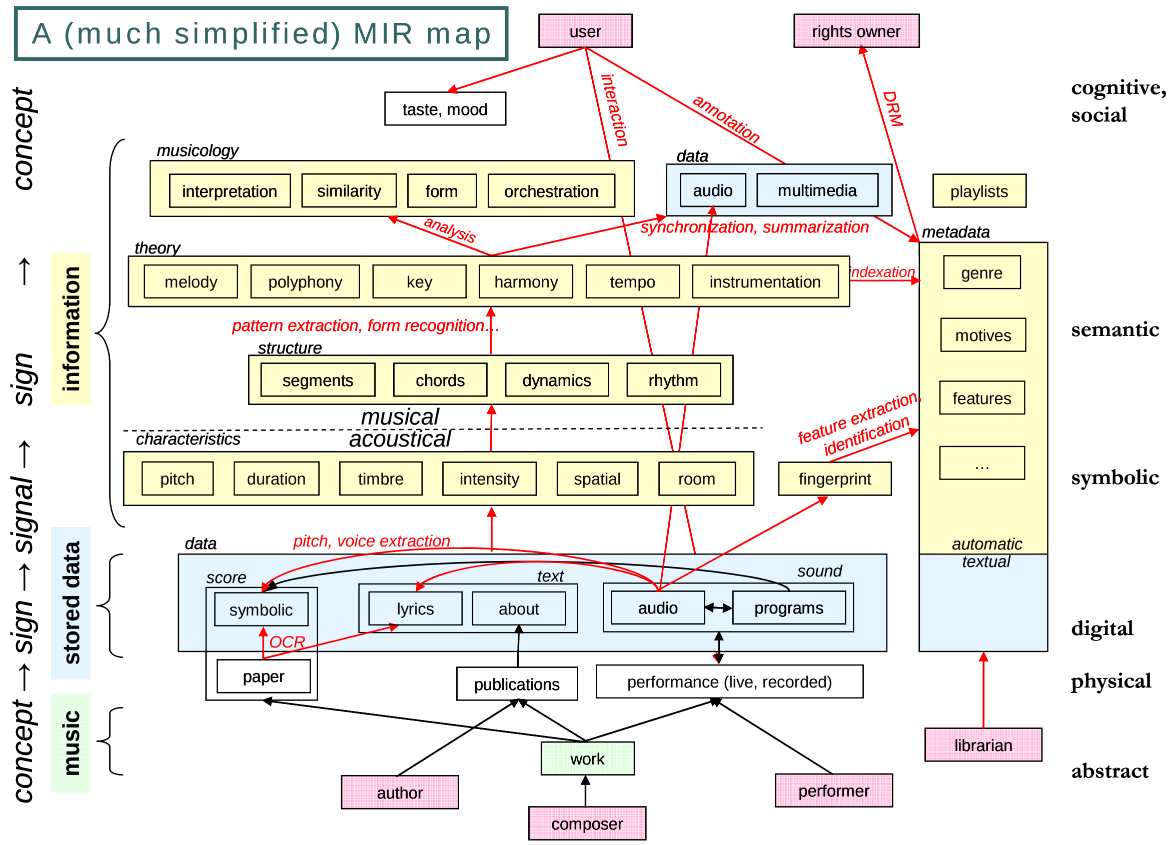

Figure 1. What is music information retrieval all about

levels of abstraction (*in* [Fingerhut 2004]). Red arrows indicate known computable derivations.

These issues are the subject of intense research and development in an emerging multidisciplinary domain, called *music information retrieval* (or MIR)[2]. The advances of science and technology which it has produced have helped reshape our imagination. As a result our current expectations of a better software environment for the musicologist and the music teacher are higher and more focused:

- The ability to interact with a work through annotation and manipulations of its score, by allowing one to focus, say on a section or a phrase, on the melody or on part of the polyphony; to highlight it visually and aurally; to alter the piece at the overall structural level or in some of its constituents (tempo, pitch relations, sound balance between instruments…) and listen to the result so as to test the coherence of the original work; to find and select individual notes, pitch sets, harmonic patterns, etc., produce an analytical reduction of the work and listen[3] to the resulting remix.
- The ability to search for like instances of an aspect under scrutiny (such as a harmonic progression) in the same and other works of the same or different style and period, by defining the desired kind and degree of similarity.
- The ability to compare students' analyses (calculations, analytical scores, rewritings…) of the same work; combine some of their proposals into a collective classroom reading of the piece; show how it stands against other known work from the past in order to build an evolutive pedagogy.

Recent and on-going projects at IRCAM have addressed some of these points, which we will now describe.

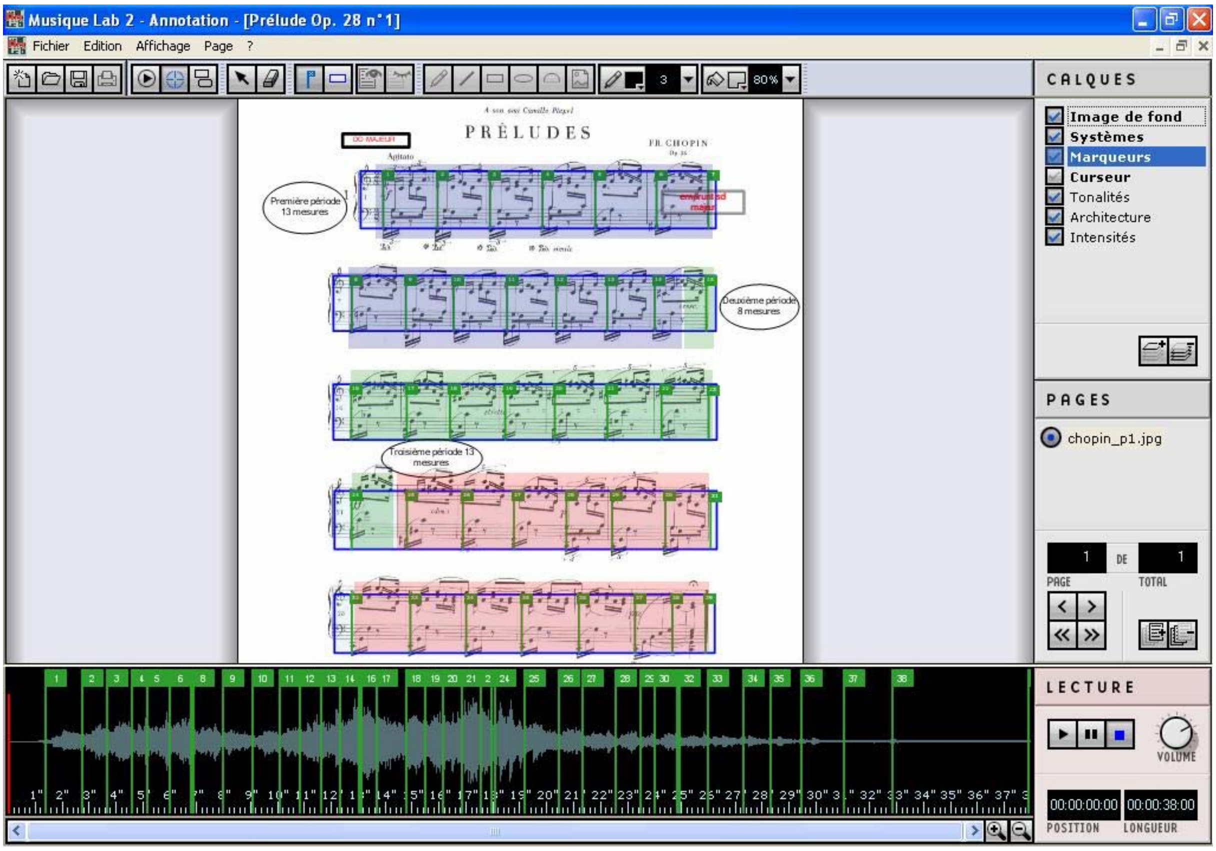

Figure 2. Annotation of Chopin's *Prélude* op. 28 no. 1

Production and annotation of rich views of musical works

Careful reading is an activity which involves, more often than not, annotation. While many tools exist to perform this action for text, IRCAM has been working on developing a toolbox for music annotation, to be used both for teaching and studying music (see Figure 2). It is part of the Music Lab 2 project on behalf of the French Ministries of Culture and Education, whose goal is the development of a suite of software applications for teaching music at the levels of high school and music conservatories (see [Puig 2005]).

The aim of this component, called ML-Annotation, is to experiment with new, manual and computer-assisted syntactic and semantic annotation and visualization paradigms of complex multimedia objects (using automatic music summaries, libraries of music annotation symbols…). As IRCAM has been active in research in these domains (in particular: audio summarization [Peeters 2002], metadata modelization and automatic extraction, semantic tagging, synchronization…)., this development aims at integrating these advances into a usable application which provides multiple views of a single work: sketches, score, recordings, comments… It is to be used in stand-alone systems as well as in distributed online digital libraries, thereby providing tools for the exploration of vast collections of such objects and their appropriation by individual users, for personal, educational and professional uses.

The source material concerning a single work (musical, but this fits many other intellectual productions) comes in a multiplicity of forms and media, static (text, hypertext, images – e.g., of musical scores, but these could be any still pictures) as well as time-dependent ones (audio and video recordings – of performance, of master classes, or of any other relevant situation). In order to provide meaning to this variety of material, several classes of non-destructive operations are needed:

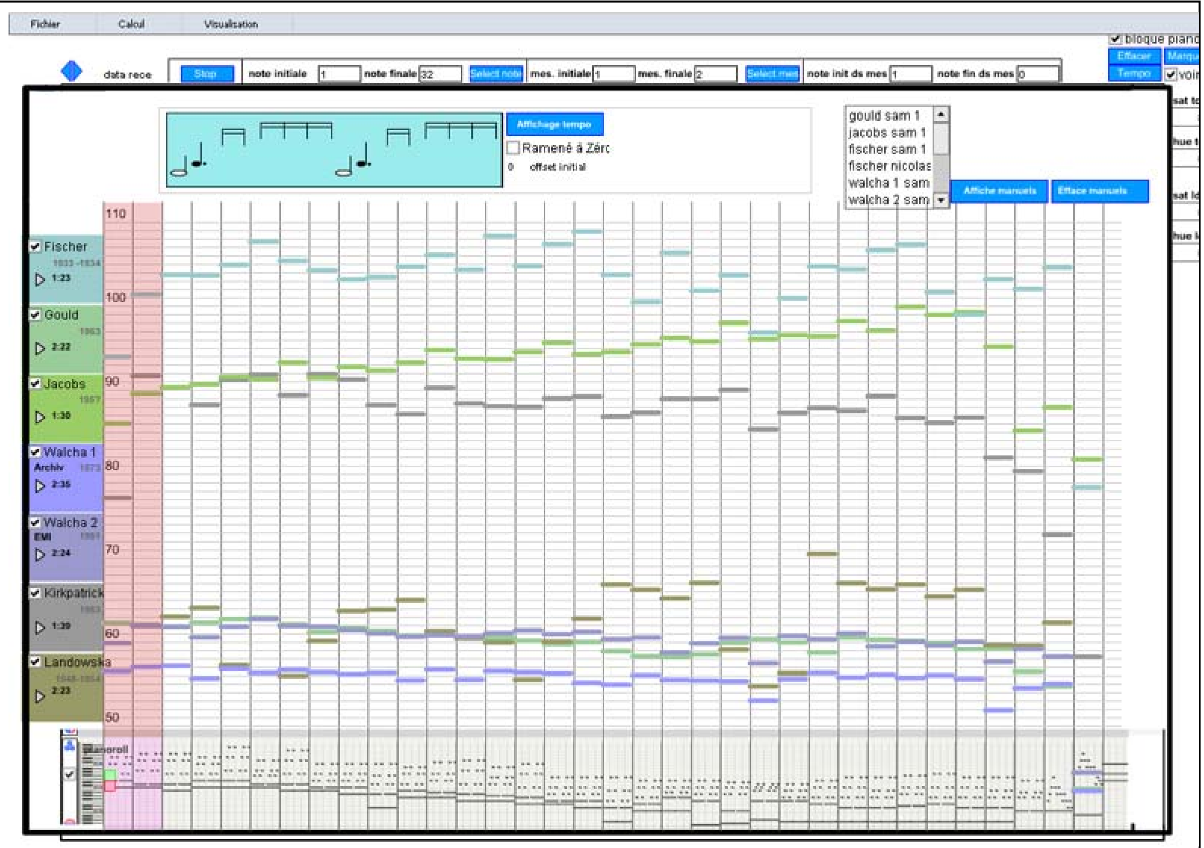

Figure 3a. Performance comparison of Bach's *First Prelude*

Bottom is piano roll. Each column is a bar, the horizontal bar represents the average tempo of each performance, color-coded according to the legend in the left margin (Walcha 1 is the slowest and most uniform, Fischer is the fastest and quite uneven.

- Synchronization, comparison, summarization…
- Tagging and time-dependent annotations.
- Visualization

for the construction of time-dependent, multi-layered interactive presentations.

Synchronization is the operation by which different documents are tagged in a way to indicate the occurrence of corresponding temporal events, even when the document is static (a text, a sketch, a musical score, a still picture) and multipage, or with documents with a different notion of time (distinct performances, e.g.). It can then used for the manual or computer-assisted comparison of documents of the same nature (audio recordings of different performances of the same type, e.g. [Peeters 2004]), as well as for multimedia visualization (e.g., score-following of a recording).

Automatic summarization allows one to provide bird's eye views of complex documents (long audio recordings, in this case) so as to ease the task of navigating into them, as well as in collections of documents. It consists in analyzing the timbre features of the audio signal, and then in selecting one or several characteristic segments of varying duration according to various criteria[4]. While this works well for popular music which usually possesses simple timbral structure, it may be adapted to some extent, by varying its parameters, to other genres of music. A separate application developed last year in the Multimedia Library, allows for the selection of "profiles" (families of parameters) deemed to be appropriate to a specific recording or collection of recordings, and perform automatic summaries of all the selected items.

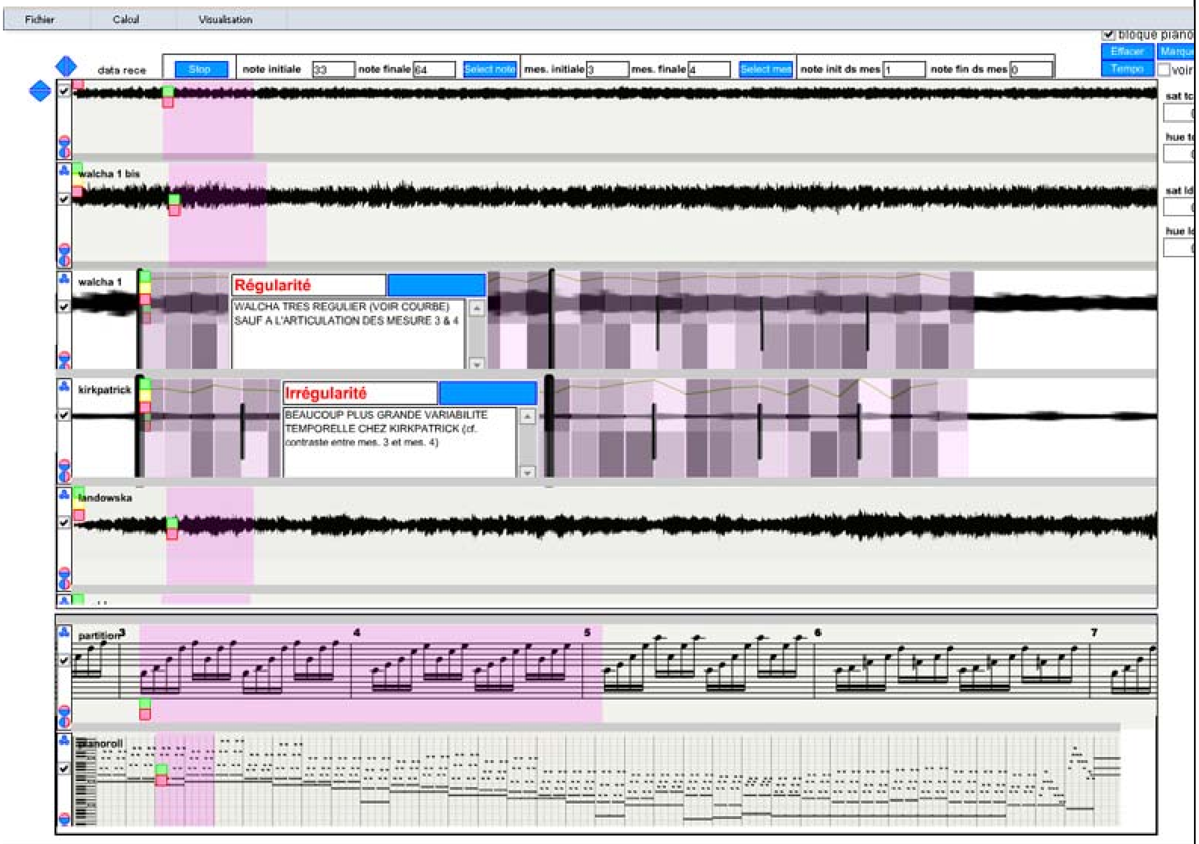

Figure 3b. Performance comparison of Bach's *First Prelude*

Any selection (here: two bars, highlighted in pink) in any representation (performance, score, piano roll) shows up automatically in the corresponding part in all other representations. Detailed indications on note tempo and intensity can be displayed and manual annotations added to the display.

Annotations are visuals (shapes, text, hypertext, symbols…of varying size and color) meant to draw the attention to specific parts of documents as they unfold in time. They may be tagged, manually or with the help of computer pattern and characteristic extraction from any of the representations of the document (symbolic, spectrograms, sonograms…), so as to attach semantics to documents. Both may be later used as indexes to searches into single documents as well as in a collection of documents.

At any step, the resulting rich document can be viewed statically (as leafing through the score) or played, allowing the user to select which kind of information shows up.

This application currently allows one to synchronize a sound file (typically: a recording of a performance of a piece of music) together with one or several images (typically: the score; but it could be any still – or moving image), and then add annotations of various shapes (rectangles, ovals, lines, images…), colors and textual content, which can appear and disappear at set times when the music plays, or stay throughout the piece.

Annotations can be grouped in layers which may selectively be shown at play time. They may represent distinct aspects of a single analysis (voice leading, melody, tonality…), or may be used to compare analyses made by different people, say students in a classroom. The score, actually a multi-page arbitrary image, is a layer by itself, can be displayed or hidden at play time. When it is shown, a cursor follows the sound as it is played, and repeats and page turns are performed as appropriate. Several different recordings (e.g., performances) can be synchronized with one single annotated score and thus compared.

Figure 4. Audio player showing the musical structure and the program notes

This process is non-destructive, as it only references the sound and image files. The annotations are kept in a separate file (in XML). It thus can be safely added on top of the existing digital music library and make full use of the audio documents it contains.

■

A related development ([Donin 2005]) has directly addressed the comparison of performances of a single specific work, the first prelude of J.-S. Bach's *Well-Tempered Clavier*, using a combination of manual annotation and digital signal processing algorithms. An audio alignment algorithm uses a MIDI file as a reference structure in order to get a precise onset estimation in the corresponding audio file. When several versions of the same classical work are to be compared, the MIDI file acts as the common "score" for the distinct sound files; the algorithm then evidences even the tiniest time variations between the interpretations.

This tool allows the user to switch between different representations. The general view (see Figure 3a), in which all performances are aligned with respect to a common piano roll, provides automatically-computed information on tempo variations. A windowing and zooming facility (see Figure 3b) allows one to concentrate on a few bars and see graphical indications of the duration and intensity of the individual notes. In this representation, textual annotations may be added anywhere, and saved together with the parameters of the view.

This experimentation will hopefully lead to the development of a generic performance comparison toolbox applicable to a wide range of pieces, and which will provide built-in additional music-savvy computations. This will open a fascinating field of empirical research in the historical and analytical study of performance.

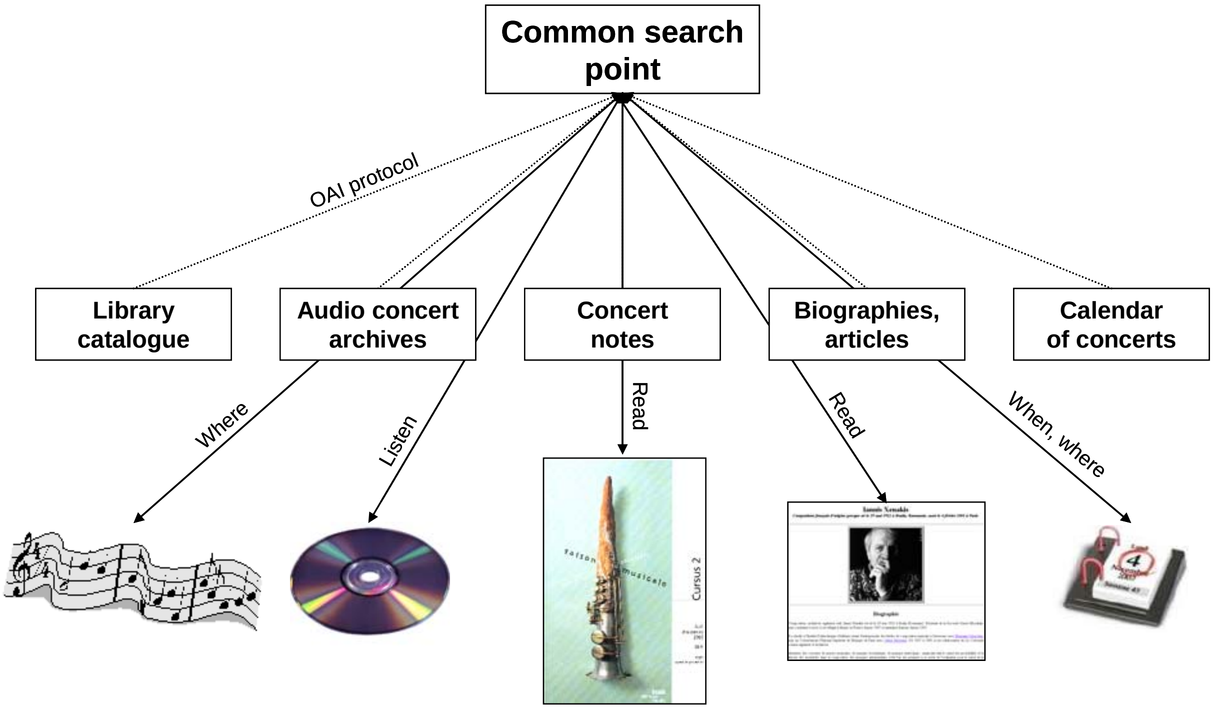

Figure 5. Sharing information sources with OAI

Navigation within individual sound documents

A music work has structure. Like a book subdivided in chapters, it can be composed of movements (acts and scenes in opera). At a deeper level, it may have a formal structure too (subject, countersubject… etc.). This is indeed true of any audio document (e.g., spoken word). While it is easy to browse through a text document, as its structure is may be apparent to the eye, and as hypertext may be used to highlight it, it is not the case for sound without visual cues. On compact discs, for instance, a table of contents on the cover allows for associating some structure with the audio contents (tracks), while current players show a one-dimensional view of the contents.

At IRCAM, we have been faced since 1995 with the need to provide our users with such mechanisms to access our online sound archives. While several emerging standards allow for describing the structure of multimedia documents (such as METS[5]) and for providing so-called hypermedia navigation capabilities "into" those objects, such as (such as SMIL[6]), this was not the case then. So we designed our own simple structured metadata schema representation of pieces and their movements, and developed a stand-alone player (see Figure 4) capable of displaying this structure and allowing the listener to move from piece to piece or movement to movement.

More recently, we have been working on the design and development of a network of servers hosting heritage sound collections (mostly spoken or sung) belonging to French institutions and which have been digitized under the aegis of the Ministry of Culture[7]. As some of the audio documents last well over an hour, it was necessary to provide tools to navigate within such document. The holding institution, the French National Archives, uses EAD[8] to describe each individual document: topical segmentation with incipits, indexation (dates, places…), which we transform automatically into a modified SMIL[9] object, including its own navigation capabilities and which can be viewed by an enhanced FLASH[10] player. These objects are integrated in the overall system and their "internal" metadata (used to

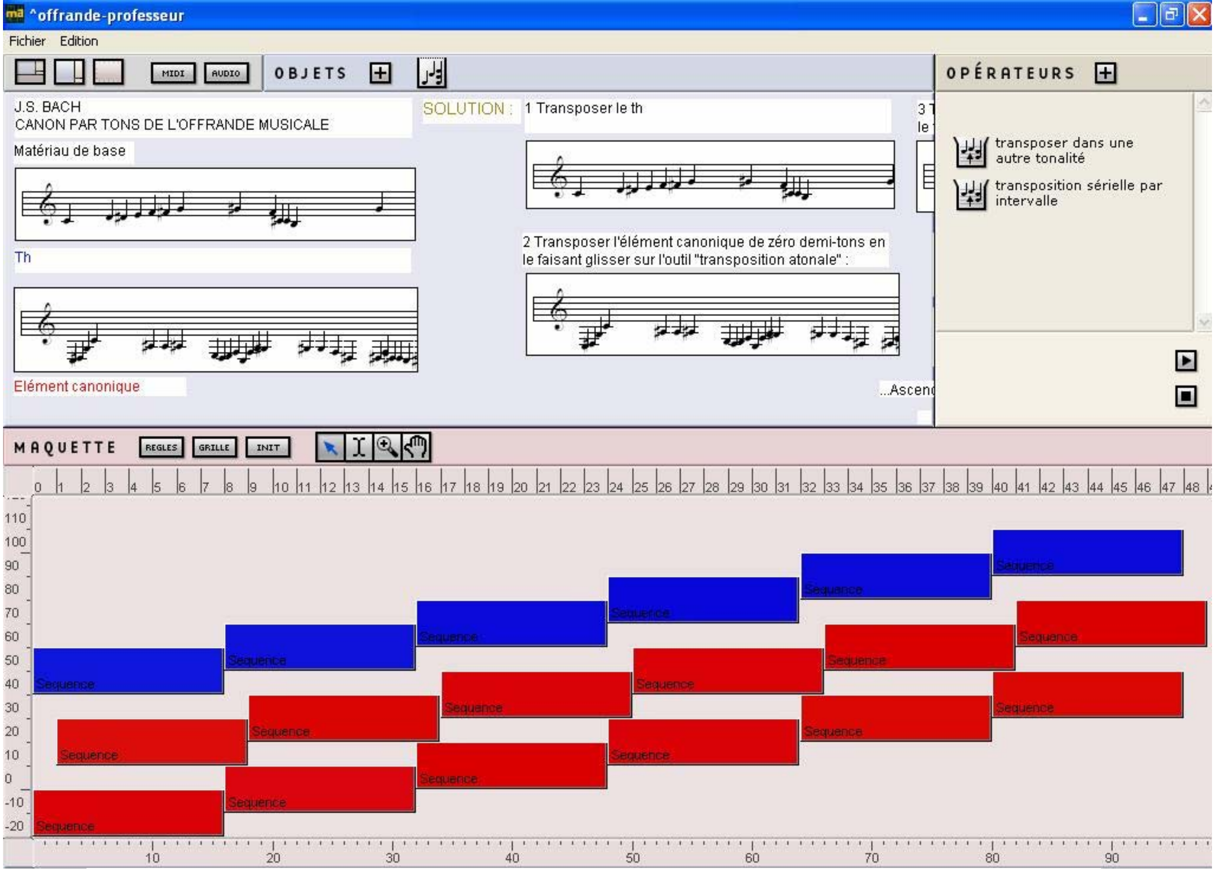

Figure 6. Maquette of J.S. Bach's *Canon 5. a 2 per Tonus* from the *Musical Offering*

describe their structure and content) is indexed along with the external metadata (used to describe the collections and the files as single elements).

Needless to say, this can be used for any kind of sound recording, including classical and contemporary music. Moreover, the annotation tool we mentioned above could be used to provide such navigation metadata.

Sharing music information sources

In addition to its catalogue of books, scores and periodical, the IRCAM Multimedia Library holds the digitized sound archives and concert notes of its public concerts, supplemented by a collection of compact discs of mostly contemporary music, both available online. A separate database, BRAHMS[11], provides biographies of hundreds of living composers and detailed information about their works. Another database contains online versions of the publications of its researchers in all its domains of inquiry (science, technology, musicology and related fields). An online calendar lists the events of its musical season, a year of concerts which culminate in its spring festival, Agora.

Faced with this multiplicity of information sources about the work and its many contexts, the question arises as to how to provide an efficient way to access all these heretofore distinct sources and to browse rapidly in their contents. This is the gist of a project currently under way at IRCAM, whose goal is to devise a single access point for a federated search through these sources, using the OAI-PMH[12] protocol to collect the metadata records from each individual source, transforming them into a common model and allowing one to search within this common pool through specific indexes as well as through full text – in the metadata records and, if available online, in the document itself[13] (see Figure 5). When the intellectual property issues are resolved, we plan to integrate automatic summaries of audio contents as a browsing aid in the online sound collections.

While the common records are less detailed than those found in each individual database, they provide enough information to satisfy most queries, and include music-specific information, such as instrumentation of musical works. Thus, upon searching, say, for a

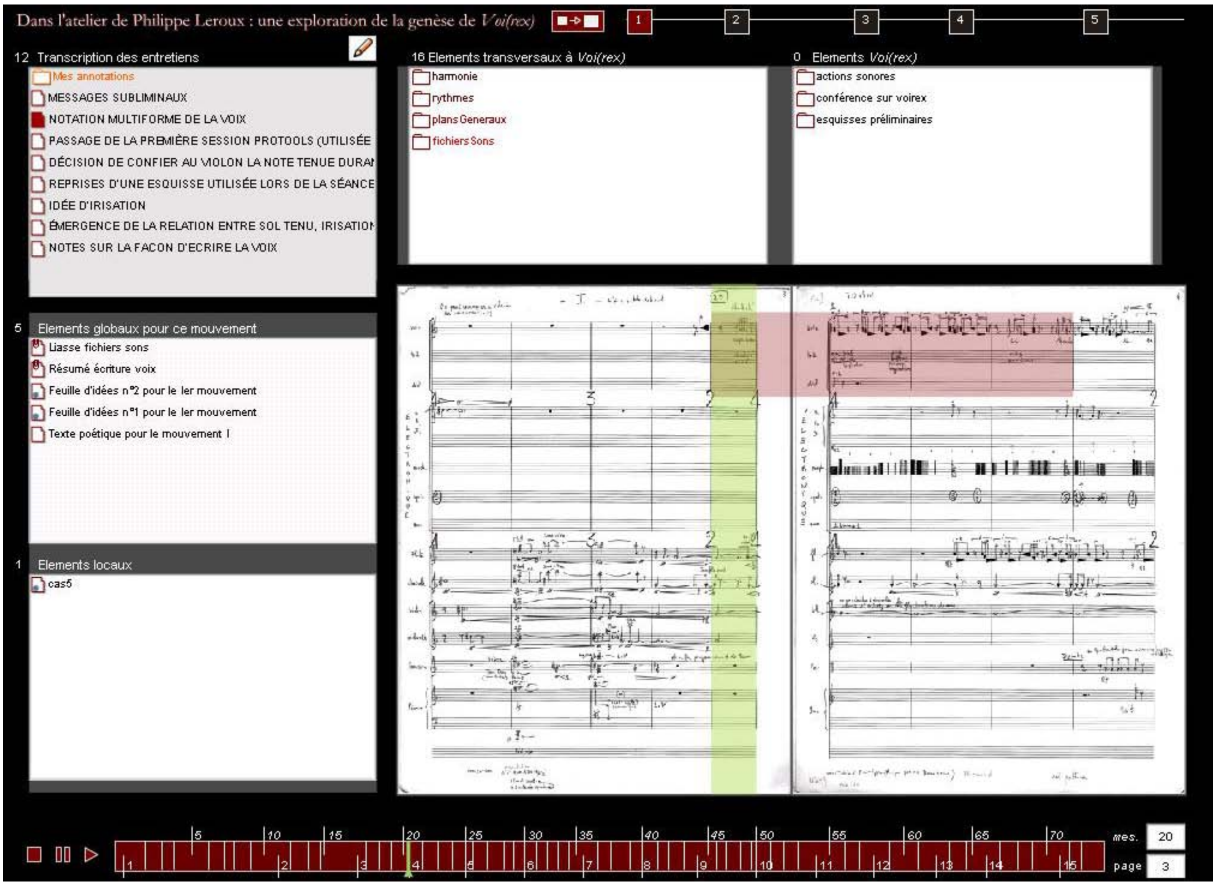

Figure 7. Genetic navigation in Leroux' *Voi(rex)*

specific music work by composer and title, this system will return the location of the (physical) scores of this piece and of monographies about it, provide online access to available recordings from past concerts together with the digitized concert notes from those events as well as to online notes about the piece from the BRAHMS database, list dates of future concerts where this piece will be played, etc. This system also allows one to access the original records as found in the specific databases, if he so wishes.

To implement this system, we have been using SDX[14], which not only allows for the indexation and retrieval of XML documents (including the use of one or several thesauri), but also comprises an OAI module, which is used to query all the relevant databases. This requires a usually minor development[15] to enable them to respond to these queries, as the protocol is relatively simple to implement.

As this system uses open standards, it can include external sources (which it clearly identifies as such), provided they respond to OAI queries. One interesting such source to include would be the Grove dictionary. Another use of this system which we envision is its extension to operate as a common portal to French contemporary music resources.

Model validation and genetic approaches

Another application in the Music Lab 2 suite, ML-Maquette, is a tool which can be used to validate a model of a musical work which results from its analysis. It is based on the OpenMusic environment [Bresson 2005], an object-oriented computer-assisted composition environment allowing for the manipulation of musical symbolic elements in order to produce music sequences.

It is based on the concept of the *maquette*, a 2-dimensional drawing-board, whose horizontal axis represents time, while the vertical one can represent any predetermined variable (e.g.: pitch, volume…). Temporal objects can be laid out and edited in this space: single notes, chords, rhythmic or harmonic sequences, envelopes… but also audio files, to which tonal and non-tonal transformations can then be applied (e.g., transposition, arpeggiation, inversion…).

While this environment may be used to compose, it can also be used to *re*compose a piece of music, starting from the constituents which a prior analysis has identified, such as its melodic elements and various transformations such as sequencing, inversion, transposition… (see Figure 6). The resulting music sequence can then be compared with the original work. This has already been used for such works as the Busoni transcription of the *Chaconne* from J.-S. Bach's *Partita* in D minor and Gérard Grisey's *Partiels*.

Once a model has been elaborated and validated, alternate parameters and combinations can then be tried interactively so as to examine different solutions to a given musical problem which the original work addresses. This is precisely the process which Bonardi alluded to in the above-mentioned paper when he writes: "*The musicologist is at the same time a listener and a composer, since analyzing a piece a music leads to 'rewriting' it*".

■

A multimedia project on a recent work by French composer Philippe Leroux (*Voi(rex)*, for voice and 6 instrumentalists, premiered 2002 at IRCAM) allowed the second author to experiment new ways of explaining music and guiding the listener using a genetic musicology approach. It involved Web technologies combined with IRCAM software.

The visualization interface aims to be a genetic navigation tool: it follows the processes adopted by the composer and provides, in addition to the score and sound recordings, the same manipulations which were used in the composition process: simulation of an OpenMusic patch, demix of a movement which had been designed in part with ProTools… and includes scanned sketches from the composer's archive (see [Donin 2006]).

Additional features allow the user to explore a database of documents – annotations and analytical texts – which help to make explicit the relations between the sketches and the score. These documents were produced in the course of an empirical musicological study which included interviews with the composer and whose goal was the reconstruction of the compositional process. Finally, the interface provides every user with means to add his own observations and hyperlinks into the database and to view them in conjunction with the score (see Figure 7).

A generic interface will be developed in order to facilitate critical editing of other contemporary music works using genetic research.

In conclusion

The potential contribution of computers for analysis can be summarily reduced to two major operating paradigms:

(i) abstraction (pattern extraction, recognition, comparison…);

(ii) synthesis (generation of sequences from a model).

It is in the second one that the above-described advances lay and address many the points in the actualized wish list of Bonardi. The hardest one, form recognition and identification, belongs to the first category and is far from being solved[16]. The higher we are in the information space shown in Figure 1, the harder it is to determine automatically (and sometimes even by hand) in the general case[17]. Contemporary compositions, which usually avoid adherence to well-known musical models from the past – and even sometimes from the present – are even harder to analyze (see Bonardi's paper).

Yet the fact that such a tool as ML-Maquette is built on an environment also used for composition, and that it can communicate and exchange analysis data with ML-Annotation[18] may herald the emergence of new tools for musicologists and thus help reshape its conceptual tools.

With the advent of an increasingly huge amount of digitized sound, we are seeing the emergence of a computer-assisted musicology of performance through the quantitative analysis of common practice making increasing use of measuring, averaging, comparison, navigation at the micro- and macro-levels of the work and the corpus, and of wider classification paradigms (e.g., not only traditional thesauri but also folksonomies). Historical musicology will probably be affected, too, as the ability to build a historically valid model of recording and listening is now increasingly possible, as well as to perform more effective stylistic comparisons of past performances. But insights could also be gained in the near-future capacity to analyze and manipulate the constituents of recorded sound of public performances (e.g., the spatial positions of the instruments, hall response, the mixing process…).

As these tools develop, they will no doubt allow musicologists to address unforeseen questions which will in turn affect the whole field of musicology. It is all the more necessary then that they be involved in their development[19].

References

[Agawu 1989] Agawu, Kofi: “Schenkerian Notation in Theory and Practice”, *Music Analysis*, 8(3), 1989, p. 275-301.

[Bonardi 2000] Bonardi, Alain: “Information Retrieval for Contemporary Music: What the Musicologist Needs”, *Proceedings of the [First] International Symposium on Music Information Retrieval*, Plymouth (Mass., USA), October 2000. Available here: <http://mediatheque.ircam.fr/articles/textes/Bonardi00a/>

[Bresson 2005] Bresson, Jean, Carlos Agon and Gérard Assayag: OpenMusic 5: “A Cross-Platform Release of the Computer-Assisted Composition Environment”, *10th Brazilian Symposium on Computer Music*. Belo Horizonte (MG, Brazil), Octobre 2005. Available here: <http://mediatheque.ircam.fr/articles/textes/Bresson05b/>

[Byrd 2002] Byrd, Donald and Michael Fingerhut: “The History of ISMIR – A Short Happy Tale”, *D-Lib Magazine* vol. 8 no. 11, November 2002. Available here: <http://www.ismir.net/texts/Byrd02.html>

[Donin 2005] Donin, Nicolas: “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*, XII/4, 2005. Forthcoming.

[Donin 2006] Donin, Nicolas, Samuel Goldszmidt and Jacques Theureau : “De *Voi(rex)* à *Apocalypsis*, fragments d'une genèse. Exploration multimédia du travail de composition de Philippe Leroux” (DVD-ROM), in *L'Inouï. Revue de l'Ircam* no. 2, 2006.

[Fingerhut 2004] Fingerhut, Michael: “Music Information Retrieval, or how to search for (and maybe find) music and do away with incipits”. IAML-IASA 2004 Congress, Oslo (Norway), August 8-13, 2004. Available here: <http://mediatheque.ircam.fr/articles/textes/Fingerhut04b>

[ISMIR 2005] Online Proceedings of the ISMIR Series of Conferences, 2000-2005. Available here: <http://www.ismir.net/all-papers.html>

[Peeters 2002] Peeters, Geoffroy, Amaury La Burthe and Xavier Rodet: “Toward Automatic Music Audio Summary Generation from Signal Analysis”, *Proceedings of the Thirs International Conference on Music Information Retrieval* (ISMIR), Paris (France), October 13-17, 2002. Available here: <http://ismir2002.ismir.net/proceedings/02-FP03-3.pdf>

[Peeters 2004] Peeters, Geoffroy: “A Large Set of Audio Features for Sound Description (similarity and classification)”, internal report, Ircam, 2004.

[Plato 1993] “PLATO reports, PLATO documents, and CERL progress reports, 1958-1993”. University of Illinois at Urbana-Champaign. Computer-based Education Research Laboratory Available here: <http://www.cbi.umn.edu/collections/inv/cbi00133.html>

[Plato 1994] Michael Szabo: “Enhancing the Interactive Classroom Through Computer Based Instruction: Some Examples from Plato”, *Computer-Mediated Communications and the Online Classroom*, vol 1. Hampton Press, New Jersey (USA), 1994. Available here: <http://www.quasar.ualberta.ca/edmedia/readingsnc/cmc/cmc.html>

[Puig 2005] Puig, Vincent, Fabrice Guédy, Michel Fingerhut, Fabienne Serrière, Jean Bresson and Olivier Zeller: “Musique Lab 2 : A three level approach for music education at school”, *Conference Proceedings [of the] International Computer Music Conference* (ICMC), Barcelona (Spain), 4-10 September 2005.

[1] From IRCAM as well, see e.g. the Music Lab 1 software package (<http://www.ircam.fr/237.html?&L=1>) and the MusicWeb consortial project (<http://musicweb.koncon.nl/>).

[2] A series of annual conferences in this domain, ISMIR, has been established in 2000. For its history, see [Byrd 2002]. Full online cumulative proceedings of almost all the papers presented at the conference are available in [ISMIR 2005]. Information on past and future conferences is available at <http://www.ismir.net>.

[3] Following Agawu’s interesting suggestion to play Schenkerian graphs [Agawu 1989].

[4] Including legal ones: as it appears, it is easier to obtain the rights to provide a 1-segment summary, i.e., an excerpt, rather than a “compilation” of excerpts.

[5] Metadata Encoding & Transmission Standard), used to encode metadata regarding objects in a digital library. See <http://www.loc.gov/standards/mets/> for details.

[6] Synchronized Multimedia Integration Language, an HTML-like language allowing for the authoring of interactive audiovisual presentations. See <http://www.w3.org/AudioVideo> for details.

[7] It is available online at the following address :<http://www.archison-culture.fr.eu.org>

[8] Encoded Archival Description, an XML schema used to encode archival finding aids. See <http://www.loc.gov/ead/> for details.

[9] Synchronized Multimedia Integration Language, an HTML-like language allowing for the authoring of interactive audiovisual presentations. See <http://www.w3.org/AudioVideo> for details.

[10] See <http://www.macromedia.com/software/flash/> for details.

[11] Base Relationnelle d’Articles Hypertextes sur la Musique du 20^{e} Siècle (relational database of hypertext articles about 20^{th} century music), available here: <http://brahms.ircam.fr/>.

[12] Open Archive Initiative for Metadata Harvesting, a protocol used to *harvest* (collect) metadata records (coded in XML) from possibly heterogeneous *repositories*, usually so as to provide a combined data store as a single search point for all harvested sources. This protocol is much “lighter” to implement and use than z39.50, but a comparison is out of the scope of this paper. See <http://www.openarchives.org/> and <http://www.loc.gov/z3950/agency> for details.

[13] Full-text search in metadata, while very powerful, is not good enough to provide relevant replies to precise queries (such as, say, instrumentation). It is thus useful to combine it with searches within specific indexes.

[14] System for Documentation in XML, an open source search engine and publishing framework for XML documents. See <http://adnx.org/sdx/> for details.

[15] Provided the database software is extensible (i.e., that its software can be extended). Where this is not the case (such as in our library system), alternative ways of exposing the metadata records can be developed.

[16] New algorithmic techniques such as unsupervised learning show promising results.

[17] For very simple music, this may be computationally feasible: many of the test corpora described in the ISMIR papers (see [ISMIR 2005]) consist of “songs”.

[18] As well as with a third program, ML-Audio, part of the same suite. See [Puig 2005].

[19] This is precisely what the IRCAM *Analysis of Musical Practices* group, headed by the second author, is currently researching. For a general description in English, see <http://www.ircam.fr/apm.html?L=1>. For online examples of results, see the French version of the page, <http://www.ircam.fr/apm.html>, and in particular the section *Pratiques contemporaines d’écoute et d’analyse musicale.*