François Pachet, Active Listening : What is in the Air ? Sony CSL internal Report, 1999.

Music Listening: What is in the Air ?

François Pachet
SONY Computer Science Laboratory Paris
6, rue Amyot
75005 Paris - France
pachet@csl.sony.fr

Abstract

The XX[th] century is full of technological inventions that made the very idea of a listening device possible, from the early gramophones to the latest portable mini disk players. What evolutions can we predict for the listening devices of the future, and how these evolutions will change the way we access and listen to music ? In this chapter, we suggest that listening devices can be greatly enhanced by providing new forms of user *controls* which provide users with semantically preserving variations. These controls are intended to allow listeners different musical perceptions on a piece of music, by opposition to traditional listening, in which the musical media is played passively by some neutral device. The objective is both to increase the musical comfort of listeners, and, when possible, to provide listeners with smoother paths to new music (music they do not know, or do not like). This chapter illustrates this idea on a few examples of active listening projects conducted at Sony Computer Science Laboratory, Paris, based on the notion of constrained exploratory space. These constrained spaces suggest that the classical boundaries between composing, listening and mixing may be redefined, thereby assigning new roles to composers, sound engineers and listeners.

1. From Buttons to Exploration

We propose the idea of exploratory listening environments, as a natural evolution in the history of musical controls. We first sketch a brief history of musical controls, and then introduce the notion of semantic-preserving musical exploratory environment.

1.1 History of Musical Controls

Each technological advance has brought with it new forms of controls. The origins of listening machines with mass-produced musical materials may be traced back to the Phonograph, invented by Thomas Edison in 1878, which used tin foil cylinders, and shortly after the Gramophone, invented by Berliner in 1888, which used flat disks. In these devices, there was no control intentionally given to the user (se, e.g. Read & Welch, 1976). There was, however, an unintentional control in the Gramophone in that the horn could be *turned* around, thereby influencing the directivity of the sound source. Electricity soon began to be used for listening devices, both with radio and with new electrically recorded disk players in the 20s. The use of electricity also introduced new controls: the *volume* button and the *treble/bass* button. Juke-boxes were introduced in 1927, allowing listeners to *select* explicitly music titles from a given catalogue of disks, using various sorts of push buttons. The next big technological advance was the invention of binaural (stereo) recording method in 1931. The corresponding control was the *panoramic* button allowing to control the amount of signal in one loudspeaker or the other. Finally, digital format for audio introduced more controls, e.g. on the equalization of sound. In all these cases, technological advances were followed by the introduction of "technical" controls, i.e. controls operating directly on the technology (see Figure 1).

Figure 1. A Phonograph (Edison, 1978, left); a Gramophone (Berliner, 1988, middle), a Rock-Ola 120-selection Juke-Box, and a Mini disk player (Sony, 1997, right). Advances in technology do not necessary imply more intelligent user control.

1.2 A Matter of Semantics

The very notion of musical control raises the issue of *semantics*. The issue of musical semantics - does music have meaning ? - has been long debated by musicologists, leading to different theories, which usually paralleled the theories of semantics for languages. One of the main distinction made by theorists is the opposition between so-called "referentialists" and "absolutists". Referentialists claim that musical meaning comes from actual references of musical forms to outside objects, i.e. music means something which is external to music itself. For instance, a particular scale in Indian music may have a reference to a particular human mood. Absolutists, e.g. Strawinsky, claim on the contrary that the meaning of music, if any, lies in music itself, i.e. in the relations entertained by musical forms together. Although these two viewpoints are not necessarily exclusive, as noted by Meyer (Meyer, 1956), they leave open much of the question of meaning. Eugene Narmour elaborated a much more precise theory of musical meaning based on the psychological notion of expectation (Narmour, 1992). In this theory, meaning occurs only when musical expectation are deceived. On the other hand, Rosen argues (Rosen, 1994) that the responsibility of preserving the meaning of a musical piece lies only in the performer itself, who has to choose carefully among a infinite set of possible interpretations which one is closest to the one "intended" by the composer.

Without committing to one particular theory of musical meaning, we can note that meaning - whatever it means - has to do with *choosing* among a set of interpretations the "right one" or the "right ones", i.e. those intended by the composer. A second remark is that the controls given by the history of sound recording technology have never had any concern about musical semantics: what does it mean to raise the sound level of a record ? to shift the signal to the left loudspeaker ? to increase the bass frequency ? Are the intentions of the composers, or even of the sound engineers, preserved in any way ?

From this remark, we suggest that "interesting" musical controls should preserve some sort of semantics of the musical material, i.e. preserve intentions, whenever possible. We argue that more meaningful controls, in the context of modern digital multimedia technology, amount to shifting from traditional button-based technology to musical exploration spaces.

1.3 Music Interactivity

As we have seen, technological buttons bear no semantics, because they are directly grounded on the technology, without any model of the music being played. But what can be such a model ?

Interesting approaches in musical interactivity are the music notation systems, in the context of annotation of music documents, as in the works of Lepain (1998), or in the Acousmograph system (INA-GRM). In these systems, the primary issue addressed is not music listening per se, but rather music *notation*, i.e. how to represent graphically a musical document (the document itself or the perception of the document), or how to infer a model of the music which can be noted or represented graphically.

Another answer may be found in the notion of *open form*, initially developed in literature (Eco, 1962), which has had much impact on music theory and composition (Stockhausen, Boulez). The idea of musical open form is that the composer does not create a ready-to-use score, but rather a set of potential performances, which can be seen as a *model* of scores, as explained by (Eckel, 1997): "Music is not any longer conceived in form of finite units but in terms of models capable of producing a potentially infinite number of variants of a particular family of musical ideas". The selection or instantiation of the actual score to be played is delegated to the performer. In recent incarnations of open form, it is the listener himself who instantiates the model, as for instance in the *Cave* (Cruz-Neira et al., 1993) or *CyberStage* (Eckel, 1997). In these cases, the user is immersed in a realistic virtual environment, and has the control on his position and movement in a virtual world. His movements are translated into variations in the musical material being heard. These approaches may be considered as radical, in the sense that the user has a great deal of responsibility in making the music. However, the issue of semantics is not directly addressed, since the model in principle is under-designed, i.e. all possible explorations are always "licit", whatever they may be. In this respect, there is a strong relation between open form virtual environments and programming languages for music composition, such as *OpenMusic* (Assayag et al., 1997), *CommonMusic* (Taube, 1991) or *Elody* (Orlarey et al. 1997). In these approaches indeed, the goal is to propose the user to explore spaces with as much freedom as possible, and not constrain the user in specific areas.

1.4 Active Listening

Active Listening refers to the idea that listeners can be given some degree of control on the music they listen to, that gives the possibility of proposing different musical perceptions on a piece of music, by opposition to traditional listening, in which the musical media is played passively by some neutral device. The objective is both to increase the musical comfort of listeners, and, when possible, to provide listeners with smoother paths to new music (music they do not know, or do not like). Active listening is thus related to the notion of open form outlined above but differs by two important aspects: 1) we seek to create listening environments for existing music repertoires, rather than creating environments for composition or free musical exploration and 2) we aim at creating environments in which the variations always preserve the original semantics of the music, at least when this semantics can be defined precisely. For us, the issue if therefore not to introduce yet another technological button in the interface of the listening device, but rather to design buttons that "make sense", thereby breaking the long tradition of technological buttons initiated by Edison.

What "sense", what "meaning" are we talking about ? How can music controls be designed to trigger semantic preserving actions ? The answer stems from the new landscape of music recording created by digital multimedia, sketched in the next section. We will then illustrate our ideas by two examples of active listening projects at Sony Computer Science Laboratory - Paris.

2. The New Facts of Multimedia

Digitalization of multimedia data has a number of technical advantages which are well known today: better sound quality, better compression, lossless copy, etc. The aim of this chapter is to show that digitalization of multimedia data also induce - even in a still potential form - a number of revolutions in the way music may be accessed and listened to by end users. We will outline three of these revolutions, which form the basis of our argumentation, focusing on the paradigmatic shifts they convey, rather than on technical aspects.

2.1 Structured Audio: Home as a Reconstruction Machine

The idea of structured audio has initially been devised to allow better compression of high quality audio. Standardization efforts like the Mpeg-4 project embody this idea, and try to make it practical on a large scale (see, e.g. the Machine listening Group of the Media lab, Sheirer et al., 1998).

The idea is simple: instead of transmitting a ready-to-listen sound, only a description of how to make the sound is transmitted. The actual sound is reconstructed at home, or at the listener's location, provided of course he/she has the right software to process this reconstruction properly. Structured audio actually extends this basic idea to include fully-fledged *scene descriptions*, that is, not only descriptions of individual sounds, but description of groups of sounds playing together to make up a piece of music. The actual technical details of scene description also include all what is needed to reconstruct a sound or piece of music rightfully, e.g. effects, adaptation to the local sound reproduction system, and so forth.

In our context, we argue that the notion of scene description opens up new doors for meaningful controls. Indeed, since the music is delivered as a "kit", lots of possibilities can be imagined to influence the way the kit is actually built, according to user preferences. Of course, these variations around how the kit should be assembled have to be “coherent”, which are precisely the matter of our work.

2.2 Meta-data and All That Jazz

The fact that musical data is now produced, coded and transmitted in a digital form has numerous and well-known advantages: better sound quality, possibility of lossless transmission and copying (thereby raising new copyright problems). An important non technical consequence is the possibility to encode not only the music itself - the digitalized sound - but also any sort of symbolic information. Such symbolic information may be used to code and transmit data on the music itself, so-called information on content, meta-data or also “bits about bits”.

Why would one want to transmit such meta-data ? The interests are obvious in the context of document indexing. If musical data is accompanied with corresponding adequate descriptions, digital catalogues can then be accessed using sophisticated query systems. Current standardization efforts like Mpeg-7 embody this idea (MPEG7, 1998), and try to define standards for describing meta-data for all sorts of multimedia documents. MPEG-7 aims for instance at making the web more searchable for multimedia content than it is today, make large content archives accessible to the public.

Here again, we would like to emphasize the conceptual rather than the technical aspects of this paradigm shift: meta-data opens also doors for imagining new listening systems in which the user may access data in a drastically different way. Instead of being a passive, neutral support, music becomes an active, self-documented knowledge base. Again, what kind of listening devices can be imagined that exploit this information ?

2.3 Size of Digital Catalogues

Digitalization of multimedia data has yet another consequence: the availability of huge catalogues of multimedia data to users. In the case of music, there is, here also, a conceptual shift which has nothing to do with the technology of large databases. The main issue raised by this technological advance is how to access huge catalogues of music, not from a technical viewpoint, but from a user’s viewpoint. Recall the juke box, invented in the late 20s: a typical juke box would contain about 120 titles, which is the size of an average user’s discotheque. Browsing through all the titles was probably part of the pleasure, and selection could be made just like at home: by choosing one item out of a collection of items, which at least the user has seen once.

Now a typical catalogue of a major company is about 50.000 items. What happens when the collection to select from is such a catalogue ? Even more terrifying, what happens if all the recorded titles become available through networks to users at home ? Estimating the total number of all recorded music is difficult, but it can be approximated to about 2 million titles (see, e.g. the size of MusicBoulevard or Amazon databases). The figure can be probably doubled to include non Western music. Every month, about 4000 new CDs are issued on the market. It is clearly impossible to apply usual techniques of music selection in this new context. What does it mean to “look for” a title when the mass of titles is so huge ?

François Pachet, Active Listening : What is in the Air ? Sony CSL internal Report, 1999.

3. Spatialization: The MusicSpace Project

The first parameter which comes to mind when thinking about user control on music is the spatialization of sound sources. We conduct a project for investigating the technical and conceptual issues related to meaningful user-control of music spatialization, called *MusicSpace*.

3.1 Motivation and Description of MusicSpace

In MusicSpace, the user can listen to pieces of music using an interface in which each instrument in the piece is represented by a graphical object (see Figure 2). Moving these objects around modifies the mixing of sound sources in the global sound. Moreover, an object representing the listener himself - avatar - is also represented in the interface, so that all the mixing parameters (volume, panoramic position, etc.) are computed according to the avatar's position. The basic system provides the possibility of 1) moving around the avatar, to induce a mixing as if the listener was moving around the actual musical setup, and 2) moving around the instruments themselves, thereby inducing a different mixing as if the listener was a sort of sound producer.

Experimentations of this basic system were conducted on average listeners and music composers. It clearly appeared that although the physical actions of moving avatar or instrument icons around in a window are very similar, the possibility of moving around listener's avatars is quite different conceptually than the possibility of moving around instruments. Indeed, moving the avatar corresponds to the action of moving oneself around a musical setting. Moving instruments correspond to a more technical view on the music - the sound engineer's view. This second possibility appeared to some users as heretic, since it practically gives users the possibility of totally changing the overall mixing of the musical piece !

The second phase of our project consisted in introducing a way of somehow constraining user actions, to avoid situations where the mixing produced is totally unrelated to the original spirit of the music (Pachet & Delerue, 1998). We proceeded by introducing a particular technique, called constraint perturbation, which precisely allows instruments to be linked together by relations that are always enforced: the system uses these constraints to propagate changes, so that the setup always remain consistent. For instance, a "related" constraint may be set between the drum and the bass, so that one of them is moved closer to the listener's avatar, the other one is moved accordingly (with the same distance ratio). On the contrary, a "balance" constraint may be set between two sound sources that should always be mutually in opposition: for instance, when the chorusing instrument is brought closer, the accompaniment is moved away. These constraints can finally be composed together to create rich environments in which users may change the instrument positions, but the constraint system ensures that the overall mixing always remain consistent with the engineer or composer constraints.

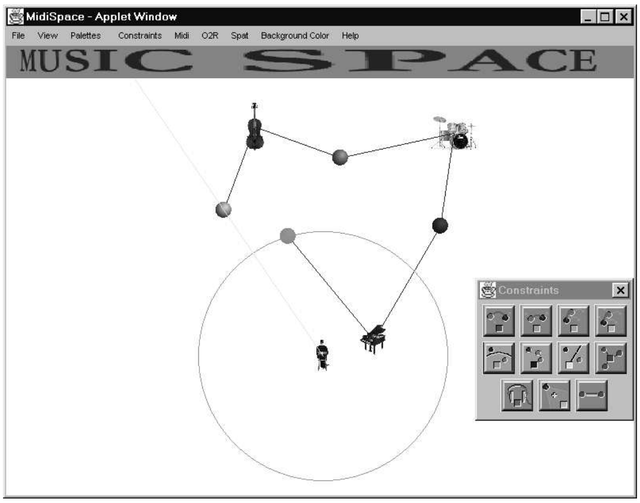

Figure 2. The interface of MusicSpace. Instruments are related by constraints. The avatar as well as instruments can be moved around by the user. The constraints embody an "automatic" sound engineer.

3.2 Exploration Space

There are two ways to interpret MusicSpace. One is to see it as an embodiment - simplistic but operational - of a sound engineer: the user may move sounds using high level, simple actions; the system "corrects" these actions by moving other sound sources according to his knowledge of sound mixing. This knowledge is explicitly represented as constraints.

The other viewpoint is to see mixing constraints as an *ontology of mixing actions,* which allows to mix in terms of properties of setups, rather than in terms of atomic actions on knobs and faders. This ontology allows to specify properties of configurations, which are guarantied to be always enforced, rather than specify explicit configurations. In this respect, constraints represent a semantics of sound source configuration, and the resulting - constrained - exploration space allows to explore various configurations without violating the spirit of the original mixing.

MusicSpace is also to be seen as an example of exploitation of "reconstructed music". As outlined in Section 2.1, future standards will deliver music by chunks, possibly transmitting sound sources separately, together with specifications on how to reconstruct the music whole from the parts. Constraints are one way of specifying this reconstruction, which nevertheless leaves room for new semantic-preserving user control. As such, it is a radically new form of Gramophone, as described in 1.1: not only does MusicSpace provide more refined controls on sound spatialization than turning the horn around, but these controls preserve the underlying intention of sound source configurations.

4. Music Catalogue Access

The issue of music delivery concerns the transportation of music in a digital format to users. Music delivery has recently benefited from technological progress in network transmission, compression of audio, and protection of digital data (Memon & Wong, 1998). These advances allow now or in the near future to deliver quickly and safely music to users in a digital format through networks, either internet or digital audio broadcasting.

Moreover, as seen in Section 2.2, digitalization of data makes it possible today to transport information on content, and not only data itself. Together, these techniques give the users, at home, access to huge catalogues of annotated multimedia data, music in particular. These techniques aim at solving the *distribution* problem, i.e. how to transport data quickly and safely to users. Paradoxically, these technological advances also raise a new problem for the user: how to choose among such huge catalogues ?

4.1 Motivation and Ideas

From the user viewpoint, accessing a large quantity of music indeed is problematic: it cannot be reduced to a simple database problem, because, by definition, users do not know precisely what they look for. The problem of choosing items is general in western societies, in which there is an ever increasing number of products available. For entertainment and specially music the choosing problem is specific, because the underlying goals - personal enjoyment and excitement - do not fall in the usual categories of rational decision making. Although understanding a user’s goals in listening to music is very complex in full generality, we can summarize the problem to two basic and contradictory ingredients: desire of repetition, and desire of surprise.

The desire of *repetition* is well known in music theory and cognition. Experimental psychology shows the importance of repetitions in music. At the melodic or rhythmic levels of music “repetition breeds content”. For instance, sequences of repeating notes create expectations of the same note to occur. At a higher level, tonal music, for instance, is based on structures that create strong expectations or the next musical events to come (for instance, a dominant seventh chord creates an expectation of a resolution). Music theorists have tried to capture this phenomenon by proposing various theories of musical perception based on expectation mechanisms (see e.g. Meyer, 1956), particularly for modeling the perception of melodies (Narmour, 1992). At the more global level of music selection, this desire of repetition tends to have people wanting to listen music that they know already (and like) or music that is similar to music they already know. For instance, a Beatles fan will most probably be interested in listening the latest Beatles bootleg containing hitherto unreleased versions of his favorite hits.

On the other hand, the desire for *surprise* is a key to understanding music, at all levels of perception. The very theories that emphasize the role of expectation in music also show that listeners do not favor expectations that are always fulfilled, and enjoy surprises and untypical musical progressions (see e.g. Smith and Melara, 1990). At a larger level, listeners want from time to time to discover new music, new titles, new bands, or new musical genres. This desire is not necessarily made explicit, but is nevertheless as important as the desire for repetition.

Of course, these two desires are contradictory, and the issue in music selection is precisely to find the right compromise between these two forces: provide users with items they already know, *and* provide them with items they do not know, but will probably like.

François Pachet, Active Listening : What is in the Air ? Sony CSL internal Report, 1999.

From the viewpoint of record companies, one goal of music delivery is to achieve a better exploitation of the catalogue. Indeed, record companies have problems with the exploitation of their catalogue using standard distribution schemes. For technical reasons, only a small part of the catalogue is actually "active", i.e. proposed to users, in the form of easily available products. More importantly, the analysis of music sales shows clearly decreases in the sales of albums, and short-term policies based on selling lots of copies of a limited number of items (hits) seem to be no longer profitable. Additionally, the sales of general-purpose "samplers" (e.g. "Best of Love Songs") are no longer profitable, either because users have already the hits in their own discotheque, or because they do not want to buy samplers in which they like only a fraction of the titles. Exploiting more fully the catalogues has become a necessity for record companies. Instead of proposing a small number of hits to a large audience, a natural solution is to increase diversity, by proposing more customized albums to users.

4.2 Approaches in Music Selection

Current approaches in music selection can be split up in two categories: 1) query systems for accessing music catalogues, and 2) recommendation systems for proposing novel titles to users. In both cases, these approaches provide sets of items to the user, which he/she has still to choose from.

Query systems address mainly database issues for storing and representing musical data. They propose means of querying musical items using some sort of semantic information. Various kinds of queries can be issued by users, either very specific (e.g. the title of the Beatles song which contains the word "pepper"), or largely under specified (e.g. "Jazz" titles).

Collaborative filtering approaches (Shardanand, and Maes, 1995) aim primarily at achieving the "surprise" goal, i.e. issue recommendations of novel titles to users, with the hope that these recommendations will be enjoyed. Collaborative filtering is based on the idea that there are *patterns* in tastes - tastes are not distributed uniformly. This idea can be implemented very simply by managing a so-called *profile* for each user connected to the service. The profile is typically a set of associations of items to grades. For instance, in the *MyLaunch* system, grades vary from 0 (I hate it) to 5 (this is my preferred item). In the recommendation phase, the system looks for all the agents having a similar profile the user's. This similarity can be computed easily by a distance measure on profiles, such as a hamming distance. Finally, the system will look for items liked by these similar agents, which are not known by the user, and recommends these items to him/her. Typical collaborative filtering systems for music are the Firefly system (Firefly, 1998), *MyLaunch* (MyLaunch, 1998), the Amazon web site (Amazon, 1998), or the similarity engine (Infoglide, 1998).

However, there are limitations to this approach. These limitations appear by studying quantitative simulations of collaborative filtering systems, using simulations techniques inspired from works on the dissemination of cultural tastes (Epstein, 1996; Cavalli-Sforza and Feldman, 1981) .

The first one is the inclination to "cluster formation", which is induced by the very dynamics of the system. The experimental results achieved so far show that such systems produce interesting recommendations for naïve profiles, but get stuck as soon as the profiles get bigger (about 120 items): eclectic profiles are somehow disadvantaged.

Another problem, shown experimentally, is that the dynamics inherently favors the creation of hits, i.e. items which are liked by a huge fraction of the population. Of course, the existence of hits is not a bad thing in itself, but hits nevertheless limit the probability of other items to "survive" in a world dominated by weight sums.

In short, collaborative filtering is a means of building similarity relations between items, based on statistical properties of groups of agents. As such, it addresses the goal of surprise, in a safe way, by proposing users items which are similar to already known ones. However, cluster formation and uneven distribution of chances to items (e.g. formation of hits) are the main drawbacks of the approach, both from the user's viewpoint (clusters from which it is difficult to escape), and the content provider's viewpoint (no systematic exploitation of the catalogue).

4.3 On-the-fly Music Program Generation

The RecitalComposer Project (Pachet et al., 1999) is based on a radically different approach to music selection: instead of proposing users sets of individual titles, we propose to build fully-fledged music programs, i.e. sequences of music titles.

There are several motivations for producing music programs, rather than unordered collections of titles. One is simply based on the recognition that music titles are rarely listened to in isolation: CDs, radio programs, concerts are all made up of temporal sequences of pieces, in a certain order. This order is most of the time significant, i.e. different orders do not produce the same impressions on listeners. In a way, the whole craft of music program selection is precisely to build coherent sequences, rather than simply select individual titles.

The second motivation is that properties of sequences play an important role in the perception of music: for instance, several music titles in a similar style convey a particular atmosphere, and create expectations for the next coming titles. As a consequence, an individual title may not be particularly enjoyed by a listener *in abstracto*, but may be the *right piece at the right time* within a sequence.

Rather than focusing on similarity of individual titles, we can exploit properties of sequences to satisfy the three goals of music selection. The proposal is therefore the following. First we build a database of titles, with content information for each title. Then we specify music programs by giving the properties or patterns we want the program to have. These properties are represented as constraints, in the sense of constraint satisfaction techniques. Finally, a constraint solver computes the solutions of the corresponding combinatorial pattern generation problem.

The problem, as we define it, is therefore to build music programs, seen as temporal sequences of titles, in order to satisfy the three goals of music selection problem: repetition, surprise, and full exploitation of the catalogue. As an example, we will take a music program for which we specify the desired properties. In the next sections, we will focus on the format of the database and the nature of constraints.

Here is a "liner-note" like description of a typical music program. The properties of the sequence may be grouped in three categories: 1) user preferences, 2) global properties on the coherence of sequences, and 3) constraints on the exploitation of the catalogue. The following example describes a music program called "Driving a Car", ideally suited for listening to music in a car:

User preferences

Note that these constraints specify global properties of the sequence, and do not specify the position of items in the sequence:

- No slow/very slow tempos (Cardinality Constraint)
- At least 30% female-type voice
- At least 30% purely instrumental pieces
- At least 40% brass
- At most 20% "Country Pop" style
- One song by "Harry Connick Jr".

Constraints on the coherence of the sequence

- Styles of titles are close to their neighbors (successor and predecessor). This is to ensure some sort of continuity in the sequence, style-wise.
- Authors are all different.

Constraints on the exploitation of the catalogue

- Contains twelve different pieces. This is to fit on a typical CD or minidisk format.
- Contains at least 5 titles from the label "Epic/Sony Music". This is a typical bias to exploit the catalogue in a particular region.

4.4 Database of Music Titles

The database required for building music programs contains content information needed for specifying the constraints. More precisely, each item is described by a set of attributes, which take their value in a predefined taxonomy. The attributes are of two sorts: technical attributes and content attributes.

Technical attributes include the name of the title (e.g. "Learn to love you"), the name of the author (e.g. "Connick Harry Jr."), the duration (e.g. "279 sec"), and the recording label (e.g. "Epic/Sony Music"). Content attribute are typical meta-data: they describe musical properties of individual titles. The attributes are the following: *style* (e.g. "Jazz Crooner"), *type of voice* (e.g. "muffled"), *music setup* (e.g. "instrumental"), *type of instruments* (e.g. "brass"), *tempo* (e.g. "slow-fast"), and other optional attributes such as the *type of melody* (e.g. "consonant"), or the main *theme* of the lyrics (e.g. "love").

In the current state of our project, the database is created by hand, by music experts (including the third author). However, it should be noted that 1) some attributes could be extracted automatically from the signal, such as the tempo, see e.g. (Scheirer, 1998) and 2) all the attributes are simple, in the sense that they do not require sophisticated musical analysis to be filled.

An important aspect of the database is that the values of content attributes are linked to each other by similarity relations. These similarity relations are used for specifying constraints on the continuity of the sequence. For instance, the preceding example contains a constraint on the continuity of styles. More generally, the taxonomies on attributes values allow to establish links of partial similarity between items, according to a specific dimension of musical content.

Some of these relations are simple ordering relations. For instance tempos can take their value in the ordered list (fast, fast-slow, slow-fast, slow). Other attributes such as *style*, take their value in full-fledged taxonomies. The taxonomy of styles is particularly worth mentioning, because it embodies a global knowledge on music which is a clear added value for the system.

Various taxonomies of musical styles have been designed, particularly by internet music retailers, such as Amazon (1998) or MusicBoulevard (1998). However, these classifications are mainly designed with a query-based approach. For example the taxonomy of styles proposed by Amazon is a tree-like classification oriented toward presentation of items in a search-oriented way. This taxonomy embodies a relation of "generalization/specialization" between styles: "Blues" is more general than "Memphis Blues". As such, it is well suited for navigating in the catalogue to find under-specified items. However, it does not represent similarities between styles, for instance, similarities between styles that have common origins, like, say, "Soul-Blues" and "Jazz-Crooner".

Conversely, we designed a taxonomy of styles representing explicitly relations of similarity between styles. Our taxonomy is a non-directed graph in which vertices are styles and edges express similarity. It currently includes 120 different styles, covering most of western music. A part of the graph is represented in Figure 3.

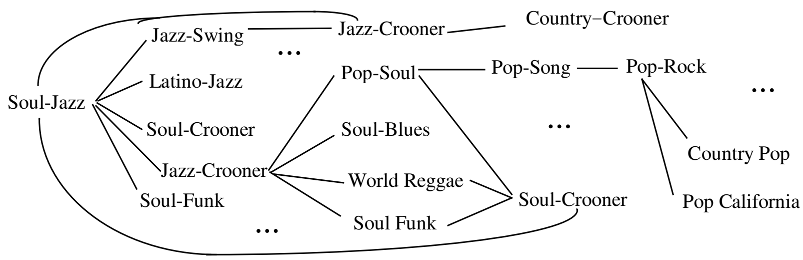

Figure 3. A part of a taxonomy of musical styles. Links indicate a similarity relation between styles. "Jazz-Crooner" is represented as similar with "Soul-Blues".

4.5 Services and Interface Issues

Computing music programs from a database and a set of constraints is shown to be a complex combinatorial problem. Constraint satisfaction techniques may be used to solve it, as explained in (Pachet & al., 1999).

The resulting technique can be used to build a number of services related to music delivery with large-scale music catalogues. We list here examples of currently built applications: automatic CD assembly, a Path Builder and a Baroque recital composer. Other applications are envisaged for set-top-boxes services and digital audio broadcasting which we do not detail here for reasons of space.

- Sampler Builder

The simplest application of this technology is a system targeted at music professionals for building music programs (so-called samplers) from a given database. In the application, the user can specify the constraints using an interface, and launch the system on a database. This system is aimed at professionals who want to express explicitly all the properties of the desired programs, and thus have full control on all the constraints.

- Progressive programs

In this scheme, the user only specifies the stylistic structure of the program: the genres of the beginning, middle and end. This may be used for instance for creating long programs for parties, in which you know in advance the structure (e.g. begin with Pop, then Rock, then Slows, etc.).

- Path across different styles

Services dedicated to average end users should allow them to express only their preferences, possibly using automatic profiling systems, and contain predefined, fixed constraints for the coherence properties and catalogue exploitation, according to predetermined ambiences or configurations. A typical configuration is a path between two titles. In this scheme, the user can specify a starting title and an ending title. The system contains hidden constraints on continuity of genres, and tempos are fixed. For instance, find a continuous path between Céline Dion's "All by myself", and Michael Jackson's "Beat it" (see Figure 4).

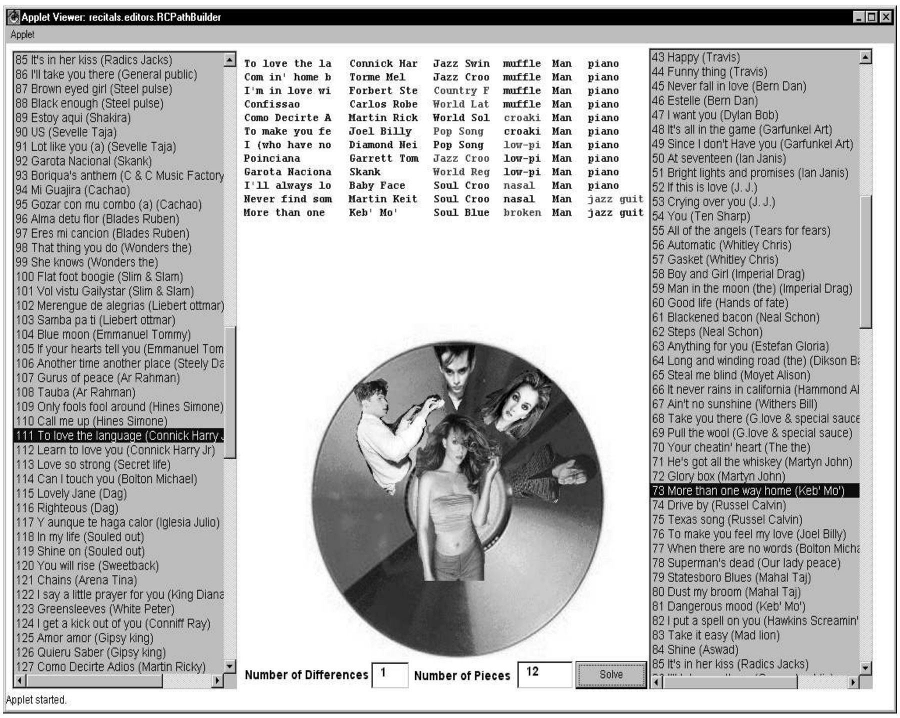

Figure 4. The PathBuilder program. The user chooses a starting and ending title, as well as a degree of tightness between successive titles.

- Specific music domains

The approach can be used to produce music programs in specific styles, by adding domain specific constraints. A prototype application dedicated to Baroque music implemented in our lab allows to build various "recitals" in the domain of Baroque harpsichord music. Baroque music is a good example of a specific domain, because recitals of Baroque music (XVII[th] century) follow rules identified by musicologists (Bukofzer, 1947), while allowing a great deal of freedom to performers. A typical rule concerning the structure of recitals is the "continuity of tempos" between consecutive pieces. More specific rules are also used, such as rules on the tonality: at this period of musical history, recitals where allowed to modulate - i.e. change tonality - only once. Other constraints concern the structure of the recital (introductory part with necessary piece types), as well as necessary alternation of piece types.

The system allows the user to create and listen to different music programs, while ensuring the consistency of these programs, according to the rules of the structure of recitals. The database contains titles with content description adapted to the domain. For instance, attributes such that "type" (e.g. "Gigue", "Chaconne", etc.), "tonality" and "density" are added to the database for describing relevant aspects of titles. The constraint system contains the constraints corresponding to the rules described above. The resulting system allows to produce a great number of different recitals, which all have the desired properties of "good" recitals, in the style of the composer's time (see Figure 5).

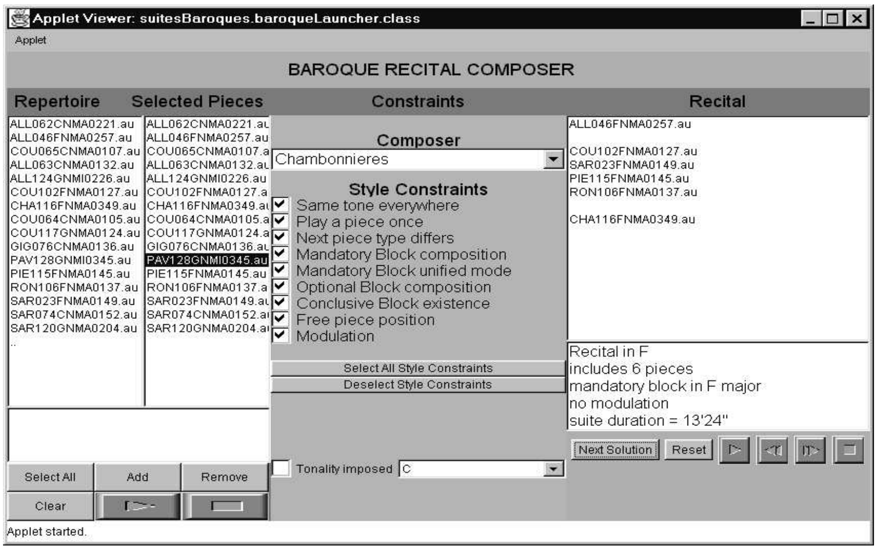

Figure 5. The interface for *BaroqueComposer*. The user can select a Baroque composer, and then a corresponding catalogue of pieces of Baroque music. He/she can then build music programs which satisfy the constraints of the Baroque style, and listen to them in the right order.

This kind of service lies between two extreme bounds: fixed order and randomness. On the one hand, a CD played in a standard fashion contains a fixed music program. On the other hand, a common feature of CD players (or Juke boxes) is the "random" selection button, which chooses at random between different CDs and between the titles of the CDs. Constraint techniques provide an intermediary degree of control between these two extremes, where the user can still express some preferences, but the system computes a program which yields properties of coherence.

François Pachet, Active Listening : What is in the Air ? Sony CSL internal Report, 1999.

4.6 Exploration Space

RecitalComposer is an enabling technology for building high-level music delivery services exploiting large-scale music catalogues. The system is based on the idea of creating explicit sequences of items, specified by their global properties, rather than computing sets of items satisfying queries. One of its main advantages over query-based or collaborative filtering approaches is that it produces “ready for use” ordered sequences of items, which satisfy the three goals of music selection, i.e. repetition, surprise, and exploitation of catalogues. It creates “coherent” music programs from user specifications, where the coherence is specified in terms of meta-data on music titles and as such can be seen as another example of “semantic” control, where the semantics is the structure of music programs. Compared to the juke box of the 20s, it allows to access much larger music catalogues with simple controls (e.. user preferences) which, once again, make sense, without requiring an *a priori* knowledge of the underlying music catalogue.

5. Conclusion

The new landscape of digital multimedia opens new doors for interactive listening environments which provide richer musical experiences. We have argued that such environments require some sort of semantic preserving systems. We have illustrated this idea with two projects currently developed at Sony CSL, in the areas of sound spatialization, and content-based music selection. In both cases, the technology of constraints is proposed for representing these “seeds of semantics”, that yield exploration spaces with meaningful controls. A lot remains to be done, in other areas of music listening and perception, but these projects already suggest that the traditional borders between composition, production and listening may have to be redefined. In particular, a question which arises is what kind of music composers will make, if they know that listeners have active listening devices at home ? If we know what technology is in the air, what music will be in our ears ?

References

Amazon Music Store web site, http://www.amazon.com, 1998.

Assayag G., Agon C., Fineberg, J., Hanappe P., “An Object Oriented Visual Environment For Musical Composition”, *Proceedings of the International Computer Music Conference*, pp. 364-367, Thessaloniki, 1997.

Bukofzer, M. *Music in the Baroque Era, from Monteverdi to Bach*, W.W. Norton & Company, 1947.

Cavalli-Sforza, L. and Feldman, M. *Cultural Transmission and Evolution: a Quantitative Approach*, Princeton University Press, 1981.

Cruz-Neira, C., Leight, J., Papka, M., Barnes, C., Cohen, S.M., Das, S., Engelmann, R., Hudson, R., Roy, T., Siegel, L., Vasilakis, C., DeFanti, T.A., Sandin, D.J., “Scientists in Wonderland: a Report on Visualization Applications in the CAVE Virtual Reality Environment”, Proc. IEEE Symp. on Research Frontiers in V.R., pp. 59-66, 1993.

Eco, U. Opera Aperta. Bompiani (Milan), 1962.

François Pachet, Active Listening : What is in the Air ? Sony CSL internal Report, 1999.

Eckel G., “Exploring Musical Space by Means of Virtual Architecture”, *Proceedings of the 8th International Symposium on Electronic Art*, School of the Art Institute of Chicago, 1997.

Epstein, Joshua M. *Growing Artificial Societies: Social Science from the Bottom Up*, MIT Press, 1996.

Firefly web site, http://www.firefly.com, 1998.

Infoglide web site, http://www.infoglide.com, 1998.

Lepain, P. Ecoute interactive des documents musicaux numériques, in *Recherches et Applications en Informatique Musicale*, Chemillier & Pachet Eds, Hermes, Paris, 1998.

Meyer, L. *Emotions and meaning in Music*, University of Chicago Press, 1956.

Memon, N, Wong, P. W. “Protecting Digital Media Content”, *Communications of the ACM*, July 1998, pp. 34-43, 1998.

MPEG7 Requirements Group, “MPEG-7 Requirements Document”, Doc. ISO/MPEG N2461, MPEG Atlantic City Meeting, October 1998.

MyLaunch web site: www.mylaunch.com, 1998.

Narmour, E. *The analysis and cognition of melodic complexity*. University of Chicago Press, 1992.

Orlarey Y., Fober D., Letz S., “Elody : a Java + MidiShare Based Music Composition Environment”, *Proceedings of the ICMC*, 1997.

Pachet, F. Delerue, O. “A Constraint-based Temporal Music Spatializer”, *ACM Multimedia Conference*, Brighton (UK), pp. 351-360, 1998.

Pachet, F. Roy, P. Cazaly, D. “A Combinatorial Approach to Content-based Music Selection”, *IEEE International Conference on Multimedia Computing and Systems*, Firenze (Italy), 1999.

Read, Oliver and Walter Welch. From Tin Foil to Stereo: Evolution of the Phonograph. Indianapolis, 1959, 2nd edition 1976.

Rosen, C. *The Frontiers of Meaning: Three Informal Lectures on Music*, Hill & Wang Pub, 1994.

Shardanand, U. and Maes, P. “Social Information Filtering: Algorithms for Automating "Word of Mouth"”, *Proceedings of the 1995 ACM Conference on Human Factors in Computing Systems*, pp. 210-217, 1995.

Scheirer, E. D. (1998). “Tempo and beat analysis of acoustic musical signals”, *Journal of the Acoustical Society of America*, 103(1): 588-601.

Sheirer, E. Väänänen, R. Huopaniemi, J. “AudioBiffs: The MPEG-4 Standard for Effects Processing”, *Proceedings of the first Digital Audio Effects Workshop*, Barcelona, pp. 159-167, November 1998.

Smith, D. Melara, R. “Aesthetic preference and syntactic prototypicality in music: ‘Tis the gift to be simple”, *Cognition*, 34, pp. 279-298, 1990.

Taube H., “Common Music: A Music Composition Language in Common Lisp and CLOS”, *Computer Music Journal*, vol. 15, n° 2, 21-32, 1991.