

THE CUIDADO MUSIC BROWSER: AN END-TO-END ELECTRONIC MUSIC DISTRIBUTION SYSTEM

François Pachet, Amaury Laburthe, Jean-Julien Aucouturier
SONY Computer Science Laboratory
6, rue Amyot 75005 Paris, France
+33 144 080 516
pachet | amaury | jj @csl.sony.fr

ABSTRACT

The IST project Cuidado, which started in January 2001, aims at producing the first entirely automatic chain for extracting and exploiting musical metadata for browsing music. The Sony CSL laboratory is primarily interested in the context of popular music browsing in large-scale catalogues. First, we are interested in human-centred issues related to browsing “Popular Music”. Popular here means that the music accessed to is widely distributed, and known to many listeners. Second, we consider “popular browsing” of music, i.e. making music accessible to non specialists (music lovers), and allowing sharing of musical tastes and information within communities, departing from the usual, single user view of digital libraries. This research project covers all areas of the music-to-listener chain, from music description - descriptor extraction from the music signal, or data mining techniques -, similarity based access and novel music retrieval methods such as automatic sequence generation, and user interface issues. This paper describes the scientific and technical issues at stake, and the results obtained, in the current state of the IST project.

1. INTRODUCTION

1.1. Existing Popular Music Access Systems

There are now many on line searchable music databases. We can classify them in the following categories.

First, purely editorial systems propose systematic editorial information on popular music, including albums track listings (CDDB, Musicbrainz), information on artists and songs (AMG and Muze). This information is created by music experts, or in a collaborative fashion (CDDB, Musicbrainz). These systems provide useful services for EMD systems, but cannot be considered as fully-fledged EMD systems *per se*, as they provide only superficial and incomplete information on music titles, supposed to exist somewhere else.

The MoodLogic browser proposes a complete solution for Popular Music access. The core idea of MoodLogic is to associate metadata to songs automatically thanks to

two basic techniques: 1) an audio fingerprinting technology able to recognize music titles on personal hard disks, and 2) a database collecting user ratings on songs, which is incremented automatically, and in a collaborative fashion. An ingenious proactive strategy is enforced to encourage users to rate songs, in order to get tokens that allow them to get more metadata from the server. Moodlogic relies entirely on metadata obtained from user ratings and does not perform any acoustic analysis of songs. However, collaborative music rating does not exhaust the description potential of music, and our Browser proposes many other types of metadata.

Other proposals have been made either for fully-fledged music browsers, or for ingredients to be used in browsers (fingerprinting techniques, collaborative filtering systems, metadata repositories, e.g. Wold et al. 1996) that we cannot cover here for reasons of space. We will describe in this paper only the parts of our project that we think are original and may contribute to address the needs of our targeted users.

1.2. The Cuidado Music Browser

The Cuidado music browser aims at developing all the ingredients of the music-to-listener chain, for a fully-fledge content-based access system. More precisely, the project covers the areas of 1) editorial metadata, 2) acoustic metadata, 3) metadata exploitation and browsing tools, 4) management and share of metadata among users. The next sections describe the most important results obtained for each of these aspects.

2. EDITORIAL METADATA

To manage collections of music titles an application must have access to many information to identify, categorize, index, classify and generally organize music titles.

We consider here two types of data as editorial metadata:

- Consensual information or facts about music titles and artists,
- Content description of titles, albums or artists.

The first category is common to already existing EMD systems and does not raise any particular problem, as this information is universal by nature. It includes for

instance: artist and songs name, albums and tracks listing, group members, date of recording for a given title, short biography for artists with date of birth, years of activity, etc.

The second category is more problematic. Content description includes such widely needed information as artist style, artist instruments, song mood, song review, song or artist genre and more generally attributes aiming at describing the intrinsic nature of the musical item at stake (artist or song). These descriptions are useful to the extent that they can be used for musical queries in large catalogues. The music browser enables to issue queries for both categories.

Furthermore, the music browser has a tool (see figure 1) devoted to editorial information management. The global architecture of the system is detailed in section 6. This tool allows editing and adding artists and/or songs properties.

2.1. Editorial metadata philosophy

Editorial metadata are associated distinctly with music titles and artists.

Artists (taken in the most general sense) are key *music identifiers* for many users: Yesterday is by “The Beatles”, and “The 5th symphony” is by Beethoven. Artists are used also for solving ambiguity: “With a Little Help from my Friends” by the Beatles, is definitely not the same tune as the version by Bruce Springsteen. The “Stabat Mater” by Pergolese is not the one by Boccherini, etc. We call these artists “primary artists” as they are most commonly used to identify music titles. These examples show that primary artists are common ways of identifying music titles but also that the role of primary artists changes with styles: in Classical music, primary artists are usually composers. In non Classical music they are usually performers. In our Browser, we introduced the notion of primary artists in a deliberate ambiguous way, to cope for Classical and non Classical music in a uniform way.

There are cases where primary artists are not enough for characterizing the identity of a piece. The “1st partita” of Bach has been recorded by Glenn Gould, and also by many other pianists, and this distinction is of course very important: not only for interpreters, but also for conductors (for orchestral pieces). In non-Classical music the need for secondary artists is also obvious, for instance to indicate that the Springsteen version of “A little help” is indeed a Beatles song.

Existing repositories of editorial information do not provide systematic schemes for accessing artists and their relations to songs. This led us to constitute a database of artists, or more generally of “Musical Human Entities” (MHE), including both performers, composers, but also groups (the Beatles), orchestra (the Berlin Philharmonic), duets (Paul McCartney & Michael Jackson). To each artist (or MHE) is associated a limited but useful set of properties in fixed ontologies: type (composer, singer,

instrumentist, etc.), country of origin, language (for singers), type of voice (for singers also), main instrument (for instrumentists). Other information concern the relation MHE entertain with each other. For instance, Paul McCartney is a *memberOf* The Beatles, and artist Phil Collins a *memberOf* the group Genesis. The Editorial MHE database may be seen more as a knowledge base than a database.

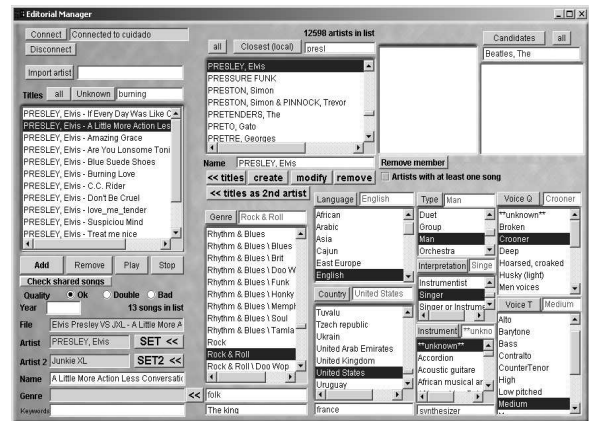


Figure 1 – The editorial data management panel

Concerning music titles, our tool enables basic editions as title name or keywords, as well as less obvious features such as title genre, primary and secondary artist introduced before.

Both artists and songs can be associated with a specific genre. Genres are badly needed for accessing music, and are as badly ill-defined. Our studies on existing taxonomies of genres have shown that there is no consensus, and that a consensus is probably impossible [4]. However, we propose here several ways to partially solve this problem. After several years of trials [20] and errors, we ended up with a simple two-level genre taxonomy consisting of 250 genres. The main property of this taxonomy is flexibility: users can classify artists or songs either in a generic way (Classical, Jazz), more precisely (Jazz / BeBop, Classical/Baroque). However, simpler taxonomies may also produce frustration, as some categories may contain artists or songs that users would consider very different. To make our taxonomy more flexible, we have introduced an optional “keyword” field, which may contain free words. These words may be entered by users to further refine their own classification perspective on artists or songs. This simple yet flexible approach has the advantage of uniformity: artists and songs are classified in the same taxonomy, allowing for various degrees of precision. For instance, The Beatles is classified in “Pop / Brit”, but Beatles songs may be classified in other genres (e.g. “Revolution 9” is “Rock / Experimental”).

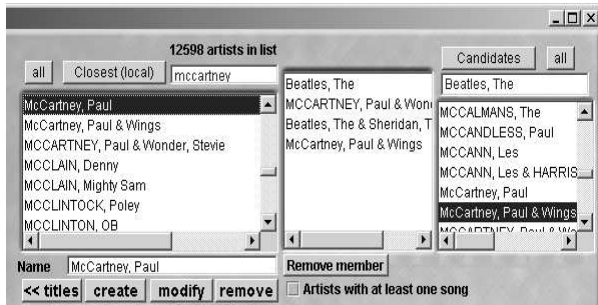


Figure 2 – the “member_of” predicate

3. ACOUSTIC METADATA

The main type of metadata that the MB proposes for songs besides editorial information is acoustic metadata, i.e. information extracted from the audio signal. The Mpeg7 standard aims at providing a format for representing these information, and a specialized audio group produces specific constructs to represent musical metadata [7]. However, music metadata in Mpeg7 refers in general to low-level, objective information that can be extracted automatically in a systematic way. Typical descriptors (called LLD for Low-Level Descriptors in the Mpeg7 jargon) proposed by Mpeg7 concern superficial signal characteristics such as Means and Variance of amplitude, spectral frequencies, spectral centroid, ZCR (Zero Crossing Rate), etc.

Concerning high-level descriptors that can be mapped to high-level perceptual categories, Mpeg7 is strictly concerned with the format for representing this information, and not the extraction process *per se*.

3.1. Extracting High-Level Music Percepts

We have conducted in the project several studies focusing on particular dimensions of music that are relevant in our context.

3.1.1. Rhythm

We have proposed a rhythm extractor [16], that is able to extract the time series of percussive sounds in music

signals of popular music. Rhythm information is a useful extension of tempo or beat, as proposed by Scheirer in [13]. However, many things remain to be done in the field of rhythm. One key issue seems to rely not so much in how to extract rhythm, but how to exploit the information: most people are unable to describe rhythm with words, and even less to produce rhythm (our attempts at designing a query by rhythm did not prove successful).

3.1.2. Energy

In [17], we have addressed another dimension of music pertaining to popular music access, the perceptual energy, i.e. whether a song is thrilling and exciting (e.g. hard rock, dance music), or relaxing and calm (e.g. a piano piece by Schumann).

We have studied the correlation of experimental measures (user tests) with a variety of signal features, such as tempo, raw signal energy, spectral analysis, the associated variances, correlations... as well as their linear combinations (using discrimination analysis) and their possible compositions with signal operators (filters, etc...). The most discriminative parameter we found is $\log_{10}(\text{var}(\text{diff}(x^2)))$, which gave a classification error of 22% on the validation set.

3.1.3. Timbre

In [2], we have proposed to describe music titles based on their global *timbral quality*. Our motivation is that, although it is difficult to define precisely music taste, it is quite obvious that music taste is often correlated with timbre. Some sounds are pleasing to listeners, other are not. Some timbres are specific to music periods (e.g. the sound of Chick Corea playing on an electric piano), others to musical configurations (e.g. the sound of a symphonic orchestra). In any case, listeners are sensitive to timbre, at least in a global manner.

We model the global “sound” of a music title as a distribution in the space of mel cepstrum coefficients (MFCC). MFCCs provide a compact representation of the signal’s spectral envelopes, which are a good correlate of the timbre. By comparing timbre distributions between titles, it is then possible to match music titles of possible very different genres based solely on their timbral color. Figure 3 shows a 3D projection of the feature space (which is originally of dimension 8), showing two distributions of mfccs, each modelled with a mixture of 3 gaussian distributions (GMM). The light-gray GMM is the timbre model of the song “The Beatles – Yesterday”, and the dark-gray GMM is the timbre model of the song “Joao Gilberto – Besame Mucho”. This two songs have a very similar “sound” (acoustic guitar and a string quartet, plus a gentle and melancholic male voice), and indeed we see that their mfcc distributions are very close.

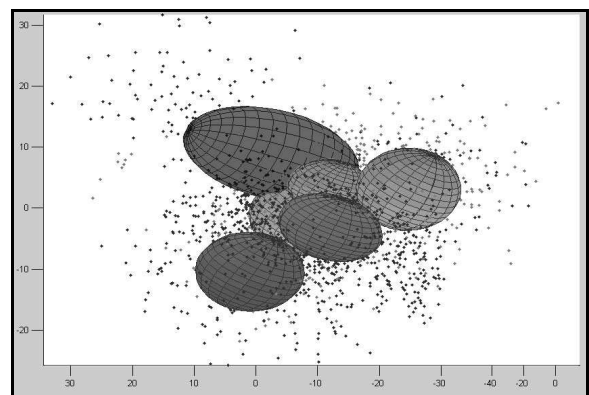


Figure 3: comparison of the timbre models of two songs: "The Beatles - Yesterday" and "Joao Gilberto – Besame Mucho"

3.2. EDS: A General Framework for Extracting Extractors

These various studies in descriptor extraction from acoustic signals have shown that the design of an efficient acoustic extractor is a very heuristic process, which requires sophisticated knowledge of signal processing, intuitions, and experience. Indeed, most approaches in feature extraction as published in the literature consist in using statistical analysis tools to explore spaces of combinations of LLD. The approaches proposed by Peeters [12] and Scheirer [14] typically fall in this category. However, these approaches are not capable of yielding precise extractors, and depend on the nature of the palette of LLD, which usually do not capture the relevant, often intricate and hidden characteristics of audio signals. Consequently, designing extractors is very expensive and hazardous.

On the other hand, user studies have shown that there is a virtually infinite number of extractors of musical attributes that could be useful in EMD systems. Different users have different needs: one – say, a jazz musician – might be interested in listening to songs which exhibit a particular chord sequence, another may be interested by the sound (“some saturated guitar with a little bit of chorus”), while another simply wants to find “funky” music for his birthday party. Even when talking about the same attribute, the definitions (i.e. in terms of pattern recognition, the training sets) vary a lot. The perception of “harmonic complexity” of a tune for instance highly depends on the musical expertise of the listener.

These experiments have given rise to a systematic approach to feature extraction, embodied in the EDS system [11]. Departing from the usual LLD approach, the idea of EDS is to automate – in part or totally – the process of designing extractors. EDS searches in a richer and more complex space of signal processing functions, much in the same way than experts do: by inventing functions, computing them on test databases, and modifying them until good results are obtained.

To reach this goal EDS uses a genetic programming engine, augmented with fine grained typing system, which allows to characterize precisely the inputs and outputs of functions. EDS uses also rewriting rules to simplify complex signal processing functions (see the example of the Perceval equality being used by EDS to simplify the expression in Figure 4). Finally EDS uses expert knowledge to guide its search, in the form of heuristics.

Typical heuristics include “do not try functions which contain too many repetition of the same operator”, or “apply twice a fft on a signal is interesting, but not 3 times”, or also “spectral coefficients are particularly useful when applied on signals in the temporal domain, possible filtered”, etc.

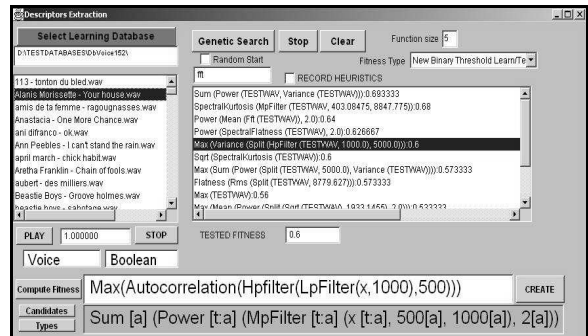


Figure 4: Screenshot of the EDS system.

The current extractors targeted by EDS are Perceptual Energy (or a refinement of the descriptor we designed by hand), discrimination between songs and instrumental, discrimination between studio and live versions of songs, harmonic complexity, etc. The ambitious goal of EDS makes it a project in itself, as it aims at capturing complex knowledge, in an expanding field. However, we think that the contribution to the MIR community is potentially important as it is a first step towards a unified vision of high level audio feature extraction.

4. SIMILARITY

The notion of similarity is of utter importance in the field of music information retrieval, and the expectation to have systems that find songs that are “similar” to one or several seed songs is now second nature. However, here again, similarity is ill-defined, and it can be of many different sorts. For instance, one may consider all the titles by a given artist as similar. And they are, of course, artist-wise. Similarity can also occur at the feature level. For instance, one may consider that Jazz saxophone titles are all similar. Music similarity can yet occur at a larger level, and concern songs in their entirety. For instance, one may consider Beatles titles as similar to titles from, say, the Beach Boys, because they were recorded in the same period, or are considered as the same “style”. Or two titles may be considered similar by a user or a community of users for no objective reason, simply because they think so.

4.1. Acoustic similarity

Feature-based similarity is trivially obtained by defining similarity measures from the metadata obtained and described above, either editorial or acoustic. Most descriptors yield implicit similarity measures that can be useful in some circumstances, e.g. similarity of tempo, of energy, or similarity based on artist relationships, etc.

search systems to innovative graphical representations of play lists. For instance, Gigabeat display music titles in spirals to reflect similarity relations titles entertain with each other. The gravitational model of SmartTuner of mzz.com, represent titles as mercury balls moving graciously on the screen, to or from “attractors” representing the descriptors selected by the user. However gracious, these interfaces impose a fixed interaction model, and assume a constant attitude of users regarding exploration: either non-exploratory - music databases in which you get exactly what you query - or very exploratory. But the users may not choose between the two, even less adjust this dimension to their wish. The current interface of the Music Browser aims at allowing users to choose between many modes of music access: explorative, precise, focused or hazardous.

5.1. Focused interfaces

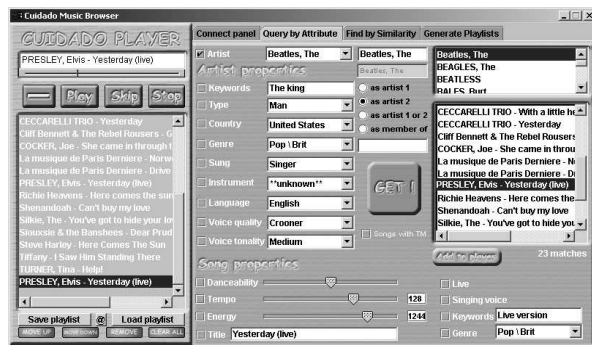


Figure 7 - Screenshot of the query panel in the Music Browser

The query panel (figure 7) is mostly dedicated to focused search in the database. In this panel users can issue queries on all available artists and songs metadata. These metadata can be editorial: artists’ names, songs’ names, voice quality, etc. as well as computed: subjective energy, tempo, etc. The result of a query is a music titles list. Then this result set can be further filtered to return only songs with fast tempo, or only songs with a male singer. This result list can be transferred to the player for listening/exporting purpose

5.2. Explorative interfaces

5.2.1. Sliding between similarities

An interesting issue resulting from the studies on feature-based and cultural similarities is the comparison between these different sorts of similarity. For instance in Figure 5, a starting title such as “Le moment de vérité” played by Ahmad Jamal, is considered by the MB as similar timbre-wise to “Humoresque Op. 20” by Schumann or “Blue and sentimental” by Hank Jones, but culturally, it is closer to “Ahmad’s blues” by Miles Davis, because of the strong relationship between these two players, captured by the web crawler. Of course, there is no grounded truth here, and all these similarities are

relevant. The next issue to solve is to aggregate these similarities, or at least propose users simple and meaningful ways of exploiting these different techniques. In [2], we have proposed an interface, the “aha slider”, which allows the user to rank the results of a query according to two possibly orthogonal types of similarity. The slider is simply a way to filter the result set of one similarity according to the values of the second similarity measure. For instance, one can ask for timbrally similar songs which are also very close according to cultural similarity (e.g. “Ahmad’s blues” by Miles Davis), or, on the contrary, filter the result set so that it only contains songs which are culturally very distant from the query (e.g. Schumann or William Sheller).

This interface attempts to give the user full control over the degree of surprise and freedom in the way the system satisfies his request. A non-exploratory behavior (e.g. culturally similar) implies that the system should return exactly the answer to the query, or an answer that is as expected as possible (same title, same artist). An exploratory behavior (e.g. culturally distant) consists in letting the system try different regions of the catalogue rather than strictly match the query.

5.2.2. Playlist Generation

An original feature introduced by the Browser is a powerful playlist generation system, based on constraint satisfaction techniques ([2] & [3]). This technique allows user to get entire music playlists from a catalogue, by specifying only abstract properties on the playlist, such as:

- the playlist should contain 12 different titles,
- the playlist should not last more than 76 minutes,
- the genre of a title should be *close* to the genre of the next title,
- the playlist should contain at least 60% of *instrumental* titles,
- the sequence should contain titles with increasing tempo, etc.

The problem of generating such playlists given a very large title catalogue with musical metadata, and a set of arbitrary constraints is a NP-hard combinatorial problem. Moreover, in the case of a contradictory set of constraints, there may not be an exact solution. An ideal system should therefore be able to generate good approximate compromises. The Cuidado Music Browser is able to generate such playlists automatically (figure 8), using a fast algorithm based on adaptive search [3].

We give here an example of a 5-title playlist with the following constraints:

- 1- Timbre continuity: the playlist should be timbrally homogeneous, and shouldn’t contain abrupt changes of textures.

- 2- Genre Cardinality: the playlist should contain 30% of Rock pieces, 30% of Folk, and 30% of Pop
- 3- Genre Distribution: the titles of the same genre should be as separated as possible.



Figure 8 - Screenshot of the playlist generation system

One solution found by the system is the following playlist:

- Rolling Stones – You Can’t always get what you want - Genre = Pop/Blues
- Nick Drake - One of these things first - Genre = Folk/Pop
- Radiohead - Motion Picture Soundtrack - Genre = Rock/Brit
- The Beatles - Mother Nature's Son - Genre = Pop/Brit
- Tracy Chapman - Talkin' about a Revolution - Genre = Folk/Pop

6. ARCHITECTURE

This section describes the general architecture of the Music Browser (Figure 9). The central element of the architecture is the metadata server. This server is a MySQL database hosted on a SQL server. The server acts both as a server for Php scripts and servlets. The MusicBrowser is implemented in Java and communicates with the MySQL database using JDBC drivers. The metadata server runs a Php server accessible over the Internet. Specific Php scripts allow client applications to fetch and submit metadata to this server.

The music browser contains four panels aimed at music title access: the player, the query panel, the similarity panel and the playlist panel.

Additionally, the browser includes Two management tools: the editorial data management tool and the extractor and computation management tool. The purpose of the computation management tool is to compute descriptors for the songs in the database as well as distance matrix. It can use any stand-alone extractor

(exe or bat files) developed by third party. The editorial metadata management tool is used to manage artists and songs properties. It provides choice lists for each property and enables basic editions such as title name or keywords, as well as title genre, primary and secondary artist, as described in section 2.1. This tool interacts on-line with our metadata server.

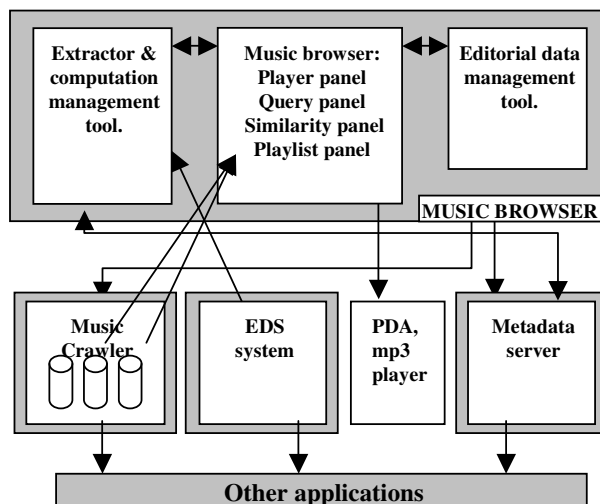


Figure 9 – General interaction of all components

Lastly, with the apparition of ad-hoc networks, users can share their data easily with other users and in a transparent way. This situation raises an issue in the management and synchronization of metadata. We describe in [18] a solution to allow both private and shared metadata to coexist in a single environment.

7. CONCLUSION

The Cuidado music browser is the first large scale, fully content-based music access system. It includes all the technologies needed to extract descriptors, create similarity relations, and make these information easily available to users. In this final stage of the project, the system is fully operational, and user tests have started to assess the usability of content information for music access. Further work concentrates on scalability issues, in particular the speed up of extraction algorithms and similarity matrix computation.

8. REFERENCES

- [1] Allamanche, E. Herre, J. Helmuth, O. Frba, B. Kasten, T. and Cremer, M. (2001) "Content-Based Identification of Audio Material Using MPEG-7 Low Level Description" in Proc. of the Int. Symp. of Music Information Retrieval.
- [2] Aucouturier, J.-J. Pachet, F. (2002) Music similarity measures: what's the use ? Proc. of the 3rd Int.

Pachet, F. Laburthe, A. Aucouturier, JJ. The Cuidado Music Browser. Proc of CBMI 03, Rennes (Fr).

- Symp. of Music Information Retrieval (ISMIR02), Paris, October.
- [3] Aucouturier, J.-J. Pachet, F. (2002) Scaling up Playlist generation, In Proc. of the IEEE Int. Conf. on Multimedia and Expo (ICME 02), Lauzanne, October.
- [4] Aucouturier, J.-J. Pachet, F. (2003) Musical Genre: a Survey, In Journal of New Music Research, 32:1.
- [5] Belkin, N. (2000) Helping people find what they don't know. *CACM* Vol. 43, N. 8, August 2000, pp. 58-61.
- [6] Cohen, W., Fan, W. (2000) Web-Collaborative Filtering: Recommending Music by Crawling The Web, 9th International World Wide Web Conference, Amsterdam.
- [7] Herrera, P, Serra, X. Peeters, G. (1999). Audio descriptors and descriptors schemes in the context of MPEG-7. Proceedings of the 1999 ICMC, Beijing, China, October.
- [8] Pachet, F. Roy, P. Cazaly. D. (2000) "A Combinatorial approach to content-based music selection". *IEEE Multimedia*, pp. 44-51, March.
- [9] Pachet, F. and Laigre, D. (2001) A Naturalist Approach to Music File Name Analysis. in Proceedings of 2nd Int. Symp. on Music Information Retrieval.
- [10] Pachet, F. Westerman, G. Laigre, D. (2001) Musical Data Mining for Electronic Music Distribution., Proceedings of WedelMusic 2001, Firenze (It).
- [11] Pachet, F. and Zils, A. (2003) "EDS: Evolving High Level Music Descriptors", Sony-CSL Tech. Report 2003-01-01.
- [12] Peeters, G. Rodet, X. (2002) Automatically selecting signal descriptors for sound classification. Proceedings of the 2002 ICMC, Goteborg (Sweden), September.
- [13] Scheirer, Eric D. (1998) "Tempo and beat analysis of acoustic musical signals", *J. Acoust. Soc. Am. (JASA)* 103:1 (Jan 1998), pp 588-601.
- [14] Scheirer, Eric and Slaney, Malcolm. Construction and evaluation of a robust multifeature speech/music discriminator. Proc ICASSP '97, pp. 1331-1334.
- [15] Wold, E. Blum, T. Keislar, D. Wheaton, J. (1996) Content-Based Classification, Search, and Retrieval of Audio, *IEEE Multimedia*, 3:3, pp. 27-36.
- [16] Zils A., Pachet F., Delerue O., Gouyon F. (2002) Automatic Extraction of Drum Tracks from Polyphonic Music Signals. Proc. of WEDEL MUSIC.
- [17] Zils A., Pachet F.,(2003) Evolving automatically high level music descriptors from acoustic signals, Submitted to ISMIR2003.
- [18] La Burthe A., Pachet F., Aucouturier JJ. (2003) Editorial Metadata in the Cuidado Music Browser: between universalism and autism. WedelMusic Conference.
- [19] Pachet F. (2003) Content Management for Electronic Music Distribution: The Real Issues. Proc. Of CACM.
- [20] Pachet F., Cazaly D. (2000). A taxonomy of musical genres. In Proc. Content-based Multimedia Information Access (RIAO), Paris, France, 2000.