in *Encyclopedia of Knowledge Management*, Schwartz, D. Ed. Idea Group, 2005.

Knowledge Management and Musical Metadata

François Pachet
Sony CSL – Paris,
6 rue Amyot, 75005 Paris
Email: pachet@csl.sony.fr

Abstract:
The explosion of digital music has created in the recent years an urgent need for powerful knowledge management techniques and tools. Without such tools, users are confronted to huge music catalogues they cannot fully exploit. The very nature of music calls for the development of specific knowledge management techniques: on the one hand, the goals of users are ill-defined, or rather, based on enjoyment rather than on clear tasks or problems to solve. On the other hand, music in Western countries has been the subject of a long tradition of formalization and knowledge which is of crucial importance for building reasonable music information systems. The article outlines the main issues on music management, and focuses on the three main types of musical metadata being currently considered: editorial, cultural and acoustic. For each of these, the main issues are stated and the most successful techniques are discussed.

1 Introduction

Is music a form of knowledge? Probably not, even if music is undoubtedly an important part of our cultural heritage. Music is not a type of knowledge, at least in first approximation, because music has no consensual, shared *meaning*. One of the main reasons why music has no meaning, as opposed to text or even pictures, is that music is not *referential*: music is made of elements (notes, chords, sounds) which do not refer to any objects or concepts outside the musical world (Meyer, 1956). Being without meaning, music is not a type of knowledge.

However, our heavily digitized society continuously produces and exploits an increasing amount of *knowledge about* music. This knowledge, also called *metadata*, has taken a growing importance in the music industry and deserves a special treatment in this encyclopaedia because of the specificities of music. On one hand, music is ubiquitous and pervasive: there are about 10 millions music titles produced by the major music labels in the Western world. Adding the music produced in non-Western world probably doubles this figure. The music industry is one of the prevalent industries in the Western world today. On the other hand, music is elusive, i.e. it is difficult to define exactly what is music (for instance, distinguishing music from ambient sounds is not always trivial). To make all this music easily accessible to listeners, it is important to describe music in ways that machines can understand. Music knowledge management is precisely about this issue: 1) building meaningful *descriptions* of music that are easy to maintain and 2) exploiting these descriptions to build efficient music access systems that help users find music in large music collections.

2 Background

The issue of building music description is the subject matter of the audio part of the Mpeg-7 standard (Nack & Lindsay, 1999). Mpeg-7 focuses only on the notion of metadata, as opposed to its predecessors (Mpeg-1, 2 and 4), and proposes schemes to represent arbitrary symbolic and numeric information about multimedia objects, such as music or movies. However, Mpeg-7 deals only with the syntax of these descriptions, and not on the way these descriptions are to be produced. Here is, for instance, an extract of a Mpeg-7 description of the music title "Blowin' in the wind" by Bob Dylan. This extract declares the name of the artist, the name of the song and its genre (here, "Folk", according to a genre classification indicated in the extract itself):

```
<?xml version="1.0" encoding="UTF-8"?>
<Mpeg7
  xmlns="urn:mpeg:mpeg7:schema:2001"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
  xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 mpeg7-smp-2004.xsd">
  <Description xsi:type="CreationDescriptionType">
    <!-- ID3 Track number -->
    <CreationInformation id="track-01">
      <Creation>
        <!-- ID3 Song Title -->
        <Title type="songTitle">Blowin' in the wind</Title>
        <!-- ID3 Album Title -->
        <Title type="albumTitle">The Freewheelin'</Title>
        <!-- ID3 Artist -->
        <Creator>
          <Role href="urn:mpeg:mpeg7:RoleCS:2001:PERFORMER"/>
          <Agent xsi:type="PersonType">
            <Name>
              <FamilyName>Dylan</FamilyName>
              <GivenName>Bob</GivenName>
            </Name>
          </Agent>
        </Creator>
      <!-- ID3 Genre  -->
      <Classification>
        <Genre href=" urn:id3:cs:ID3genreCS:v1:80"><Name>Folk</Name></Genre>
      </Classification>
    </CreationInformation>
  </Description>
</Mpeg7>
```

Figure 1. A Mpeg-7 extract for describing information about a music title.

The first step toward music knowledge management is probably music identification. Robust audio fingerprinting techniques have been developed recently to identify music titles from the analysis of possibly distorted sources, such as radio broadcasts, or direct recordings from cell phone microphones (Cano et al., 2002). Audio fingerprinting is not a knowledge management technique *per se*, but is a prerequisite to build music collections. This technique has received considerable attention in the last years, and today very robust solutions have been designed and implemented in real world systems, such as the MoodLogic Music Browser.

To give a concrete idea of typical music descriptions used in musical knowledge management systems, let us give here three examples and their related use.

Several companies produce and exploit so-called *editorial* musical metadata (for instance AllMusicGuide (Datta, 2002) or MusicBrainz (http://www.musicbrainz.org). This information typically relates to song and albums (e.g. track listing of albums) but also include information on artists (biographies, periods of activities) and genres. A typical scenario of use is the display in a popular music player of an artist's biography and genre, when a title is played. When a title is played, an identification mechanism produces the identity of the title and artist, and a query is made to AllMusicGuide to retrieve more information, e.g. the biography of the artist, or the photograph of the album the title comes from.

Another popular application of musical metadata is *query by humming*. Query by humming consists in letting users sing or hum a melody, and retrieves the songs whose melody match the input (Birmingham et al., 2002). Technically, query-by humming is one instance of music information retrieval systems. In terms of knowledge management, this application makes use of the analysis of melodies from the audio signal and the sung inputs, so they fall in the category of acoustic descriptors as described below.

Finally a popular view on music knowledge management is *collaborative filtering*, as used in music portals such as Amazon. Collaborative filtering makes intensive use of user profiles, and exploits similarity or patterns in large databases of profiles. Technically, collaborative filtering is one instance of so-called *cultural descriptors*, as we will see below.

The three examples are deliberately chosen to represent three types of information: editorial, cultural and acoustic. These three types of information cover actually the whole range of techniques for music knowledge management. The next section reviews in more details each of these types of information and highlights the main technical issues related to each of them.

3 Three types of Musical Metadata

Although there is a virtually infinite number of Musical metadata that can be thought of concerning the description of music, we propose here to classify all of them in only three categories: editorial, cultural and acoustic. This classification is based on the nature of the process that leads to the elaboration of the metadata.

3.1 Editorial Metadata

Editorial metadata refers to metadata obtained, literally, by the editor. Practically this means that the information is provided manually, by authoritative experts. Examples of editorial metadata in music range from album information (e.g. the song "Yellow Submarine" by the Beatles appears on the Album "Revolver" issued in the UK) to administrative information such as the dates of recording, the composers or performers. Because editorial metadata covers a wide range of information, from adminstratrivia to historical facts, it is difficult to define precisely its scope other than by stating how it was produced.

Editorial metadata is not necessarily objective. For instance, the All Music Guide editorial metadata portal (Datta, 2002) provides information about artist biographies,

which may be biased by cultural factors. In particular, genre information – seen as editorial metadata, i.e. entered by human experts - is known to be particularly subjective.

Technically, the tasks of organizing editorial metadata raises specific challenges, such as:

- Providing a consensual view on subjective editorial information. For instance, agreeing on a taxonomy of musical genres.
- Coping with the evolving nature of music. New artists, new genres, new events occur all the time in music. The organization of an editorial information system must be able to cope with these changes efficiently.
- Organizing the human effort into clear and distinct roles, such as editorial management and data enterers.

There is another distinction one can make concerning editorial metadata, which concerns the nature of the human source: editorial metadata as produced in AllMusicGuide is prescriptive: the information is decided by one well-defined expert or pool of experts.

Editorial metadata can also be produced in a non-prescriptive manner, using a collaborative scheme, i.e. by a community of users. In this case, both the nature of the information provided and the management techniques differ.

A typical example of this "collaborative editorial" information is the CDDB effort. (www.cddb.com) CDDB is a database of "track listing", i.e. the information, for each music album produced, of the songs contained in the album. Surprisingly, this track listing information is not systematically present in CD albums, and it is precisely the role of CDDB to fill this gap. The identification technique used is very simple and relies on a hashing code produced by the number of tracks and their exact durations. This signature uniquely identifies most of the albums. To the signature is associated the track listing information. Such editorial information is, however, not prescriptive, and is on the contrary produced by a collaborative effort. When a user fetches a track listing information for a given album, it is retrieved automatically from the CDDB database (provided the media player used has a license with CDDB). If the album is not recognized, then the user can input the information himself, and thus contribute to the database content.

Another example of such an approach is MoodLogic (www.moodlogic.com). The Moodlogic approach consists in building a database of song "profiles" from rating of users. This database is used to classify and recommend music, and is integrated in various music management tools such as music browsers. When a song is added to a user's collection, a fingerprinting technique identifies the song and fetches the corresponding metadata in the MoodLogic database. If the song is not present in the database, the user is asked to rate the song. This approach has proven to be scalable, as the Moodlogic database now contains profiles for about one million titles. The nature of the information entered is quite different, however, than the information present in prescriptive systems such as AllMusicGuide: Moodlogic includes information such as genres, mood, perceived energy, etc.

It is important to stress again here that these information are considered in our context as editorial – more precisely as collaborative editorial – because of the way they are provided. However, we will see that this kind of information can be used in a totally different context, in particular to produce acoustic metadata.

3.2 Cultural Metadata

Cultural information or knowledge is produced by the environment or culture. Contrarily to editorial information, cultural information is not prescribed or even explicitly entered in some information system. Cultural information result from an analysis of emerging patterns, categories or associations from a source of documents.

A common method of obtaining cultural information is collaborative filtering (Cohen and Fan, 2000). In this case, the source of information is a collection of user profiles.

However, user profiles are a relatively poor source of information, and there are many other cultural information schemes applicable to music. The most used sources of information are web search engines like Google, music radio programs, or purely textual sources such as books or encyclopaedia. The main techniques used borrow from natural language processing, and are mostly based on co-occurrence analysis: for a given item of interest (say an artist or a genre), co-occurrence techniques allow to associate to this item other items which are "close" in the sense that they appear often close to each other. Co-occurrence can be based on closeness of items in a web page, or by neighbouring relations in music playlists. The main difficulty in this approach is to derive a meaningful similarity relation from the co-occurrence information. Approaches such as (Pachet et al. 2001) or (Whitmann & Lawrence, 2002) give details on the actual language processing techniques used and the evaluation of results. The typical information that can be obtained from these analysis are:

- Similarity distance between musical items such as artists or songs. Such similarities can be used in music management systems such as music browser, or music recommendation systems.
- Word associations between different word categories. For instance, a co-occurrence technique described in (Whitmann & Lawrence, 2002) indicates which most common terms are associated with a given artist. The same technique can also be used to infer genre information; by computing the co-occurrence between an artist name (say, "the Beatles") and different genre names (say "Pop", "Rock", "Jazz", etc.). In this case, the resulting information may also be called *genre*, as in the editorial case, but editorial genre and cultural genre will most of the time not coincide (see **Error! Reference source not found.** for a discussion).

3.3 Acoustic Metadata

The last category of music information is acoustic metadata. Acoustic here refers to the fact that this information is obtained by an analysis of the audio file, without any reference to a textual or prescribed information. It is intended to be a purely objective information, pertaining to the "content" of the music.

A typical example of musical acoustic information is the tempo, i.e. the number of beats per second. Beat and Tempo extraction have long been addressed in the community of audio signal processing and current systems achieve today excellent performances (Sheirer, 1998). Other, more complex rhythmic information can also be

extracted, such as the metric structure (is it a ternary rhythm, like a waltz, or binary rhythm?), or the rhythm structure itself.

Besides rhythm, virtually all dimensions of music perception are subject to such extraction investigation: percussivity (is a sound percussive or pitched), or instrument *recognition*, (Herrera et al., 2002), or perceived energy (Zils and Pachet, 2003), or even mood (Liu et al., 2003). The results of these extractions are very disparate, and today no commercial application exploits these descriptors. But the robustness of these descriptors will likely greatly improve in the coming years, due to the increase of attention these subjects have attracted recently.

These preceding examples are *unary* descriptors: they consist of one particular value for a whole title and do not depend on other parameters e.g. the position in the music title. Non-unary descriptors are also very useful to describe music and manage large music collections. Melodic contour, or pitch extraction can be used for instance for query-by-humming applications (Birmingham et al., 2001). At a yet higher level, music *structure* can be inferred from the analysis of repetitions in the audio signal (Peeters et al., 2002), leading to applications such as automatic music summaries.

The issue of representing in a standardized manner all these metadata is addressed by the audio part of the Mpeg-7 standard (Nack & Lindsay, 1999). However, Mpeg-7 focuses on the syntax of the representation of these descriptors, and it is quite obvious that the success of the standard heavily depends on the robustness of the corresponding extractors.

One major problem this endeavour has to deal with is that there is rarely any "music grounded facts", except for trivial information. Building a grounded facts databases is therefore one of the main difficulty in acoustic descriptor design. Information obtained from collaborative editorial sources, such as MoodLogic (see 3.1) can, paradoxically, prove very valuable in this context.

Another issue is that although there is a lot of formal knowledge about music structure (tonal music in particular), this knowledge is rarely adapted to perceptive problems. For instance, taxonomies of genres, or taxonomies of instruments are not directly useable for building ground truth databases, because they are not based on perceptive models: depending on the playing mode, context, etc. a clarinet can sound very close to a guitar and very different from another clarinet.

3.4 DISCUSSION

Because of the wide diversity of music knowledge types, there is a growing concern about the evaluation and comparison of these metadata. Indeed, the exploitation of large-scale music collections is possible only if these metadata are robust. But what does it mean exactly to be robust?

There are different types of evaluations in our context, some of which do not raise any particular problems. For instance, the evaluation of acoustic descriptors targeting consensual, well defined music dimensions (such as tempo or instrument recognition on monophonic sources) do not usually raise any particular issues. The evaluation of acoustic similarities is more problematic, as the elaboration of a ground truth reference is itself a hard task (Aucouturier & Pachet, 2004).

However, the most complex evaluation task is probably the comparison of metadata across different categories. For instance, comparing acoustic similarity with cultural similarity is not a well-defined problem. Indeed, cultural metadata can be used to train machine-learning algorithms to produce acoustic metadata or similarities. In this case, the comparison is simple to do, but misleading, since the cultural similarities are known to be based not only on acoustic features. On the other hand, comparing two similarity measures obtained from different sources (e.g. Berenzweig, 2003) produces results that are hard to interpret or exploit.

Another important consequence of this diversity of sources of metadata is that complex information dependency loops can be created that eventually produce meaningless musical knowledge, at least to non-informed users. The example of genre is, to this respect, emblematic, as genre can be produced by any of our three categories of approaches:

- *Editorial genre* is a genre prescribed by an expert, say, the manager of a label, or the team of AllMusicGuide. In this case, the Beatles can be described as "Pop-Sixties".
- *Cultural genre* is extracted from an analysis of textual information such as the web. Depending on the source used, the Beatles can be described, culturally, as, say "Pop" (versus "Jazz" and "Classical).
- Finally, *acoustic genre* can be extracted too, using audio signal processing techniques (see, e.g. Tzanetakis et al., 2001). It is important to note that acoustic genre will entirely depend on the learning database used for building the extractor. This database usually comes either from editorial or cultural information sources.

These intricate dependencies of information call for a better realization, by users, of the implications and meanings of the metadata they are provided with for managing their collections. Instead of trying to artificially compare or fit these different sources of knowledge about music, a simpler and more efficient strategy is probably to find simple ways to explain to users what each of them is doing.

4 Future Trends

The representation of musical knowledge, as represented by metadata, is a blooming field. From the early experiments in beat tracking to the industries of metadata, many results have been obtained and problems solved. More are being addressed with promising results, such as the separation of sources in polyphonic recordings, which will bring new descriptions to music management systems.

Important directions concerning the future of music knowledge in this context are:

- The invention of new music access modes. So far, the main use of music metadata has been for implementing efficient music query systems. Metadata can also be used to create new music access modes, for instance integrating performance and music access. Preliminary works have been proposed, such as concatenative synthesis (Musaicing, Zils & Pachet, 2001), which exploit metadata to create new music, and not only to listen to songs.

- More subjective measures of user interests. So far, work on evaluation has focused on objective measures. However, users accessing large-scale music collections are often animated by desires such as the quest for discovery or the pleasure of partially controlled browsing. Music access systems would clearly benefit from measures of interestingness combining possibly contradictory similarity relations together.

5 Conclusion

While music itself is not a form of knowledge, musical knowledge is needed to manage large-scale music collections. We have discussed a classification of musical metadata into three basic categories based on the nature of the process leading to the creation of the metadata, and their potential uses. These three categories may intersect, at least superficially, and it is important to understand the possibilities and limits of each of these categories to make full use of them. It is very likely that future applications of music content management will make increasing use of such metadata, and conversely will exert pressure for the creation of new music metadata types.

6 References

Aucouturier, J.-J. and Pachet F. (2004) Improving Timbre Similarity: How high is the sky?. Journal of Negative Results in Speech and Audio Sciences, 1(1).

Berenzweig, Adam, Daniel Ellis, Beth Logan, Brian Whitman. (2003) "A Large Scale Evaluation of Acoustic and Subjective Music Similarity Measures." In Proceedings of the 2003 International Symposium on Music Information Retrieval. 26-30 October, Baltimore, MD.

Birmingham, Dannenberg, Wakefield, Bartsch, Bykowski, Mazzoni, Meek, Mellody, and Rand, ``MUSART: Music Retrieval via Aural Queries," in *ISMIR 2001 2nd Annual International Symposium on Music Information Retrieval*, Bloomington: Indiana University, (2001), pp. 73-82.

Cano, P. Batlle, E. Kalker, T. and Haitsma, J. (2002) A review of algorithms for audio fingerprinting. In International Workshop on Multimedia Signal Processing, US Virgin Islands, December. http://www.iua.upf.es/mtg.

Cohen, W., Fan, W. (2000) Web-collaborative filtering: recommending music by crawling the Web Computer Networks: The International Journal of Computer and Telecommunications Networking Volume 33, Issue 1-6.

Datta, D. (2002) Managing Metadata. Proc. of International Symposium on Music Information Retrieval 2002, Paris.

Herrera, P., Peeters, G., Dubnov, S. (2002) Automatic classification of musical instrument sounds, *Journal of New Music Research*, 31(3).

Liu, D. Lu. L, Zhang, H.-J. (2003) Automatic Mood Detection form Acoustic Music Data. Proceedings of ISMIR 2003, Washington, USA.

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

Nack, F. Lindsay, A. (1999) Everything you wanted to know about Mpeg-7: Part 2, IEEE Multimedia, 6(4), 1999.

Pachet, F., Westerman, G. and Laigre, D Musical Data Mining for Electronic Music Distribution. Proceedings of the 1st WedelMusic Conference, 2001

Peeters, G. La Burthe, A. Rodet, X (2002) Towards automatic music audio summary generation from signal analysis, In *Proceedings of the International Conference on Music Information Retrieval (ISMIR02)*, Ircam.

Scheirer, E. (1998) *Tempo and beat analysis of acoustic musical signals*. JASA, 103(1):588--601.

Sheirer, E. (2002) About this Business of Metadata. Proc. of International Symposium on Music Information Retrieval 2002, Paris.

Tzanetakis, G. Essl, G., Cook, P. (2001) Automatic Musical Genre Classification of Audio Signals. Proceedings of ISMIR 2001 Bloomington, Indiana (USA).

Whitman, Brian and Steve Lawrence (2002). "Inferring Descriptions and Similarity for Music from Community Metadata." In "Voices of Nature," Proceedings of the 2002 International Computer Music Conference. pp 591-598. 16-21 September 2002, Göteborg, Sweden.

Zils, A. & Pachet, F. (2003) Extracting Automatically the Perceived Intensity of Music Titles. Proceedings of the 6th COST-G6 Conference on Digital Audio Effects (DAFX03), September. Queen Mary University, London.

Zils, A. and Pachet, F. (2001) Musical Mosaicing. Proceedings of DAFX 01, December. University of Limerick.

7 Concept Definitions

7 concepts with their definitions

Acoustic metadata: metadata obtained from an analysis of the audio signal.

Fingerprinting: technique to associate a single – and small – representation of an audio signal that is robust to usual audio deformations. Used for identification.

Cultural metadata: metadata obtained from the analysis of corpora of textual information, usually from the Internet or other public sources (radio programs, encyclopedias, etc.)

Editorial metadata: metadata obtained manually, by a pool of experts. Typically AMG.

Prescriptive metadata: produced by a single expert or group of experts.

Tonal music: music following the rules of tonality, i.e. based on scales. Usually opposed to atonal music such as serial music (based on the principle that all notes must be used with the same frequencies), spectral music (based on the nature of sounds rather than on pitches), minimalism, etc.

Timbre: a dimension of music which is defined by negation: timbre is not pitch nor dynamics, and is everything else. Timbre defines the texture of the sound, and allows to differentiate between different instruments playing the same note (pitch) at the same volume.