Content Management for Electronic Music Distribution: What Are the Issues?

François Pachet

Sony CSL-Paris, pachet@csl.sony.fr

Abstract:

EMD (Electronic Music Distribution) refers to the transportation and distribution of digital music. The profusion of EMD systems spreading in the PC and Internet world makes it difficult to have a global understanding of the real issues at stake, to compare and assess EMD technologies and services. We present a survey of existing approaches by stressing on the nature of the underlying *content management technologies*. We show that there is a recurring compromise to achieve between two approaches: a hard way and an easy way. We illustrate these different approaches on real world EMD examples.

The Dream of Electronic Music Distribution

Although the representation of audio data in digital format has been devised a long time ago, the possibility to store and manipulate such representation with good sound quality for music titles is much more recent. The emergence of efficient audio compression technologies such as MP3 has brought about the possibility of easily transporting and broadcasting music data across networks: Napster had 80 millions registered users at its peak [3] and 1.76 million songs were downloaded in November 2000. Another consequence of the power of audio digitalization is that the granularity of music distribution has shifted from music albums to music titles. Electronic Music Distribution (EMD) usually refers to the technical issues of transporting music data across networks, copy protection and copyrights management. However, there is much more to EMD than telecommunication and protection. A major challenge of EMD is to allow the shift from a mass-market approach to a personalized distribution approach. Providing this digital link between music and people is not a trivial task for several reasons, and it still remains a dream more than reality.

First, size. Estimations based on major label catalogues yield a total of 10 millions titles restricting ourselves to published, occidental, popular music. The number of Internet users is about 500 millions in 2000, according to a Nua survey. Traditional mass-market distribution consists in distributing only a small fraction of music titles (hits) to a large number of people: the fraction of so-called "active" titles in major label catalogues is estimated to about 1%. The EMD dream is primarily about proposing personalized distribution schemes that make more titles available to more people.

Second, EMD touches upon our intimate relationship with music. Browsing music is different from browsing a traditional digital library: we don't want to simply "access" or "find" music, as we would, e.g. for bibliographical references. Users do not always know how to specify what they look for (the *language mismatch problem* created by ontologies users do not understand, see [1]); nor do they always even *know* what they look for. The design of an EMD system requires therefore that we know more about what users want to do with music.

However, EMD systems seem to abound, in a large number of incarnations: Digital Audio Broadcast, CD-on-demand, music downloading, music streaming, Internet radios, music file management systems, music servers including peer-to-peer communication systems, content-based music retrieval systems. From a technical point of view, these systems differ mainly in the nature of the inputs and outputs they connect together (e.g. servers to CDs for on-demand CDs, to amplifiers for Internet Radios, etc.). EMD should be able in principle to handle all kinds of music sources and destinations.

In which respect do these systems achieve the EMD dream? For us the answer lies in *content management*: Only content management will provide an efficient link between listeners and music. We describe here the most important issues underlying music content management: from identification to content-based music search and retrieval and user interfaces. We exhibit an opposition between a "hard way" requiring brute force and sophisticated technology which provide objective information but are costly to develop and maintain, and the "easy way" approaches based on statistical analysis of superficial data, which are straightforward to implement but less reliable.

Content Management: The Fundamental Technology of EMD

Music Title Identification

How can a system identify music titles? In the simplest case, identification information is added to the music data itself, for instance through ID tags in Mpeg files. The ISRC (International Standard Recording Code) was developed by ISO (ISO 3901) to identify sound and audio-visual recordings. ISRC is a unique identifier of each recording that makes up an album. Unfortunately it is not followed by all music production companies, and is hardly used in unofficial music sources, so the majority of existing digital music files do not contain any built-in identification.

Worse, music data may not carry any external reference information: this is the case of Hertzian radio for instance. In this case, identification can be done either the hard way, by analyzing the signal, typically a portion of the music title, and matching it against a database of prerecorded music signals. This task is addressed by technologies such as Broadcast Data Systems (US) or MediaControl (Germany), and is used typically by copyrights management companies to infer radio play lists. The techniques used to perform the identification range from pattern matching to more elaborate statistical methods based on characterization of the evolution of spectral behaviors. The easy way consists in using external information on the titles when available. External information can be as simple as file names, with the difficulty that names are even less standardized: an artist such as “The Beatles” may be catalogued as “The Beatles”, “Beatles, The”, or any other combination. Other, more reliable external information can be exploited: The Emarker system exploits the geographical and temporal location of a user listening to a radio and requesting a song, and then queries a large database containing all radio stations programs. The approach is lighter, no signal processing is required, and can scale-up to recognize virtually any number of titles. Of course, it works only for titles played on official radio stations.

Music Genre

The most prominent information about a music title is probably its *genre*. Music distributors and retailers have long created and used genre classifications. However, the study of these classifications [6] shows that there is hardly any convergence: terms are not consensual (“Easy listening” in one classification is called “Variety” in another), and, worse, taxonomy structures do not match: “Rock” for instance denotes different songs in different classifications. Additionally, music classifications have been designed mostly for music albums, and are not directly usable for music titles: a *Pop-Rock* album by the Beatles may contain titles in many different genres: from *Country-Folk* “Rocky Raccoon” to the *Symphonic Easy Listening* “Good Night”.

A genre classification can be made by hands, by experts. These classifications have the advantage of containing expert knowledge, and of being relatively consistent. Their main disadvantage is that they are not easy to update (update must be done by hand and usually by the same experts), and not always readable because terms, even coined by professionals, are rarely consensual (what do you mean by “Zouk-Love”?). Classifications can also be built automatically, by an analysis of usage. Proposals to emerge genre classifications have been made based on collaborative systems [2], as well as data mining techniques as described below.

Music Similarity, the Easy Way

The main task of music content management is to extract relevant similarities between music titles. Similarities can be of various sorts. One may consider all the titles by a given artist as similar. And they are, artist-wise. Similarity can occur at the feature level: one may consider that Jazz saxophone titles are all similar. Similarity can yet occur at a larger level, and concern songs in their entirety: one may consider Beatles titles as similar to titles from the Beach Boys, because they were recorded at the same period. Or two titles may be considered similar by a user for no objective reason, simply because he/she thinks so.

Non-objective types of similarity may be extracted the easy way, at least to some extent, by two main classes of techniques: collaborative filtering (CF) and data mining.

Collaborative filtering is a general technique to infer patterns in taste within communities of users. The technique was originally put forward by Pattie Maes for a general-purpose recommendation agent [9]. The use of CF for music recommendation is now widespread, and most of Internet music retailers (e.g. Amazon, CDNow, MyLaunch) use it to provide recommendations to their customers. The basic idea of CF is to make personalized recommendations based on similarities in user profiles. The repeated logs of each user to the system progressively build a profile of his/her taste. The profile can be as simple as the titles selected, the list of the CDs bought, or more subtle preference rankings, as proposed for instance by MoodLogic.

Although technical evaluations of musical collaborative filtering have been performed (the Jaboom team (http://www.jaboom.com/); [2]), the nature of the music similarity exhibited by CF is difficult to characterize. CF-based similarity typically comes from culturally grounded affinities. For instance, most of the people who like the Beatles will probably also like the Beach Boys, and, generally speaking, the Pop music of the sixties. The interesting property of CF is that these relations will be exhibited easily (the easy way!) without human intervention and without complex signal processing. Nonetheless, the technique has several drawbacks, too. First, the similarities are not complete, and will address only titles that were actually rated by many users. Second, there are limitations to CF in the nature of the recommendations. Only strong patterns in communities are actually propagated, so eclectic profiles do not gain much from CF, because they are not statistically close to a large enough population of profiles: the more specialized the profile, the less interesting strong patterns will be for the corresponding user.

Collaborative filtering is a particular case of data mining technique, focusing on databases of user profiles. Other data mining techniques can be used to infer similarities, such as *co-occurrence analysis*. This technique consists in checking when two or more titles appear together in different contexts, and building a distance function based on these co-occurrences. These techniques can be used to infer automatically clusters of related titles, as well as genre taxonomies as shown in our studies [7]. The taxonomy has the great advantage of being done entirely automatically, and is easy to update. However, it is difficult – at least directly - to assign a label (such as Rock or Jazz) that would make sense for users.

Music Features, the Hard Way

Music Classification yields only one feature of music: genre. A music title has many other features. The Mpeg7 standard aims at providing a basis for representing all common features for audio-visual documents and will be finalized in 2001. Music metadata in Mpeg7 refers in general to low-level, objective information that can be extracted automatically [8] such as energy level, or spectral information. Extraction of higher-level features is a primary issue and one can distinguish, here alsohuman-based approaches and automatic approaches.

Several attempts have been made at extracting metadata *manually*, by teams of musical experts (MongoMusic or MoodLogic). Humans are very good at categorizing but not very good at sharing ontologies. The difficult part is to determine the right level of qualification: descriptors must be sufficiently consensual without being too obvious. To manually describe a large number of individual titles is a huge and risky enterprise. Less ambitious projects can however be conducted, for instance to describe artists, who are usually consistent in their musical genre. For instance, it is reasonable to say that Wes Montgomery has produced “Jazz Guitar” music and Madonna “Pop Song”. Although there can be some notable exceptions, this information can be used as a starting point to browse large catalogues.

Attempts at extracting automatically high-level music features from audio signals have focused on specific issues such as fundamental frequency [4], beat extraction and tempo induction [10] or segmentation. Some of these technologies are now mature enough to be actually exploited, for instance by the MuscleFish tool [11] which proposes a software suite of audio extractors, or MusicGenome (www.musicgenome.com). This fascinating field of musical feature extraction is just beginning. The Cuidado European funded project (www.ircam.fr/cuidado) aims precisely at developing systematically extractors for low-level and high-level musical features, in the context of Mpeg7 and will provide a first systematic approach at high-level musical feature extraction.

Query Systems

Most existing EMD systems follow a traditional query-answer scheme: the system provides a *set* of titles, possibly sorted by weights corresponding to how they satisfy the query.

However, music titles are usually not listened to individually, but in sequence, for instance in a radio program, a concert or a CD. These sequences usually have some global properties that make them consistent or interesting such as continuity or thematic consistency. These properties are easier to state than properties of items, because they do not require the knowledge of specific ontologies: everybody understands what *continuous* means. We proposed in [5] to address the music retrieval problem from the sequence viewpoint, and showed that this approach allows users to access music in a much simpler and more efficient way. Additionally, the approach avoids the language mismatch problem inherent to metadata access: metadata is used only internally by the system to build the sequence, and not explicitly by the user. Figure 5 shows *PathBuilder*, a prototype developed at CSL that builds a music path between two music titles selected by the user. The path is as continuous as possible, and continuity is defined by a weighted sum of similarity measures on a set of music features (genre, voice type, tempo, etc.).

Figure 1. The PathBuilder system creates music compilations from a starting and ending titles, by computing a path using musical metadata. Metadata is shown in columns. Blue means continuity, red means discontinuity.

User Interfaces

Content management technologies have to be integrated in highly interactive user interfaces to provide useful services, and to enable learning at both sides (user and system). Many user interfaces can be found on the Internet, from straightforward feature-based search systems (MongoMusic) to innovative graphical representations of play lists. For instance, Gigabeat display music titles in spirals to reflect similarity relations titles entertain with each other. Departing from traditional play list interfaces, the gravitational models of SmartTuner of mzz.com, or MoodLogic, represent titles as small mercury balls which move graciously on the screen, to or from “attractors” representing the descriptors selected by the user. These interfaces impose a fixed interaction model, and assume a constant attitude of users regarding exploration: either non-explorative - music databases in which you get exactly what you query - or very exploratory, usually based on collaborative filtering techniques. But the users may not choose between the two, even less adjust this dimension to their wish.

PersonalRadio, a prototype for set-top-box music services developed at CSL, addresses explorativeness explicitly. Figure 2 shows an interface of PersonalRadio, with a slider ranging between two extreme values (conservative to exploratory). Depending on the position of this slider, the music selection proposed is either conservative, exploratory or anything in between. Explorativeness is also represented as the color of title names, ranging from blue to red.

User studies of PersonalRadio reveal interesting behaviors. While some users react negatively towards exploration in the beginning of their interaction, in the long run they tend to systematically shift to exploratory modes. This can be explained by the fact that most users quickly exhaust their capacity in issuing explicit queries: it is only once well-known artists and hits are queried, in a non exploratory mode, that the desire for novelty pops up, and that such a feature appears as useful.

Figure 2. PersonalRadio. When the exploratory slider is moved on the right, music programs contain titles farther away from the request. Here, the genre based request is "Jazz guitar". Titles in red are more distant from this request than titles in blue.

Conclusion

We have surveyed EMD applications by focusing on content management technologies as a key ingredient. These techniques, including title identification, genre classification, feature extraction, similarity extraction, music retrieval and user profiling, are necessary to ensure an efficient mapping from large music catalogues to users and eventually make possible 1-to-1 music distribution. Whether we follow the hard or easy way, through brute force or through statistical analysis of superficial data, there is still a long way to go to achieve the EMD dream. New problems will arise when these technologies are mature, for instance concerning the legal status of metadata: Can an artist prevent someone to create and distribute metadata about his music? Many institutions now tend to favor open source and patent-clear approaches to multimedia management (see the open source streaming techniques developed by the Xiphophorus or Icecast projects): in this context, should metadata, also, be free?

References

1. Belkin, N. (2000) Helping people find what they don't know. *CACM* Vol. 43, N. 8, August 2000, pp. 58-61.
2. Cohen, W., Fan, W. (2000) Web-Collaborative Filtering: Recommending Music by Crawling The Web, 9th International World Wide Web Conference, Amsterdam.
3. Lam, Calvin. Tan, Bernard (2001) The Internet is Changing the Music Industry. Communications of the ACM, 44:8, pp. 62-68, August, 2001.
4. Lepain, Philippe (1999) Polyphonic Pitch Extraction from Musical Signals, Journal of New Music Research, 28:4.
5. Pachet, F. Roy, P. Cazaly. D. (2000) "A Combinatorial approach to content-based music selection". IEEE Multimedia, pp. 44-51, March 2000.
6. Pachet, F. Cazaly, D. (2000) "A Classification of Musical Genre", Content-Based Multimedia Information Access (RIAO) Conference, Paris.
7. Pachet, F. Westermann, G. Laigre, D. (2001) Data Mining for Electronic Music Distribution. Proceedings of WedelMusic, Firenze, 2001.
8. Philippe, Pierrick (2000) Low-level musical descriptors for MPEG-7, *Signal Processing: Image Communication*, 16 181-191.
9.

10. Shardanand, U. and Maes, P. (1995) Social Information Filtering: Algorithms for Automating "Word of Mouth". Proceedings of the 1995 ACM Conference on Human Factors in Computing Systems, pp. 210-217.
11. Scheirer, Eric D. (1998) “Tempo and beat analysis of acoustic signals”, JASA, 103(1).
12. Wold, E., Keislar, T., Wheaton, J. (1996) Content-Based Classification, Search, and Retrieval of Audio, IEEE Multimedia, 3:3, pp. 27-36.