Aspects of Music Information Retrieval

Will Meurer

School of Information at

The University of Texas at Austin
Abstract

This paper outlines the complexities of music as information and gives an overview of Music Information Retrieval (MIR). The technologies and systems currently available or in production within the realm of MIR are also discussed, as are major issues, concerns, and solutions pertaining to the development and advancement of this field. Finally, some possibilities are given for the near future of MIR, and a simple approach for system development is described as a suggestion for future MIR research and development.
Aspects of Music Information Retrieval

1. Introduction

The study of Information Retrieval (IR) covers more than just the realm of text-based information. Researchers have spent over half a century focusing on text indexing and retrieval; most recently, powerhouses in the IR world are becoming more well-known to the public—namely, Web search engines like Google and Yahoo. Now that IR has accomplished so much in the realm of text retrieval, this discipline can expand to a focus on other types of information such as images, video, and music. Music Information Retrieval (MIR) is already advancing daily, and those involved are deep into research within this domain.

To say MIR is even close to the point of success that text retrieval is would be more than an overstatement. Much of the research has been conducted during the past decade and a half, and music remains a complication for MIR researchers (Downie, 2003a). Though technology is continually improving and much advancement has been made in MIR, the public domain has yet to see systems beyond those which use bibliographic data for retrieval purposes. To improve in this area, more IR and music professionals must become aware of the concerns and implications of this domain in research. This paper reviews fundamental aspects, concerns, technology, and goals within the realm of MIR.

2. MIR Overview

Researchers cannot simply leap into MIR; a high learning curve exists, beyond even the knowledge of text IR professionals. Music is a language in itself. It requires a
deep understanding of theoretical, analytical, and practical approaches in its various environments before researchers can learn how and why it may be accessed. Readers may ask, “Why do I need MIR when I have Google and Amazon?” To counter, why does one need image retrieval? Why does one use different search engines? Information retrieval systems are tailored to specific purposes. Music is a complex and dynamic form of information, and, therefore, users do need MIR for complete but simple access. Consequently, MIR faces many challenges in order to provide this access.

2.1 Facets

Music differs from text in many ways, but the most implicative difference, as far as IR is concerned, is that it is multifaceted whereas text is comparatively simplistic. MIR researchers must analyze and build systems based on their overall performance and value to the targeted users. Because there are so many different needs originating among various user groups, MIR systems must include coverage of multiple aspects of music. Stephen Downie (2003a), a noted MIR professional, identifies seven facets of music and notes that each has a significant effect on how systems can retrieve music. Downie’s seven facets are pitch, temporal, harmonic, timbral, editorial, textual, and bibliographic (Figure 1).

First, pitch is the melodic material. Included in this are specific notes, intervals (distances between notes), key signature and change of key. Second, the temporal facet includes any tempo (speed), rhythmic, or time-based information. Third, the harmonic facet consists of three parts: 1) whether music has two concurrent pitches (polyphony) or one (monophony), 2) what type of harmony exists, such as Major or Minor chords,
and 3) the pattern from chord to chord—similar to the interval aspect of pitch. Fourth, the timbral facet is any aural sound difference due to orchestration and instrumentation. For example, the change in sound a full orchestra makes as opposed to that of a piano arrangement is considered timbral information. Music scholar Jan LaRue (1992) suggests timbre also might include some aspects of pitch: tenors in a men’s choir singing high pitches in their high-voice, or falsetto, as compared to them singing lower pitches in their normal range, for example. Fifth, the editorial facet includes any
markings on the score that may act as “performance instructions” (Downie, 2003a). This includes dynamic markings, like soft ($p$) or loud ($f$), articulation, bowings for strings, and fingerings. Confusingly, some temporal markings like *ritardando* and *accelerando*, and timbral markings like pedaling and mutes, may be grouped with the editorial facet as well. Sixth, the textual facet is any text sung or spoken as part of a musical performance. Seventh, the bibliographic facet may include a piece’s title, composer, publisher, date, orchestra, soloist(s), and any other available information about the piece (Downie, 2003a).

2.2 Representations and Formats

Music information has many different representations and formats, as well as many facets. It exists in visual (musical scores, manuscripts) and aural (digital music) representations (Huron, 2000). Visually, music information exists in the symbolic formats of Common Music Notation (CMN) (Figure 1) and nonstandard types, like guitar tablature (McClane, 1996). Aurally, music is stored in many formats. Some such forms include records, cassettes, compact disks, MP3s and digital audio formats.

![Figure 2](image-url) Handel's Fugue in B minor in CMN, a symbolic notation
In addition to the two basic representations, visual and aural, early music information research was devoted to new representations of music as text. This format was developed for easy music access, manipulation, and production using computer systems that needed simplified versions of music for easier processing. Simply put, melodies, harmonies, and other aspects of the work were translated into text based on a specified algorithm. This coded the music so that it could be stored in a machine-readable format and therefore retrieved for evaluation and production (McClane, 1996). Downie (2003a) refers to such systems as Analytic/Production MIR systems. By large, MIR systems index music by text-coding if not in original aural or visual formats. There are also hybrid representations of music, such as those consequential of recording (Figure 2) or of the necessity for standard transport between devices such as Music Instrument Digital Interface (MIDI) (McClane, 1996). Though MIDI originally served this purpose, it is used now as a format of storage and retrieval in MIR systems (Burnand, 2004; McClane, 1996). Generally, MIR systems use and serve more than one representation and facet.
2.3 Uses and User Groups

The multitude of music listeners demands extensive MIR capability. Those with a more than a casual need for MIR systems are: music students, performers, composers, and conductors; music therapists; musicologists; music librarians and library patrons; audio engineers; scholars; researchers, and; intellectual property lawyers (Downie, 2003a; Downie, 2001). Each group understands and uses music information differently, so the needs of each user group are distinctly unique. A music student, for example, would have a need for access to bibliographic information, instrumentation (timbral facets), and score visuals, if possible, and could access those by means of most any of Downie’s facets of the music. An intellectual property lawyer may have no musical training and therefore needs solely bibliographic material access, by means of bibliographic queries only.

MIR may be of help in other ways as well. According to a study of the Google Answers service, a human-answered offering from Google, users had a hard time even constructing music-related queries (Bainbridge, Cunningham & Downie, 2003). These results suggest that if MIR systems could aid users in structuring their queries, the general public may have a greater interest in MIR research.

3. Challenges in MIR

The first papers in IR devoted to music were published as early as the 1950’s, and yet professionals still consider it an “emerging discipline” (Downie, 2003a). Why? As described above researchers must first learn all aspects of music for any MIR needs to
make sense. Once they have tackled the learning curve, there are further complications in the actual application of MIR.

Two obstacles are subjectivity and versioning. When a soloist performs a rest longer in duration than is written or an arranger transposes a piece down two semitones for the sake of orchestration, how does a system recognize that these secondary versions are the same as the primary versions? This “plasticity” of music, as Downie calls it, affects how MIR systems handle a work as a separate entity (Downie, 2003b).

Specialization may continue to complicate the MIR discipline as a whole, however. Downie (2003b) identifies three standardization issues in MIR research:

1. No standard collection of music against which each team could test its techniques;
2. No standardized sets of performance, and;
3. No standardized evaluation metrics.

Before specialization can begin within MIR, there must first be some sort of agreement of quality control and assessment methods within the community. Music copyright issues are strict, even for searching purposes. Therefore, collections for use in the public domain, even for the purpose of system evaluation, may be difficult to build. Fortunately, HNH Hong Kong International, Ltd. is working with MIR researchers, Downie and colleagues, to provide “research access to HNH’s entire catalogue of Classical, Jazz, and Asian digital recordings” for international MIR research (Downie, 2003b). A standard database is a step forward, but once this is implemented, researchers must develop a system, similar to that in text IR, that defines how MIR system
performance is evaluated and furthermore what performance levels are acceptable. Text IR has such evaluation standards—for example, the Cranfield model (Cleverdon, 1967). MIR differs from text IR and ultimately may require different parameters for evaluation. Even widely-used evaluation metrics such as recall and precision may not apply to all aspects of MIR, at least in the same fashion (Downie, 2003). For example, relevance may be more difficult to judge in MIR because of the various formats and facets and how they interplay with result sets. A user will often not specify whether he or she wants audio files or visual representations. Should the system be judged on those parameters if they are not specified in the query.

Downie (2001) identifies other major issues in MIR: lack of bibliographic control and interaction between interested disciplines. The vast professions working in and around MIR carry many discipline-specific data. Without communication in place among these groups, MIR systems will remain limited in scope and slow to improve. Downie (2001) has implemented the “Music Information Retrieval Annotated Bibliography Website Project” (www.music-ir.org) which serves to diminish these issues. This website provides a common area for professionals from any discipline, with an interest in MIR, to contribute to and abide by the resources and decisions made available for research. It serves as an informational environment for all projects and research in MIR.

Environments and content also present a challenge and affect the way systems are designed. The main section of a music library, for example, would not be the best place for aural queries because these are traditionally low-noise environments. Aural
results would require headphones at each station as well, in this case. Therefore, system
designers must take into account the environment within which a system may be used.
Similarly, the content of the database is worth consideration. A MIR system can be used
to find any of the abovementioned representations and formats, and a system must be
designed around those found in its database. For example, someone searching through
a database of MP3s would not find it much help to search for the publisher of the score.
These physical and digital environmental factors will be even more noticeably
important as MIR systems are more widely implemented.

4. Current MIR Technology

Like other areas of IR, MIR is advancing, and researchers have lately built
systems to solve these complexities. A variety of systems for the retrieval of music have
addressed a number of complications identified. The following are examples of some
such systems as well as other technologies and projects of interest to the MIR
community.

4.1 Indexing and N-Gramming

A major part of any IR system is the index. In text IR, deciding what to index is
clear: among the options are abstracts, bibliographic material, and full text. For aural
query systems in MIR, commonly known as Query-by-Humming (QBH) systems
(Doraisamy & Rüger, 2002), indexing the music in a way that makes it accessible by
aural query is a complex issue. Music requires a great amount of storage space in an
audio format, like MP3. Similarly, indexing large amounts of audio files could be quite
expensive to process and store. Therefore, MIR indexes extract thematic melodies and
translate them into text-based representations, which are then matched according to the aural query submitted by the user. A system then only process the audio query, transforms it into text code, and compares this to the index. If the index was in audio format, or another format that needs translation, each entry in the index would have to be translated to compare it with the query.

Deciding what part of the music is “thematic” and how it must be indexed is another issue. Musart, an aural query music retrieval system, is one example of how to extract thematic material for indexing and ultimately compare it with search strings (Birmingham et al., 2003). Musart does this automatically according to a specified algorithm and can translate from multiple formats into multiple formats, allowing users to view information in a variety of ways. Musart finds common passages by finding repeated patterns in a work and records the most likely thematic melodies or riffs. Medina and Smith (2001) describe another technique for theme-finding based on musical concepts of a theme. Many developers use the concept of n-grams for querying and indexing. This is not a concept new to IR; it has been used previously in text IR systems (Heaps, 1978). Downie (Downie & Nelson, 2000) first began using n-gramming, defining them as “musical words”. In this case N-grams are small snippets of a melody. Using n-gramming, a MIR system breaks down a query into several small chunks and compares those chunks to indexed n-grams of thematic material. When a match occurs, that work is returned in the result set (Downie & Nelson, 2000). N-gramming not only provides quicker searching, but also acts as a fault-tolerance mechanism.

4.2 Aural Queries
How do those who are not musically inclined query music databases easily? An aural query is one answer. Aural query or QBH systems have a lot of potential. There are numerous examples of these currently available or in development (e.g. Birmingham et al., 2003; Doraisamy & Rüger, 2003; Downie & Nelson, 2000, Uitdenbogerd & Zobel, 2002). These systems take user input for a query string in the form of a melodic line or in some cases a polyphonic input. Once the system parses the query string, in many cases using n-gramming, it attempts to match the interval sequences with those in the database. The Musart system can capture melody, harmony, and rhythm, adding to the complexity and versatility of QBH systems (Birmingham et al., 2003).

4.3 Fault Tolerance of N-gramming and Aural Queries

Most recently developers’ concerns have focused on the quality control of such query systems. Researchers have recently begun to use recall and precision for the evaluation of MIR systems. Although Downie (2003b) questions the validity of such measures for MIR evaluation, some researchers are finding that these measures do apply (e.g. Uitdenbogerd & Zobel, 1999). Research has afforded measures for fault tolerance in aural queries. By nature, n-gramming is somewhat fault-tolerant because it can match any part of a query to a similar part in the database (Downie, 2000). For example, a query that is six notes long may be distorted between the fifth and sixth notes. Using n-gramming a query may be broken down into just the first three or four pitches, affording three usable intervals to match with the indexed n-grams. Others have taken this n-gram fault tolerance further and included pitch assumption
techniques that estimate the pitch intended by the searcher (Birmingham et al., 2003; Sorsa & Huopaniemi, 2001). Researchers have also begun to test these fault-tolerance capabilities in an effort to relay common shortcomings of QBH systems (Doraisamy & Rüger, 2002; Huopaniemi & Sorsa, 2001).

4.4 Polyphonic Focus

A thematic melody is a highly memorable aspect of a musical work, arguably the most memorable (Birmingham et al., 2003). Music as a whole is not monophonic as are the melodies humans mostly likely remember. This spurs an interest to implement systems with polyphonic capabilities. Since most queries will be monophonic and databases contain chiefly polyphonic music, this may affect the quality of results. Doraisamy and Rüger (2002) evaluated monophonic queries against a full-music index of polyphonic music and found the results “promising”. Querying a database of polyphonic music using polyphonic queries is another concern. Harmonies are more complex as inputs and in use as matching techniques. If a user performs a C Major chord with C as the root (bottom note) and the desired work has a C Major chord with E as the root instead, this may cause a mismatch in the database or no match at all. Some algorithms can overcome this by flattening the chord tones in the query (Birmingham et al., 2003), which enables a more likely match. This may not be suited for larger datasets where more matches could be made because larger datasets yield larger result sets. Another issue with this is that flattening a harmony loses the shape of the original chord. Shape is an aspect of timbre, and anytime one loses a note in the chord, a significant change occurs in the music.
4.5 Fusing the Representations and Formats

One way to solve the issue of multiple levels of music ability and needs is to synchronize the data in multiple formats. Arifi, Clausen, Kurth, and Muller (2003) began in this vein by synchronizing three formats: Score, MIDI, and PCM. Using this technology a MIR system can search through any of the synchronized formats concurrently, providing fast searching and results in various forms. Furthermore, a system using similar technology could tailor results specific to users; aligning MIR with the principles of Personalized Information Retrieval.

5. Concerns

Researchers and developers have certainly made advancements toward an understanding of the complexities of MIR. Still, there are issues with which researchers have not yet dealt. The following are young or undiscovered topics in MIR research.

5.1 Another Facet?

Jan LaRue (1970) argues that there are five aspects of music which guide music analysis: sound, harmony, melody, rhythm, and growth. LaRue’s aspects closely parallel those of Downie (2003a), with some exception (Table 1). LaRue includes Downie’s bibliographic, editorial, and textual facets as part of sound, harmony, melody, and rhythm. The aspect of music that does not truly parallel any of Downie’s facets and may be added as MIR matures is growth. Growth is an aspect of music that describes the continuation methods and overall movement within a piece of music. For example, a climax is a part of growth. Though the subjectivity of this may be beyond the scope of current MIR, it may prove to be very useful to some user groups, as a means of analysis.
and search. Some automatic structural analysis may also be advantageous to parse music for basic structure, such as sonata, song, and rondo forms.

Table 1 Comparison of LaRue's aspects to Downie's facets

<table>
<thead>
<tr>
<th>LaRue</th>
<th>Downie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound</td>
<td>Timbral</td>
</tr>
<tr>
<td>Harmony</td>
<td>Harmonic</td>
</tr>
<tr>
<td>Melody</td>
<td>Pitch</td>
</tr>
<tr>
<td>Rhythm</td>
<td>Temporal</td>
</tr>
<tr>
<td>Growth</td>
<td>???</td>
</tr>
</tbody>
</table>

5.2 Interfaces and Usability

Though MIR research is still considered to be in its infancy (Downie, 2003a), those involved must not overlook those advancements currently made in other areas of IR. Specifically, interface design does not seem to be a high priority in MIR. One of the few groups in MIR that has looked into new interface possibilities is Pampalk, Dixon, and Widmer (2003). Though they developed an interface that presents multiple views of a result set simultaneously, it remains fairly unfriendly (Figure 3). As IR continues to move to a more user-centered design approach, MIR systems must also include a well-designed interface.
6. Recommendations

Music’s diverse representations and formats give rise to a number of opportunities to exploit technology in favor of its users. Olson and Downie (2003) touch on the possibilities during the construction of an online digital score collection. They mention content-based search features, searching of the symbolic content, optical music recognition—a technique likely akin to optical character resolution by which musical symbols may be parsed when scanned—and distinguishing between versions and variations of a work (Olson & Downie, 2003). To spur the progression of MIR development, researchers must take advantage of the many uses and representations and learn to build environments of music information through which many may explore.

Most of the abovementioned technology that is currently available or in development stages is well-directed to solve many issues in MIR. There is a focus issue,
however. Not one of these technologies is currently in widespread use. These systems are still being researched and developed, tested and retested, or waiting for proper measures to test the ability of systems. The best move for MIR is to break into the public sector. Allow the technology to be seen and used by music libraries or retail music stores. Once the public eye sees how advanced MIR is and once they understand the implications for the future of this technology, they will push vendors and research forward with desire for these products.

7. Conclusions

As the demand for MIR increases, music information research will grow more in size and complexity. MIR researchers are currently creating technologies that, however primitive, have promise. Hurdles in the profession, such as copyright laws and standardization among the interested disciplines, may prevent MIR from progressing at the rate other IR research has. Though MIR professionals have yet to make interface design and usability a main focus, these will inevitably direct advancement and user-acceptance. MIR systems will remain largely in the research arena, however, until researchers begin to offer practical implementation of the technology currently available.
References


