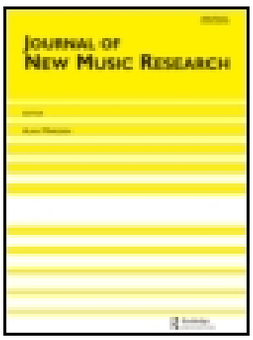

Journal of New Music Research

ISSN: 0929-8215 (Print) 1744-5027 (Online) Journal homepage: http://www.tandfonline.com/loi/nnmr20

Interdisciplinary Research Issues in Music Information Retrieval: ISMIR 2000–2002

Joe Futrelle & J. Stephen Downie

To cite this article: Joe Futrelle & J. Stephen Downie (2003) Interdisciplinary Research Issues in Music Information Retrieval: ISMIR 2000–2002, Journal of New Music Research, 32:2, 121-131

To link to this article: http://dx.doi.org/10.1076/jnmr.32.2.121.16740

Published online: 09 Aug 2010.

Journal of New Music Research
2003, Vol. 32, No. 2, pp. 121–131

0929-8215/03/3202-121$16.00

Interdisciplinary Research Issues in Music Information Retrieval: ISMIR 2000–2002

Joe Futrelle and J. Stephen Downie

Graduate School of Library and Information Science, University of Illinois, Champaign, IL, USA

Abstract

Music Information Retrieval (MIR) is an interdisciplinary research area that has grown out of the need to manage burgeoning collections of music in digital form. Its diverse disciplinary communities, exemplified by the recently established ISMIR conference series, have yet to articulate a common research agenda or agree on methodological principles and metrics of success. In order for MIR to succeed, researchers need to work with real user communities and develop research resources such as reference music collections, so that the wide variety of techniques being developed in MIR can be meaningfully compared with one another. Out of these efforts, a common MIR practice can emerge.

1. Introduction

Music Information Retrieval (MIR) is a rapidly growing interdisciplinary research area encompassing computer science and information retrieval, musicology and music theory, audio engineering and digital signal processing, cognitive science, library science, publishing, and law. Its agenda, roughly, is to develop ways of managing collections of musical material for preservation, access, research, and other uses. In this way it resembles traditional library science, and indeed, libraries have historically led the development of music collections. The idea of applying automatic information retrieval (IR) techniques to music dates back to the 1960's (Kassler, 1966). But in particular, MIR has grown recently out of an explosion of interest in networked collections of musical material *in digital form*, precipitated by the development of compression technologies such as mp3, online services such as Napster, advances in optical musical recognition (OMR), and the ever-plummeting costs of digital storage and bandwidth. In this sense MIR is closely related to Digital Libraries.

As in other interdisciplinary fields, discourse in MIR is impeded at disciplinary boundaries by unfamiliar jargon, differing methodology, and even philosophical and ethical differences. To understand the field, it is currently necessary to acquire at least a cursory understanding of each of the disciplines, and MIR researchers are undertaking this as they begin to develop a common practice (Downie, 2001a).

This paper investigates MIR's interdisciplinary communities and research issues by surveying the proceedings of the International Symposia on Music Information Retrieval (ISMIR 2000 and 2001) and the International Conference on Music Information Retrieval (ISMIR 2002). The ISMIR series is an attempt to gather together all of the disciplines and research areas pertinent to MIR. ISMIR can also claim to be the only conference series exclusively devoted to the advancement of MIR research. For these reasons, we believe the proceedings of ISMIR provide a representative "snapshot" of the major issues comprising MIR research and development. Based on these proceedings, and with reference to components of their supporting literatures, this paper will characterize the field, outline the major research communities involved in the field, assess the state of the art in each community, identify coverage gaps, and propose a research agenda aimed at addressing those gaps.

2. What is MIR research?

What are MIR researchers trying to build? What problems are they trying to solve? There is broad consensus that the rapidly-increasing availability of digital music requires attention (Durey & Clements, 2001; Hoos, Renz, & Görg, 2001; Kornstädt, 2001; Yang, 2001). But few effective retrieval techniques exist for digital music collections of any size, and consequently only a small fraction of MIR research investi-

Accepted: 21 March, 2003

Correspondence: Joe Futrelle, Graduate School of Library and Information Science, University of Illinois, Champaign, IL 61820, USA. Tel.: +1 217 265 0296, E-mail: futrelle@uiuc.edu

gates scaling existing techniques to large collections (Lemström & Perttu, 2000; Jang, Chen, & Kao, 2001; Haitsma & Kalker, 2002). Instead, most MIR research is focused on basic research into IR techniques.

Developing effective IR techniques for music is challenging because of the wide variety of ways music is produced, represented, and used (Smiraglia, 2001; Hemmasi, 2002; Downie, 2003). Basic research in MIR can be categorized roughly by the kind of music representation employed. Table 1 shows the major categories of representations employed and the kinds of MIR research being applied to them.

In addition to the variety of music representations, their complexity presents a problem as well. Like language, music in virtually all of its representations contains difficult-to-extract layers of significance, including harmony, polyphony, and timbre. Even the most robust representations still require sophisticated processing techniques to extract some of these features from musical material, and developing these techniques is an active area of MIR research. This area is often called "content-based" MIR, to distinguish it from more traditional digital and pre-digital approaches based on manually-produced metadata of bibliographic and related varieties.

As basic MIR research begins to produce results, it raises the questions of what kind of MIR systems can be built, what their user interfaces would be, and what teir institutional or economic contexts would be. These questions have been most comprehensively investigated in a set of relatively large testbed projects, including VARIATIONS (Dunn & Mayer, 1999; Dunn, 2000; Dunn, Davidson, & Isaacson, 2001), MELDEX (Bainbridge, Nevill-Manning, Witten, Smith, & McNab, 1999; Bainbridge, 2000), and the Levy Sheet Music Collection (Choudhury, DiLauro, Droettboom, Fujinaga, & Harrington, 2000). Other work in this area involves issues such as Intellectual Property rights (Levering, 2000; Chiariglione, 2002).

3. MIR communities

Understanding an interdisciplinary field requires understanding the disciplines involved and the variety of research interests they represent. Table 2 summarizes the research

Table 1. Music Representations in MIR.

Representation	Description	Research
Symbolic	Notation (scores, charts), Event-based recordings (MIDI), Hybrid Representations	Matching, Theme/Melody Extraction, Voice Separation, Musical Analysis
Audio	Recordings, Streaming Audio, Instrument Libraries	Sound/Song Spotting, Transcription, Timbre/Genre Classification, Musical Analysis, Recommendation Systems
Visual	Scores	Score Reading ("Optical Music Recognition")
Metadata	Cataloging, Bibliography, Descriptions	Library Testbeds, Traditional IR, Interoperability, Recommendation Systems

Table 2. MIR Communities.

Community	Type of Institution(s)	Typical Research Areas
Computer Science, Information Retrieval	Academic, Commercial	Representation, Indexing, Retrieval, Machine Learning, User Interface Design
Audio Engineering, Digital Signal Processing	Academic, Commercial	Compression, Feature Detection, Pitch Tracking, Machine Learning, Classification, Playlist Generation, Musical Analysis
Musicology, Music Theory	Academic	Representation, Musical Analysis
Library Science	Libraries, Academic	Representation, Metadata, User Studies, Classification, Intellectual Property Rights, User Interface Design
Cognitive Science, Psychology, Philosophy	Academic	Representation, Perception, User Studies, Ontology
Law	Government, Legal Profession, Academic	Intellectual Property Rights

Table 3. MIR Research Areas.

Research Area	Description
Representation	How should musical material be represented in digital form? What aspects of music are critical to represent for the purpose of building music collections? At what level of granularity can we represent music? What kinds of representation are the most efficient? How can markup languages be applied to music? (Chen, 2000; Good, 2000; Hoos et al., 2001; Lindsay & Kim, 2001; Maidín & Cahill, 2001; Dannenberg & Hu, 2002; Lartillot, 2002; Pardo & Birmingham, 2002; Song, Bae, & Yoon, 2002)
Indexing	How can database indexing techniques be applied to musical material so it can be retrieved effectively and efficiently? (Downie, 1999; Chen, 2000; Pickens, 2000; Hsu & Chen, 2001; Jin & Jagadish, 2002; Pienimäki, 2002; Yang, 2002)
Retrieval	What kinds of queries can we perform on indexed collections of musical material? How can the performance of these queries be evaluated and improved? (Hsu & Chen, 2001; Lemström, Wiggins, & Meredith, 2001; Doraisamy & Rüger, 2002; Pickens, Bello, Monti, Crawford, & Dovey, 2002)
User Interface Design	How can user interfaces be built which enable users to effectively find and use digital musical material from a collection? (Fernström & Maidín, 2001; Kornstädt, 2001; MacMillan, Droettboom, & Fujinaga, 2001; Baumann & Klüter, 2002; Rauber, Pampalk, & Merkl, 2002; Vinet, Herrera, & Pachet, 2002)
Recommendation	How can systems be built which automatically select music for listeners? How can playlists be generated that meet users' individual needs? (Aucouturier & Pachet, 2002; Logan, 2002; Pauws & Eggen, 2002)
Compression	How can audio be encoded more efficiently? What are the implications for MIR of various emerging compression technologies? (Lindsay & Kim, 2001; Haitsma & Kalker, 2002)
Feature Detection	How can distinguishing features of music be detected from audio signals? How can these techniques be applied to MIR systems? (Foote, 1999; Ismirli, 2000; Logan, 2000; Nam & Berger, 2001; Tzanetakis, Essl, & Cook, 2001; Heittola & Klapuri, 2002; Kim & Whitman, 2002; Whitman & Smaragdis, 2002)
Classification/ Machine Learning	What kinds of classification techniques and schemes can/should be applied to digital music collections? How can we deduce or induce aggregate musical features of collections, so that they can be organized for retrieval? What are the most efficient and effective ways of representing these aggregate features? (Herrera, Amatriain, Batlle, & Serra, 2000; Reiss, Aucouturier, & Sandler, 2001; Aucouturier & Pachet, 2002; Ellis, Whitman, Berenzweig, & Lawrence, 2002; Whitman & Smaragdis, 2002)
Musical Analysis	How is a musical composition organized? How is it similar to, or different from, other pieces of music? How can MIR systems meet the needs of musicologists and musicians? (Bonardi, 2000; Larson, 2000; Cope, 2001; Kornstädt, 2001; Cooper & Foote, 2002; Dannenberg & Hu, 2002; Scheirer, 2002; Vinet et al., 2002)
Summarization	How can a musical work be automatically summarized? What aspects of a musical work belong in its summary? (Cooper & Foote, 2002; Hirata & Matsuda, 2002; Peeters, La Burthe, & Rodet, 2002)
Metadata	What kinds of descriptive or contextual information about musical material can and/or should be managed by an MIR system, and how should such metadata be represented? Can we make a "Semantic Web" for music? (cf. Representation) (Choudhury et al., 2000; Lindsay & Kim, 2001; Smiraglia, 2001; Chiariglione, 2002; Datta, 2002; Hemmasi, 2002; Scheirer, 2002)
User Studies	What kinds of MIR capabilities do users need? How do users search for musical material, and why? What would be the ideal MIR system for a given user community? (Itoh, 2000; Selfridge-Field, 2000; Uitdenbogerd, 2000; McPherson & Bainbridge, 2001; Downie & Cunningham, 2002; Kim & Belkin, 2002)
Intellectual Property Rights	Who owns musical material? Under what conditions? Under what arrangements can digital libraries of musical material and owners of IP rights for musical material peacefully coexist? What are the appropriate incentives for content providers to make digital music available? (Levering, 2000; Chiariglione, 2002)
Perception	How do people perceive music? How can music perception inform the design of MIR systems? What is music? How is musical similarity perceived? (cf. Ontology) (Perrot & Gjerdigen, 1999; Huron, 2000; Scheirer, Watson, & Vercoe, 2000; Hofmann-Engl, 2001; Aucouturier & Pachet, 2002; Hofmann-Engl, 2002)
Epistemology/ Ontology	What is music? What is a musical composition? What is the relationship between different representations of "the same" piece? How do improvised aspects of music relate to composed aspects with respect to collections of musical material? (Smiraglia, 2001; Baumann & Klüter, 2002; Geekie, 2002; Hemmasi, 2002)

communities involved in MIR along with their typical home institutional settings and typical areas of research.

Table 3 describes major research areas in MIR, along with basic research questions and representative papers in those areas.

3.1 Computer science, information retrieval

Of course, virtually all MIR research employs techniques from computer science. But there is an important subset of MIR research whose origins can be traced back to the Information Retrieval research on bibliographic text retrieval systems in the early 1960s. This ongoing research emphasizes techniques for locating items in a collection or index that match a query, rather than techniques for analyzing aggregate properties of collections of items (e.g., data mining). For an overview of traditional IR see Baeza-Yates and Ribeiro (1999). MIR research based on traditional IR is typically aimed at supporting a scenario in which users know characteristics of the music they desire, and use an MIR system to locate musical material that most closely matches those characteristics. Downie (1999, 2003) calls this a "locating" MIR system.

In traditional IR, a query on a collection can be thought of as a fragment or reduced form of the desired item from the collection. For text collections, the query is often a word that occurs in the desired document. MIR researchers have taken some pains to devise MIR systems which fit this model, including Downie (1999) and Doraisamy and Rüger (2002) who reduce symbolic musical material to n-grammed sequences of intervals, which can then be indexed using inverted files. Other research on locating MIR systems using other traditional IR strategies includes probabilistic modeling (Pickens, 2000) and approximate string matching (Lemström, et al., 2001). Some work also addresses IR issues such as relevance and ranking (Uitdenbogerd, 2000). Most of this research is based on symbolic music representations, but it has also been applied to audio which is pre-processed and converted to symbolic sequences of audio feature classes as in Aucouturier and Sandler (2001) and Batlle and Cano (2000).

A great deal of attention has been paid to so-called "Query by Humming" (QBH) systems, which retrieve pieces based on melodic fragments sung by the user. QBH systems typically combine melody extraction from an audio query with a locating MIR system to match the melody against a target database (Bainbridge, 2000; Haus & Pollastri, 2001; Pauws, 2002). Research on other retrieval modalities has recently focused mostly on automatic playlist generation (Aucouturier & Pachet, 2002; Logan, 2002; Pauws & Eggen, 2002), although a small number of studies also explore browsing (Rauber et al., 2002; Vinet et al., 2002) and recommendation (Uitdenbogerd & Van Schyndel, 2002). A few studies examine alternative processing models such as clustering (Jang et al., 2001) and peer-to-peer (Wang, Li, & Shi, 2002).

3.2 Audio engineering, digital signal processing

A major category of MIR research concerns audio representations of music (i.e., recordings, audio streams, or live performances). Techniques used in this area date back to the pioneering works in Digital Signal Processing (DSP) of the late 1960s (Gold & Rader, 1969), and to following decades of research on speech recognition and audio compression. MIR researchers are now investigating the application of these techniques to MIR. For a good overview of audio information retrieval techniques which includes MIR, see Foote (1999). The techniques are applied to an interrelated set of problems:

- What are the most important features of audio representations of music for MIR, and how can they be extracted from audio? (e.g., melodies, harmonies, instrument timbres, etc.)
- Given a set of features extracted from audio, what techniques can be used to understand the relationships between those features in an audio collection?
- How can we use audio to perform structural analyses of music, and how can these be used to improve MIR systems?

A number of audio features have been used in MIR research. Virtually every audio MIR system uses some kind of frequency-domain transformation of the signal, such as the Fast Fourier Transform (FFT) (Brigham, 1988) or its potentially more musically-relevant derivative, mel-frequency cepstral coefficients (MFCC) (Logan, 2000; Cooper & Foote, 2002). A number of other features are used, including time-domain autocorrelation (mostly used for pitch tracking) and wavelet transforms (Tzanetakis et al., 2001), but most of the features employed are statistics computed from the FFT. The reason frequency-domain transformations are so prevalent is the primacy of periodicity in the perception of musical aspects such as pitch, timbre, and rhythm.

Once a set of features is selected and can be reliably extracted from audio, the problem is essentially one of multivariate analysis, in which each piece of music in a collection can be conceptualized as a vector in n-dimensional feature space. Traditional multivariate techniques as well as probabilistic machine-learning techniques such as Hidden Markov Models (HMM) and neural networks can be applied to identify salient features and perform data reduction, often through classification. A good overview of these and other classification techniques for audio can be found in Herrera et al. (2000).

3.3 Musicology, music theory

The study of music is an important application area for MIR, and thus drives much of MIR research. Musicology and music theory are ancient fields of inquiry which have been transformed by computational techniques (Bel & Vecchione, 1993) and promise to be further transformed by ready access

to large digital music collections. Musicology-related MIR research includes work on computational music analysis by Cope (2001), Barthélemy and Bonardi (2001), and Dannenberg and Hu (2002). Other work involves MIR systems specialized for musicologists, such as those of Bonardi (2000) and Kornstädt (2001). Selfridge-Field (2000) has discussed how the information needs of musicologists and music theorists differ from those of other kinds of users. An interesting tension exists in MIR between musical analysis which concerns the nature of music *per se* and is often qualitative as in Larson (2000), and approaches that attempt to empirically demonstrate improved retrieval performance and thus rely on quantitative techniques such as statistical analysis and machine learning. Cope's work with computational musical analysis (2001) and algorithmic composition (1992) bridges these two different perspectives by using computational techniques to divine aspects of music, such as style, which have traditionally been investigated qualitatively or with exhaustive manual effort as in Van der Merwe (1989). There is clearly some middle ground that remains unexplored, since aspects of music that have currently only been characterized by musicologists may yet prove useful in the design of MIR systems.

3.4 Library and information science

Libraries and library scientists are involved in MIR as part of their ongoing effort to cope with rapidly growing multimedia collections. Libraries face all of the issues raised by MIR, from basic research questions such as how to represent and index musical material, to applied information technology issues such as integrating traditional bibliographic systems with advanced metadata and MIR tools, to policy issues such as how to manage intellectual property rights for the producers and users of music collections. Of particular importance to MIR are a number of testbed projects being undertaken at academic libraries and digital library research facilities, including Indiana's Digital Music Library projects (Dunn & Mayer, 1999; Dunn, 2000; Dunn et al., 2001), the University of Waikato's MELDEX digital library (Bainbridge et al., 1999; Bainbridge, 2000; McPherson & Bainbridge, 2001), and the Levy sheet music collection at Johns Hopkins (Choudhury et al., 2000). These testbeds integrate a variety of MIR tools with significant music collections in order to explicitly address issues such as usability, scale, and multi-modal access to musical works (e.g., linking scores with recordings). They also support the application of MIR tools and collections to specialized use cases such as music theory education.

Testbeds are a good way to begin to evaluate who the potential users of MIR systems are and what features most interest them. Preliminary user studies such as McPherson and Bainbridge (2001) indicate a trend away from speculative user requirements analysis such as that found in, for example, Bonardi (2000) and Blandford and Stelmaszewska (2002) towards empirically grounded approaches and techniques. The user modeling methods put forward by Rolland (2001) suggest that future MIR systems can be tailored to meet the needs of a variety of user communities.

3.5 Cognitive science, psychology, philosophy

A small subset of MIR research concerns the implications of music perception on the design of MIR systems. Research efforts range from models of music perception such as those discussed by Huron (2000), Dannenberg (2001) and Hofmann-Engl (2001) to epistemological analysis of music information such as Smiraglia (2001). Much research has been done on music perception in psychology, music psychology and cognitive science (Deliáege & Sloboda, 1997; Cook, 1999). There also have been notable efforts in both music philosophy (Adorno, 1973) and cultural studies (Attali, 1985; McClary, 1991) to characterize how music is understood as a social and cultural phenomenon. Significantly, however, MIR researchers have so far rarely adopted work in these areas as a basis for MIR studies, although some recent work shows promise (Hofmann-Engl, 2002).

3.6 Law

High-profile cases such as Napster demonstrate that MIR systems are being developed in an uncertain regulatory environment with enormous economic stakes, and legal issues will continue to be important to MIR researchers until this situation changes. Issues such as the Digital Millennium Copyright Act (Levering, 2000), intellectual property rights management, and researcher access to music databases (Byrd, 2001; Downie, 2001b) will be important to the field indefinitely. To a large extent, copyright law is a policy issue rather than a technical issue, but legal issues dramatically affect the priorities of commercial and non-commercial agencies funding MIR research, and thus are of critical importance (Chiariglione, 2002). Increasingly, MIR research on metadata includes or focuses primarily on the use of metadata for IP rights management (Datta, 2002; Scheirer, 2002), and is informed by the perspectives of commercial organizations such as publishing companies and electronics manufacturers.

4. Critical analysis of coverage gaps in MIR research

MIR's newness and multi-disciplinary constituency make the field strong on innovation and basic research. However, there are three major coverage gaps in MIR research, considered as whole:

1. There are no commonly accepted means of comparing the efficacy of retrieval techniques;
2. There have been few rigorous attempts to study potential users of MIR systems to find out what they need; and,

3. MIR research has failed to significantly address music outside of the common-practice Western music canon.

These problem areas are interrelated in that meaningful evaluation of retrieval techniques must be grounded in a significant understanding of user requirements and musical culture. Together, they represent a risk that the techniques developed by MIR researchers will fail to effectively meet the needs of user communities. This makes them more critical than other coverage gaps, such as lack of effort on scaling existing techniques, or relative inattention to the latest digital library technologies, that represent technical problems that can be solved without rethinking MIR research practice and methodology.

4.1 Difficult to compare techniques

Research into MIR techniques rarely presents results that can easily be compared with results of other research. There are a variety of reasons for this. Some studies, for example Spevak and Polfreman (2001) and Rolland (2001), do not present any evaluation results at all. Other studies do not adequately describe their evaluation methodology: for example Yang (2001) reports "90% retrieval accuracy" without explaining what constitutes "accurate" retrieval. Still others, such as Nishimura, Hashiguchi, Takita, Zhang, and Goto (2001) adopt evaluation metrics without explaining why one metric was chosen instead of another equally applicable one. But by far the most troubling problem with MIR evaluation is that many studies evaluate MIR techniques against very small music collections, often the personal collections of the authors, with seemingly no concern about problems of bias or lack of generality. So far, the most sophisticated attempts to rigorously compare a variety of MIR techniques are being done in limited domains as in Hsu and Chen (2001) and Uitdenbogerd (2000).

We believe this inconsistency largely arises from MIR's interdisciplinary nature. Evaluation metrics that are well understood in one field, such as precision and recall in traditional IR, are new and unfamiliar to other fields such as audio engineering. In addition, there are no community-wide music collections against which researchers can cross-evaluate a wide variety of different techniques, a problem which the community is eager to address (Byrd, 2001; Downie, 2001b). ISMIR 2002 included *The Panel on Evaluation Frameworks* (Downie, 2002b) aimed at developing a community-wide consensus on establishing standardized testbeds, tasks and evaluation metrics. The resultant collection of white papers on the topic, solicited from members of the major MIR constituencies, is further evidence of the growing concern that MIR research needs a stronger, more rigorous, evaluation paradigm (Downie, 2002a).

4.2 Few attempts to assess user requirements

Jef Raskin's talk at ISMIR 2001 about how to make computer systems more usable was notable in that it suggested using theoretical models of users to guide user interface design choices rather than involving users in the design process (Raskin, 2001). This emphasis on basic research over application to, and involvement with, users is common in MIR research, and may result from the influence of the computer science and audio engineering communities.

Already, MIR is beginning to emphasize certain areas of research without having identified user communities and evaluated whether the techniques developed will meet the needs of those communities. In particular, Query-by-Humming has become the dominant retrieval paradigm in the MIR community, despite the fact that there is no evidence cited that users prefer this modality, and even some that suggests that they do not (McPherson & Bainbridge, 2001). In the ISMIR 2001 proceedings, 10 out of 43 papers, posters, and talks (23%) concerned QBH systems, and this was little changed in 2002, when 13 out of 58 (22%) papers focused on QBH. Instead of empirically-derived user needs assessments, QBH papers typically begin with speculations that such a system would be useful, such as "singing is naturally used as input" (Haus & Pollastri, 2001), or with anecdotal evidence such as

> The potential utility of such systems is attested to by music librarians, who report that library patrons often hum or whistle a phrase of music and ask them to identify the corresponding musical work (Smith & Medina, 2001).

Until recently, even MIR research focusing on usability has rarely involved user studies. For instance, Kornstädt (2001) presents a graphical user interface apparently tailored for the needs of musicologists, but cites neither research into what kinds of tools musicologists need nor any evaluation of the system by musicologists. Although analysis of user behavior has received more attention recently from researchers such as Downie and Cunningham (2002), Kim and Belkin (2002), and Baumann and Klüter (2002), these kinds of studies remain exceptions. User interface research, now undertaken most often as an afterthought to research into retrieval techniques, is clearly under-emphasized, especially since retrieval interfaces may have to incorporate complex audio strategies such as those explored by Fernström and Ó Maidín (2001) and Rauber et al. (2002).

To some extent, this is a chicken-and-egg problem: MIR researchers cannot study user behavior without complete system implementations which include fully-developed interfaces and large, comprehensive collections of music with which users can spontaneously interact. As testbed projects continue to develop, they will be in the best position to analyze their own users (e.g., McPherson and Bainbridge, 2001 and Dunn, 2000), which should provide valuable guidance to basic researchers. However, it still remains to be seen whether or not real user needs match the interests or technological capabilities of the many disciplines currently involved in MIR research.

4.3 Undue emphasis on Western music

Another significant coverage gap in MIR concerns non-Western music. The music used in MIR studies is predominantly common-practice Western music. This is particularly problematic for symbolic MIR systems, which tend to depend on representations derived from common-practice Western music notation. Notable exceptions include Linardis, Politis, Kotopoulous and Alygizakis (2001), who describe a retrieval system based on Byzantine neumatic notation, and Lee, Downie, and Renear (2002), who propose an XML representation of traditional Korean music, and Geekie (2002), who attempts to characterize Carnatic ragas as MIR entities. Audio MIR systems are presumably more flexible than symbolic MIR systems because the audio representations and features they employ are presumably more culturally neutral, but no audio MIR research has specifically investigated this hypothesis.

Addressing the undue emphasis on common-practice Western music in MIR research requires, finally, a radical rethinking of MIR research practice. Assumptions commonly made by MIR researchers about music – that it has melodies, that its rhythm is metrical, and that it can be arbitrarily re-contextualized – must be replaced by provisional sets of assumptions resonant with the cultural milieus of real user communities.

5. First principles and a MIR research agenda

To best overcome the gaps in current MIR research, we believe that MIR research must embrace as *sine qua non* the following three principles:

1. MIR systems are developed to serve the needs of particular user communities.
2. MIR techniques are evaluated according to how well they meet the needs of user communities.
3. MIR techniques are evaluated according to agreed-upon measures against agreed-upon collections of data, so that meaningful comparisons can be made between different research efforts.

To realize these first principles, the following steps must be taken before an MIR technique is developed or an MIR system is implemented:

1. MIR research must begin by assessing existing MIR systems (broadly defined to include both digital and traditional formats), including libraries, music retailers, on-line media merchants, and individual collections. The implementation and real-world use of these systems must be evaluated to establish baselines of usability upon which new MIR systems must improve. These evaluations must be systematic and empirical and involve the participation of both users and maintainers of existing systems, rather than being based on the opinions and speculations of MIR researchers.
2. The investigation of existing MIR practice must identify distinct user communities and investigate what they need from proposed MIR systems. As MIR researchers have already pointed out, the musicological community has, for example, quite different requirements for MIR systems than other communities. Future MIR research must explicitly identify which community's needs it is attempting to address.

Once MIR techniques and systems have been developed, they must be empirically validated. To this end, MIR researchers must take the following steps:

3. They must standardize their evaluation measures. Retrieval accuracy and system effectiveness should be measured using clearly delineated, agreed-upon methodologies and reported consistently across studies. To this end, MIR researchers must share music collections, so that a variety of techniques can be applied to the same collection and results replicated or refuted by independent research teams. The development of a set of "universal" test collections as called for by the ISMIR 2001 "resolution" (see http://music-ir.org/mirbib2/resolution) is an important step in this direction.
4. Finally, MIR systems must prove themselves usable and useful to the user communities for which they were designed. This must be established through rigorous user studies and analysis of use cases in both controlled and real-world settings.

If the MIR research community embraces these principles and this research agenda, future MIR systems will better provide real users with the tools, features, and ease-of-use they need to get the most out of rich and comprehensive collections of music in digital form.

Acknowledgements

An earlier version of this paper appeared in the proceedings of ISMIR 2002 (Futrelle & Downie, 2002). The authors wish to thank the anonymous reviewers of the ISMIR 2002 committee for their helpful suggestions on improving the original draft of this paper. J. Stephen Downie also wishes to thank the Andrew W. Mellon Foundation for its financial support of *The Music Information Retrieval Annotated Bibliography* and the *MIR Evaluation Frameworks* projects (http://music-ir.org).

References

Adorno, T.W. (1973). *Philosophy of modern music*. New York: Seabury Press.

Attali, J. (1985). *Noise: The political economy of music*. Minneapolis: University of Minnesota Press.

Aucouturier, J.-J., & Pachet, F. (2002). Music similarity measures: What's the use? *Third International Conference on Music Information Retrieval*. Paris, France. pp. 157–163.

Aucouturier, J.-J., & Sandler, M. (2001). Using long-term structure to retrieve music: Representation and matching. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 1–2.

Baeza-Yates, R., & Ribeiro, B.d.A.N. (1999). *Modern information retrieval*. New York: ACM Press.

Bainbridge, D. (2000). The role of music ir in the new zealand digital library project. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Bainbridge, D., Nevill-Manning, C.G., Witten, I.H., Smith, L., & McNab, R.J. (1999). Towards a digital library of popular music. *Fourth ACM Conference on Digital Libraries*. Berkeley, California.

Barthélemy, J., & Bonardi, A. (2001). Figured bass and tonality recognition. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 129–136.

Batlle, E., & Cano, P. (2000). Automatic segmentation for music classification using competitive hidden markov models. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Baumann, S., & Klüter, A. (2002). Super-convenience for non-musicians: Querying mp3 and the semantic web. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 297–298.

Bel, B., & Vecchione, B. (1993). Introduction. *Computers and the Humanities*, *27*.

Blandford, A., & Stelmaszewska, H. (2002). Usability of musical digital libraries: A multimodal analysis. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 231–237.

Bonardi, A. (2000). Ir for contemporary music: What the musicologist needs. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Brigham, E.O. (1988). *The fast fourier transform and its applications*. Englewood Cliffs, New Jersey: Prentice-Hall.

Byrd, D. (2001). Music-notation searching and digital libraries. *ACM/IEEE Joint Conference on Digital Libraries*. Roanoke, VA. 239–246.

Chen, A.L.P. (2000). Music representation, indexing and retrieval at nthu. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Chiariglione, L. (2002). Technology and art – putting things in context. *Third International Conference on Music Information Retrieval*. Paris, France. 238–240.

Choudhury, G.S., DiLauro, T., Droettboom, M., Fujinaga, I., Harrington, B., & MacMillan, K. (2000). Optical music recognition system within a large-scale digitization project. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Cook, P.R. (1999). Music, cognition, and computerized sound: An introduction to psychoacoustics.

Cooper, M., & Foote, J. (2002). Automatic music sumarization via similarity analysis. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 81–85.

Cope, D. (1992). Computer modeling of musical intelligence in emi. *Computer Music Journal*, *16*, 69–83.

Cope, D. (2001). Computer analysis of musical allusions. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 83–84.

Dannenberg, R.B. (2001). Music information retrieval as music understanding. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 139–142.

Dannenberg, R.B., & Hu, N. (2002). Pattern discovery techniques for music audio. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 63–70.

Datta, D. (2002). Managing metadata. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 249–251.

Deliáege, I., & Sloboda, J.A. (1997). *Perception and cognition of music*. Hove, East Sussex: Psychology Press.

Doraisamy, S., & Rüger, S.M. (2002). A comparative and fault-tolerance study of the use of n-grams with polyphonic music. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 101–106.

Downie, J.S. (1999). Evaluating a simple approach to music information retrieval: Conceiving melodic n-grams as text.

Downie, J.S. (2001a). Music information retrieval annotated bibliography website project, phase i. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 5–7.

Downie, J.S. (2001b). Whither music information retrieval: Ten suggestions to strengthen the mir research community. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 219–222.

Downie, J.S. (Ed.). (2002a). *The mir/mdl evaluation project white paper collection* [On-line]. Available: http://music-ir.org/evaluation/wp.html.

Downie, J.S. (2002b). Panel on music information retrieval evaluation frameworks. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 303–304.

Downie, J.S. (2003). Music information retrieval. *Annual Review of Information Science and Technology*, *37*, pp. 295–340.

Downie, J.S., & Cunningham, S.J. (2002). Toward a theory of music information retrieval queries: System design implications. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 299–300.

Dunn, J.W. (2000). Beyond variations: Creating a digital music library. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Dunn, J.W., Davidson, M.W., & Isaacson, E.J. (2001). Indiana university digital music library project: An update. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 137–138.

Dunn, J.W., & Mayer, C.A. (1999). Variations: A digital music library system at indiana university. *Fourth ACM Conference on Digital Libraries*. Berkeley, California.

Durey, A.S., & Clements, M.A. (2001). Melody spotting using hidden markov models. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 109–117.

Ellis, D.P.W., Whitman, B., Berenzweig, A., & Lawrence, S. (2002). The quest for ground truth in musical artist similarity. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 170–177.

Fernström, M., & Ó Maidín, D. (2001). Computer-supported browsing for mir. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 9–10.

Foote, J. (1999). An overview of audio information retrieval. *ACM Multimedia Systems*, *7*, pp. 2–11.

Futrelle, J., & Downie, J.S. (2002). Interdisciplinary communities and research issues in music information retrieval. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 215–221.

Geekie, G. (2002). Carnatic *ragas* as music information retrieval entities. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 257–258.

Gold, B., & Rader, C.M. (1969). *Digital processing of signals*. New York,: McGraw-Hill.

Good, M. (2000). Representing music using xml. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Haitsma, J., & Kalker, T. (2002). A highly robust audio fingerprinting system. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 107–115.

Haus, G., & Pollastri, E. (2001). An audio front end for query-by-humming systems. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 65–72.

Heittola, T., & Klapuri, A. (2002). Locating segments with drums in music signals. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 271–272.

Hemmasi, H. (2002). Why not marc? *Third International Conference on Music Information Retrieval*. Paris, France. pp. 242–248.

Herrera, P., Amatriain, X., Batlle, E., & Serra, X. (2000). Towards instrument segmentation for music content description: A critical review of instrument classification techniques. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Hirata, K., & Matsuda, S. (2002). Interactive music summarization based on gttm. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 86–93.

Hofmann-Engl, L. (2001). Towards a cognitive model of melodic similarity. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 143–151.

Hofmann-Engl, L. (2002). Panel on similarity in music. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 306–308.

Hoos, H.H., Renz, K., & Görg, M. (2001). Guido/mir – an experimental musical information retrieval system based on guido music notation. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 41–50.

Hsu, J.-L., & Chen, A.L.P. (2001). Building a platform for performance study of various music information retrieval approaches. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 153–162.

Huron, D. (2000). Perceptual and cognitive applications in music information retrieval. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Ismirli, O. (2000). Using a spectral flatness based feature for audio segmentation and retrieval. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Itoh, M. (2000). Subject search for music: Quantitative analysis of access point selection. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Jang, J.-S.R., Chen, J.-C., & Kao, M.-Y. (2001). Miracle: A music information retrieval system with clustered computing engine. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 11–12.

Jin, H., & Jagadish, H.V. (2002). Indexing hidden markov models for music retrieval. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 20–24.

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

Kim, J.-Y., & Belkin, N.J. (2002). Categories of music description and search terms and phrases used by non-music experts. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 209–214.

Kim, Y.E., & Whitman, B. (2002). Singer identification in popular music recordings using voice coding features. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 164–169.

Kornstädt, A. (2001). The jring system for computer-assisted musicological analysis. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 93–98.

Larson, S. (2000). Searching for meaning: Melodic patterns, combinations, and embellishments. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Lartillot, O. (2002). Integrating pattern matching into an analogy-oriented pattern discovery framework. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 269–270.

Lee, J.H., Downie, J.S., & Renear, A. (2002). Representing korean traditional musical notation in xml. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 263–264.

Lemström, K., & Perttu, S. (2000). Semex – an efficient music retrieval prototype. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Lemström, K., Wiggins, G.A., & Meredith, D. (2001). A three-layer approach for music retrieval in large databases. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 13–14.

Levering, M. (2000). Intellectual property rights in musical works. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Linardis, P., Politis, D., Kotopoulous, H., & Alygizakis, A. (2001). Musical information retrieval for delta and neumatic systems. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 23–24.

Lindsay, A., & Kim, Y. (2001). Adventures in standardization, or, how we learned to stop worrying and love mpeg-7. *Second*

International Symposium on Music Information Retrieval. Bloomington, IN, USA. pp. 195–196.

Logan, B. (2000). Mel frequency cepstral coefficients for music modelling. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Logan, B. (2002). Content-based playlist generation: Exploratory experiments. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 295–296.

MacMillan, K., Droettboom, M., & Fujinaga, I. (2001). Gamera: A structured document recognition application development environment. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 15–16.

McClary, S. (1991). *Feminine endings: Music, gender, and sexuality*. Minneapolis: University of Minnesota Press.

McPherson, J.R., & Bainbridge, D. (2001). Usage of the meldex digital music library. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 19–20.

Nam, U., & Berger, J. (2001). Addressing the "same but different – different but similar" problem in automatic music classification. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 21–22.

Nishimura, T., Hashiguchi, H., Takita, J., Zhang, J.X., Goto, M., & Ryuichi, O. (2001). Music signal spotting retrieval by a humming query using start frame feature dependent continuous dynamic programming. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 211–218.

Ó Maidín, D., & Cahill, M. (2001). Score processing for mir. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 59–64.

Pardo, B., & Birmingham, W. (2002). Encoding timing information for musical query matching. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 267–268.

Pauws, S. (2002). Cubyhum: A fully operational query by humming system. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 187–196.

Pauws, S., & Eggen, B. (2002). Pats: Realization and user evaluation of an automatic playlist generator. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 222–230.

Peeters, G., La Burthe, A., & Rodet, X. (2002). Toward automatic music audio summary generation from signal analysis. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 94–100.

Perrot, D., & Gjerdigen, R.O. (1999). *Scanning the dial: An exploration of the factors in the identification of musical style*. Society for Music Perception and Cognition. Evanston, IL 88.

Pickens, J. (2000). A comparison of language modeling and probabilistic text information retrieval approaches to monophonic music retrieval. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Pickens, J., Bello, J.P., Monti, G., Crawford, T., Dovey, M.J., Sandler, M., & Byrd, D. (2002). Polyphonic score retrieval using polyphonic audio queries: A harmonic modeling approach. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 140–149.

Pienimäki, A. (2002). Indexing music databases using automatic extraction of frequent phrases. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 25–30.

Raskin, J. (2001). Making machines palatable. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 39–40.

Rauber, A., Pampalk, E., & Merkl, D. (2002). Using psychoacoustic models and self-organizing maps to create a hierarchical structuring of music by sound similarity. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 71–80.

Reiss, J., Aucouturier, J.-J., & Sandler, M. (2001). Efficient multidimensional searching routines for music information retrieval. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 163–171.

Rolland, P.-Y. (2001). Adaptive user modeling in a content-based music retrieval system. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 27–30.

Scheirer, E.D. (2002). About this business of metadata. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 252–254.

Scheirer, E.D., Watson, R.B., & Vercoe, B.L. (2000). On the preceived complexity of short musical segments. *International Conference on Music Perception and Cognition.* Keele, UK.

Selfridge-Field, E. (2000). What motivates a musical query? *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Smiraglia, R.P. (2001). Musical works as information retrieval entities: Epistemological perspectives. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 85–91.

Smith, L., & Medina, R. (2001). Discovering themes by exact pattern matching. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 31–32.

Song, J., Bae, S.Y., & Yoon, K. (2002). Mid-level music melody representation of polyphonic audio for query-by-humming system. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 133–139.

Spevak, C., & Polfreman, R. (2001). Sound spotting – a frame-based approach. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 35–36.

Tzanetakis, G., Essl, G., & Cook, P. (2001). Automatic musical genre classification of audio signals. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 205–210.

Uitdenbogerd, A.L. (2000). Music ir: Past, present, and future. *International Symposium on Music Information Retrieval*. Plymouth, MA, USA.

Uitdenbogerd, A.L., & Van Schyndel, R. (2002). A review of factors affecting music recommender success. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 204–208.

Van der Merwe, P. (1989). *Origins of the popular style: The antecedents of twentieth-century popular music*. Oxford Oxfordshire: Clarendon Press.

Vinet, H., Herrera, P., & Pachet, F. (2002). The cuidado project. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 197–203.

Wang, C., Li, J., & Shi, S. (2002). A kind of content-based music information retrieval method in a peer-to-peer environment. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 178–186.

Whitman, B., & Smaragdis, P. (2002). Combining musical and cultural features for intelligent style detection. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 47–52.

Yang, C. (2001). Music database retrieval based on spectral similarity. *Second International Symposium on Music Information Retrieval*. Bloomington, IN, USA. pp. 37–38.

Yang, C. (2002). The macsis acoustic indexing framework for music retrieval: An experimental study. *Third International Conference on Music Information Retrieval*. Paris, France. pp. 53–62.