Is Music mining relevant to support Music listening ?

F. ROUSSEAUX

***Abstract*--Now, when on-line music is seducing listeners and threatens to eliminate CDs, the industry of music recording faces both the need of creating new on-line services adapted to this new requirement, and giving up an object which traditionally concentrated its know-how.**

So, understanding the functioning of the disk as a semiotic object becomes urgent, since it cannot be reduced to a simple support for some (abstract) contents. The method is understood as a form of reverse engineering from the analysis of practices effective even now and will support an activity of requirements engineering destined to lovers of on-line music.

Our aim is to define a model for a computer-based co-operative musical listening system, able to assist users when choosing recorded music and elaborating their own modalities for describing titles.

The research program is led with the help of the CNRS, in partnership with the IST CUIDADO European project team, as well as the MPEG7 content descriptor standardization groups.

This investigation involves, besides the IRCAM own resources, an ad-hoc consortium grouping together requirements engineering, computational linguistics, musicology, cognitive sciences and data browsing assisted systems researchers.

***Index Terms*--Music mining, electronic music distribution, knowledge acquisition, machine learning, MPEG7.**

INTRODUCTION : WHAT ABOUT MACHINES HELPING US TO DEVELOP MUSICAL AFFINITIES?

IT is well known that our machines have plenty of free time. They could listen in on music, silently and at length, instead of us, in order to elaborate recommendations and suggestions for our own listening.

So, while it is pleasant to listen to a desired music, chosen by ourselves, the obligation of listening to a given music with the only aim of determining its affinities is a painful task.

This form of analytical listening, for mere selection, is not only highly unpleasant, but also very time-consuming, since disk tracks cannot be browsed like the pages of a book, and sound abstracts are not envisaged by the techniques of today.

In fact, this is the reason why we greatly appreciate the recommendations of expert friends, and, while there are so many musical spectacles and music records offered, "by word of mouth" information is so difficult to replace in some cases. Once more, why not to implement computerized assistance systems, which could find out new customized investigation tracks, or propose us relevant plans for navigation through the musical worlds?

In order to switch to a kind of artificial listening, it is perhaps not so bad to understand better natural listening, even if other flashes of inspiration could also become deciding factors.

Manuscript received December 19, 2001. This work was supported in part by IST project CUIDADO (IST-1999-20194).

Francis Rousseaux (francis.rousseaux@ircam.fr) is a member of the LERI (Laboratoire d'Etudes et de Recherches en Informatique, Université de Reims Champagne Ardenne, Boulevard de la Paix, 51100 Reims) and is presently sent on secondment by the CNRS to the IRCAM (Institut de Recherche et de Coordination Acoustique/Musique, Place Stravinsky, 75004 Paris, France).

AN INQUIRY ON NATURAL LISTENING

In order to start our investigation, let us try to initiate a trip throughout the world of natural musical listening. This will happen under high protection, avoiding any confusion, even if this reduces our speech to its most essential form.

A. Listening tracked through Literature and External World

Laziness is sometimes a good teacher: it recommends to neglect the first requirement of philosophy (i.e. the suspension of already available theses for allowing the search by oneself), while hoping that relevant answers are already available.

So, any lover of music is all ears when literary or philosophical texts refer to music. Lots of authors try to explain music and its listening, but only a few really assume this world, instead of giving us a ghost image of it. Aristotle himself said : "the aim of music is to be listened to rather than to be reasoned about".

The remaining ones, having the deepest insight, recommend, following Husserl, a reduced listening, able to extract the sound objects from the strait jacket of the Aristotle's causes and human desires. Presently, this group is of no help for us, because they cannot easily indicate us the fixed points where we could connect our artificial systems.

Then, let us try to give a definition of musical listening on the basis of the tracks left in the world.

We shall admit that the track of music listening in the world (its *world correlate*, in the philosophers' language) is more than diffuse: the potentiometer of the radio set was switched by a man zapping to a new channel, the record-dealer's door was opened by a man searching in a hurry the record he loved so much and the same man falls on his piano for striking some notes not so different from the music mode of the hit song just listened to.

It is difficult to identify fixed points inside such a mess. On the contrary, going to the other end of the spectrum, there is our own body (or the soul, to avoid philosophical quarrels).

In fact, we may admit, without any major risks, that the correlate of musical listening mainly concerns the listener's own body, which, *inter alia,* facilitates neither the observation task, nor the artificial system intrusion strategies "within the

loop".

If we consider that our own body is the main mediator for listening (remember Guillaume de Machaut, who asserted that "music, is what makes us dance and sing"), we must recognize that this is a place hidden from theories (with the possible exception of phenomenology). When we notice that most theoreticians are reducing our own body to the brains, we are sure that more attractive proposals on this subject will not be available before long.

Finding mediations other than our own body, which are sufficiently limited to condense the tracks of listening … this requires a trial!

B. *From our own body to the social body*

To sum up: sometimes, the listener is moved by music (if the emotion is pleasant, he will try to develop an elective affinity with this work, which becomes a reference), whilst in most cases the result is only tedium and indifference. When listening, the person shows his/her taste, or simply a form of "musical mood" in a spontaneous action: switching to another channel on the radio set, renewing the repertoire if the person is a conductor, or … extemporizing under the shower.

However, this carefree character is not general. Some persons are traditionally led to produce in concrete terms the everyday aspects of musical life. Here are the composer who receives a specific order, the professor who corrects a musical dictation, the radio compere who explains to the manager that his broadcast audience is dropping, the record-dealer who organizes his shop, the producer who tries to defend the consistency of a label on the market, all of these interacting actors strive to achieve projects, and therefore, create dynamic devices able to produce links between musical sounds and acoustic data, or between the various representations of music. All of them use a terminology and a language suited to the circumstances, and more or less specialized for a given profession.

There would be places for a privileged observation of language descriptions of the musical listening, and these places would typically coincide with co-operative musical projects, where actors co-operate and are trying to arrive at a common set of descriptors which makes explicit and optimally expresses each one's aims.

But these places present the disadvantage of being business-oriented, and it is very probable that the mediations engaged here are quite particular, which could take us far away from the general public practices.

Now, to be able to see the daily practice of recorded music lovers, and to get a model, we should try to find a mediation place which is both common and central. Have a good chance at hunting!

THE RECORD AS A SPONTANEOUS MEDIATOR OF MUSICAL LISTENING

Recorded music is dramatically accumulating throughout the world: it is distributed to many of us on digitized compact disks (CD), stored and classified on the shelves of major distributors, concentrated and labeled in the databases of the producers and copyright managers. Supply is overabundant, and is even increasing, with specialized products.

The traditional center of this world is a specific material object, the CD, a worthy successor of vinyl records, both as a support of contents and as a distinguished semiotic object.

C. *The CD as a semiotic object*

In fact CDs are typical supports of the recorded music, but also work as supports of an immediate recalling, since the CD packaging bears images, texts and icons characterizing the musical contents enough to avoid the need of really playing the full score.

Through this mnemonic assistance, the CD sleeve prescribes a practical and efficient organization. We need these colored packages, when we, as music lovers, try to find out a true successor of the CD which has just expired, and must discriminate among a number of concurrent proposals.

Moreover, particular collections are carefully looked at by buyers, who like get a visual knowledge of record sleeves, having a particular and differentiating character when choosing own music.

This is at the very center of the tradition which concerns us: the presentation of the CDs, as well as the organization of their contents, prescribe the practical set-up of shelves, at the dealer and also at the collector. This gives a material track of the recorded music which is preferred.

Because the CD also offers, besides the restitution of contents, a classification of the contained titles through the notion of album or compilation, and yet, through the notion of collection or label, so, a way towards the classification of the objects themselves.

To sum up, once again: the CD is the material support allowing the restitution of musical contents at request (if you own the proper device and have acquired the object). This object is the very heart of the music industry since many years (production, distribution, consumption), and its influence is not only commercial.

As for the production, the release of a CD means the end of a significant step in the life of an artistic project (here we have the notion of *album*). Traditional organization uses a calibration of the product in terms of quantity and quality.

For the distribution, the classification of the CDs on the distributors' shelves needs a set, more or less consistent, of classification and cataloguing criteria, forming a sufficiently robust taxonomy. Moreover, the CDs are accompanied by some musicological notices in the wide sense (artistic / cultural / historical hints), offering a direct access (without listening) to contents, via a description.

Finally, as for the consumption, the taxonomy of genres proposed by the distribution is naturally offered for the personal re-use managing the own objects which are now the bought CDs. Notices become then the symbolic place where life emotions and experiences are stored, instituted somewhat as places for culture standardization.

D. *Can the CD be the center of a cultural exchange model?*

At this stage, we shall consult the "requirements" pages of

our information system engineering manuals. It is recommended, in order to insert an artificial system in a human activity, to begin by setting up a functional model of the human activities to be eased. And in the case of a co-operative listening man-machine system, the same manuals would recommend to begin by a model of committed activities, at the cognitive level.

The proposed method would be a reverse engineering schema from the observed real practices, and would arrive at a model allowing to see clearly the central place of the CD in cognitive and semiotic exchanges. This model would encapsulate some terminological references understood by everybody, and would play the role of conceptual graph labels.

To be more direct - the buyer of a CD buys this CD because he likes it; and he puts it in his collection because "it" is compelled to him. He is far from representing to himself the manner generated and maintained by his own musical elective affinities as a course through a conceptual graph with terminological labels.

After this, he learned, often unconsciously, to recognize himself, or rather to "re-find himself", into a maze of polymorphous representations centered on the CD, to be investigated in an intelligent and spontaneous manner. By the way, he validates the semiotic representations and a terminology, up to discovering standard values.

To come back to our models: to create them, we need, at least, some meta-models describing human cognition. We have therefore to venture into a very abstract work, which could leave us alone, the pencil between our teeth, while some well-bred engineers will look at us somewhat suspiciously.

Let us avoid potential traps and follow our inquiry.

E. Music on-line: don't throw out the baby with the bath water

The apparition of the Web and network multi-media computers allows, at request, the access to digitized music, without the old constraint of material supports. Now, the technical devices of potential customers already open this theoretical possibility.

One thing is sure: the CD as a set of titles (where the vinyl read technology offered only one sequence, the CD technology allows in fact a direct access to titles, now perceived as a set) is not a robust concept for the distribution of digital music by the Web. Obviously, the consumers do not want to pay for titles they do not listen to, and it happens they focus on "hit songs" and never listen to all the titles of a CD. They want to listen to precise titles, or ad-hoc sequences, to be determined depending on the moment's mood.

Certainly, this will arise some strategic, juridical or practical problems for the major actors of concerned markets, for the simple reason that, even if the CD market is structured by the consumption, the organization of the music through the CDs is largely determined by the production.

But the trend is easy to identify and the supply will follow. Now, we have to anticipate the realization of possible conditions for practices yet emergent in order to imagine and define them, and build suitable tools. With e-commerce, the lion's share will be won by pioneers. A twofold question arises: "Why to do it? " (what is the aim of future users?) and "With which means? " (which tools are needed?).

To sum up, let's say that the support (here the CD) is for contents, which are also represented, and, as for all representations, some details are hidden: to get rid of it opens a new way of recording, but requires money and time to build a new representation of the contents.

Let us return now to our CD buyer. When he/she receives friends who will listen to this music, a sequence is built on the basis of the listening to one title, extracted from a record chosen among his/her own collection. After listening to this piece of music, and all comments, criticisms and questions, generally, you have to listen to another title extracted from a different record, but coming from the same collection.

With music distributed through the Web, the procedure is quite different - you have to listen to a title chosen among the set of titles available on the market, in this situation, every one can choose, the host as well as the guests.

Here, the mediator is no longer the CD, but a semiotic device including programs to be defined and specified, which are, at the same time, a man-machine interface, a set of functionalities and a level of performance.

It is not impossible that our user finds a way through these new semiotic objects and practices, which will finally run without his/her knowledge.

With a better look, we are now in a situation rather strange and uncomfortable - the present listening practice of the general public are largely ruled by the compact disk, which will be replaced by practices requiring services yet to be defined. As these services are not yet existing, potential users would describe their requirements by using the old fashioned terms of the disk age.

So, we need to pass skillfully, through modeling and requirements engineering, from the CD buyer to the user of on-line services, while believing that the usage will remain the same. To follow …

In any case, we will need meta-models describing the major features of human activities viewed as cognitive representations. This will remain the direction of our exploration.

LISTENING VIEWED THROUGH A COGNITIVE SCIENCE MODEL

Acoustic data are normally characterized by a number of parameters, instantaneous and time stamped. These parameters are naturally measurable, because their very nature of parameters depends on - we often use these acoustic data to describe what is listened to.

However, the sound which becomes a musical element is first perceived as a singularity, that any description could never exhaust, although we need to come near to it by refining the approximations - more precise perceptive metaphors, more specialized cognitive representations, or even new semantic organizations.

So, when the matter is music, the perception is coupled with an irreducible claim to be able to categorize the musical character of the sound.

Cognitive science experts agree, more or less, that any cognitive activity should be modeled as the production of representations : inscription, classification and categorization. These models seem to be suitable for the case of musical listening, and it is certainly very interesting for us to adapt them.

F. Inscription as representation

We shall accept that an inscription is a representation, otherwise, we shall reject from the field of cognition the sensors, risking so the impossibility to explain how the cognitive process are related to the objects of the external world. However, this type of representation must be considered as elementary, not to be mixed up with the type of representation that a conscious agent can experience.

Because all sensitive variations through time and space correspond to a new representation in terms of inscription. Obviously, such a number of representations cause almost immediately an overcharge of status memory. Consequently, this type of representation cannot explain how a cognitive agent can build a synthetic internal status for the multiplicity of projections built through the continuity of time and spatial variation.

Inscription is a representation based on sensor automatisms. Except the case of a failure, a sensor will react only when the situation is that designed when it was produced – a sensor can feel only what was foreseen for it, thus ensuring a stable functioning of the whole, as a distinctive technical information bearer. Consequently, the inscription functions as an extension logical principle. i.e. only a stimulus of the same type can cause the same reaction. Because of the automated operations during the inscription, the transformations undergone by this type of representation are limited, and if the reaction can vary, it is contained within the range of parameters. With respect to classification, the inscription serves as an input.

Inscription leads to the acquisition of an information element, and we can now consider that artificial inscription is now well established, due to digitalized records, basic extraction procedures and signal processing.

G. Classification

Classification is a more elaborated representation, which approximates any singular item to a particular, that is, a conceptual entity comparable to a generalized form of one or more items. This cognitive model is infinitely problematical, and constitutes a place of abysmal reflection, called the "problem of the intentional character".

From a semiotic point of view, this type of representation is the prototype of what is a representation, without any doubt, because any representation is generalizing, and the classification is the very type of generalization.

Present cognitive research offers a number of models showing the production and causal emergence of classes. For some ones, the process is linear and based on the classic concept of efficient causality. For others, the process is parallel and calls for non-linear causal models, even chaotic and dissipative. Here must be inserted the control and attractor parameters. Concrete entities are always individuals, but they are of various types in the sense that their functional roles are different.

This does not prevent to simulate a classification operation, which appears so simple when the problem of intentional character is hidden in the *a priori* part of the system: this is the situation of neural networks (not able to generate classes without learning) or the symbolic classification systems. To simplify, there are two possible uses of these networks: either their aim is to build classes (inductive systems, used to build large significant taxonomies), or class attachments are supplied to them in order to produce the classification procedures. In both cases, the systems acquire an artificial intentional character, either at the design, or through interactions with the users.

In the case of the artificial classification dedicated to musical listening, one may try to build libraries of generic classes (independent of assigned finalities) associated with classification procedures. This method is both naïve and legitimate - naïve because that are so few chances that the classification will not be prescribed, and finalized during the next step of categorization (characterization of the classes by using more general models), and so, the significant generic character is limited, but legitimate because it is easier to adapt these class and algorithm libraries (re-use) than to perpetually create them. However, one should be aware not to consider such re-use libraries as interpretation sources, which would systematically bias the creativity of the models, due to practical considerations.

H. Categorization

Categorization asks for a representation more precise than classification, or *a fortiori* of inscription. Inscription and classification of stimuli are basic cognitive operations, while the handling of the results coming from classification, by defining a specific relationship to actions - the categorization, needs complex cognitive agents.

The categorization is finalized, oriented to the deployment of an action and its resolution (just a musical metaphor!).

To categorize, the agent must be able to handle classes as objects, which implies the use of a semiotic form. So, the categorization can assume the form of propositions, schemas, prototypes, names, morphs, attractors or analogies. In philosophy, categorization is usually of a conceptual type, but we have reasons to think that categorization is not necessarily and immediately of a conceptual type. All these forms of categorization are semiotic, some ones are formally symbolic and others formally indicial.

A categorization organizes the classes into a categorical system, called *ontology*. It often occurs that this systematization creates a link between categorization and conceptualization. However, we ignore yet if any categorization is necessarily a conceptualization.

The categorization is a very elaborated modeling process, finalized through an adaptation to the living world which would require to invoke many domains of modeling to represent the various kinds of knowledge: praxiological,

normative, epistemic, didactic, ipseical.

Thus, cognitive sciences propose us models for explaining the consequences of our cognitive activities in connection with our action strategies. The praxiological level would describe the action schemas as such (listening leads to dancing or to an improved form of playing an instrument). The normative level integrates, coordinates and valorizes the action strategies, aiming at global success (the project of a listener, composer, performer or producer appears and grows). The epistemic level models the conditions and circumstances of a reliable perception (listening as an inquiry in a world of belief). The didactic level models the learning and time evolving functions (listening as knowledge acquisition), while the ipseical level models the self-consciousness that a cognitive agent can develop (listening as participation in the subject's building).

I. Which cognitive model for co-operative man-machine listening?

At this stage of our investigation, what can we assert on the man-machine co-operative system that we are specifying? Which are the representations (inscription, classification, categorization) accessed by the artificial part of the system?

At one end of the spectrum, the artificial inscription is arrived at for the whole set of digitized titles. At the other end, which could be the role of a computerized system at the categorization level, among the model attractors constituted by the *praxis, nomos, episteme, didaskein* and other *ipseities* ?

The problem, as mentioned above, is that the natural categorization is essentially obtained through the mode of the action, which is out of an artificial system's reach, unless the action is in fact an interaction with the system.

However, listening implies the desire of hearing more, i.e. a desire of continuation within an experience, and the specification of a new object to be experienced. So, the requirement of continuity generates a requirement of rupture, the succession (cognitive difference) determines a variation (difference of type) and the need of consistency through the succession is assimilated to the need of an alteration presenting some similarities or analogies with the preceding part.

Classification is found between inscription and categorization. What about artificial classification? The intentional character is simulated by the specification of the classes/classification procedures, in a hybrid manner - the classes are specified through the intuition of what will be useful for the categorization, while the procedures are specified through the intuition of what is envisaged with respect to the inscription. During a cycle of the project, the choice of the classes effectively participating in the constitution of categories could be selected, specified, parameterized, edited.

This indicates that the computerized system for co-operative musical listening could be modeled as an artificial system able to assist users in the choice of recorded music and the elaboration of their own modalities of title description, if a suitable requirements engineering work is dedicated to on-line music lovers.

The key question is then that of categorizing musical titles , and its expression in the field of cognitive sciences could be: "How to define a computerized system for assisting music categorization?".

TOWARDS A MUSICAL LISTENING MAN-MACHINE CO-OPERATIVE SYSTEM

Now we have to precise how the artificial listening system is inserted in its user's activity of musical elective affinities deployment.

As already asserted, the proper character of human listening is to prescribe its object and cognitive modalities, i.e. to define what will be listened to and in which conditions. This means that the desire of listening also depends on what has just be listened to. The process is not necessarily conceptual, or even rational.

J. An assistance system for data browsing

Therefore, the matter is to give the users search motors and browsers with a performance allowing to select specific titles at request - through filtering based on multi-criteria specification or by analogy from examples and similarity criteria.

The first possibility is to derive differentially proposals from examples, here the assisted description justifying the example as such would be a relevant functionality of an assistance system proposing the titles. If necessary, the search specification may include some elements of the user's profile: the acquisition of a user typology pointing at the ontology of the musical listening domain is a compulsory requirement.

When the user tries to elaborate sequences, it may happen that knowledge elements of a different order are extracted, in relationship with the notion itself of sequence, for which an associated ontology has to be produced. What is a musical sequence? Why this is not a mere list?

By the way, let's say that parts of titles could be viewed as autonomous titles during the man-machine interactions, and this could add an incremental dimension to requests – this form of analogy will make the system more powerful and expressive, and everybody will be able to refine the examples while creating a customized form of investigation.

This differential possibility can give way, in particular when no example is used, to a simple description, equivalent to multi-criteria search in a semi-structured database.

A further possibility is to require of users that they qualify their activity in order to assist the learning process.

K. From the living world to ontological models

We have to repeat that only models that can run on Turing machines (computers) are calculable formal models, i.e. conceptual networks, prescribed by computational linguistics for having the required properties of calculability, and ultimately pointing at classes of objects with suitable calculation procedures.

One must keep in mind that this ontology* (the name which is presently in fashion for such conceptual dependency networks, somewhat indicating that analytic philosophy thinks that mainland philosophy is out of date) is subject to constraints of various origins, and all of them must be fulfilled.

Because in the design of an ontology, categories are to be reduced to their conceptual description.

As for artificial listening, man-machine interactive (MMI) systems have not only to supply functional input to the system, but also a semiotic view of this conceptual representation, so that a place "to live in" is created, even if this requires some effort of learning, accepted by the user if the gain is obvious.

Through the bias of man-machine interface systems, users must therefore move to this ontology. Once again, this "bias" is not a hollow formula, but really corresponds to a modeling activity which compensates what was lost when ontology was reduced from everyday life categories (the cognitive categories are considered to be able to model perfectly living and feeling) to conceptual and rational categories. The art to attain this result through rational means (MMIs are, and are only, computer programs) has the appearance of magic, but in fact, this can be only conceived as an art of subverting human dialogue to produce the illusion of a man-machine dialogue. By the way, this art is without any doubt a key dimension for re-thinking the epistemological status of computer science.

Besides, the ontology must also point at calculable classes, i.e. having classification procedures which must terminate, the calculation time being reasonably short. Here, the effects of performance are not only quantitative, but sometimes also qualitative - a very fast approximate procedure, but with results that can be refined through feed-back, may be preferred to a slower procedure claiming better results).

Let's add that there must be no confusion between this requirement of calculability and a requirement of withdrawal on a single objective source. First, because it is pure non-sense to think that the feeling induced by an artistic object is a property of the object (e.g. the symbolic character of the wedding golden rings would be hidden somewhere in gold atoms).

Consequently, some of the classes pointed at by a domain's ontology will necessarily fall out of what can be derived from the sound signal or its acoustic properties. To simplify, this aspect will be considered as participating in the definition of the user's profile, or something like that. Some characteristics could be calculated (for instance, from the user's behavior, since this manifestation can be observed through the interactive interfaces), while others would be kept as man-machine system meta-data.

On the contrary, some classes pointing at objective notions will be difficult to calculate from the rough signal (with the present state-of-the-art of science and technology) and will be better also viewed as meta-data (e.g. aspects connected to organs). Here we have a heavy discrimination to be applied, through mutual agreement between all project's actors.

L. Acquisition of an ontological model for musical listening

But users must "live in" the model, in the sense that here their needs are satisfied. Let's say at once that nobody believes now in needs absolutely pre-existing to their satisfaction – new requirements appear while old ones are either satisfied, or simply eliminated by more powerful new requirements (a need is satisfied if it is annihilated). Requirements engineering would then be equivalent to building an ontology with sufficient theoretical power to attract, if classes can be calculated, users for a long time.

So, the only solution is to propose valid taxonomies for efficiently categorizing the titles at our disposal. For this, we have only to describe user requirements in terms of abstract descriptors, while taking into account the practical possibilities for title concrete identification. But this point certainly deserves an explanation, since it is connected to some deep-rooted difficulties.

Let's imagine the process as if it was sequential, to demonstrate that this is totally false. Through requirements engineering techniques, one may hope to acquire the knowledge of user requirements - the future users of digitized music from the Web - with respect to descriptors useful for categorizing the music. But these abstract descriptors must also be mutually consistent, and, for that, they must constitute the basis of a knowledge acquisition model both complete and discriminatory - this is the domain's ontology. Moreover, they must be realistic, and so associated with operational procedures ending up to a capability of reasonably determining the value that corresponds to every descriptor proposed per individual title.

Let's imagine then the following sequential procedure: Elicitation of needs in terms of abstract descriptors useful to describe music -> Elimination of unrealistic descriptors -> Modeling an ontology from these descriptors.

The sequential character of the procedure would immediately lead to the following problems:

- the paradox of user oriented design - the possibility of a user contribution to the specification will radically change the requirements expressed soon after the realization; this is an even more difficult situation when precisely the requirements are breaking with the tradition. A variant - a paradox of user typologies: we may suppose that there are several user types, but the typology is also emergent, and so we want the new typology, and not the old one; how to produce a typology readable through emergent criteria?
- the paradox of requirements driven design - it is quite impossible to find out realistic descriptors without a clever modification built in the acquisition of user criteria.

Proposals for facing these paradoxes:

- to produce the user typology at the same time as the descriptors, following the order of pertinence produced on a super-set, which is itself produced by a group of experienced and creative users (a questionnaire allows us to validate the representative character of the old order

* The ontology is termed "Descriptor scheme" in the IST CUIDADO project, while categories and classes are called "High Level Descriptors" and "Low Level Descriptors", respectively.

users and to set up a correlation);
- to explain the new typology in the terms of the old one, through data analysis methods;
- to identify pilot users (including possibly the creative aspects of learning during the normal operation of the system) and a group of experienced users, the results having to be compared with prescriptions of expert pools structured by themes (markets, laws, domain's science).

INDUSTRIAL ORGANISATION OF THE IST CUIDADO PROJECT

CUIDADO is the IST-1999-20194 project, linked to Key Action "KA3", related to the Action line "Media representation and access: new models and standards".

CUIDADO means "Content-based Unified Interfaces and Descriptors for Audio/music Databases available Online".

The project started on 01-Jan-2001, and it will last 36 months.

M. Project Abstract

Information overload, inability to quickly browse through audio, poor added-value to music via Internet distribution, keyword dictatorship, inability to search for similarities among sounds: these are music consumer complaints addressed by CUIDADO.

It aims at developing content-based technologies using MPEG7 output. Building re-usable modules for audio feature extraction, statistical indexing, database management, networking and constraint based navigation, CUIDADO target two pilot applications:

- The Music Browser features musical paths and automatic compilations according to user's tastes, search for music similarities, learning systems based on user's personal choices. It also serve as a monitoring system for Web music files,
- The Sound Palette involves music creators for developing an authoring tool in an existing professional audio environment taking full advantage of the extracted audio features for innovative retrieval, editing and processing.

N. Objectives

Develop content-based audio modules and applications using the MPEG7 Media representation standard. The project covers the analysis process (extraction of descriptors), the navigation process (retrieval methods and interfaces implemented in a leading database system with Web interoperability), up to the creative process (consuming and authoring tools) involving content creators and consumers in each stage.

The project address both the audio (samples) and the music domain (titles) since high level descriptors for music (style, rhythm) should rely on robust lower level audio descriptors (pitch, energy or spectral features) in order to cover a wider range of applications.

This approach answers the needs of record labels and copyright societies for information management methods for both marketing and protecting their contents. CUIDADO is also a first attempt to go beyond content retrieval with an authoring system using content features for professional musicians and studios.

O. List of Participants

- Institut de Recherche et de Coordination Acoustique Musique - Ircam, France
- Artspages International As, Norway
- Ben-Gurion University of the Negev, Israel
- Creamware Datentechnik Gmbh, Germany
- Oracle Iberica, S.R.L., Spain
- Sony France S.A., France
- Universitat Pompeu Fabra, Spain

P. Description of the Work

The project involve music creators and consumers, on-line quality control and regular specification update according to MPEG7 inputs (WP1) throughout :

Two Pilot applications (WP3) :

- The Music Browser includes sound similarity search, constraint based music selection, personal keyword use and user's profiling; one version is tied to Web music monitoring and another to Web music sales and customized radios;
- The Sound Palette has an online version with retrieval and processing features scalable to the network performances and an off-line version which includes content-based editing/processing in an existing professional authoring environment.

Re-usable technical modules (WP2) :

- Audio features extraction will provide content descriptors to the other modules, first derived from current Instrument Timbre Description proposed in MPEG7 by the CUIDAD Group (Esprit 28793) and then from signal analysis and perceptual studies following MPEG7 Multimedia Description Scheme;
- Statistical music indexing will derive high-level descriptors (such as music style or genre) from previous task audio descriptors and MIDI files;
- Database management system will be optimised for large audio & descriptors databases using Java beans and MPEG7-DDL (or XML);
- Distant interoperability will use Corba Network Computer Architecture; Sound processing and networking modules for raw synthesis and fast listening over the network will be packaged with existing secured music transaction systems (IST project RAA and SDMI);
- Music sequence generation applies constraints on music titles/audio descriptors.

User validation, standardization and dissemination (WP4) :

- User groups for requirements, creativity and validation will gather music creators and consumers but also record labels, sound archives and copyright societies using a feed-back loop methodology;
- Standardization in MPEG7 and dissemination to content providers is a key issue.

CONCLUSION

While the progressive disappearance of the semiotic support that the compact disk was for the general public of music lovers becomes patent, we have to describe musical titles in terms of contents as soon as possible. For this, the assistance of an artificial system for music selection, running in a co-operative mode, is needed.

This system shall offer means of translating the user's desire into descriptors, and, besides, build an operational basis for the description in terms of a set of titles, the identification of which requires a reasonable delay and offers a fair image of initial desires.

We have tried to present the multi-disciplinary search to be mobilized for arriving at the design and implementation of such a system, following the initial action of the IST CUIDADO project.

ACKNOWLEDGMENT

This work was supported in part by the IST project CUIDADO (IST-1999-20194).

REFERENCES

[1] Alain Bonardi and Francis Rousseaux, *Interagir avec un contenu opératique : le projet d'opéra virtuel interactif Virtualis,* Revue d'Interaction Homme-Machine (RIHM), numéro spécial "Interaction et documents", September 2001

[2] Frederic Brentano, *Psychologie du point de vue empirique*, Aubier, Paris, 1944

[3] Alain Cardon, *L'informatique, science ou technologie ? Un formidable défi pour l'IA !* , Bulletin de l'AFIA n°42, July 2000

[4] E. Diday, Y. Kodratoff, T. Brito and M. Moulet, *Induction symbolique-numérique à partir de données,* Editions Cepadues, 2000

[5] John Fodor, *The Modularity of Mind*. Cambridge, MA: MIT Press, 1982

[6] Edmund Husserl, *Recherches logiques***,** Paris**,** Presses universitaires de France, 1959

[7] M. Jarke, C. Rolland, A. Sutcliffe and R. Domges (eds*), The NATURE of Requirements Engineering*, Shaker Verlag, Aachen, 149-174, 1999

[8] Y. Kodratoff, *Applications de l'apprentissage automatique et de la fouille de données,* Actes du colloque Extraction et Gestion des Connaissances, 2001

[9] *MPEG7, context and objectives,* International Organization for Standardization, report ISO/IEC JTC1/SC29/WG11, October 1998

[10] Jean-Guy Meunier, *Categorial structure of Iconic Languages,* Theory and Psychology, vol. 8, n° 6, p. 805-827, 1998

[11] Jean-Guy Meunier, *Représentation en Sciences Cognitives,* RSSI, vol. 19, n° 2-3, p. 83-105, 1999

[12] Alan Newell, *The Knowledge Level,* Journal of Artificial Intelligence 18, 1982

[13] Anne Nicolle and V. Saint-Dizier de Almeida *Vers un modèle des interactions langagières,* In Analyse et simulation des conversations : de la théorie des actes du discours aux systems multi-agents, InterEditions Lyon, 1998

[14] Anne Nicolle, Pierre Beust and Vincent Perlerin*, Un analogue de la mémoire pour un agent logiciel interactif,* In Cognito, to be printed in 2002

[15] François Pachet, *A Taxonomy of Musical Genres,* RIAO, Paris, April 2000

[16] Danièle Pistone and Jean-Pierre Mialaret (ed.), *Analyse musicale et perception,* Cahiers de l'Observatoire Musical Français n° 1, 1994

[17] Danièle Pistone (ed.), *Musique et style. Méthodes et concepts,* Cahiers de l'Observatoire Musical Français n° 3, 1995

[18] François Rastier & al., *Sémantique pour l'analyse,* Masson, Paris, 1994

[19] Colette Rolland and N. Prakash, *From Conceptual Modelling to Requirements Engineering*, Special Issue of Annals of Software Engineering on "Comparative Studies of Engineering Approaches for Software Engineering", 10, 151-176, 2000

[20] Francis Rousseaux, *Une contribution de l'intelligence artificielle et de l'apprentissage symbolique automatique à l'élaboration d'un modèle d'enseignement de l'écoute musicale,* cahier du LAFORIA n° 80, February 1990

[21] John Searle, *The rediscovery of the Mind*, MIT Press, 1992

[22] Ludwig Wittgenstein, *Investigations philosophiques,* Gallimard, 1961

[23] Manuel Zacklad and Francis Rousseaux, Modelling Co-Operation in the Design of Knowledge Production Systems : The MADEINCOOP Method - An example in the field of C3I systems, Computer Supported Cooperative Work, The Journal of Collaborative Computing 1-22, Kluwer Academic Publishers, the Netherlands, 1996