Toward a Robust and Large-Scale
Music Information Retrieval System:
A Multidisciplinary Approach to Development and Evaluation

A Proposal Submitted by

J. Stephen Downie, Ph.D.,
Assistant Professor
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

For Application to

The NCSA Fellows Programme

2000-2001

ABSTRACT

The problems associated with the creation, deployment, and evaluation of robust, large-scale, and content-based Music Information Retrieval (MIR) systems are far from trivial. Music information is inherently multi-faceted, multi-representational (i.e., can be represented in many different ways), and space-intensive. The dizzyingly complex interaction of Pitch, Temporal, Harmonic, Timbral, Editorial, Textual, and Bibliographic facets, make music information difficult to store, and then retrieve, in any robust, large-scale, and comprehensive manner (i.e., this complex interaction is the “MIR problem”). The goal of this NCSA Fellows proposal is to draw upon the multi-disciplinary expertise of those at NCSA (and around UIUC) in dealing with other large-scale and complex information types. Through the implementation of a multi-disciplinary approach to MIR, I intend to identify, and then formally evaluate, the possible MIR application of those pre-existing supercomputing solutions currently being developed and implemented at NCSA.
Introduction
The problems associated with the creation, deployment, and evaluation of robust, large-scale, and content-based, 1 Music Information Retrieval (MIR) systems are far from trivial. Music information is inherently multi-faceted, multi-representational (i.e., can be represented in many different ways), and space-intensive. The dizzyingly complex interaction of Pitch, Temporal, Harmonic, Timbral, Editorial, Textual, and Bibliographic facets make music information difficult to store, and then retrieve, in any robust, large-scale, and comprehensive manner (i.e., this complex interaction is the “MIR problem”). The goal of this NCSA Fellows proposal is to draw upon the multi-disciplinary expertise of those at NCSA (and around UIUC) in dealing with other large-scale and complex information types. Through the implementation of a multi-disciplinary approach to MIR, I intend to identify, and then formally evaluate, the possible MIR application of those pre-existing supercomputing solutions currently being developed and implemented at NCSA.

Background
Automating access to music information through the use of digital computers has intrigued musicologists, computer scientists, librarians and music lovers alike. Each has his own purpose in mind and thus there seems to be as many approaches to developing MIR systems as there are users and researchers. A half-hour’s perusal of the back issues of Computing in Musicology (Hewlett and Selfridge-Field, eds.) will bring this fact to the fore. Some have designed complex suites of computer tools (e.g., David Huron’s (1991) "Humdrum" [http://dactyl.som.ohio-state.edu/Humdrum/] to analyze all the varied facets of music. Others have tried to automate the thematic catalogue by including incipit or thematic extracts as part of a bibliographic record (e.g., RISM (1997)[http://www.rism.harvard.edu/rism/]; also, Huron and Kornstaedt’s "Themefinder" [http://musedata.stanford.edu/databases/Themefinder/]). Still others have explored the idea of using sophisticated approximate string matching techniques (e.g., McNab et al.’s (1996; 1997) "Meldex" [http://www.nzdl.org/cgi-bin/gw?c=meldex&a=page&p=coltitle]; Prechelt & Typke’s (1998) "Tuneserver" [http://wwwipd.ira.uka.de/tuneserver/]). All of these approaches are similar in that each has some kind of shortcoming. For example, the more powerful analytic systems can be very difficult to use; incipit and thematic indexes leave out large amounts of music that might be of interest; and, approximate string matches can be computationally expensive without necessarily giving better results.
The recently published *Melodic Comparison: Concepts, Procedures, and Applications* (Selfridge-Field Hewlett and Selfridge-Field 1998) [http://musedata.stanford.edu/publications/cm/idx11.html] and the upcoming *International Symposium on Music Information Retrieval* (October 2000) are indicative of the growing interest in MIR issues. However, in the recently drafted literature review for my Ph.D. dissertation on MIR (Downie 1999), I noted that, despite the growing interest in MIR issues, the formal literature remains sparse, particularly where the evaluation of MIR systems is concerned. It appears that one reason for the sparseness of the MIR literature is that the various MIR research teams are operating autonomously. I hope that the opportunity afforded by the NCSA Fellows program will contribute to the formation of a more cohesive, multidisciplinary, MIR research and evaluation programme by providing all interested parties at NCSA, GSLIS, UIUC—and, eventually around the world—the chance to contribute their expertise, insights, and achievements to resolving the MIR problem.

**Central Questions**

At the heart of this NCSA Fellows proposal resides a rather simple, yet non-trivial, pair of questions:

1. "How is music information **similar** to other types of information?"; and,
2. "How is music information **dissimilar** to other types of information (i.e., unique)."

I believe that answer to the second question is to be determined by a thorough examination of the first. Furthermore, by apprehending the answers to these two questions, it becomes possible to determine which computational techniques are most appropriately applied to the MIR problem. For those aspects of music information that prove unique, novel approaches will have to be developed. However, for those aspects of music information that are found to have strong similarities with other types of information, the prudent route to pursue would be the application of techniques already shown to be efficient for those other types of information. Simply put, why reinvent the wheel?

---

1 That is, having the ability to pose music queries musically.
Strengths and Weakness of Music Information Analogues
In my recently completed dissertation work (Downie 1999a), I showed that an extraordinarily simplistic representation of melodic information, in the form of interval-only n-grams, could be treated, for the purposes of MIR, as though the n-grams were text (i.e., melodic n-gram = “word”). By making the assumption that my melodic n-grams were “words,” I treated my entire collection of 9354 folksongs as though it were yet another text collection. I then proceeded to evaluate the informetric properties and retrieval characteristics of the collection using traditional text-based techniques (e.g., normalized recall, normalized precision, etc.). My "MIR engine" was Salton's famous text retrieval programme, SMART (using tf*idf as the "term" ranking method) (see Salton 1989). The results of the experimental evaluations were very positive. Thus, I added a bit of data toward answering Question #1; that is, in some simplistic respects, music information and text can be said to be similar. This finding in turn suggests that, within some limited domains, the MIR problem can be addressed though the application of traditional text-based IR techniques.

Notwithstanding the success of the above approach, it is obvious that the "music-as-text" analogy has some severe limitations. First, what does a given melodic n-gram actually mean? Second, my dissertation work threw away all harmonic, rhythmic, timbral, textual, bibliographic, and editorial information in order to make the "n-gram equals word" analogy work. These aspects of music information must be represented within any large-scale MIR system of the future. Therefore, there still much work to be done in identifying and evaluating:

1. other potentially useful analogies; and,

2. the unique aspects of music information.

Outstanding Questions
During my life as a Ph.D. student, I gave numerous seminars, conference papers, and poster presentations about my research. Almost without fail, interested parties would offer up a common list of comments and

---

3 Downie (1999b) contains the front matter, table of contents, and bibliography of the thesis. Downie and Nelson (2000) presents the principal findings of the IR experiments and evaluations of the thesis. Both are found within the supporting materials accompanying this proposal.
suggestions, all of which were/are valid; however, time constraints forced these important aspects of MIR research to be put off to another day for resolution. Five of the more resonant comments were:

A. “The MIR problem is the same as a gene-sequence IR problem.”
B. “The MIR problem is the same as the "knowledge discovery" / "data mining" problem.”
C. “The heart of the MIR is the problem of scale. What can we do about representing, storing, and retrieving the seemingly infinite amount of music information? There are googol-bytes of recorded CDs, mp3s, musical scores, etc. to which we want access but how will we do this without sacrificing their inherent complexities for the sake of efficiency?”
D. “What about music that is not notated? What methods are available to extract the necessary information directly from recordings? What about non-traditional, non-Western, styles of music?”
E. “A real sticking point with MIR is the interface issue: How are naive users going to interact with a MIR system when they have limited computer and music skills?”

Addressing the Outstanding Questions via the NCSA Fellows Programme
For my proposed NCSA Fellows research project, I would like acknowledge and address the important questions outlined above. I would also like to use these questions as a means to show that my MIR research programme and the various regions of expertise found at NCSA have some very strong commonalities. If accepted as a Fellow, I would call upon each of the following NCSA members for guidance and advice. Here is how I see my outstanding questions and their expertise fitting together:

Dr. Jakobson (Question A): The CompBio Research Group appears to be the ideal source for advice on the "MIR-as-gene-sequence-IR" analogy. More specifically, the Bioinformatics & Protein Structure Subgroup's work on recognizing patterns in genomic sequence data is precisely the type of prior research from which to begin examining the strengths and weaknesses of the gene-sequence approach to MIR.

Dr. Welge (Question B): The techniques used in Knowledge Discovery in Databases (KDD) can have real utility within the realm of MIR. Both problems involve large-sets of multi-faceted data and the extraction of not-readily-apparent, embedded attributes, and commonalities. Having
reviewed Dr. Welge’s slides on KDD (http://www.ncsa.uiuc.edu/edu/course98/slides/week5/welge/) I can see some non-trivial applications of the KDD clustering and visualization techniques with in the domain of MIR. For example, it would be of great use to naive and expert users alike to be able partition large-scale collections of music into meaningful sub-groups of interest and to see how those sub-groups relate to one another.

**Dr. Folk** (Question C): As stated previously, music information is both multi-faceted and space-intensive. Having reviewed several of the HDF5 presentations available on the web (e.g., http://www.ncsa.uiuc.edu/alliance/alliance98/Proceedings/poster/folk-hdf5/hi/), I see that HDF5 is designed to deal just such data. The development of a robust HDF5 format for MIR should be investigated before creating a novel representation solution.

**Dr. Bagar** (Question D): I had the pleasure of meeting briefly with Dr. Bagar on 22 Feb. 2000. From our conversation, I became aware of others on campus who, in Dr. Bagar's experience, have made great strides in signal processing, the application of which might find utility in extracting music information from digital audio recordings. Dr. Bagar's overlapping expertise in sonification, signal processing, audio issues, audio displays, and, of course, music, would make him a invaluable resource upon which to draw.

**Dr. Patterson** (Question E): I would like to draw upon Dr. Patterson's expertise in the development and implementation of novel and immersive HCI solutions. Without some radical innovations in present MIR interface design, I fear that all the work done addressing the first four questions would be in danger of becoming irrelevant.

**Implications and Concluding Remarks**
The problems associated with MIR system development and evaluation are intellectually intriguing in their own right. Notwithstanding the worthiness of the intellectual aspects of MIR, the successful development of robust, large-scale MIR systems has important social and commercial implications. According to Wordspot (2000), an Internet consulting company which tracks queries submitted to Internet search
engines, the search for music—specifically, the now-popular mp3 format—has usurped the traditional search for sex-related materials as the most popular quest. At this moment, not one so-called “mp3 search engine”, is doing anything more than merely indexing the textual meta-data supplied by the creators of the files. It is not exaggerating to claim that a successful, commercially-based MIR system has to the potential to garner vast sums of revenue, given the recent success of the text-based WWW search-engines and directories. Beyond the commercial implications, the implementation of robust MIR systems would create significant added value to the huge collections of underused music currently being warehoused in the world’s libraries by making the entire corpus of music readily accessible. This accessibility would be highly beneficial to musicians, scholars, students, and general members of the public, alike

**Budget Overview**

I would like to apply for appointment as a NCSA Fellow for the full calendar year of 21 August 2000 to 20 August 2001. During my fellowship period, I would like to request:

- salary funding for the two summer months of 2001 (to devote my time exclusively to the experimental evaluations)
- RA funding for one 50% equivalent graduate student for 11 months (to support me throughout each phase of the project)
- stipend funding for sundry material costs (e.g., photocopying, books, presentation materials, etc. with an emphasis on the necessary materials required to bring me “up-to-speed” on the various supercomputing solutions being investigated.)
- stipend funding to acquire necessary programmer time on a task-by-task basis (acquiring programmer time can be problematic, particularly when the actual programming needs are yet to be determined; in the past, I have had great success keeping down costs by acquiring necessary programming on a task-by-task basis, as this prevents all of the funds being spent simply to retain a potentially under-utilized programmer)
- The detailed budget form is attached.
Research Programme Overview

Mission
• To establish the University of Illinois at Urbana-Champaign as the world’s preeminent locus of MIR research, development, and evaluation.

Goals
• To establish a truly multidisciplinary approach to the MIR problem
• To extend the application of supercomputing techniques within the realm of fine arts, performing arts, and humanities information problems
• To leverage the support of NCSA and UIUC’s Office of the Vice-Chancellor for Research to attract international, federal, state, university, and corporate research funding
• To integrate multidisciplinary MIR research with multidisciplinary undergraduate and graduate pedagogy in keeping with UIUC’s commitment to IT education (i.e., bring together interested students from music, library science, commerce, computer science, engineering, and liberal arts and sciences, etc., as contributors, research assistants, as well as, classroom participants)

Objectives
• Formation of a multidisciplinary team of MIR researchers
• Identification of MIR sub-problems amenable to pre-existing supercomputing solutions
• Formal evaluation of pre-existing supercomputing solutions with an eye toward identifying the boundaries, or breaking points, of those solutions
• Identification of MIR sub-problems requiring novel solutions

Research Calendar

Phase 0 (Spring-Summer 2000) Pre-Fellowship Preparatory Work
• Establish contacts within NCSA and across campus
• Bring self “up-to-speed” on work at NCSA in possible solution areas (e.g., HDF5, data mining, gene-sequence IR, etc.)
Phase I (Fall Term 2000): Cross-Pollination, Team Building, and Brainstorming
- Conduct a series of colloquia in various locations (e.g., NCSA, GSLIS, School of Music, etc.) with three goals:
  1. to explicate the MIR problem to those expert in other domains
  2. to exploit the expertise of those in other domains (i.e., garner from them possible applications of their expertise to the MIR problem)
  3. to foster a meaningful, multidisciplinary, and ongoing dialogue among parties interested in the MIR problem
- Establish formal mechanisms of communication between interested parties (e.g., a schedule of meetings, brown-bags, email-lists, etc.) with the object of brainstorming a list of possible solutions

Phase II (Spring Term 2001): Option Selection and Prototyping
- From the brainstorming, identify an interim short-list of the more promising pre-existing supercomputing applications and techniques
- Classify potential solutions:
  1. Supercomputing dependent: those solutions that require constant application of supercomputing hardware and methods to implement a feasible MIR system or system component
  2. “Run-once” supercomputing dependent: those solutions that require supercomputing hardware and methods only to build and maintain—but not to implement—a feasible MIR system or system component (e.g., building of special indices, signal processing, informetric modeling, data conversion, etc.)
- Prototype the MIR application of the selected solutions in order to establish the feasibility of the various approaches
- Finalize short-list to a select few—1, 2, or perhaps, 3—of the most promising solutions (N.B. at this juncture, I believe a slight preference should be given to the “Run-once” solutions as such solutions could allow for the building of portable demonstration systems)
• During the Spring 2001 term I am slated to teach both LIS250MSI *Music and Sound as Information*, and our Doctoral research seminar, LIS490; I plan on incorporating within these courses the findings-to-date of the project, as well as, guest speakers from the multidisciplinary MIR community at NCSA and UIUC.

**Phase III (Summer 2001): Experimentation and Evaluation**

- Determine the formal criteria for success, failure, and indecision (N.B. these will be dependent upon the solutions chosen for evaluation)
- Conduct formal, hypothesis driven, experimental evaluations of the selected solutions based upon the criteria determined (N.B. experimental evaluations will be simulation-based without the need for human subjects)
- Explicate results with an eye toward determining the limitations (i.e., boundaries, or breaking points) of the evaluated pre-existing supercomputing solutions to the MIR problem
- Disseminate results and build upon findings.

---

4 LIS250 MSI: *Music and Sound as Information* (Spring 2001) Course description:
The advent of multimedia computing has awakened a growing interest in the representation, manipulation, and use, of both music and sound as part of our everyday lives. This introductory course will examine the wide variety of methods used to create, record, represent, modify, and present music and sound information. Basic acoustics, major notation schemes, and formats such as streaming audio, mp3, WAV, and MIDI, are explored with an eye toward learning how music and sound fit into our information universe. Prerequisite: LIS201 or LIS202 and junior or senior standing, or permission of the instructor. Students should have basic computer skills.
Bibliography


Recent and Current Projects

- The most important of my recent research projects has been the successful completion of my Ph.D. dissertation (completed 16 July 1999; defended 31 August 1999; deposited 20 September 1999).

- While completing the dissertation, I undertook to situate myself at the center of MIR research. To this end, I proposed, and had accepted, the first full-day workshop exclusively devoted to MIR issues ever held in North America. The *Exploratory Workshop on Music Information Retrieval*, was held as part of the SIGIR ’99 Annual Conference, at Berkeley, California, 19 August 1999. MIR researchers from such diverse locations as France, Canada, New Zealand, United States, England, Norway, and Italy convened to inform one another about their widely varied approaches to the MIR problem. Another heartening aspect of this workshop was the multidisciplinary backgrounds of the participants: library science, musicology, industry, computer science and engineering. I also presented a well-received poster at the conference, based upon my dissertation findings.

- The connections established at the SIGIR ’99 workshop led me to become programme chair for the *International Symposium on Music Information Retrieval* to be held 23-25 October 2000 at Plymouth, MA. I am working principally with the MIR researchers at the University of Massachusetts, Center for Intelligent Information Retrieval to bring this NSF-sponsored event about. We are currently looking into publishing the proceedings as a book (CIIR has strong connections with Kluwer). Marvin Minsky of MIT has agreed to be our keynote speaker. Besides being programme chair, I intend to submit an invited paper on the topic of MIR system evaluation.

- Other attempts at establishing connections have also borne fruit. I attended the Annual Meeting of the American Society for Information Science (ASIS) this past October-November. From the invited panel presentation I gave on the state of the art in MIR development and evaluation I garnered:
  1. a Ph.D. applicant who wishes to study MIR issues with me at GSLIS;
  2. an invitation, which I have accepted, to write a literature review for the 2001 edition of the *Annual Review of Information Science and Technology* (ARIST) on the topic of MIR; and,
  3. an invitation, which I have accepted, to be guest editor of special MIR issue of the *Journal of the American Society for Information Science* (JASIS) [date of issue is yet to be determined]
References

Dr. Linda C. Smith, Professor and Associate Dean (External thesis examiner)
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
Phone: (217) 333-7742
Email: smith@alexia.lis.uiuc.edu

Dr. Robert Wood, Professor (Second thesis reader)
Faculty of Music,
Talbot College, University of Western Ontario
London, Ontario, Canada
Phone: (519) 661-2111 x85335
E-mail: r.w.wood@uwo.ca

Dr. Michael Nelson, Associate Professor (Thesis supervisor)
Faculty of Information and Media Studies
Graduate Programs in Library and Information Science
Middlesex College, University of Western Ontario,
London, Ontario, Canada
Phone: (519) 661-2111 x88507
E-mail: mnelson@julian.uwo.ca