IMPORTANT
PLEASE READ

Brief Citation:

Downie, J. Stephen. 2003. Music information retrieval. *Annual Review of Information Science and Technology 37*: 295-340. Available from http://music-ir.org/downie_mir_arist37.pdf

Full Citation:

Downie, J. Stephen. Music information retrieval (Chapter 7). In *Annual Review of Information Science and Technology 37*, ed. Blaise Cronin, 295-340. Medford, NJ: Information Today, 2003. Available from http://music-ir.org/downie_mir_arist37.pdf

FYI: The URLs in the document should be "hot" and usable while reading in Acrobat Reader.

Comments and feedback: jdownie@uiuc.edu

CHAPTER 7

Music Information Retrieval

J. Stephen Downie
University of Illinois at Urbana-Champaign

Introduction

Imagine a world where you walk up to a computer and sing the song fragment that has been plaguing you since breakfast. The computer accepts your off-key singing, corrects your request, and promptly suggests to you that "Camptown Races" is the cause of your irritation. You confirm the computer's suggestion by listening to one of the many MP3 files it has found. Satisfied, you kindly decline the offer to retrieve all extant versions of the song, including a recently released Italian rap rendition and an orchestral score featuring a bagpipe duet.

Does such a system exist today? No. Will it in the future? Yes. Will such a system be easy to produce? Most decidedly not.

Myriad difficulties remain to be overcome before the creation, deployment, and evaluation of robust, large-scale, and content-based Music Information Retrieval (MIR) systems become reality. The dizzyingly complex interaction of music's pitch, temporal, harmonic, timbral, editorial, textual, and bibliographic "facets," for example, demonstrates just one of MIR's perplexing problems. The choice of music representation—whether symbol-based, audio-based, or both—further compounds matters, as each choice determines bandwidth, computation, storage, retrieval, and interface requirements and capabilities. Overlay the multicultural, multiexperiential, and multidisciplinary aspects of music and it becomes apparent that the challenges facing MIR research and development are far from trivial.

Consider the sheer magnitude of available music facing MIR researchers: 10,000 new albums are released and 100,000 works registered for copyright each year (Uitdenbogerd & Zobel, 1999). Notwithstanding the intrinsic intellectual merits of MIR research problems, the successful development of robust, large-scale MIR systems will also have important social and commercial implications. According to Wordspot (2001), an Internet consulting company that tracks queries submitted to Internet search engines, the search for music—specifically, the now-ubiquitous MP3 format—has displaced the search for sex-related materials as the most popular retrieval request. Yet at this moment, not one of the so-called “MP3 search engines” is doing anything more than indexing the textual metadata supplied by the creators of the files. It is not an exaggeration to claim that a successful, commercially based, MIR system has the potential to generate vast revenue. In the U.S. alone, 1.08 billion units of recorded music (e.g., CDs, cassettes, music videos, and so forth), valued at $14.3 billion, were shipped to retailers in 2000 (Recording Industry Association of America, 2001). Vivendi Universal, parent company of Universal Music Group, recently bought MP3.com, a popular Internet-based distributor of MP3 files, for $372 million (Welte, 2001). Beyond the commercial implications, the emergence of robust MIR systems will create significant added value to the huge collections of underused music currently warehoused in the world’s libraries by making the entire corpus of music readily accessible. This accessibility will be highly beneficial to musicians, scholars, students, and members of the general public alike.

A growing international MIR research community is being formed, drawing upon multidisciplinary expertise from library science, information science, musicology, music theory, audio engineering, computer science, law, and business. Through an examination of the multidisciplinary approach to MIR, this chapter identifies and explicates the MIR problem spaces, historic influences, current state-of-the art, and future MIR solutions. The chapter also outlines some of the major difficulties that the MIR community faces as MIR research and development grows and matures into a discipline in its own right.

Facets of Music Information: The Multifaceted Challenge

Over the years, I have found it useful to conceive of music information as consisting of seven facets, each of which plays a variety of roles in defining the MIR domain. These facets are the pitch, temporal, harmonic, timbral, editorial, textual, and bibliographic facets. Due to the intricacies inherent in the representation of music information, what follows is not a facet analysis in the strict sense because the facets are not mutually exclusive. For example, the term *adagio* when found in a score could be placed within both the temporal and editorial facets, depending on context. The harmonic facet, likewise, chiefly derives from the interplay of the pitch and temporal facets. The difficulties that arise from the complex interaction of the different music information facets can be labeled the “multifaceted challenge.”

Pitch Facet

Pitch is “the perceived quality of a sound that is chiefly a function of its fundamental frequency in—the number of oscillations per second” (Randel, 1986, p. 638). The graphical representation (e.g., ♪, 𝅗𝅥, ω, I, etc.) in which pitch is represented by the vertical position of a note on the staff is familiar to most. Note names (e.g., A, B, $C^{\#}$), scale degrees (e.g., I, II, III …VII), solfège (e.g., do, ré, mi … ti) and pitch-class numbers (e.g., 0, 1, 2, 3 ... 11) are also among the many methods of representing pitch.

The difference between two pitches is called an interval. Intervals can be represented by the signed difference between two pitches as measured in semitones (e.g., -8, -7 ... -1, 0, +1 ... +7, +8, etc.) or by its tonal quality as determined by the location of the two pitches within the syntax of the Western theoretical tradition. For example, the interval between A and $C^{\#}$ is called a Major 3rd, whereas the aurally equivalent distance between A and $D^{\flat}$ is a Diminished 4th. Melodies can be considered sets of either pitches or intervals perceived as being sequentially ordered through time.

The notion of key is included here as a subfacet of pitch. The melodic fragment EDCEDC (e.g., “Three Blind Mice”) in the key of C Major is considered to be musically equivalent to BAGBAG in the key of G Major.

That is to say, their melodic contours (i.e., the pattern of intervals) are perceived by listeners to be equivalent despite the fact that their absolute pitches are different.

Temporal Facet

Information concerning the duration of musical events falls under the temporal facet. This includes tempo indicators, meter, pitch duration, harmonic duration, and accents. Taken together these five elements make up the rhythmic component of a musical work. Rests in their various forms can be considered indicators of the duration of musical events that contain no pitch information. Temporal information poses significant representational and access problems. Temporal information can be absolute (e.g., a metronome indication of MM=80), general (e.g., *adagio, presto, fermata*), or relative (e.g., *schneller, langsamer*). Temporal distortions are sometimes encountered (e.g., *rubato, accelerando, rallentando*). Because the rhythmic aspects of a work are determined by the complex interaction of tempo, meter, pitch, and harmonic durations, and accent (whether denoted or implied), it is possible to represent a given rhythmic pattern many different ways, all of which yield aurally identical results. Some performance practices, in which it is expected that the player(s) will deviate from the strict rhythmic values noted in the score (e.g., in Baroque, Jazz), give rise to added complexities, similar to those caused by the temporal distortions mentioned above. Thus, representing temporal information for retrieval purposes is quite difficult indeed.

Harmonic Facet

When two or more pitches sound at the same time, a simultaneity, or harmony, is said to have occurred. This is also known as polyphony. The absence of polyphony is called monophony (i.e., only one pitch sounding at a time). Pitches that align vertically in a standard Western score are creating harmony. The interaction of the pitch and temporal facets to create polyphony is a central feature of Western music. Over the centuries music theorists have codified the most common simultaneities into several comprehensive representational systems, based upon their constituent intervals or pitches and the perceived function of those intervals or pitches within contexts of the works in which they appear. Theorists have also codified the common sequential patterns of simultaneities found within Western tonal music. Although it is beyond the scope of this chapter to examine in detail the complex realm of Western harmonic theory and praxis, it is important to note that an individual harmonic event can be denoted by a combination of the pitches or interval(s) it contains and the scale position of its "root," or fundamental, pitch. A chord, like that sounded when a guitar is strummed, is an example of an harmonic event. Sequences of chords, or harmonic events, can be represented by chord names. The very common harmonic sequence, or progression, in the key of C Major, [C+ F+ G+ C+] is here represented by the note name of the fundamental pitch of each chord. The "+" denotes that each chord contains the intervals of Major 3rd and Perfect 5th as measured from the fundamental note. Another method of representing this harmonic progression, that generalizes it to all major keys, is to indicate the scale degree of the root of the chord using Roman numeral notation: I-IV-V-I.

Simple access to the codified aspects of a work's harmonic information can be problematic because its harmonic events, although present in the score, are not usually denoted explicitly in one of the ways described above. Exceptions to this are the inclusion of chord names or chord symbols in most popular sheet music, and the harmonic shorthand, called *basso continuo*, or figured bass, commonly found in music of the Baroque period. The matter is further complicated by the fact that the human mind can perceive and consistently name one of the codified simultaneities, despite the presence of extra pitches called non-chord tones. Even with the absence, or delay, of one or more of the chord's constituent pitches, most members of Western societies can still consistently classify the chord.

Timbral Facet

The timbral facet comprises all aspects of tone color. The aural distinction between a note played upon a flute and upon a clarinet is caused by the differences in timbre. Thus, orchestration information, that is, the designation of specific instruments to perform all, or part, of a work, falls under this facet. In practice orchestration information, although really part of the timbral facet, is sometimes considered part of the bibliographic facet. The simple enumeration of the instruments used in a composition is usually included as part of a standard bibliographic record. This information has been found to assist in the description, and thus the identification, of musical works.

A wide range of performance methods also affects the timbre of music (e.g., *pizzicatti*, mutings, pedalings, bowings). Here the border between timbral and editorial information becomes blurred, as these performance methods can also be placed within the editorial facet. The act of designating a performance method that affects timbre is editorial; the aural effect of the performance of the chosen method is timbral. Timbral information is best conveyed in an audio, or signal-based, representation of a work. Accessing timbral information through a timbral query (e.g., playing a muted trumpet and asking for matches) requires advanced signal processing capabilities. A simpler, yet less precise, method would be to access timbral information through some type of interpretation of the editorial markings. This possible solution would, of course, be subject to the same difficulties associated with representing editorial information, which are discussed next.

Editorial Facet

Performance instructions make up the majority of the editorial facet. These include fingerings, ornamentation, dynamic instructions (e.g., *ppp, p, ...f, fff*), slurs, articulations, *staccati*, bowings, and so on. The vagaries of the editorial facet pose numerous difficulties. One difficulty associated with editorial information is that it can be either iconic (e.g., -, 3, !), or textual (e.g., crescendo, diminuendo), or both. Furthermore, editorial information can also include the parts of the music itself. The writing out of the harmonies from the *basso continuo*, also known as the "realization of the figured-bass," is an editorial act. *Cadenzi* and other solos, originally intended by many composers to be improvised, are frequently realized by the editor. Lack of editorial information is yet another problem to be considered. Like the *basso continuo*, where the harmonies are implied, nearly all composers prior to Beethoven—and many since—have simply assumed that the performers were competent to render the work in the proper manner without aid of editorial information. In many cases, the editorial discrepancies between editions of the same work make the choice of a "definitive" version of a work for inclusion in a MIR system very problematic.

Textual Facet

The lyrics of songs, arias, chorales, hymns, symphonies, and so on, are included in the textual facet. Libretti, the text of operas, are also included. It is important to note that the textual facet of music information is more independent of the melodies and arrangements that are associated with it than one would generally believe. A given lyric fragment is sometimes not informative enough to identify and retrieve a desired melody and vice versa (Temperley, 1993). Freely interchanging lyrics and music is a strong tradition in Western music. A good example of this phenomenon is the tune, "God Save the Queen." Known to citizens of the British Commonwealth as their royal anthem, this simple tune is also known to Americans as their republican song, "America," or "My Country 'tis of Thee." Many songs have also undergone translation into many different languages. Simply put, one must be aware that a given melody might have multiple texts and that a given text might have multiple musical settings. It is also important to remember the existence of an enormous corpus of music without any text whatsoever.

Bibliographic Facet

Information concerning a work's title, composer, arranger, editor, lyric author, publisher, edition, catalogue number, publication date, discography, performer(s), and so on, are all aspects of the bibliographic facet. This is the only facet of music information that is not derived *from* the content of a composition; it is, rather, information, in the descriptive sense, *about* a musical work. It is music metadata. All of the difficulties associated with traditional bibliographic description and access also apply here. Howard and Schlichte (1988) outline these problems along with some of their proposed solutions. Temperley (1993) is another important work tackling this difficult subject.

Why Is MIR Development So Challenging?

The multifaceted challenge, unfortunately, is not the only problem facing MIR research. Developers and evaluators must constantly take into account the many different ways music can be represented (i.e., the "multirepresentational challenge"). Music transcends time and cultural boundaries, yet each historic epoch, culture, and subculture has created its own unique way of expressing itself musically. This wide variety of expression gives rise to the "multicultural challenge." Comprehending and responding to the many different ways individuals interact with music and MIR systems constitutes the "multiexperiential challenge." Maximizing the benefits of having a multidisciplinary research community while minimizing its inherent drawbacks represents MIR's "multidisciplinarity challenge." For another informative overview of the difficulties facing MIR research, I recommend Byrd and Crawford (2002).

The Multirepresentational Challenge

With the exception of the bibliographic facet, each of the aforementioned facets can be represented as symbols, as audio, or both. Symbolic representations include printed notes, scores, text, and myriad discrete computer encodings, including Musical Instrument Digital Interface (MIDI), GUIDO Music Notation Format, Kern, and Notation Interchange File Format (NIFF). Audio representations include live performances and recordings, both analog and digital (e.g., LPs, MP3 files, CDs, and tapes). The choice of representations, whether they be symbolic or audio, is predicated on a mixture of factors including desired uses of the systems, computational resources, and bandwidth. Symbolic representations tend to draw upon fewer computational and bandwidth resources than do audio representations. For example, a 10-second snippet of music represented in stereophonic CD-quality digital audio requires approximately 14 megabits of data to be processed, transmitted, or stored. Under the simplest of symbolic representations, the same musical event could be represented in as few as eight to 16 bits. However, because the vast majority of listeners understand music solely as an auditory art form, many MIR developers see the inclusion of audio representations, despite their inherent consumption of resources, as absolutely necessary.

The pragmatics of simple availability (or nonavailability) of particular representations is also influencing design decisions. For example, many researchers limit themselves to using music in the MIDI, CD, and/or MP3 formats because it is relatively easy to build collections of these types using Web spidering techniques. Intellectual property issues also create availability difficulties for system developers. The 1998

Sonny Bono Copyright Term Extension Act has created a situation where "virtually all sound recordings are protected until the year 2067" (Haven Sound, 2001). Under the terms of this law, building a multirepresentational database that integrates royalty-free public domain scores (e.g., Bach, Beethoven) and MIDI files with readily accessible audio recordings (e.g., CD or MP3 files) might become impossible for all but the very well financed. An academic developer might, for example, have a collection of public domain MIDI files and scores representing the keyboard works of the Baroque period but cannot provide a more robust, multi-representational set of access methods because of the financial and administrative costs associated with obtaining copyright clearances for the requisite MP3 representations. Levering (2000) provides a summary of intellectual property law as it pertains to the development of digital music libraries and MIR systems. Extensive information about locating and using public domain music can be found at http://www. pdinfo.com/.

The Multicultural Challenge

Music information is, of course, multicultural. However, a cursory review of the extant MIR literature could lead one to the erroneous conclusion that the only music worth retrieving is tonal Western classical and popular music of the last four centuries (i.e., music based on what is known as "Common Practice"). I believe the bias toward Western Common Practice (CP) music has three causes. First, there are many styles of music for which symbolic and audio encodings are not available, nonstandard, or incomplete. Improvised jazz, electronic art music, music of Asia, and performances of Indian ragas all are examples. Likewise, we do not yet have comprehensive recording sets of African tribal songs nor Inuit throat music. Acquiring, recording, transcribing, and encoding music are all time-consuming and expensive activities. For some musics, whole new encoding schemes will also have to be developed. Thus, it is pragmatically more expedient to build systems based upon easier-to-obtain, easier-to-manipulate, CP music. Second, developers are more familiar with CP music than with other styles, and thus are working with that which they understand. Third, I believe that developers wish to maximize the size of their potential user base and therefore have focused their efforts on CP music because it arguably has the largest transcultural audience. Bonardi (2000) provides an informative overview of the shortcomings of CP representations and the problems musicologists experience as they work with non-CP materials.

The Multiexperiential Challenge

Music ultimately exists in the mind of its perceiver. Therefore, the perception, appreciation, and experience of music will vary not only across the multitudes of minds that apprehend it, but will also vary within each mind as the individual's mood, situation, and circumstances change. Music can be experienced as an object of study, either through scores or through the deliberate audition of recordings, as is the case with many music students, music lovers, and musicologists. Sometimes these same music experts will relegate their objects of study to the background during housework and "listen" to them only at a subconscious level. Soundtrack recordings are listened to by many as an *aide memoire* to re-invoke the pleasurable experience of going to the cinema or theater. Music can be experienced as a continuation of familiar traditions with the singing of nursery songs, hymns, camp songs, and holiday carols being prime examples. Music is experienced by some as a means of religious expression, sublime or ecstatic, through such genres as plainsong, chants, hymns, masses, and requiems. David Huron (2000) suggests that music has drug-like qualities. He contends that users seek out not specific melodic or harmonic experiences, but actual physical and emotional alterations. The seeking out of a certain kind of energetic euphoria that one might associate with hip-hop or acid-house music is a case in point.

The seemingly infinite variety of music experience poses two significant hurdles for MIR developers. First, it raises the problems of intended audience and intended use. Which set of users will be privileged and which set of uses addressed? Even if it were possible to somehow encode, query, and retrieve the drug-like effects of the various pieces of music within an MIR database, would such a system also support the analytic needs of the musicologist?

Second, the multiexperiential problem prompts questions about the very nature of music similarity and relevance. For the most part, the notion of similarity for the purposes of retrieval has been confined to the codified, and relatively limited, areas of music's melodic, rhythmic, harmonic, and timbral aspects. Thus, music objects that have some intervals, beats, chords, and/or orchestration in common are deemed to be "similar" to some extent, and hence are also deemed to be potentially "relevant" for the purposes of evaluation. For background information on the importance of, and the controversies surrounding, the notion of relevance in the traditional IR literature, Schamber (1994) is an excellent resource. For an explication of relevance issues as they pertain to MIR, Byrd and Crawford (2002) is highly recommended.

Computing in Musicology (Hewlett & Selfridge-Field, 1998) has devoted an entire volume to issues surrounding melodic similarity. For those interested, I recommend the complete volume. Several of the articles stand out as exemplary explorations of some of the fundamental concepts in MIR research. Selfridge-Field (1997) provides an excellent overview of the myriad problems associated with MIR development. Crawford, Iliopoulos, and Raman (1998) review the amazing variety of string-matching techniques that can be used in MIR. Howard (1998) discusses an interesting procedure for sorting music *incipits*. Cronin (1998) examines U.S. case law pertaining to copyright infringement suits along with analyses and explications for the decisions made by the courts on what constitutes music similarity.

In what ways, however, do we assess the similarity of a user's *experience* of one piece with others in a collection? How is a desired mood or physiological effect to be considered "similar" to a particular musical work? How would we modify an "experiential" similarity measure as the mood and perceptions of the individual users change over time? How do we adjust our relevance judgments under this scenario of ever-shifting moods and perceptions? Perhaps some combinations of melodic, rhythmic, harmonic, and timbral similarities do play a significant role in the similarity of experiences, and further research will uncover how this is so. Given the undeniable importance of music's experiential component, it is possible that future MIR systems will need to incorporate some type of biofeedback mechanism designed to assess the physiological responses of users as retrieval options are presented to them. Although the idea of having users biometrically "plugged in" to MIR systems sounds fanciful, we must remember that the experiential component of music directly shapes our internal conception of similarity and our internal conception of similarity, in turn, determines our relevance judgments. In short, to ignore the experiential aspect of the music retrieval process is to diminish the very core of the MIR endeavor; namely, the retrieval of relevant music objects for each query submitted. The creation of rigorous and practicable theories concerning the nature of experiential similarity and relevance is the single most important challenge facing MIR researchers today.

The Multidisciplinarity Challenge

The rich intellectual diversity of the MIR research community is both a blessing and a curse. MIR research and development are much stronger for having a wide range of expertise being brought to bear on the problems: audio engineers working on signal processing, musicologists on symbolic representation issues, computer scientists on pattern matching techniques, librarians on bibliographic description concerns, and so on. However, this diversity presents some serious difficulties that threaten to hinder MIR research and development.

The heterogeneity of disciplinary worldviews is particularly problematic. Each contributing discipline brings to the MIR community its own set of goals, accepted practices, valid research questions, and generalizable evaluation paradigms. Of these, the variance in evaluation paradigms is most troubling. To compare and contrast the contributions of the different MIR projects being reported in the literature is difficult at present because the research teams are evaluating their approaches using such a wide variety of formal and ad hoc evaluation methods. Complexity analyses, empirical time-space analyses, informetric analyses, traditional information retrieval (IR) evaluations, and algorithmic validations are but a few of the evaluation techniques employed.

It is worth noting that, for a research area that contains "information retrieval" in its name, the number of published works actually drawing upon any of the formal IR evaluation techniques is strikingly low. By "formal IR evaluation" is meant studies of the kind usually performed within the discipline of information retrieval as described by Keen (1992), Korfhage (1997), Tague-Sutcliffe (1992), and most definitively by Harter and Hert (1997). Projects described in Downie (1999), Foote (1997), and Uitdenbogerd and Zobel (1999) are among the very few that report results using the traditional IR metrics of precision and recall. Even among these three, each has taken a slightly different analytic approach: Foote uses average precision, Downie uses normalized-precision and normalized-recall, whereas Uitdengoberd and Zobel use 11-point recall-precision averages and precision-at-20 measures.

Why are the IR evaluation techniques not being widely accepted, and when they are applied, why not in a consistent manner? The lack of familiarity among members of the various domains with traditional IR evaluation techniques, and their associated metrics, is one reason. Another reason is the lack of standardized, multirepresentational test collections: intellectual property issues are one of the serious problems inhibiting their creation. Notwithstanding the absence of test collections, no standardized sets of queries, or relevance judgments, exist either: the MIR community has yet to arrive at a consensus concerning what constitutes a typical set of queries, and, as explained previously, the relevance question remains unresolved.

Communications are also problematic in MIR's multidisciplinary environment. Language and knowledge-base problems abound, making it difficult for members of one discipline to truly appreciate the efforts of the others. For example, when signal processing experts present their works replete with such abbreviations as FFT (Fast Fourier Transform), STFT (Short Time Fourier Transform), and MFCC (Mel-Frequency Cepstral Coefficients), their fellow experts will have no difficulty in understanding them for these are, in fact, rather rudimentary signal processing concepts. However, for most musicologists, comprehending these terms and the underlying concepts they represent will require hours of extra study. Similarly, to a signal processing expert, the enharmonic equivalence of $G^{\#}$ and $A^{\flat}$ is generally seen as a distinction without a difference. To a musicologist, however, it is common knowledge that this equivalence is not necessarily one of absolute equality for the choice of note name can imply the contextual function of the pitch in question. Communication matters are made worse because the MIR literature has no disciplinary "home base": no official MIR society, journal, or foundational textbook exists through which interested persons can acquire the basics of MIR. With the exception of a handful of small panels, workshops, and symposia (discussed later), most researchers are presenting their MIR results to members of their own disciplines (i.e., through discipline-specific conferences and publications). The MIR literature is thus difficult to locate and difficult to read. A fragmented and basically incomprehensible literature is not something upon which a nascent research community can expect to build and sustain a thriving, unified, and respected discipline.

Representational Completeness and MIR Systems

McLane's chapter in the 1996 *Annual Review of Information Science and Technology,* entitled "Music as Information," is a superlative review of the many Music Representation Languages (MRLs) that have been developed or proposed for use in MIR systems (McLane, 1996). A thorough technical comparison of the attributes of five of the historically most important MRLs can be found in Selfridge-Field (1993–1994). Selfridge-Field describes, in easy-to-understand tabular form, how the facets of music information are (or are not) represented in the MuseData, Digital Alternative Representation of Music Scores (DARMS), SCORE, MIDI, and Kern MRLs. *Beyond MIDI: The Handbook of Musical Codes* (Selfridge-Field, 1997) is an excellent resource for deeper exploration of MRL issues.

It is not the purpose of this chapter to evaluate the relative merits of individual MRLs. What is of interest, however, is the role "representational completeness" plays in the creation of various MIR systems. Inspired by McLane (1996), I define the degree of "representational completeness" by the number of music information facets (and their subfacets) included in the representation of a musical work, or corpus of works. A system that includes all the music information facets (and their subfacets), in both audio and symbolic forms, is "representationally complete."

In general, MIR systems can be grouped into two categories: *Analytic/Production* MIR systems and *Locating* MIR systems. The two types of MIR systems can, in general, be distinguished by (1) their intended uses, and (2) their levels of representational completeness. Of the two, Analytic/Production systems usually contain the more complete representation of music information. If one considers a high degree of representational completeness to be *depth,* and the number of musical works included to be *breadth,* then Analytic/Production MIR systems tend toward depth at the expense of breadth, whereas Locating

MIR systems tend toward breadth at the expense of depth. Working descriptions of the two types of MIR systems are given next.

Analytic/Production MIR Systems

Intended users of Analytic/Production MIR systems include such experts as musicologists, music theorists, music engravers, and composers. These MIR systems have been designed with the goal of being as representationally complete as possible, especially with regard to the symbolic aspects of music. For the most part, designers of such systems wish to afford fine-grained access to all the aforementioned facets of music information, with the possible downplaying of the bibliographic facet. Fine-grained access to music information is required by musicologists to perform detailed theoretical analyses of, for example, the melodic, harmonic, or rhythmic structures of a given work, or body of works. Engravers need fine-grained access to assist them in the efficient production of publication-quality musical scores and parts. Composers make use of fine-grained access to manipulate the myriad musical elements that make up a composition. Because of the storage and computational requirements associated with high degrees of representational completeness, Analytic/Production systems usually contain far fewer musical works than Locating MIR systems.

Locating MIR Systems

Locating MIR systems have been designed to assist in the identification, location, and retrieval of musical works. Text-based analogs include online public access catalogs (OPACs); full-text, bibliographic information retrieval (FBIR) systems, like those provided by the Dialog collection of databases; and the various World Wide Web search engines. Intended users are expected to have a wide range of musical knowledge, ranging from the musically naïve to expert musicologists and other musically sophisticated professionals. For the most part, users wish to make use of the musical works retrieved, either for performance or audition, rather than analyzing or manipulating the various facets of the music information contained within the system. Thus, the objects of retrieval can be considered to be more coarsely grained than those associated with Analytic/Production MIR systems. Because the objects of retrieval are more coarsely grained, access points to music information have been traditionally limited to various combinations of select aspects of the pitch, temporal, textual, and bibliographic facets. Recent research advances, however, suggest that access to the timbral and harmonic (i.e., polyphony) facets should become more common in the near future. The following section will help clarify the principal characteristics of a Locating MIR system.

Uses of a Locating MIR System

Some queries in the field of music are text-based and parallel those in other fields. The bibliographic and textual facets of music information can be used[1] to answer the following queries:

- List all compositions, or all compositions of a certain form, by a specified composer
- List all recordings of a specified composition, or composer
- List all recordings of a specified performer
- Identify a song title given a line of lyrics, or vice versa

A good review of the role the computer has played in improving retrieval from textual catalogs of musical scores and discographies is found in Duggan (1992). She points out, for example, that the Online Computer Library Center (OCLC) contains catalog records for 606,000 scores and 719,000 sound recordings, and the Music Library CD-ROM published by SilverPlatter contains more than 408,000 records for sound recordings. However, the ability to store some searchable representation of the music itself provides the user with the capability of answering queries beyond those served by a Machine Readable Cataloging (MARC) format bibliographic catalog:

- Given a composer, identify by the first few bars each of his or her compositions, or compositions of a certain type

This type of query has traditionally been answered by means of printed *incipit* indexes, typically simple listings of the beginning bars of the scores in a particular collection. Edson (1970) is a good example of a printed *incipit* index. Composer-specific thematic catalogues, such as *Bach-Werke-Verzeichnis* (J. S. Bach) (Schmieder, 1990) or *The Schubert Thematic Catalogue* (Deutsch & Wakeling, 1995), also have a rich tradition of use.

- Given a melody, for example the tune of a song or the theme of a symphony, identify the composition or work

This type of query has traditionally been answered by thematic indexes to musical compositions. Barlow and Morgenstern (1949) is an example; their book contains a few bars of one or more themes from 10,000 musical compositions, arranged by composer. A "Notation Index" in the back of the book permits the user to look up a sequence of six to eight notes, transposed into the key of C, as an alphabetical listing of transposed "themes" to identify the composition in which it occurs. Consider just how incomplete a representation of a given work is provided by the "Notation Index"; it contains only a minimalist representation of the pitch facet. Missing from this representation is all key, harmonic, temporal, editorial, textual, timbral, and bibliographic information. The *National Tune Index* (Keller & Rabson, 1980) offers two similarly minimalist representations of musical *incipits*: scale degree (represented by number) and interval-only sequence (represented by signed integers). Lincoln's (1989) index of Italian madrigals also contains an interval-only (signed integers) representation of the *incipits* it contains. The index developed by Parsons (1975) reduces the degree of representational completeness to an extreme. His index represents musical *incipits* as strings of intervals using text strings containing only four symbols—*, R, U, and D—where "*" indicates *incipit* beginning, "R" for note Repeats (interval of 0 semitones), "U" for Up (any positive interval), and "D" for Down (any negative interval).

Representational Incompleteness and Locating MIR Systems

Obviously, such incomplete representations would have very limited use in an Analytic/Production MIR system. However, as locating tools, these minimal representations have shown themselves to have surprising merit. In fact, it is the incompleteness of their music representations that makes them effective as access tools. By limiting the amount of information contained in the representation of the *incipits*, these indexes also reduce the need for the user to come up with more representationally complete queries. Thus, the musically naïve information seeker can use these representations with relatively few opportunities for introducing query errors. Furthermore, should an error be introduced, it is less likely to result in an identification or retrieval failure. So, for the purposes of identification, location, and retrieval, that is to say, for the essential functions of Locating MIR systems, it is not necessary, nor desirable, to have representational completeness.

This conclusion is supported by McLane (1996, p. 240), who commented on Locating MIR systems:

> Both the choice of view from a representation of music and the degree of completeness of a work's representation depend on the user's information needs. Information retrieval is an interactive process that depends on the knowledge of the user and the level of complexity of the desired information. In the case of the need for the simple identification of a musical work where bibliographic information is not unique enough, one may limit the view to a subjective one involving a relatively small subset of the notated elements of the work, often the pitches of an opening melodic phrase. The representation of pitches will be in a form that the user is likely to expect and be able to formulate a query using the same terminology, or at least one that is translatable into the form of the representation.

I have concluded that representational completeness is not a prerequisite for the creation of a useful Locating MIR system. However, why is it that music information tends to be reduced to simplistic representations of the pitch facet for retrieval purposes? Why not use simple representations of the rhythm facet, or perhaps, the timbral facet? McLane's (1996) comments and the decisions by Barlow and Morgenstern (1949), Parsons (1975), Keller and Rabson (1980), and Lincoln (1989) to represent only the pitch facet, and that only simplistically, were not arbitrary. Pyschoacoustic research has shown the contour, or shape, of a melody to be its most memorable feature (Dowling, 1978; Kruhmhansl & Bharucha, 1986). Thus, any representation that highlights a work's melodic contour (i.e., sequences of intervals) while filtering out extraneous information (e.g., exact pitches, rhythmic patterns) should, in theory, increase the chances for the successful identification, location, and retrieval of a musical work.

More Uses of Locating MIR Systems

Some Locating MIR systems are best considered automated replications of *incipit* and thematic indexes: The *RISM* (1997) database and Prechelt and Typke's (2001) *Tuneserver* are good examples. Other systems, like the *MELDEX* systems discussed by McNab and colleagues exploit the information found in some machine-readable "full-text" representation of the music to overcome the limitations of *incipit* and thematic indexes (McNab, Smith, Bainbridge, & Witten, 1997; McNab, Smith, Witten, Henderson, & Cunningham, 1996). Here "full-text" is used in the sense that melodic information is not arbitrarily truncated (as it is in *incipit* and thematic indexes). For example, Parsons' (1975) index contains no melodic string longer than fifteen notes. The greatest advantage to extending the traditional *incipit* and thematic indexes to include full-text information is that memorable music events can occur anywhere within a work and many potential queries will reflect this fact (McNab, Smith, Witten et al. 1996; Byrd & Crawford, 2002). Thus, when full-text access is made possible, a Locating MIR system should also satisfy the following queries:

- In which compositions can we find the following melodic sequence anywhere in the composition?
- Which composers have used the following combination of instruments in the orchestration of a passage?
- What pieces use the following sequence of simultaneities? Which pieces use the following chord progression?

As MIR research progresses, and issues of aural and experiential similarity are addressed, we should add two important types of queries to this list:

- Which compositions “sound” like, or are in the same style as, this piece?
- Which compositions will induce happiness (or sadness, or stimulation, or relaxation)?

Development and Influence of Analytic/Production MIR Systems

Although this review focuses primarily on Locating systems, it is important to acknowledge the valuable contributions that Analytic/ Production research has made to their development. Many early MIR researchers saw the development of Analytic/Production MIR systems as a computer programming language problem; their work laid the foundation for much of present-day MIR research. I believe the honor of the earliest published study in the domain of MIR research in its modern sense belongs to Kassler for his 1966 article, noteworthy for its title, “Toward Musical Information Retrieval.” Kassler (1970) describes the MIR language he and others developed to analyze the works of Josquin des Prez. Another early work in the field (Lincoln, 1967) has been credited by Lemström (2000) for laying out the general framework of modern computerized music input, indexing, and printing.

Over the years, many others have contributed to the retrieval language aspect of MIR system research and development. Sutton (1988) developed a PROLOG-based language called MIRA (Music Information Retrieval and Analysis) to analyze Primitive Baptist hymns. A Pascal-like language called SML (Structured Music Language) was developed by Prather and Elliot (1988). McLane (1996) reports, however, that none of these languages has found general acceptance. He provides an explanation for this development, citing Sutton (1988, pp. 246–247), “the literature seems to show … that scholars interested in specific musical topics have found it more useful to develop their own systems.”

The late 1980s saw some important doctoral theses completed. Rubenstein (1987) extended the classic entity-relation model to include two novel features: hierarchical ordering and attribute inheritance. These features allowed Rubenstein to propose the creation of representationally complete databases of music using the relational database model. The extraordinary number of entities required to realize his model meant that an operational system was never implemented. Rubenstien's proposal to exploit the performance-enhancing characteristics of A-tree indexes to speed up searching is worth noting, for it is one of the first instances in the early literature in which the use of indexes instead of linear scanning is explicitly suggested for music.

McLean's (1988) doctoral research attempted to improve retrieval performance by creating a representationally complete encoding of the score. He concluded that a variety of sequential and indexed-based searches would be part of the necessary set of database-level services required for the creation of useful Analytic/Production MIR systems. Other than a brief discussion of the usefulness of doubly linked lists, how he would implement such indexing schemes is unclear.

Page (1988) implemented an experimental system that afforded access to both rhythmic and melodic information. Although he also mentioned that some type of indexing would improve retrieval performance, his system used a query language based on regular expressions. The musical data were searched using specially designed Finite State Automata. Items of interest were retrieved via a single-pass, linear traversal of the database.

An important goal of Page's doctoral thesis was to map out the necessary components of a musical research toolkit. Many of today's Analytic/Production systems are best thought of as suites of computer tools. Each tool is designed to address one of the many processes involved in the creation and use of an MIR system. Tools include encoding computer programs, extraction, pattern matching, display, data conversion, analysis, and so on.

David Huron's *Humdrum Toolkit* is an exemplar of this type of work (http://www.music-cog.ohio-state.edu/Humdrum/). His collection of more than 50 interrelated programs is designed to exploit the many information-processing capabilities found in the UNIX operating system. Taken together, these tools create a very powerful MIR system in which "queries of arbitrary complexity can be constructed" (Huron, 1991, p. 66). Interest in his system is high, and courses on its use are regularly offered. Huron (1991, p. 66) best describes *Humdrum*'s flexibilities:

> The generality of the tools may be illustrated through the *Humdrum pattern* command. The pattern command supports

> full UNIX regular-expression syntax. Pattern searches can involve pitch, diatonic/chromatic interval, duration, meter, metrical placement, rhythmic feet, articulation, sonorities/chords, dynamic markings, lyrics, or any combination of the preceding as well as other user-defined symbols. Moreover, patterns may be horizontal, vertical, or diagonal (*i.e.*, threaded across voices).

Like most things in life, all of this power comes at a price. Kornstädt (1996, pp. 110–111) provides a fine example of how *Humdrum*'s Unix-style command-line interface

> minimizes the number of potential users. For example, in order to search for occurrences of a given motive and to annotate the score with corresponding tags, the user has to construct the following command:
>
> extract –i'**kern' HG.kern | semits -x | xdelta -s = | patt -t Motive1 -s = -f Motiv1.pat | extract –-i**'patt' | assemble HG.krn
>
> The construction of such a command requires a substantial facility in the use of UNIX tools.

The naïve user ever managing to formulate such a search statement is hard to imagine. That *Humdrum* in its original incarnation was intended for use by musically sophisticated users who needed analytic power more than they needed syntactic simplicity must be stressed. Such users would be motivated to take the time to learn its methods. However, Kornstädt (1998) and his colleagues have gone on to develop a Web-based, user-friendly, Locating system, built upon *Humdrum* technology, called *Themefinder* (http://www.themefinder.org/).

MAPPET (Music Analysis Package for Ethnomusicology) was another collection of programs designed to assist in the encoding, retrieval, and analysis of monophonic music (Schaffrath, 1992b). The ESsen Associative Code (ESAC), a simple alpha-numeric scheme containing pitch and duration information, is used to represent the melodies. Melodies were first manually parsed into their constituent phrases; phrase determination in vocal music is not ambiguous, this process was relatively easy and consistent (Schaffrath, 1992a). The phrases were then ESAC encoded; and each encoded phrase was placed "on its own line in one field of a relational (AskSam) database" (Schaffrath, 1992a, p. 66). There were fields containing title, key, meter, and text information, as well as fields derived from the melodic information, such as mode, pitch profiles, and rhythmic profiles. MAPPET's ANA(lysis) and PAT(tern) software subcomponents could be used to translate an analyst's complex search criteria (e.g., intervallic, scale degrees, and rhythmic patterns) into AskSam queries. Detailed explanations of *MAPPET* and its use in the retrieval of monophonic information can be found in Schaffrath (1992b). Camilleri (1992) used MAPPET to analyze the melodic structures of the *Lieder* of Karl Collan.

The Essen databases of ESAC-encoded melodies are the primary source for the "McNab" collection, which forms the heart of the original MELDEX system (McNab, Smith, Bainbridge, et al., 1997; McNab, Smith, Witten, et al., 1996). Some 7,700 of McNab's 9,400 melodies come from Schaffrath's Essen collection and the remaining 1,700 were taken from the Digital Tradition collection (Greenhaus, 1999). The "McNab" collection was used for our own evaluations (Downie, 1999; Downie & Nelson, 2000; Nelson & Downie, 2001). Pickens (2000) and Södring and Smeaton (2002) have also made use of this collection.

Other examples of the many researcher toolkits available include MODE (Musical Object Development System) (Pope, 1992), the LIM Intelligent Music Workstation (Haus, 1994), and *Apollo* (Pool, 1996).

Revisiting the Facets of Music Information: Affording Access

Pitch and Temporal Access

The *Répetoire International des Sources Musicales*, Series A/II, *Music Manuscripts after 1600* database is the official title of what is generally known as the RISM database. The RISM database is one of the oldest and most ambitious of all MIR systems (McLean, 1988; Howard & Schlichte, 1988). It is an automated thematic index of gargantuan proportions. Originally conceived in the late 1940s as an attempt to catalog more than 1.5 million works, the RISM developers were quick to realize the need for automation (Howard & Schlichte, 1988). Now in its fourth edition, the database contains bibliographic records for more than 200,000 compositions by more than 8,000 composers (RISM, 1997). The RISM database is available on CD-ROM and via the Internet (http://www. RISM.harvard.edu/RISM/Welcome.html). The number of indexed access points is remarkable. The "Music Incipit" index is of most interest, as it contains pitch and duration information. *Incipits* are encoded using Brook's alpha-numeric *Plaine and Easie Code* (Brook & Gould, 1964). This is a very simple encoding scheme originally designed for use on typewriters with pitch denoted alphabetically and duration numerically. Howard and Schlichte (1988, p. 23) provide the following example of the *Plaine and Easie Code incipit* for Mozart's *Il core vi dono* from *Cosi fan tutte*:

%F-4$bB@3/8#'8C.6.3§, B'C&/,8A'D6(-)D/,8G'8.C,6B/8F

The ability afforded by the RISM database to search the *incipits* moved "music bibliography into a new realm" (Duggan, 1992, p. 770). Significant problems remain, however, with accessing the *incipit* information found within the RISM database. First, the *incipits* are entered into the MARC records exactly as shown above. This means that each *incipit* is indexed as one long, rather incomprehensible, "word." Second, because of the way the *incipit* is represented in the index, queries must also be posed using *Plaine and Easie*. Third, bringing together works that contain the same melody transposed into different keys is impossible because exact pitch names are used, not intervals. Fourth, searching on pitch or rhythm exclusively is impossible for one would have to know exactly which values to wildcard along with their exact locations. Fifth, and finally, an *incipit* can be represented in several, equally valid, ways, which puts the onus on users to frame their melodic queries in multiple ways (RISM, 1997).

The advent of multimedia personal computing prompted rising interest in the development of prototype Locating MIR systems. Fenske (1988) briefly describes a project at OCLC, led by Drone, called *HyperBach, a Hypermedia Reference System*. This system is also described by Duggan (1989, p. 88) as having "search access from Schmieder number and music entered through a MIDI interface and keyboard synthesizers." These descriptions represent the extent of information available about the *HyperBach* system. Hawley (1990) also developed a limited system that used a MIDI keyboard as the query interface to find tunes whose beginnings exactly matched the queries. Ghias, Logan, Chamberlin, and Smith (1995) developed a more sophisticated prototype system that incorporated autocorrelation methods for pitch tracking and where input was converted to melodic contours for matching against a 183-song database.

Any discussion of accessing the pitch and temporal (i.e., rhythm) facets of music must include the MELDEX system developed at the University of Waikato, New Zealand (McNab, Smith, Bainbridge, et al., 1997; McNab, Smith, Witten, et al., 1997; Bainbridge, Nevill-Manning, Witten, Smith, & McNab, 1999). Now part of the New Zealand Digital Library, MELDEX represents the clearest picture of how a large-scale, robust, and comprehensive Locating MIR system will look in the future (http://www.nzdl.org/musiclib/). The original collection of roughly 10,000 folksongs (based upon a combination of the Essen and Digital Traditional collections) has been enhanced with a second collection of roughly 100,000 MIDI files pulled from the World Wide Web by a spider. Of the monophonic, symbol-based, Locating retrieval systems currently in use, the MELDEX system is the gold standard.

Listing some of the central research and design features of the MELDEX project provides an overview of this project in particular and elucidates the central research and development trends of the MIR literature in general:

- Search modes, which include "query-by-humming"
- Application of Mongeau and Sankoff's (1990) string matching framework in recognition of the need for fault tolerance
- Related to the previous point, the conception of melodic retrieval as a contiguous-string retrieval problem and not a traditional IR indexing problem
- Search options, which range from basic intervallic contour, such as Parsons (1975), to exact match with or without the use of rhythm

- Recognition of scalability issues: that dynamic programming techniques increase search accuracy but at considerable computation cost when compared with the special modification of dynamic programming by Wu and Manber (1992).
- Implementation of browsing capabilities including the automatic creation of thematic thumbnails
- Use of multiple representations including graphic scores, audio files, and MIDI for both browsing and feedback purposes

Three projects, Downie (1999), Pickens (2000), and Uidenbogerd and Zobel (1998, 1999) have two interesting features in common. First, all three evaluated the retrieval effectiveness of interval-only, monophonic representations using melodic substrings, called n-grams (see Downie and Nelson [2000] for a description of the n-gramming process). Second, each project was influenced by the techniques and evaluation methods of traditional text-based IR. A number of factors limit the comparability of these projects, however. For example, Pickens evaluated probabilistic and language-based models; Downie, a vector-space model; and Uitdenbogerd and Zobel, a variety of methods. Tseng (1999), Doraisamy and Rüger (2001), and Södring and Smeaton (2002) present three additional projects using n-grams. Because the tokens created by n-gramming have many properties in common with word tokens, the use of n-grams allows traditional IR techniques to be employed. Notwithstanding the differences in techniques, it is important to note that all six of these teams have found intervallic n-grams to have significant merit as a retrieval approach. Melucci and Orio's (1999) melodic segmentation research is also inspired by the idea of applying traditional IR text retrieval methods.

Rolland, Raskinis, and Ganascia (1999) provide an overview of Rolland's Melodiscov approach to pitch and rhythm searching. Rolland's research is noteworthy as he tested his methods on a *corpus* that included transcriptions of improvised jazz, a particularly difficult genre with which to work. Jang, Lee, and Kao (2001) continue to develop their SuperMBox system, which provides fault tolerant searches via microphone input. Sonoda's ECHO system (Sonoda & Muraoka, 2000) is also designed to accept sung inputs and is tolerant of errors in rhythm and pitch. Related to this line of research is Smith, Chiu, and Scott (2000), who are developing an interface that takes spoken input to construct more accurate rhythm queries. Byrd (2001) reports on work that applies the pitch contour ideas of Parsons (1975) to rhythm. This work was done to allow the same kind of flexibility to rhythm searches as contours afford melodic searches.

Researchers at National Tsing Hua University in Taiwan have an impressive record investigating melodic, chordal, and "query-by-rhythm" approaches along with various indexing schemes (Chen & Chen, 1998). Chen (2000) briefly outlines the work of this productive group. It has implemented an evaluation platform called Ultima with an eye toward establishing consistent comparisons between retrieval techniques, both theirs and those of others (Hsu & Chen, 2001).

Harmonic Access

The harmonic facet of music information provides several challenges for MIR. One problem is the automatic disambiguation of melodic material from the harmonies that underpin it (e.g., accompaniment) or of which it is a part (e.g., contrapuntal music). The identification or extraction of melody from polyphonic sources is a classic figure and ground problem (Byrd & Crawford, 2002). Early, yet still important, work in this area comes from research into the creation of automatic accompaniment programs to allow a computer to "accompany" the performances of live musicians in real time (Bloch & Dannenberg, 1985; Dannenberg, 1984).

Uitdenbogerd and Zobel (1998, 1999) explored a variety of techniques to extract the melody from a collection of roughly 10,000 polyphonic MIDI files. The most notable aspect of this research was the use of listeners to assess the output of the different methods. Bello, Monti, and Sandler (2000) are developing a set of methods that can take audio input and extract monophonic melodic information as well as transcribe polyphonic sources. Durey and Clements (2001) apply audio retrieval "wordspotting" techniques to the problem of melody extraction from collections of audio files. Von Schroeter, Doraisamy, and Rüger (2000) examine polyphonic audio input and polyphonic Humdrum encodings to locate recurring themes. Meek and Birmingham (2001) have developed a Melodic Motive Extractor (MME) designed to locate and extract recurring themes from collections of MIDI files; they compared the test results with those indexed by Barlow and Morgenstern (1949). Barthelemy and Bonardi

(2001) are attempting to extract harmonic and tonal information automatically from scores through the information contained in the figured bass.

The second problem in dealing with the polyphonic aspect of the harmonic facet is searching. Polyphonic searching is particularly difficult because the search space is multidimensional, but the query can be either monophonic or polyphonic. Lemström's MonoPoly algorithm (Lemström & Perttu, 2000; Lemström & Tarhio, 2000) uses bit-parallel techniques to locate monophonic sequences efficiently in polyphonic databases. Huron's (1991) Humdrum system can be used to perform monophonic, polyphonic, and harmonic progression searches. Dovey (1999) developed an algorithm capable of either monophonic or polyphonic searches through polyphonic music. He has extended his work by formalizing his polyphonic search methods as a regular expression language (Dovey, 2001a). Meredith, Wiggins, and Lemström (2001) also deal with pattern matching and induction in polyphonic music. Pickens (2000) explores techniques for both monophonic and "homophonic" (i.e., simultaneity) extraction. Doraisamy and Rüger (2001) use n-grams of both interval and rhythmic information in conjunction with traditional IR techniques to search polyphonic music with promising preliminary results. Clausen, Engelbrecht, Meyer, and Schmitz (2000) have also adapted and extended traditional IR techniques to the polyphonic searching problem, again with promising results.

Timbral Access

Explicit access to timbral information within the context of MIR is not as well developed as other aspects. Musclefish (http://www.musclefish.com/) has developed several commercial products based upon their audio retrieval research. One product, called Soundfisher, can be used over a collection of audio files to locate similar sounds (http://www.soundfisher.com/). Another product, Clango, is designed to identify and then retrieve metadata about music audio files as they are being played (http://www.clango.com/). Cano, Kaltenbunner, Mayor, and Batlle (2001) are working on the identification problem with noisy radio broadcasts as their domain of interest. Foote (2000) has implemented an audio-based identification system called Arthur that performs its identification using the dynamic structure (i.e., loudness and softness) of the input. Foote has also mounted a limited demonstration system that conducts audio similarity searches (http://www.fxpal.com/people/foote/musicr/doc0.html). Nam and Berger (2001) build upon the work of Foote to generate their approach to audio-based genre classification. Fujinaga and MacMillan (2000) use genetic algorithms and a k-NN classifier to perform a real-time recognition of orchestral instruments. Liu and Wan (2001) report satisfactory results using a limited set of timbral features to classify instrument sounds in the traditional categories of brass, woodwind, string, keyboard, and percussion. Tzanetakis, Essl, and Cook (2001) exploit timbral information as part of their automated approach to classification and genre identification of audio files. Batlle and Cano (2000) use Competitive Hidden Markov models to perform automatic segmentation and classification of music audio. Rauber and Frühwirth (2001) employ Self-Organizing Maps (SOMs) to cluster music based on audio similarity.

Although these systems utilize highly sophisticated signal processing technologies, it is important to note that these are holistic approaches to identification. The input audio is treated as an indivisible entity, and access to, or identification of, say, the bassoon part in an orchestral piece is not yet practical. The timbral search engine work at the Institute de Recherche et Coordination Acoustique/Musique (IRCAM) in Paris does illustrate, however, that timbral-specific searches are possible in theory (http://zappa.ircam.fr/php3/php.exe/cuidad/Timbre.html).

For more information on the complexities of timbre identification and searching, consult Martin (1999), which is the seminal work in this area. Another foundational work is Scheirer (2000). Hererra, Amatriain, Batlle, and Serra (2000) provide a comprehensive review of the different techniques being suggested for the automatic classification of instruments from audio files along with discussion of the feasibility of each. Foote (1999), Kostek (1999), and Tzanetakis and Cook (2000) are all excellent introductions to the techniques used in signal processing and audio information retrieval.

Editorial, Textual, and Bibliographic Access

XML and other structural markup languages are being put forward as a means of enhancing MIR (Good, 2000; MacLellan & Boehm, 2000; Roland, 2000; Schimmelpfennig & Kurth, 2000). One implication of this line of work is that the editorial components of music can be explicitly tagged and thus retrieved. Navigation through a piece, or a set of pieces, via the hyperlinks that can be constructed using structural markup languages is another potential benefit of this development stream. For more information on the hypertextual navigation of musical works, consult Blackburn (2000), Blackburn and DeRoure (1998), or Melucci and Orio (2000).

Choudhury et al. (2000) and Droettboom et al. (2001) report on the large-scale digitization project being undertaken on the Lester Levy Collection of Sheet Music at Johns Hopkins University. Using the Optical Music Recognition (OMR) technology they developed for the project, symbolic representations are created in both MIDI and GUIDO formats. Lyrics and metadata information are also captured and stored for eventual retrieval. In conjunction with the Levy project, the developers of GUIDO (Hoos, Renz, & Görg, 2001) are enhancing the system's search capabilities to take advantage of the high level of representational completeness while at the same time exploiting the power of probabilistic search models.

Query-by-singing is starting to supplement the more traditional query-by-humming methods. Milan-based Haus and Pollastri (2001) and the University of Michigan-based Museart project (Mellody, Barstch & Wakefield, 2002; Birmingham et al., 2001) are two groups that suggest the promise of lyric searches based upon singing rather than text input. The Milan team uses the modeling of singer errors to recognize sung input. The Museart work is based on the characteristics of sung vowels.

Pachet and Laigre (2001) developed a set of analytical tools designed to interpret, classify, and identify song titles based upon the names of the files that contain them. This activity is necessary for increasing the automation of bibliographic control because many pieces are inconsistently labeled. Smiraglia (2001) explores epistemological perspectives to outline the need for, and the difficulties associated with, interlinking all extant versions and derivatives of individual musical works. Allamanche et al. (2001) use audio processing techniques to classify and identify input streams of possibly unlabeled music audio so that the appropriate metadata can be associated with the analyzed works.

DiLauro, Choudhury, Patton, Warner, and Brown (2001) have implemented techniques for automating name authority control over the digitized scores of the Levy Collection using XML, Library of Congress authority files, and Bayesian probability methods. Dovey (2001b) aims to integrate his content-based retrieval system with a traditional online music catalog via the Z39.50 protocol. The most ambitious work being done on improving bibliographic access to electronic audio-visual material, and thus music files, is the MPEG-7 (Moving Picture Experts Group) project (http://mpeg.telecomitalialab.com). This ISO/IEC (International Organization for Standardization/International Electrotechnical Commission) standardization project has created a flexible yet comprehensive method of describing the contents of multimedia files. Music-specific components to the standard include melodic contour and timbral descriptions (Allemanche et al., 2001; Lindsay & Kim, 2001). An overview of the standard is available (http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm).

Concluding Remarks on the Future of MIR

In this chapter, I have outlined the many challenges facing MIR research both as an intellectual endeavor and as a newly emerging discipline. These challenges are not insignificant; and it is obvious that much work remains to be done. I am, however, increasingly confident that the future of MIR research and development is bright. The growing number of researchers interested in MIR issues appears to be reaching a critical mass. For example, as of August 2001, the music-ir@ircam.fr mailing list had over 350 subscribers (http://www.ircam.fr/listes/archives/music-ir/maillist.html). Since 1999, at least two symposia, two workshops, and three panel sessions exclusively devoted to MIR issues have been conducted. These meetings represent the first steps in overcoming the disciplinary fragmentation noted earlier. I suggest interested readers take the time to explore the links provided in the following list of recent MIR events, with an eye toward uncovering the many worthwhile papers that space did not permit to be included in this review.

Recent MIR Workshops, Panels, and Symposia

- The Exploratory Workshop on Music Information Retrieval, ACM SIGIR 1999, Berkeley, California, USA (http://mir.isrl.uiuc.edu/mir/mir_workshop.pdf)

- Workshop on Music Description, Representation and Information Retrieval, Digital Resources in the Humanities 1999, London, UK (http://www.kcl.ac.uk/humanities/cch/drhahc/drh/abst502.htm)
- Notation and Music Information Retrieval in the Computer Age, International Computer Music Conference 2000, Berlin, Germany (http://www.icmc2000.org/intro/workshops/index.html#boehm)
- Digital Music Libraries—Research and Development, Joint Conference on Digital Libraries 2001, Roanoke, Virginia, USA (http://www.acm.org/jcdl/jcdl01/panels.html)
- New Directions in Music Information Retrieval, International Computer Music Conference 2001, Havana, Cuba, (http://benares.centrotemporeale.it/~icmc2001/papers)
- First International Symposium on Music Information Retrieval (ISMIR 2000) Plymouth, Massachusetts, USA (http://ciir.cs.umass.edu/music2000)
- Second International Symposium on Music Information Retrieval (ISMIR 2001) Bloomington, Indiana, USA (http://ismir2001.indiana.edu)

The recent establishment of large-scale, well-funded, and multidisciplinary MIR research projects is another indication of a promising future for MIR. Within the United States, three national projects and one international cooperative project are under way. In Europe, two multinational music delivery projects with strong MIR components are being conducted. All six are important and influential projects from which significant contributions have been, and will continue to be, made. Many of the authors cited in this review are working in conjunction with one or more of these projects.

Major MIR Research Projects

- Digital Music Library (DML) Project, Indiana University Bloomington, National Science Foundation, National Endowment for the Humanities (http://dml.indiana.edu)

- Online Music Recognition and Searching project (OMRAS), King's College, London, and University of Massachusetts at Amherst, Joint Information Systems Committee (UK), National Science Foundation (http://www.omras.org/)
- MuseArts Project, University of Michigan, National Science Foundation (http://musen.engin.umich.edu/musearts.html)
- Levy Project: Adaptive Optical Music Recognition, National Science Foundation, Institute for Museum and Library Services, Levy family, Johns Hopkins University (http://mambo.peabody.jhu.edu/omr)
- Content-based Unified Interfaces and Descriptors for Audio/music Databases available Online (CUIDADO), IRCAM, Oracle, Sony CSL, Ben Gurion University-Beer Shiva, artspages, creamw@re, Universitad Pompeu Fabra-Barcelona (http://www.cuidado.mu/)
- Web DELivering of MUSIC (WEDELMUSIC) Università degli Studi di Firenze, IRCAM, Fraunhofer Institute for Computer Graphics, Artec Group, and others (http://www.wedelmusic.org/)

I am encouraged that issues pertaining to relevance and experiential similarity are beginning to be addressed by various research teams: Byrd and Crawford (2002), Hofmann-Engl (2001), and Uitdenbogerd and Zobel (1998, 1999). Work by Chai and Vercoe (2000) and Rolland (2001) on the application of user modeling in the retrieval process also indicates a growing awareness of the limitations of current, system-based, matching practices and relevance assessments. On the evaluation front, other indicators of developing strength are evident. Stemming from briefing documents submitted to the participants of ISMIR 2000 and 2001 (Crawford & Byrd, 2000; Downie, 2000, 2001a), the participants of ISMIR 2001 ratified a resolution proposed by Huron, Dovey, Byrd, and Downie (2001) calling for the creation of standardized multirepresentational test collections, queries, and relevance judgments. The resolution and its current list of signatories can be found at http://music-ir.org/mirbib2/ resolution. The establishment of annual MIR competitions modeled after the Text REtrieval Conferences (TREC) (http://trec.nist.gov) is one proposed mechanism through which evaluations could be standardized.

Whatever shape a formal evaluative framework for MIR takes, it should reflect not only the traditional IR paradigms but also the goals and aspirations of the many other disciplines that comprise MIR research. It is apparent that novel definitions of relevance, new evaluation metrics, and new measures of success will have to be designed to address the needs of MIR research explicitly.

The problems associated with a lack of an intellectual "home base" for MIR research are being addressed. The organizers of ISMIR 2002 (http://ismir2002.ircam.fr) established an exploratory committee to investigate the relative merits of affiliating with one of the large research organizations (e.g., the Association for Computing Machinery [ACM], the Institute of Electrical and Electronics Engineers [IEEE], the American Society for Information Science and Technology [ASIST]) or creating an independent International Society for Music Information Retrieval (ISMIR). The Mellon Foundation has provided funding for the MIR Annotated Bibliography Project (http://music-ir.org), which is striving to bring a level of bibliographic control to the highly fragmented MIR literature (Downie, 2001b).

Much of the research discussed in this review is preliminary and exploratory because MIR is still in its infancy. Many intriguing yet-to-be-investigated questions remain within the MIR domain. For example, no rigorous and comprehensive studies in the MIR literature examine the human factors involved in MIR system use. Other than one exploratory report (Downie, 1994), I know of no literature explicitly investigating the information needs and uses of MIR system users.

To recap the central themes of this review, I see future MIR research as confronting 10 central questions:

- Which facets of music information are essential, which are potentially useful, and which are superfluous to the construction of robust MIR systems?
- How do we integrate non-Western, non-CP music into our systems?
- How do we better conjoin the various symbolic and audio representations into a seamless whole?

- How do we overcome the legal hurdles impeding system development and experimentation?
- How do we capture, represent, and then exploit the experiential aspects of music?
- What does “relevance” mean in the context of MIR?
- How do we maximize the benefits of multidisciplinary research while minimizing its drawbacks?
- What do “real” users of MIR systems actually want the systems to do?
- How will “real” users actually interact with MIR systems?
- How will we know which MIR methods to adopt and which to abandon?

I cannot predict which combinations of present-day and yet-to-be-developed MIR approaches will ultimately form the basis of the MIR systems of the future. I can predict, however, with absolute certainty, that some of these systems will rival the present-day Web search engines, both in size and general success. I can also predict, again with absolute certainty, that these MIR systems will fundamentally alter the way we experience and interact with music.

Endnotes

1. Adapted from Tague-Sutcliffe, Downie, and Dunne (1993).

Bibliography

Allamanche, E., Herre, J., Hellmuth, O., Fröba, B., Kastner, T., & Cremer, M. (2001). Content based identification of audio materials using MPEG-7 low level description. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 197–204. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/allamanche.pdf.

Bainbridge, D., Nevill-Manning, C. G., Witten, I. H., Smith, L. A., & McNab, R. J. (1999). Towards a digital library of popular music. *Proceedings of the 4th ACM International Conference on Digital Libraries*, 161–169.

Barlow, H., & Morgenstern, S. (1949). *A dictionary of musical themes*. London: Ernest Benn.

Barthelemy, J., & Bonardi, A. (2001). Figured bass and tonality recognition. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 129–136. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/barthelemy.pdf.

Batlle, E., & Cano, C. (2000). Automatic segmentation using competitive hidden Markov models. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/posters/batlle.pdf.

Bello, J. P., Monti, G., & Sandler, M. (2000). Techniques for automatic music transcription. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/bello_abs.pdf.

Birmingham, W. P., Dannenberg, R. B., Wakefield, G. H., Bartsch, M., Bykowski, D., Mazzoni, D., et al. (2001). MUSART: Music retrieval via aural queries. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 73–81. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/birmingham.pdf.

Blackburn, S. (2000). *Content based retrieval and navigation of music*. Unpublished doctoral dissertation, University of Southampton, U.K. Retrieved February 7, 2002, from http://www.ecs.soton.ac.uk/~sgb97r/phdthesis.pdf.

Blackburn, S., & DeRoure, D. (1998). A tool for content based navigation of music. *Proceedings of the 6th ACM International Conference on Multimedia*, 361–368. Retrieved February 7, 2002, from http://www.mmrg.ecs.soton.ac.uk/publications/archive/blackburn1998/html.

Bloch, J. B., & Dannenberg, R. B. (1985). Real-time accompaniment of polyphonic keyboard performance. *Proceedings of the 1985 International Computer Music Conference (ICMC 1985)*, 279–290.

Bonardi, A. (2000). IR for contemporary music: What the musicologist needs. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/invites/bonardi_invite.pdf.

Brook, B. S., & Gould, M. J. (1964). Notating music with ordinary typewriter characters (a plaine and easie code system for music). *Fontes Artis Musicae, 11*, 142–159.

Byrd, D. (2001). Music notation searching and digital libraries. *Proceedings of the 1st ACM/IEEE Joint Conference on Digital Libraries*, 239–246.

Byrd, D., & Crawford, T. (2002). Problems of music information retrieval in the real world. *Information Processing & Management, 38*, 249–272.

Camilleri, L. (1992). The *Lieder* of Karl Collan. *Computing in Musicology, 8*, 67–68.

Cano, P., Kaltenbunner, M., Mayor, O., & Batlle, E. (2001). Statistical significance in song-spotting in audio. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 3–5. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/posters/cano.pdf.

Chai, W., & Vercoe, B. (2000). Using user models in music information retrieval systems. *Proceedings of the 1st Annual International Symposium on Music

Information Retrieval (ISMIR 2000). Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/posters/chai.pdf.

Chen, A. L. P. (2000). Music representation, indexing and retrieval at NTHU. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/invites/chen_invite.pdf.

Chen, J. C. C., & Chen, A. L. P. (1998). Query-by-rhythm: An approach for song retrieval in music databases. *Proceedings of the 8th International Workshop on Research Issues in Data Engineering: Continuous-Media Databases and Applications*, 139–146.

Choudhury, G. S., DiLauro, T., Droettboom, M., Fujinaga, I., Harrington, B., & MacMillan, K. (2000). Optical music recognition within a large-scale digitization project. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/choudhury_paper.pdf.

Clausen, M., Engelbrecht, R., Meyer, D., & Schmitz, J. (2000). PROMS: A Web-based tool for searching in polyphonic music. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/clausen_abs. pdf.

Crawford, T., Iliopoulos, C. S., & Raman, R. (1998). String-matching techniques for musical similarity and melodic recognition. In W. B. Hewlett & E. Selfridge-Field (Eds.), *Computing in Musicology: Vol. 11. Melodic similarity: Concepts, procedures, and applications* (pp. 73–100). Cambridge, MA: MIT Press.

Crawford, T., & Byrd, D. (2000). Background document for ISMIR 2000 on music information retrieval evaluation. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/evaluation.html.

Cronin, C. (1998). Concepts of similarity in music-copyright infringement suits. In W. B. Hewlett & E. Selfridge-Field (Eds.), *Computing in musicology, vol. 11: Melodic similarity: Concepts, procedures, and applications* (pp. 187–209). Cambridge, MA: MIT Press.

Dannenberg, R. (1984). An on-line algorithm for real-time accompaniment. *Proceedings of the 1984 International Computer Music Conference*, 193–198. Retrieved February 7, 2002, from http://www-2.cs.cmu.edu/~rbd/papers/icmc84accomp.pdf.

Deutsch, O. E., & Wakeling, D. R. (1995). *The Schubert thematic catalogue.* New York: Dover.

DiLauro, T., Choudhury, G. S., Patton M., Warner, J. W., & Brown, E. W. (2001). Automated name authority control and enhanced searching in the Levy Collection. *D-Lib Magazine, 7*(4). Retrieved February 7, 2002, from http://www.dlib.org/dlib/april01/dilauro/04dilauro.html.

Doraisamy, S., & Rüger, S. (2001). An approach towards a polyphonic music retrieval system. *Proceedings of the 2nd Annual International Symposium on*

Music Information Retrieval (ISMIR 2001), 187–193. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/doraisamy.pdf.

Dovey, M. (1999). An algorithm for locating polyphonic phrases within a polyphonic musical piece. *Proceedings of the AISB'99 [Artificial Intelligence and Simulation of Behaviour] Symposium on Musical Creativity*, 48–53.

Dovey, M. (2001a). A technique for regular expression style searching in polyphonic music. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 187–193. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/dovey.pdf.

Dovey, M. (2001b). Adding content-based searching to a traditional music library catalogue server. *Proceedings of the 1st ACM/IEEE Joint Conference on Digital Libraries*, 249–250.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*, 341–354.

Downie, J. S. (1994). The MusiFind musical information retrieval project, phase II: User assessment survey. *Proceedings of the 22nd Annual Conference of the Canadian Association for Information Science*, 149–166.

Downie, J. S. (1999). *Evaluating a simple approach to music information retrieval: Conceiving melodic n-grams as text*. Unpublished doctoral dissertation, University of Western Ontario, London, Ontario, Canada. Retrieved February 7, 2002, from http://music-ir.org/~jdownie/downie_thesis.pdf.

Downie, J. S. (2000). Thinking about formal MIR system evaluation: Some prompting thoughts. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR2000)*. Retrieved February 7, 2002, from http://music-ir.org/~jdownie/downie_mir_eval.html.

Downie, J. S. (2001a). The music information retrieval annotated bibliography project, phase I. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 5–7. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/posters/downie.pdf.

Downie, J. S. (2001b). Whither music information retrieval: Ten suggestions to strengthen the MIR research community. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 219–222. Retrieved February 7, 2002, from http://music-ir.org/~jdownie/mir_suggestions.pdf.

Downie, J. S., & Nelson, M. (2000). Evaluation of a simple and effective music information retrieval method. *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 73–80.

Droettboom, M., Fujianga, I., MacMillan, K., Patton, M., Warner, J., Choudhury, G. S., et al. (2001). Expressive and efficient retrieval of symbolic music data. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 173–178. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/droettboom.pdf.

Duggan, M. K. (1989). CD-ROM, music libraries, present and future. *Fontes Artis Musicae*, *36*, 84–89.

Duggan, M. K. (1992). Electronic information and applications in musicology and music theory. *Library Trends, 40,* 756–780.

Durey, A. S., & Clements, M. A. (2001). Melody spotting using hidden Markov models. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 109–117. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/durey.pdf.

Edson, J. S. (1970). *Organ-preludes: An index to compositions on hymn tunes, chorales, plainsong melodies, Gregorian tunes and carols*. Metuchen, NJ: Scarecrow Press.

Fenske, D. (1988). Online Computer Library Center. *Directory of computer assisted research in musicology 1988*, 30–31. Menlo Park, CA: Center for Computer Assisted Research in the Humanities.

Foote, J. (1997). Content-based retrieval of music and audio. In C.-C. J. Kuo, S. F. Chang, & V. N. Gudivada (Eds.), *Proceedings of SPIE Vol. 3229. Multimedia storage and archiving systems II* (pp.138–147). Bellingham, WA: SPIE Press. Retrieved February 7, 2002, from http://www.fxpal.com/people/foote/papers/spie97.pdf.

Foote, J. (1999). An overview of audio information retrieval. *Multimedia Systems, 7*(1), 2–11. Retrieved February 7, 2002, from http://www.fxpal.com/people/foote/papers/acm98.pdf.

Foote, J. (2000). Arthur: Retrieving orchestral music by long-term structure. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/foote_paper.pdf.

Fujinaga, I., & MacMillan, K. (2000). Realtime recognition of orchestral instruments. *Proceedings of the International Computer Music Conference (ICMC 2000)*, 141–143. Retrieved February 7, 2002, from http://gigue.peabody.jhu.edu/~ich/research/icmc00/icmc00.timbre.pdf.

Ghias, A., Logan, J., Chamberlin, D., & Smith, B. C. (1995). Query by humming: Musical information retrieval in an audio database. *Proceedings of the ACM International Multimedia Conference & Exhibition 1995*, 231–236.

Good, M. (2000). Representing music using XML. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/posters/good.pdf.

Harter, S. P., & Hert, C. A. (1997). Evaluation of information retrieval systems: Approaches, issues, and methods. *Annual Review of Information Science and Technology, 32*, 3–91.

Haus, G. (1994). The LIM intelligent music workstation. *Computing in Musicology, 9*, 70–73.

Haus, G., & Pollastri, E. (2001). An audio front-end for query-by-humming systems. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 65–72. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/haus.pdf.

Haven Sound. (2001). *Music recordings and public domain*. Retrieved February 7, 2002, from http://www.pdinfo.com/record.htm.

Hawley, M. (1990). The personal orchestra, or audio data compression by 10000:1. *Computing Systems*, *3*, 289–329.

Herrera, P., Amatriain, X., Batlle, E., & Serra, X. (2000). Towards instrument segmentation for music content description: A critical review of instrument classification techniques. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/herrera_paper.pdf.

Hewlett, W. B., & Selfridge-Field, E. (Eds.). (1998). *Computing in musicology: Vol. 11. Melodic similarity: Concepts, procedures, and applications*. Cambridge, MA: MIT Press.

Hofmann-Engl, L. (2001). Towards a cognitive model of melodic similarity. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 143–151. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/hofmann-engl.pdf.

Hoos, H., Renz, K., & Görg, M. (2001). GUIDO/MIR—An experimental musical information retrieval system based on GUIDO music notation. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 41–50. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/hoos.pdf.

Howard, J. B. (1998). Strategies for sorting melodic incipits. In W. B. Hewlett & E. Selfridge-Field (Eds.), *Computing in musicology: Vol. 11. Melodic similarity: Concepts, procedures, and applications* (pp. 119–128). Cambridge, MA: MIT Press.

Howard, J., & Schlichte, J. (1988). Repertoire international des sources musicales (RISM). In W. B. Hewlett & E. Selfridge-Field (Eds.), *Directory of computer assisted research in musicology 1988* (pp. 11–24). Menlo Park, CA: Center for Computer Assisted Research in the Humanities.

Hsu, J.-L., & Chen, A. L. P. (2001). Building a platform for performance study of various music information retrieval approaches. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 153–162. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/hsu.pdf.

Huron, D. (1991). Humdrum: Music tools for UNIX systems. *Computing in Musicology*, *7*, 66–67.

Huron, D. (2000). Perceptual and cognitive applications in music information retrieval. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/invites/HuronAbstract.pdf.

Huron, D., Dovey, M., Byrd, D., & Downie, J. S. (2001). *Indicate your support for the ISMIR 2001 resolution on the need to create standardized MIR test collections, tasks, and evaluation metrics for MIR research and development*. Retrieved February 7, 2002, from http://music-ir.org/mirbib2/resolution.

Greenhaus, D. (1999, Spring). About the digitial tradition. *The Mudcat Café*. Retrieved February 7, 2002, from http://www.mudcat.org/DigiTrad-blurb.cfm.

Jang, R. J.-S., Lee, H.-R., & Kao, M.-K. (2001). Content-based music retrieval using linear scaling and branch-and-bound tree search. *Proceedings of IEEE International Conference on Multimedia and Expo (ICME 2001)*. Retrieved February 7, 2002, from http://neural.cs.nthu.edu.tw/jang/research/paper/2001icme/2001icme-linearScaling.pdf.

Kassler, M. (1966 Spring-Summer). Toward musical information retrieval. *Perspectives of New Music, 4*, 59–67.

Kassler, M. (1970). MIR: A simple programming language for musical information retrieval. In H. B. Lincoln (Ed.), *The computer and music* (pp. 299–327). Ithaca, NY: Cornell University Press.

Keen, E. M. (1992). Presenting results of experimental retrieval comparisons. *Information Processing & Management*, *28*, 491–502.

Keller, K., & Rabson, C. (1980). National tune index: 18th century secular music. New York: University Music Edition.

Korfhage, R. R. (1997). *Information storage and retrieval*. New York: John Wiley and Sons.

Kornstädt, A. (1996). SCORE-to-*Humdrum*: A graphical environment for musicological analysis. *Computing in Musicology, 10*, 105–130.

Kornstädt, A. (1998). THEMFINDER: A Web-based melodic search tool. In W. B. Hewlett & E. Selfridge-Field (Eds.), *Computing in musicology: Vol. 11. Melodic similarity: Concepts, procedures, and applications*, 231–236. Cambridge, MA: MIT Press.

Kostek, B. (1999). *Soft computing in acoustics: Applications of neural networks, fuzzy logic and rough sets to musical acoustics: Studies in fuzziness and soft computing*. New York: Physica-Verlag.

Kruhmhansl, C., & Bharucha, J. (1986). Psychology of music. In D. M. Randel (Ed.), *The new Harvard dictionary of music* (pp. 669–670). Cambridge, MA: Belknap Press.

Lemström, K. (2000). *String matching techniques for music retrieval*. Helsinki, Finland: University of Helsinki.

Lemström, K., & Perttu, S. (2000). SEMEX: An efficient retrieval prototype. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/lemstrom_paper.pdf.

Lemström, K., & Tarhio, J. (2000). Searching monophonic patterns within polyphonic sources. *Proceedings of the 6th Conference on Content-based Multimedia Information Access (RIAO 2000)*, 1261–1279. Retrieved February 7, 2002, from: http://www.cs.helsinki.fi/u/klemstro/publications/RIAO00.ps.

Levering, M. (2000). Intellectual property rights in musical works: Overview, digital library issues and related initiatives. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/invites/levering_invite.pdf.

Lincoln, H. B. (1967). Some criteria and techniques for developing computerized thematic indices. In H. Heckmann (Ed.), *Elektronische Datenverarbeitung in der Musikwissenschaft* (pp. 57–62). Regensburg, Germany: Gustave Bosse Verlag.

Lincoln, H. B. (1989). *The Italian madrigal and related repertories: Indexes to printed collections, 1500–1600*. New Haven, CT: Yale University Press.

Lindsay, A., & Kim, Y. (2001). Adventures in standardization, or how we learned to stop worrying and love MPEG-7. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 195–196. Retrieved February 7, 2002, from http://music-ir.org/~jdownie/lindsay-kim.pdf.

Liu, M., & Wan, C. (2001). Feature selection for automatic classification of musical instrument sounds. *Proceedings of the 1st ACM/IEEE Joint Conference on Digital Libraries*, 247–248.

MacLellan, D., & Boehm, C. (2000). MuTaTeD'll: A system for music information retrieval of encoded music. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/posters/boehm.pdf.

Martin, K. D. (1999). *Sound-source recognition: A theory and computational model*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA. Retrieved February 7, 2002, from http://sound.media.mit.edu/~kdm/research/papers/kdm-phdthesis.pdf.

McLane, A. (1996). Music as information. *Annual Review of Information Science and Technology 31*, 225–262.

McLean, B. A. (1988). The representation of musical scores as data for applications in musical computing. Unpublished doctoral dissertation, State University of New York at Binghamton.

McNab, R. J., Smith, L. A., Bainbridge, D., & Witten, I. H. (1997, May). The New Zealand Digital Library MELody inDEX. *D-Lib Magazine*. Retrieved February 7, 2002, from http://www.dlib.org/dlib/may97/meldex/05witten.html.

McNab, R. J., Smith, L. A., Witten, I. H., Henderson, C., & Cunningham, S. J. (1996). Towards the digital music library: Tune retrieval from acoustic input. *Digital Libraries '96, Proceedings of the ACM Digital Libraries Conference*, 11–18.

Meek, C., & Birmingham, W. P. (2001). Thematic extractor. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 119–128. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/meek.pdf.

Mellody, M., Barstch, M. A., & Wakefield, G. H. (2002). *Analysis of vowels in sung queries for a music information retrieval system*. Manuscript submitted for publication.

Melucci, M., & Orio, N. (1999). Music information retrieval using melodic surface. *Proceedings of the 4th ACM Conference on Digital Libraries*, 152–160.

Melucci, M., & Orio, N. (2000). SMILE: A system for content-based musical information retrieval environments. *Proceedings of the 6th Conference on Content-based Multimedia Information Access (RIAO 2000)*, 1261–1279. Retrieved

February 7, 2002, from http://133.23.229.11/~ysuzuki/Proceedingsall/RIAO2000/Friday/103CO2.pdf.

Meredith, D., Wiggins, G. A., & Lemström, K. (2001). Pattern induction and matching in polyphonic music and other multi-dimensional datasets. *Proceedings of the 5th World Multi-Conference on Systemics, Cybernetics and Informatics (SCI2001), 10*, 61–66. Retrieved February 7, 2002, from http://www.cs.helsinki.fi/u/klemstro/publications/SCI2001-final.ps.

Mongeau, M., & Sankoff, D. (1990). Comparison of musical sequences. *Computers and the Humanities, 24*, 161–175.

Nam, U., & Berger, J. (2001). Addressing the "same but different—different but similar" problem in automatic music classification. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*. Retrieved March 9, 2002, from http://ismir2001.indiana.edu/posters/nam.pdf.

Nelson, M., & Downie, J. S. (2001). Informetric analysis of a music database: Distribution of intervals. In M. Davis & C. S. Wilson (Eds.), *Proceedings of the 8th International Conference on Scientometrics and Informetrics (ISSI 2001)*, Vol. 2 (pp. 477–484). Sydney, Australia: Bibliometrics and Informetrics Research Group.

Pachet, F., & Laigre, D. (2001). A naturalist approach to music file name analysis. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 51–58. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/pachet.pdf.

Page, S. D. (1988). *Computer tools for music information retrieval*. Unpublished doctoral dissertation, Oxford University, UK.

Parsons, D. (1975). *The directory of tunes and musical themes*. New York: Spencer Brown.

Pickens, J. (2000). A comparison of language modeling and probabilistic text information retrieval approaches to monophonic music retrieval. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/pickens_paper.pdf.

Pool, O. E. (1996). The *Apollo* project: Software for musical analysis using DARMS. *Computers in Musicology, 10*, 123–128.

Pope, S. T. (1992). MODE and SMOKE. *Computing in Musicology, 8*, 130–134.

Prather, R. E., & Elliot, R. S. (1988). SML: A structured musical language. *Computers and the Humanities, 24*, 137–151.

Prechelt, L., & Typke, R. (2001). An interface for melody input. *ACM Transactions on Computer-Human Interaction, 8*(2), 133–149. Retrieved February 7, 2002, from http://wwwipd.ira.uka.de/~prechelt/Biblio/Biblio/tuneserver_tochi2001.pdf.

Randel, D. M. (ed.). (1986). *The new Harvard dictionary of music*. Cambridge, MA: Belknap Press.

Rauber, A. & Frühwirth, M. (2001). Automatically analyzing and organizing music archives. *Research and Advanced Technology for Digital Libraries, Proceedings* *of the 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001)*, 402–414. Retrieved February 7, 2002, from http://www.ifs.tuwien.ac.at/ifs/research/pub_pdf/rau_ecdl01.pdf.

Recording Industry Association of America. (2001). *Recording Industry Association of America's 2000 yearend statistics*. Washington, DC: RIAA. Retrieved February 7, 2002, from http://www.riaa.com/pdf/year_end_2000.pdf.

RISM. (1997). *Répertoire international des sources musicales: International inventory of musical sources. Series A/II, Music manuscripts after 1600* [CD-ROM database]. Munich, Germany: K. G. Saur Verlag.

Roland, P. (2000). XML4MIR: Extensible markup language for music information retrieval. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/roland_paper.pdf.

Rolland, P.-Y., (2001). Adaptive user-modeling in a content-based music retrieval system. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/posters/rolland.pdf.

Rolland, P.-Y., Raskinis, G., & Ganascia, J.-G. (1999). Musical content-based retrieval: An overview of the Melodiscov approach and system. *Proceedings of the 7th ACM International Multimedia Conference*, 81–84.

Rubenstein, W. B. (1987). *Data management of musical information*. Unpublished doctoral dissertation, University of California, Berkeley.

Schaffrath, H. (1992a). The *ESAC* databases and *MAPPET* software. *Computing in Musicology, 8*, 66.

Schaffrath, H. (1992b). The retrieval of monophonic melodies and their variants: Concepts and strategies in computer-aided analysis. In A. Marsden & A. Pople (Eds.), *Computer representations and models in music* (pp. 95–105). London: Academic Press.

Schamber, L. (1994). Relevance and information behavior. *Annual Review of Information Science and Technology, 29*, 3–48.

Scheirer, E. D. (2000). *Music-listening systems*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge. Retrieved February 7, 2002, from http://sound.media.mit.edu/~eds/thesis/.

Schimmelpfennig, J., & Kurth, F. (2000). MCML: Music contents markup language. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass. edu/music2000/posters/schimmelpfennig.pdf.

Schmieder, W. (1990). *Bach-Werke-Verzeichnis*. Wiesbaden, Germany: Breitkopf & Härtel.

Selfridge-Field, E. (1993–1994). Optical recognition of music notation: A survey of current work. *Computing in Musicology, 9*, 109–145.

Selfridge-Field, E. (Ed.). (1997). *Beyond MIDI: The handbook of musical codes*. Cambridge, MA: MIT Press.

Smiraglia, R. (2001). Musical works as information retrieval entities: Epistemological perspectives. *Proceedings of the 2nd Annual International Symposium* *on Music Information Retrieval (ISMIR 2001)*, 85–92. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/smiraglia.pdf.

Smith, L. A., Chiu, E. F., & Scott, B. L. (2000). A speech interface for building musical score collections. *Proceedings of the 5th ACM Conference on Digital Libraries*, 165–173.

Södring, T., & Smeaton, A. (2002). Evaluating a melody extraction engine. *Proceedings of the 24th BCS-IRSG European Colloquium on IR Research*. Retrieved February 7, 2002, from http://www.compapp.dcu.ie/~tsodring/ research/papers.

Sonoda, T., & Muraoka, Y. (2000). A WWW-based music retrieval system: An indexing method for a large melody database. *Proceedings of the International Computer Music Conference (ICMC 2000)*, 170–173. Retrieved February 7, 2002, from http://www.sonoda.net/papers/sonoda-icmc2000.pdf.

Sutton, J. B. (1988). *MIRA: A PROLOG-based system for musical information retrieval and analysis*. Unpublished master's thesis, University of North Carolina, Chapel Hill.

Tague-Sutcliffe, J. (1992). The pragmatics of information retrieval experimentation, revisited. *Information Processing & Management*, *28*, 467–490.

Tague-Sutcliffe, J., Downie, J. S., & Dunne, S. (1993). Name that tune: An introduction to musical information retrieval. *Proceedings of the 21st Annual Conference of the Canadian Association for Information Science*, 204–216.

Temperley, N. (1993). The problem of definitive identification in the indexing of hymn tunes. In R. D. Green (Ed.), *Foundations of music bibliography* (pp. 227–239). Binghamton, NY: Haworth Press.

Tseng, Y.-H. (1999). Content-based retrieval for music collectors. *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99)* 176–182.

Tzanetakis, G., & Cook, P. (2000). Audio information retrieval (AIR) tools. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/papers/tzanetakis_paper.pdf.

Tzanetakis, G., Essl, G., & Cook, P. (2001). Automatic musical genre classification of audio signals. *Proceedings of the 2nd Annual International Symposium on Music Information Retrieval (ISMIR 2001)*, 205–210. Retrieved February 7, 2002, from http://ismir2001.indiana.edu/pdf/tzanetakis.pdf.

Uitdenbogerd, A. L., & Zobel, J. (1998). Manipulation of music for melody matching. *Proceedings of the 6th ACM International Conference on Multimedia*, 235–240. Retrieved February 7, 2002, from http://www.acm.org/sigmm/MM98/electronic_proceedings/uitdenbogerd/index.html.

Uitdenbogerd, A. L., & Zobel, J. (1999). Matching techniques for large music databases. *Proceedings of the 7th ACM International Multimedia Conference*, 57–66.

Von Schroeter, T., Doraisamy, S., & Rüger, S. (2000). From raw polyphonic audio to locating recurring themes. *Proceedings of the 1st Annual International Symposium on Music Information Retrieval (ISMIR 2000)*. Retrieved February 7, 2002, from http://ciir.cs.umass.edu/music2000/posters/shroeter_ruger.pdf.

Welte, J. (2001, May 23). MP3.com: A dream deferred. *Business 2.0*. Retrieved February 7, 2002, from http://www.business2.com/articles/web/0,1653,15733,FF.html.

Wordspot. (2001). *WordSpot search engine word usage covering August 27th through September 3rd, 2001*. Retrieved February 7, 2002, from http://www.wordspot.com.

Wu, S., & Manber, U. (1992). Fast text searching allowing errors. *Communications of the ACM, 35*(10), 83–91.