Can Humans Benefit from Music Information Retrieval?

Frans Wiering

Department of Information and Computing Sciences, Utrecht University,
PO Box 80 089, NL-3508 TB Utrecht, Netherlands
frans.wiering@cs.uu.nl

Abstract. In the area of Music Information Retrieval (MIR), great technical progress has been made since this discipline started to mature in the late 1990s. Yet, despite the almost universal interest in music, technology that helps people find the song they have in their head is not that widely used. There seems to be a mismatch between the assumptions researchers make about the users' music information needs, and the actual behaviour of a public that to begin with may not even treat music as information Therefore, the emphasis of MIR research should be more on the emotional, social and aesthetic meaning of music to regular, untrained people. MIR applications could greatly benefit from using the results of recent research into the spontaneously-developed musical competence of untrained listeners.

Keywords: music information retrieval, musical similarity, musical content, music psychology, user interfaces

1 Introduction

Music Information Retrieval (MIR) has been defined by Stephen Downie as 'a multidisciplinary research endeavor that strives to develop innovative content-based searching schemes, novel interfaces, and evolving networked delivery mechanisms in an effort to make the world's vast store of music accessible to all' [6]. Among the contributing disciplines are computer science, information retrieval, audio engineering, digital sound processing, musicology and music theory (the latter two are generally separated in the US, but not in Europe), library science, cognitive science, psychology, philosophy and law [8]. Many researchers are motivated by their personal interest in music and therefore tend to use their own audio collection as a testbed. They often do not have advanced musical knowledge, for which they usually turn to the music professionals in the community, generally scholars and librarians.

The professional viewpoint of music is also much in evidence in an important tradition of computer-supported musicology that already emerged early in the history of computing. Indeed, this is where the expression 'Musical Information Retrieval' was first used around 1965 [15]. As a discipline, MIR has been maturing since the late 1990s. Since 2000, the yearly ISMIR conference has played a key role in this development by providing a platform where the representatives of different disciplines meet. Virtually all ISMIR papers are available online at http://www.ismir.net/all-papers.html. The most recent single-publication overview of MIR is [19].

After briefly introducing some central notions in MIR (section 2), I will discuss some interrelated issues in section 3 that I consider relevant for the future development of MIR. Hopefully, these emerge as research challenges that the Adaptive Multimedia Retrieval community—which is not listed in [8]—is interested in helping to solve.

2 Central Notions in MIR

Rather than attempting an overview of MIR, I will discuss in this section some central notions in MIR: the representation of musical content, musical similarity, retrieval methods and MIR systems. User related issues will be mentioned in passing, but I will treat these more fully in section 3.

2.1 Metadata-Based and Content-Based Approaches

Two main approaches to MIR can be discerned: metadata-based and content-based. In the former, the issue is mainly to find useful categories for describing music, distinguishing for example between different recordings of the same composition, or between artist-as-creator and artist-as-performer, or to organise the myriad of genre descriptions that exist for music. These categories are expressed in text. Hence, existing text-based retrieval methods can be used to search those descriptions, and can also be applied to another important feature of the musical content: the texts of vocal music.

The more challenging approaches in MIR are thus the ones that deal with the actual musical content, e.g. pitch, melody and rhythm. The first problem one encounters is that musical content comes in two formats: sound and notation. It would seem that the most natural one to select is sound. Humans in general have well-developed abilities to extract features from a musical signal: they can distinguish pitches, melodies and harmonies, rhythms and beat patterns, they can identify instruments, and at times they are strongly moved by the emotions these features evoke. Extracting these features from audio and using them to let people retrieve the music they like seems the obvious thing to do. This however has proved to be very difficult. Only monophonic transcription, the detection of 'pitch events' in a single melody, is now considered a solved problem, even though engineering problems remain. Transcription of polyphony, music in which several pitch events may occur at the same time, is still very much an unsolved issue. Interesting recent methods that do not depend on precise pitch detection are harmonic matching [22] and chroma-based matching [3].

Many researchers, especially those from the professional domain, have opted for the second approach, notation, using 'symbolic' representations of music. These involve encodings of musical scores in one of the many available encoding systems [4, 23]. Even though mainly meant for performers, music notation can be said to model music perception to a certain extent. For example, one or more notes usually correspond to one perceived pitch event. However, the 'chunking' of melodies that takes place in early perception and that is very important for the mental representation of music [24] is generally not made explicit in notation. Also, the output of audio transcription is often in a format that shares a number of features with notation. Therefore, methods developed for searching music notation are likely to be relevant to audio as well.

2.2 Musical Similarity

In Information Retrieval, we want to find documents that match a user's information need, as expressed in a query. In content-based MIR, this aim is usually described as finding music that is similar to a set of features or an example. Note that by equating user need to similarity, some imaginable needs are ruled out, notably surprise. Musical similarity is thus a central issue in MIR, but it is not a clearly-defined concept. There are several reasons for this.

First, many interrelating features are involved in the perception of musical similarity. I have already mentioned melody, rhythm, harmony and instruments as features. Similarity in one such feature does not necessarily lead to perception of overall similarity. For example: if the pitches of two melodies are the same but the rhythm is very different, listeners may not consider these melodies as similar.

Second, there are many different types of musical similarity. Similarity can for example be said to exist between:

- two different performances played from the same notation, for example Beethoven's Fifth Symphony;
- varied repetitions of the melody in a strophic song, for example Frank Sinatra's *My Way*;
- different performances of same pop song, for example Frank Sinatra's and Sid Vicious's performances of *My Way*;
- works created by the same composer;
- music that has the same function, for example dances such as the waltz (which has several subtypes);
- music belonging to the same genre, for example Jazz or Gregorian chant;
- music originating from same culture: for example Western, Indian, or medieval music;
- music that contributes to one's social identity, for example Hip hop or Death Metal;
- music that displays the same atmosphere (romantic) or emotion (love).

Most listeners deal with these types of similarity as a matter of course. In addition, there are forms of similarity that belong to other kinds of users, notably music industry and music professionals [25, 27]. Music industry may be interested in music that plagiarizes other music, or in music that can be sold to an audience with a known musical taste. Music professionals may be interested in finding music of similar difficulty, ensemble composition, style, or where similarity is an indication of 'musical influence.'

2.3 Retrieval Methods

In response to the just-described multiplicity of music similarity, many different computational methods for measuring it have been devised (surveyed in [25] and [27]). The most generic forms of similarity are generally best served by audio-based methods such as self-organising maps of music [20]. Very specific matching is also possible in the audio domain: in audio fingerprinting different, possibly degraded, broadcasts of the same recording can be identified [12]. Chroma-based matching [3] is able to trace similarity in musical content between closely-related compositions.

Similarity measures that act on symbolic representation are generally based on string matching, set comparison, or probabilistic methods. These are generally suitable for very specific tasks such as finding different instances of the same work, melodic variation and music based on same harmonic sequence. Methods that address the middle level of similarity (musical style, specific genre) are still very weakly developed. The reason why this area is so hard to address is probably that such forms of similarity involve quite a number of musical characteristics. How these interrelate to create a sense of style or genre is not sufficiently understood yet.

2.4 MIR Systems

Many MIR systems are described in [25]; however, few of these can be described as mature, functional, end-user oriented systems. No doubt the most popular one is Pandora (http://www.pandora.com), which however relies on matching human feature annotations. Systems like Themefinder (http://www.themefinder.org) and Musipedia (http://www.musipedia.org) are string-based engines that provide access to large collections of melodies and seem to have a well-defined audience. At Utrecht University, Rainer Typke built the Orpheus engine (http://give-lab.cs.uu.nl/orpheus), which is first of all a research prototype for testing weight flow distances such as the Earth Mover's Distance as similarity measures for music [26]. Orpheus is able to search large collections efficiently and with good results. It is currently being developed into a user-centred framework called Muugle (http://give-lab.cs.uu.nl/muugle) [2]. Designing, building and evaluating *usable* music retrieval systems, especially for untrained listeners, will certainly be an important goal for MIR research in the next few years.

3 Some Problems in MIR

For the sake of argument, I would like to present here a deliberately simplistic view of mainstream MIR, which is based on elements presented in the above account. Obviously, it does not do justice to the subtlety and originality of much research being done in the area, but it will help to clarify some issues that I believe to be crucial to the future of the discipline. This view of MIR is as follows.

In MIR research, is assumed that music is represented by its information content, that users wish to search for musical information, that such a search can be expressed as a musical query, but that users don't succeed because they are bad at expressing their information need. My criticism on this view is that it considers music primarily as information. First, this view assumes untrained users think of music as information, and furthermore underestimates the importance of other user needs besides information (such as mood, emotion, exploration and surprise); second, that it does not sufficiently take spontaneously-developed musical skills of listeners into account. Paying more attention to these factors may result in development of new forms of interaction and the integration of MIR techniques in more general applications. In the following sections I will examine these issues one by one.

3.1 Is Music Information?

In the context of Information Retrieval, the word 'information' can be taken to mean 'structured data that is suited to enhance a person's knowledge of the world'. Information in this sense is best exemplified by functional prose that has 'aboutness'. Scholarly papers, newspaper articles, computer manuals and travel guides clearly fall in this category: they are about some aspect of the world. Outside the domain of textual documents, diagrams, news broadcasts and documentary films can be said to be about something. On the other hand, the aboutness of literary writing, poetry for example, is problematic, because the factual subject matter is usually not the most important reason why one is interested in the work in question. Therefore one cannot separate the content of the work from the way it is expressed without losing some of its essence. By extension this lack of aboutness applies to other art forms as well, including music.

Yet considering music, and specifically music notation, as information [17] makes much sense as a professional view of music. Digitally encoded music notation can be queried for such things as recurrent melodies or chords, or processed to make statements about its structure or authorship. In this way it can surely contribute to one's knowledge of the musical world. However, it is hard to imagine how musical content and expression could be separated. It is for example unclear how one could embed the information from one piece in another other than by quotation. In addition, musical information tells us very little about the world outside music that is not subjective. Music possesses only a weak aboutness [17], and is thus rather 'pseudo-information'. Generally, meaning in music is a very problematic concept; it does however merit further exploration, as it is at the same time very clear that music is very meaningful to so many people.

Music has often been compared to language, most recently by Fitch [7]. As a starting point he takes Hockett's thirteen design principles of language [14]. Most of these, such as *rapid fading* and *cultural transmission*, apply to music as well. But music lacks precisely those features that in language support referentiality—the fact that it can refer to objects and events—such as *displacement* (language can refer to things that are not present) and *arbitrariness* (no fixed relation between sounds and things). The feature of *interchangeability* (one can say anything one can understand) is shared by music only to a limited extent in that most people are able to pick up certain basic skills, for example to sing simple melodies, without formal training. Yet music is more than speech without meaning. It has an obvious affective and aesthetic power that makes it 'a-referentially expressive,' one of the nine design principle that Fitch distinguishes for music.

How does this power of music work? One common explanation goes as follows. Basic musical features can often be subdivided in ones that create tension and ones that create relaxation. For example, dissonance is experienced as a tension that is resolved by the consonance that (usually) follows it. Narmour [18] refines this notion by describing music as patterns of implication and realisation. Such patterns can be manipulated by creating an expectation that is not (completely) fulfilled, for example by using a consonance other than the expected one or even by following the first dissonance with another dissonance. For an example see Fig. 1. Such patterns work at many levels and create in the listener's mind a sequence of responses that (if the composer has done his job properly) only comes to a complete relaxation at the end of the piece. These responses are probably triggered by processes such as endorphin production that also play an important role in the sensation of emotion [21]. As musical patterns are very complex and diverse, one's emotional response to music can be very rich and meaningful. Finally, such a-referential meaning is easily connected to that of language or images that coincide with the music, as in song and film. This connection is strong enough to recreate, by means of music only, the referential meaning that originally resided in the other medium.

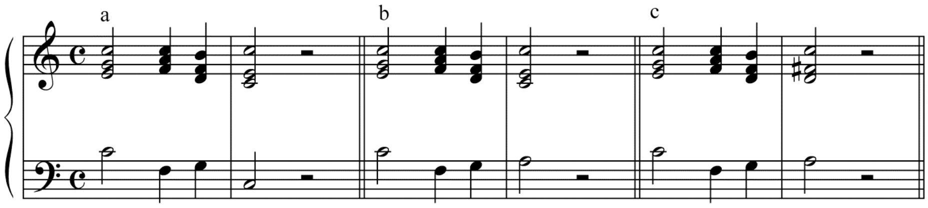

Fig. 1. Three different continuations of the same chord sequence. In each example, the first three chords are identical. These create a harmonic tension that is fully resolved in (a) by a normal ending on the 'tonic' triad: the piece could end here. In (b) another, related, consonant triad is substituted: we expect the piece to be continued. A dissonant though related chord is introduced in (c): there is a strong implication of continuation towards a resolution of this dissonance. Most untrained listeners are capable of intuitively appreciating these differences.

A-referential meaning can also be said to be present in the social functions of music, such as mother-infant bonding [7]; synchronising movement, most notably in dance; supporting activities, for example rituals, work, sports and shopping; and finally creating group coherence and identity. One evolutionary explanation of music is precisely that it makes humans function better as social animals.

If, then, the meaning of music is fundamentally a-referential, this has important consequences for MIR. The reason why most people will search for music is probably not that they want to enhance their knowledge of music by finding specific musical information, but that they search for a meaningful musical experience that satisfies their emotional or social needs. MIR systems that treat music first of all as information may not be very helpful in this scenario. One may even wonder whether a discipline named 'Music Information Retrieval' implies already in its name a perspective that is useful for professionals but marginalises other, far more numerous uses.

A Case Study: Query By Humming. The viewpoint of music as information even emerges in a particular MIR strategy, Query By Humming (QBH) that has been widely researched since it was first described in [9] with the aim of providing a 'natural way of querying.' The procedure for QBH is generally as follows: a user hums (or whistles or sings) a melodic query. The system matches the query against the musical items in a database and returns a ranked list of matching melodies. Figs 2. and 3 show an example QBH interface.

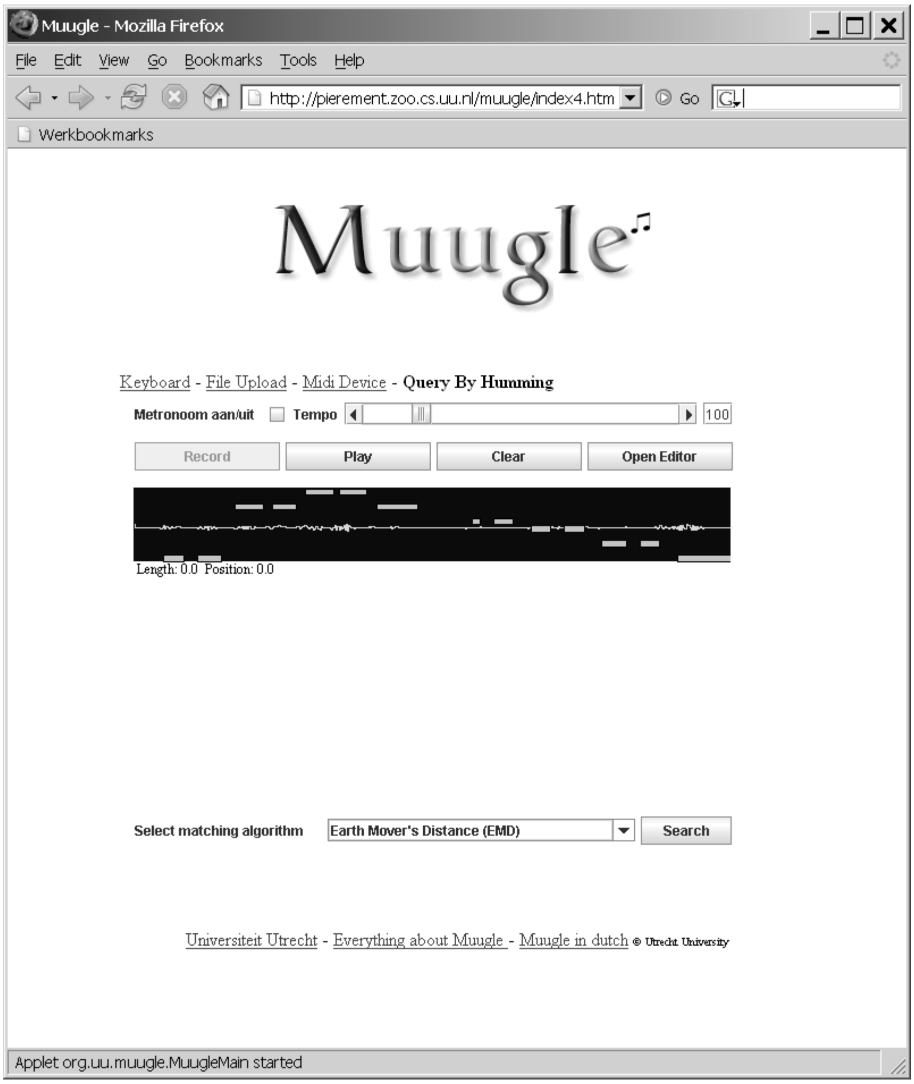

Fig. 2. Muugle's QBH interface (http://give-lab.cs.uu.nl/muugle): query interface with audio signal and transcription.

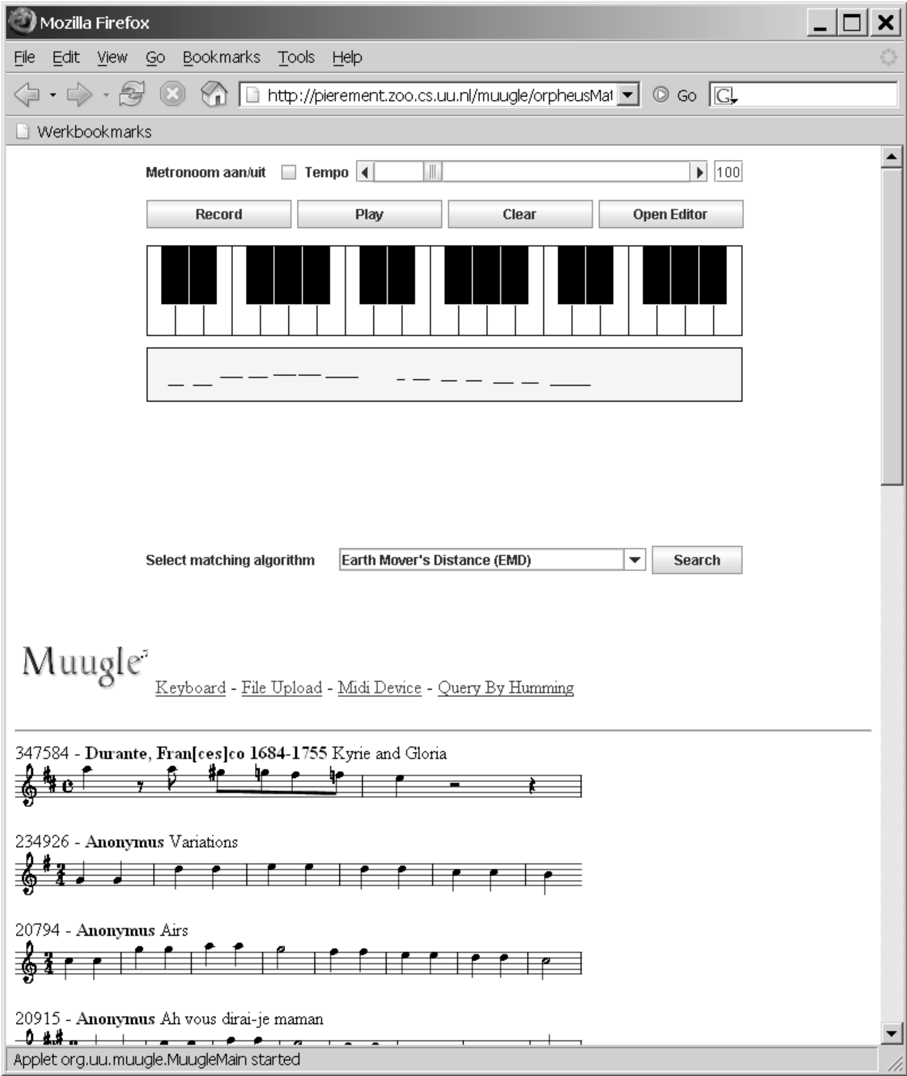

Fig. 3. Muugle's QBH interface: query transcription and result list.

Implementation issues aside, there seem to be three problems in this approach:

1. Users have considerable difficulties in generating a satisfactory query. It is often observed that they 'cannot sing in tune' or even 'cannot remember music correctly.' In fact, the task is not at all natural to them: not only must they perform music, but they must perform it exactly correct. In other words, QBH considers music as 'interchangeable,' while it is not. To give an analogy: we are all good at face recognition, but generally not at drawing them.
2. Only one type of query is possible: melody. Research into 'ecological' query formulation has shown that users wish to be able to use all sorts of sound

production, including tapping, lyrics etc. [16]. Even then, many factors remain unavailable for querying, most notably harmony and instrumentation, though people certainly have a mental representation of these.
3. Does QBH satisfy a widely-felt user need? There are situations in which it does, for example if one wishes to identify and maybe acquire a song one remembers. Often, however, people will have an 'experience need,' composed of taste, expertise, mood/emotion, and/or situation/context. The need is then generally satisfied by a set of pieces that meet the requirements, not a list in ranked order or one specific item.

All three problems relate to the fact that, despite its claim of naturalness, QBH treats music as information, so from a professional viewpoint. In this view, users need musical information that can be expressed in one particular musical dimension, melody, and possess the active musical competence to express this need. It seems to be no coincidence that QBH applications are rather unsuccessful in attracting large user communities, whereas a service like Pandora that does not presuppose a specific information need and active musical skills, is far more successful in this respect. However, Pandora relies exclusively on manual annotation of songs, not on content-based retrieval. I believe that finding techniques for extracting a-referential meaning from musical data and exploring its potential for users in similar services is a major research challenge for the future.

3.2 How Musical Are Humans?

One reason why MIR may concentrate on professional approaches to music is that historically the focus of attention in both musicology and music psychology used to be on the production of music by composers and performers. Only recently the study of listeners' 'passive' competence has begun to receive similar attention. Peretz [21] gives an interesting overview of passive musical competence and the problems of testing it without presupposing professional skills. In her view, 'the ordinary adult listener is a musical expert, although s/he may be unaware of it.' Most people acquire this expertise by being exposed to music, but a surprising percentage of the population (around 50% in the UK and USA) have received some musical instruction as well. Genuine amusicality (tone-deafness, the inability to distinguish between pitches) is genetically determined and occurs in about 4% of the population. It is not the result of lack of motivation or training. Non-trained listeners generally possess the following musical abilities, among other things:

- they are able to distinguish subtle stylistic differences;
- they are equally good at learning songs as professionals;
- they can identify out-of-key notes;
- they recognize patterns of implication and realisation [1];
- they are able move to music (as in tapping the beat of music).

3.3 Interaction with MIR Applications

Non-trained listeners thus possess considerable passive musical competence, which most likely enables them to have very concrete ideas about music that satisfies their musical needs. This, and the fact that they generally have only limited active abilities, may lead to different requirements for interaction with MIR applications. Three of these are examined here briefly and in some respects speculatively.

Emotion Retrieval. Emotional meaning is an important drive for people to listen and probably to search for music. A problem is the subjectivity of the emotional response to music. However, recent research by Lesaffre et al. indicates that affective/emotive and structural descriptors are correlated [16]. The authors could demonstrate among other things a very strong correlation between the appraisal descriptor tender-aggressive and the structural descriptor loudness (soft-hard). Such correlations were used for retrieving music on the basis of affective value in an application that users were satisfied with.

In [16], descriptors are assigned to complete pieces. My vision of the (very distant) future is that these descriptors will be created by content-based techniques for short fragments of music—maybe from implication-realisation patterns—and will be combined in a sort of path through emotion space. These paths can then be compared, so that pieces with a similar 'emotional development' can be retrieved. Such methods were already announced for video retrieval [13].

Output Presentation. The problem with MIR output is that numerous items may be retrieved, but that in the auditory domain, one can inspect these only one at a time. This is time-consuming and also a hard task for musical memory. The obvious answer is mapping the items to the visual domain. The standard solution is to put the items in a list consisting of metadata and/or snippets of music notation (see for example Themefinder and Muugle). There are a number of interesting and more intuitive alternatives, presenting a 2D or even 3D music space, for example Pampalk's Islands of music [20], Van Gulik's dynamic playlists [11], and Goto's Musicream [10]. The last-mentioned interface features three taps, from which streams of musical items emerge. The user can pick items from the streams, listen to these, and organise them.

What is still unsatisfactory about these solutions is that items are usually represented as points, disks, or by means of text labels. Can they be given some more meaningful representation? Untested ideas that recently emerged in a brainstorm include representing music as objects with a certain shape, colour, texture or movement. Users would define their own associations between visual and musical features, and musical objects would be adapted accordingly. Another option would be to present music (or at least a musical query) as a face. Faces are very individual, and in addition capable of expressing a wide range of emotions. Naturally it remains to be seen if such ideas are viable at all.

Retrieval in Context. Experiencing music is more than finding and then playing the right piece of music. I therefore expect a limited use of music search engines as separate applications. It seems more likely that they will be integrated in environments that allow different ways of exploring music and be partly or completely hidden to the end user. For example: a digital archive of some repertory (e.g. folksongs, piano music) would facilitate metadata queries, following of links between works, and the creation of new relations between them by means of retrieval tools.

Such an application may involve ways of accessing a wider context, involving for example textual documents, musical instruments, performance locations, and social functions. Music could also be part of integrated virtual cultural experience. The role of a music retrieval engine in such an environment may range from choosing appropriate background music, to supplying the right items to reconstructed events (plays, ceremonies) in which music played a role. Finally, music tends to be stereotyped if it acts in a supportive role, and it might be possible to use those stereotypes to retrieve corresponding events. For example, retrieving emotional scenes from mainstream Hollywood films might be quite feasible using the accompanying music.

4 Conclusion

The title of this paper asks whether humans can benefit from Music Information Retrieval. The problem lies in the word 'information.' I have tried to argue that considering music as information represent a limited view of music. Such a view is certainly appropriate for professionals (and industry—but I have barely touched upon their interests), who can gain a lot from these technologies. However, considering other humans I believe it is better not only to drop the I-word—and henceforth use the term 'Music Retrieval'—but also to rethink the area from a non-trained user's perspective. This may help researchers to concentrate on needs that better correspond to the 'experience of music,' to take the personal profile and especially the 'listening competence' of non-trained users as a starting point for designing interaction with the system, and finally to design interesting applications in which Music Retrieval plays an invisible but essential role in letting people experience the richness of 'the world's vast store of music' [6].

Acknowledgments. I thank Stéphane Marchand-Maillet for inviting me to present this paper as a keynote at the 4th International Workshop on Adaptive Multimedia Retrieval in Geneva, 2006. Hermi Schijf did a great job in correcting my English and identifying passages that needed further clarification. I would like to thank my colleagues and students at the Department of Information and Computing Sciences of Utrecht University for the stimulating discussions from which many ideas for this paper have emerged.

References

1. Bigand, E., Poulin-Charronnat, B.: Are We “Experienced Listeners”? A Review of the Musical Capacities That Do Not Depend on Formal Musical Training. Cognition 100 (2006) 100–130
2. Bosma, M., Veltkamp, R.C., Wiering, F.: Muugle: A Music Retrieval Experimentation Framework. In: Proceedings of the Ninth International Conference on Music Perception and Cognition, Bologna 2006, 1297–1303
3. Casey, M.: Audio Tools for Music Discovery. [online] http://www.methodsnetwork.ac.uk /redist/pdf/casey.pdf [accessed 30 November 2006]
4. Castan, C. Music Notation Formats. [online] http://www.music-notation.info [accessed 30 November 2006]
5. Downie, J.S.: Music Information Retrieval. Annual Review of Information Science and Technology 37 (2003) 295–340
6. Downie, J.S.: The Scientific Evaluation of Music Information Retrieval Systems: Foundations and Future. Computer Music Journal 28:2 (2004) 12–23
7. Fitch, W.T.: The Biology and Evolution of Music: A Comparative Perspective. Cognition 100 (2006) 173–215
8. Futrelle, J., Downie, J.S.: Interdisciplinary Communities and Research Issues in Music Information Retrieval. In: Proceedings ISMIR 2002, 215–221
9. Ghias A., Logan, J., Chamberlin, D., Smith, B.C.: Query by Humming: Musical Information Retrieval in an Audio Database. In: Proceedings of the ACM International Multimedia Conference and Exhibition 1995, 231–236
10. Goto, M., Goto, T.: Musicream: Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces. In: Proceedings ISMIR 2005, 404–411
11. Gulik, R. van, Vignoli, F., Wetering, H. van de: Mapping Music in the Palm of Your Hand: Explore and Discover Your Collection. In: Proceedings ISMIR 2004, 409–414
12. Haitsma, J., Kalker, T.: A Highly Robust Audio Fingerprinting System. Proceedings ISMIR 2002, 107–115
13. Hanjalic, A: Paradigm Shifts in Video Content Analysis Needed: The Why's and How's of Generic VCA Solutions. Dagstuhl Seminar on Content-Based Retrieval, April 2006
14. Hockett, C.F.: Logical Considerations in the Study of Animal Communication. In: W.E. Lanyon & W.N. Tavolga (Eds.), Animal Sounds and Communication. Washington, DC: American Institute of Biological Sciences, 1960. Cited after [E]
15. Kassler, M.: Toward Musical Information Retrieval. Perspectives of New Music 4:2 (1966) 59–67
16. Lesaffre, M., Leman, M., De Voogdt, L., De Baets, B., De Meyer, H., Martens, J.-P.: A User-Dependent Approach to the Perception of High-Level Semantics of Music. In: Proceedings of the Ninth International Conference on Music Perception and Cognition, Bologna 2006, 1003–1008
17. McLane, A.: Music as Information. Annual Review of Information Science and Technology 31 (1996) 225–262
18. Narmour, E.: The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model. University of Chicago Press, 1990.
19. Orio, N. Music Retrieval: A Tutorial and Review. Foundations and Trends in Information Retrieval 1 (2006) 1–90
20. Pampalk, E., Dixon, S., Widmer, G.: Exploring Music Collections by Browsing Different Views. Proceedings ISMIR 2003, 201–208
21. Peretz, I.: The Nature of Music from a Biological Perspective. Cognition 100 (2006) 1–32

22. Pickens, J., Bello, J.P., Monti, G., Sandler, M., Crawford, T., Dovey, M., Byrd, D.: Polyphonic Score Retrieval Using Polyphonic Audio Queries: A Harmonic Modeling Approach. Journal of New Music Research 32 (2003) 223–236
23. Selfridge-Field, E. (Ed.): Beyond MIDI: The Handbook of Musical Codes. MIT Press, 1997
24. Snyder, B.: Music and Memory: An Introduction. MIT Press, 2001
25. Typke, R., Wiering, F., Veltkamp, R.C.: A Survey of Music Information Retrieval Systems. In: Proceedings ISMIR 2005, 153–160
26. Typke R., Wiering F., Veltkamp R.C.: Transportation Distances and Human Perception of Melodic Similarity. Musicae Scientiae 10 (2006), forthcoming
27. Veltkamp, R.C., Wiering, F., Typke, R: Content Based Music Retrieval. In: B. Furht (Ed.), Encyclopedia of Multimedia, Springer 2006, 96–105