PERGAMON

Information Processing and Management 38 (2002) 249–272

INFORMATION
PROCESSING
&
MANAGEMENT

www.elsevier.com/locate/infoproman

Problems of music information retrieval in the real world ☆

Donald Byrd [a,*], Tim Crawford [b]

[a] *Center for Intelligent Information Retrieval, Department of Computer Science, University of Massachusetts, Amherst, MA, USA*

[b] *Music Department, Kings College, London, UK*

Abstract

Although a substantial number of research projects have addressed music information retrieval over the past three decades, the field is still very immature. Few of these projects involve complex (polyphonic) music; methods for evaluation are at a very primitive stage of development; none of the projects tackles the problem of realistically large-scale databases. Many problems to be faced are due to the nature of music itself. Among these are issues in human perception and cognition of music, especially as they concern the recognizability of a musical phrase. This paper considers some of the most fundamental problems in music information retrieval, challenging the common assumption that searching on pitch (or pitch-contour) alone is likely to be satisfactory for all purposes. This assumption may indeed be true for most monophonic (single-voice) music, but it is certainly inadequate for polyphonic (multi-voice) music. Even in the monophonic case it can lead to misleading results. The fact, long recognized in projects involving monophonic music, that a recognizable passage is usually not identical with the search pattern means that approximate matching is almost always necessary, yet this too is severely complicated by the demands of polyphonic music. Almost all text-IR methods rely on identifying approximate units of meaning, that is, words. A fundamental problem in music IR is that locating such units is extremely difficult, perhaps impossible. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Information retrieval; Searching; Music; Audio; MIDI; Notation

---

☆ This material is based on work supported in part by the Digital Libraries Initiative, Phase 2, under NSF grant IIS-9905842, by the UK Joint Information Systems Committee under project code JCDE/NSFKCL, and by the National Science Foundation, Library of Congress and Department of Commerce under cooperative agreement number EEC-9209623. Any opinions, findings and conclusions or recommendations expressed in this material are the authors' and do not necessarily reflect those of the sponsors.

[*] Corresponding author.

E-mail addresses: dbyrd@cs.umass.edu (D. Byrd), tim.crawford@kcl.ac.uk (T. Crawford).

PII: S0306-4573(01)00033-4

> This work contains about 10,000 themes...we feel that we have compiled a fairly complete index of themes, not only first themes, but every important theme, introduction, and salient rememberable phrase of the works included.
>
> –Barlow and Morgenstern, *A Dictionary of Musical Themes* (1948, p. xi).

1. Introduction

The first published work on music information retrieval (music IR), by Michael Kassler and others, dates back to the mid-1960s. Kassler (1966, 1970) and his colleagues were well ahead of their time, and for many years thereafter, very little was done; but now, interest in music IR is exploding. A paper on music IR (Bainbridge, Nevill-Manning, Witten, Smith, & McNab, 1999) won the best paper award at the Digital Libraries '99 conference, and almost every recent SIGIR, Digital Libraries, Computer Music, or Multimedia conference has had one or more papers on music retrieval and/or digital music libraries (see, for example, Downie & Nelson, 2000; Lemström, Laine, & Perttu, 1999; Tseng, 1999; Uitdenbogerd & Zobel, 1998). Furthermore, the first major grant for music-IR research, to the present authors, was recently funded (Wiseman, Rusbridge, & Griffin, 1999; OMRAS, 2000), and the First International Symposium on Music Information Retrieval (ISMIR, 2000) was held just last fall. But everything published to date reports on specific projects: no general discussion of the problems researchers need to solve has appeared. This paper attempts to fill that gap.

To put things in perspective, music IR is still a very immature field: much of what follows is necessarily speculative. For example, to our knowledge, no survey of user needs has ever been done (the results of the European Union's HARMONICA project (HARMONICA, 1999) are of some interest, but they focused on general needs of music libraries). At least as serious, the single existing set of relevance judgements we know of (Uitdenbogerd, Chattaraj, & Zobel, in press) is extremely limited; this means that evaluating music-IR systems according to the Cranfield model that is standard in the text-IR world (see, for example, Sparck Jones & Willett, 1997, pp. 4–59, 171) is impossible, and no one has even proposed a realistic alternative to the Cranfield approach for music. Finally, for efficiency reasons, some kind of indexing is as vital for music as it is for text; but the techniques required are quite different, and the first published research on indexing music dates back no further than five years. Overall, it is safe to say that music IR is decades behind text IR.

For another sort of perspective, nearly all music-IR research we know of is concerned with mainstream Western music: music that is not necessarily tonal and not derived from any particular tradition ("art music" or other), but that is primarily based on notes of definite pitch, chosen from the conventional gamut of 12 semitones per octave. In this paper, we maintain that bias. Thus, we exclude music for ensembles of percussion instruments (not definite pitch), microtonal music (not 12 semitones per octave), and electronic music, i.e., music realized via digital or analog sound synthesis (if based on notes at all, often not definite pitch, and almost never limited to 12 semitones per octave).

Music IR is cross-disciplinary, involving very substantial elements of music and of information science. It also involves a significant amount of music perception and cognition. We wanted this paper to be intelligible to readers with whatever background, but found it impractical to avoid assuming a fair amount of knowledge of information science and some knowledge of music.

2. Background

2.1. Basic representations of music and audio

There are three basic representations of music and audio: the well-known *audio* and *music notation* at the extremes of minimum and maximum structure, respectively, and the less-well-known *time-stamped events* form in the middle. Numerous variations exist on each representation. All three are shown schematically in Fig. 1, and described in Fig. 2.

The "Average relative storage" figures in the table are for uncompressed material and are our own estimates. A great deal of variation is possible based on type of material, mono vs. stereo, etc., and – for audio – especially with such sophisticated forms as MP3, which compresses audio typically by a factor of 10 or so by removing perceptually unimportant features.

"Convert to left" and "Convert to right" refer to the difficulty of converting fully automatically to the form in the column to left or right. Reducing structure with reasonable quality (convert to left) is much easier than enhancing it (convert to right).

It is often helpful to compare music and text; this is particularly true here because text also comes with varying amounts of explicit structure, though that is seldom recognized in the IR literature (see Fig. 3).

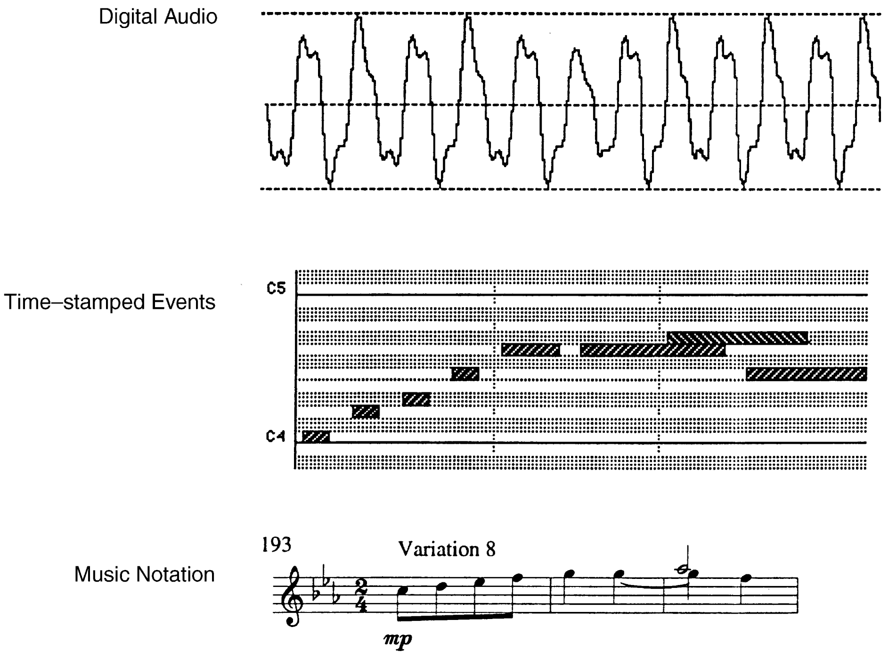

Fig. 1. Basic representations of music (schematic).

Representation	*Audio*	*Time-stamped Events*	*Music Notation*
Common examples	CD, MP3 file	Standard MIDI File	sheet music
Unit	sample	event	note, clef, lyric, etc.
Explicit structure	**none**	**little (partial voicing information)**	**much (complete voicing information)**
Avg. rel. storage	2000	1	10
Convert to left	-	easy	OK job: easy

Convert to right	1 note/time: pretty easy;
	2 notes/time: hard;
	other: very hard
		OK job: fairly hard	-

Ideal for	music bird/animal sounds sound effects speech	music	music

Fig. 2. Basic representations of music.

Explicit structure	**minimum**	**medium**	**maximum**
Music representation (and examples)	Audio (CD, MP3)	Events (Standard MIDI File)	Music Notation (sheet music)
Text representation (and examples)	Audio (speech)	ordinary text	text with markup (HTML)

Fig. 3. Text vs. music.

While musical notation is invaluable for many applications of music IR, notation of complex music is very demanding: divergencies in interpretation and inconsistencies of application often frustrate attempts at its computational treatment (see Byrd, 1984, 1994).

2.2. Music perception and music IR

As we have said, we concern ourselves here with music based on definite-pitched notes. Nearly all music familiar to Western ears is built up out of notes somewhat as text is built up out of characters or words; notes are much closer to characters than to words, but there is less similarity than might appear. We will return to this analogy.

The four basic parameters of a definite-pitched musical note are generally listed as:
pitch: how high or low the sound is, the perceptual analog of frequency,
duration: how long the note lasts,
loudness: the perceptual analog of amplitude,
timbre or tone quality.

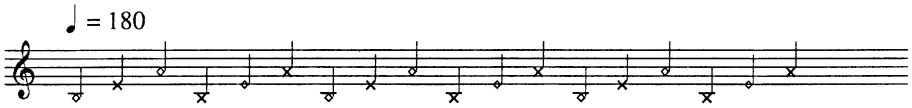

Fig. 4. Wessel's streaming illusion.

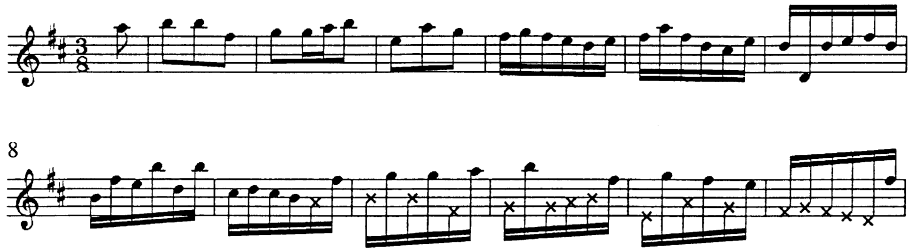

Fig. 5. Telemann: Fantasie no. 7 in D major, I.

But human beings hear music in a "non-linear" way: studies in music perception and cognition reveal many subtle and counterintuitive aspects, and these parameters are not nearly as cleanly separable as might at first appear. To cite a simple example, very short notes are heard as being less loud than otherwise identical longer notes. And when a group of notes is heard in sequence as a melody, the effects of perception can be very unobvious. For example, changing timbre can turn a single melodic line into multiple voices and vice versa. Pierce (1992) devotes an entire chapter to "Perception, Illusion, and Effect" in music. In one very striking illusion he describes (pp. 211–212), due to David Wessel, a series of notes all played in similar timbre sounds like a melody composed of repetitions of a sequence of three notes going up (Fig. 4). But if alternate notes are played in very dissimilar timbres (say, diamond-shaped notes as brass and x-shaped notes as organ), it sounds like two interleaved melodies each composed of repetitions of a sequence of three notes going down.

Such *streaming* effects can be produced by changing tempo (i.e., speed of performance, affecting both note durations and onset times) as well as changing timbre. McAdams and Bregman (1979, p. 659) describe "a repeating six-tone series of interspersed high and low tones" that, when played at a moderate tempo, produces one perceptual stream, while at a fast tempo, "the high tones segregate perceptually from the low tones to form two streams." These examples may sound very artificial, but the idea – using differences of timbre, register (pitch), or anything else to turn single-note-at-a-time passages into perceptual streams – has been known to composers for centuries. It is exploited frequently in idiomatic keyboard music (e.g., Chopin, Bach) and string music (e.g., Bach's music for unaccompanied violin as well as the virtuoso music of Paganini and others). The most dramatic examples are in works such as Telemann's Fantasies for unaccompanied flute, written well over 200 years ago: of course the flute is an instrument that can play only one note at a time [1] and therefore can produce multiple streams only by exploiting perceptual phenomena. In Fig. 5, from his Fantasie no. 7 in D major, I, Telemann produces the effect of imitative

---

[1] This is not strictly true: techniques exist with which a flutist can play multiple simultaneous notes, but they are rarely used.

counterpoint. The first four measures are treated as a fugue subject, with a second entrance of the subject consisting of the notes with x-shaped heads. Nor is this just an effect for the score reader: in a competent performance, the second entrance is quite audible.

If a music-IR system were to operate only in a single highly structured representation – that is, music notation – these effects might be less of a problem. But most systems will need to operate in other representations. Besides, musical queries are likely to be based on a listener's recollection, and thus subject to error caused by such perceptual and cognitive effects. The implications of such problems have been discussed previously by McNab, Smith, Witten, Henderson, and Cunningham (1996) and by Uitdenbogerd and Zobel (1998).

For example, consider the fact that wide skips of pitch may not be heard as such: listeners' perceptual systems may remove octaves. On paper, the opening motif of Beethoven's Piano Sonata in B-flat, Op. 106 (the "Hammerklavier") has one of the widest ranges of any melody we know of: four octaves and a fifth, some 53 semitones (Fig. 6 is the way it appears in Barlow & Morgenstern, 1948 *Dictionary*). But the wide range is due almost entirely to two huge jumps, marked A and B in the figure. Jump A, two octaves and a major third, would sound nearly the same if it was reduced by an octave, while – to the authors' ears – the alleged two-octave jump B does not sound like a jump at all, but rather a change of texture: this is evident in Fig. 7, the full score. In fact, more-or-less any combination of octave transpositions of the three segments of the motif leaves it instantly recognizable, though rhythm undoubtedly plays a role in this.

It is not easy to imagine an algorithmic way to handle this problem; pitch perception is far more subtle than it appears at first, and complex textures and wide register changes are among the factors that affect it. But the octave seems to be a basic human perceptual unit (Deutsch, 1972), a fact that both music theory and composers' practice have acknowledged for centuries, and our problem might be sidestepped by viewing pitches as octave plus pitch class (C, Bb, etc.), and melodic intervals as number of octaves plus modulo-12 interval. Then we could give the number of octaves less weight, and rely more on other factors – rhythm is an obvious candidate – to rank matches. In fact, the index that occupies over 100 pages of Barlow and Morgenstern gives only pitch classes and completely ignores octaves (and therefore melodic direction). This is surely going too far, but it illustrates the point that a note's register is generally less important than its pitch class.

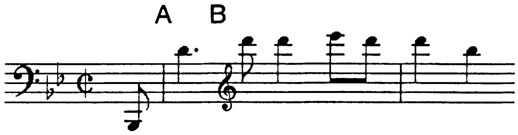

Fig. 6. Barlow and Morgenstern, after Beethoven.

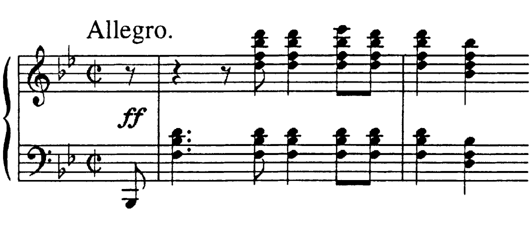

Fig. 7. Beethoven: Piano Sonata in B-flat, Op. 106, I.

2.3. Monophony, polyphony, and salience

Some music is *monophonic*, that is, only one note sounds at a time. Examples include unaccompanied folksongs and Gregorian chant. However, the vast majority of mainstream Western music is *polyphonic*: multiple notes sound at a time. As we shall see, the presence of polyphony makes music IR far more difficult. Note that in monophonic pieces like the Telemann example that employs streaming effects, the complications of polyphony are still possible, albeit in a limited way.

One complication in music IR that is largely a result of polyphony is the issue of *salience*, that is, how significant in perceptual terms an element of the music is, be it a note, chord, melody, or whatever. We will say more about salience later.

2.4. Music retrieval and the four parameters of notes

Two papers on music IR and the evaluation of musical similarity that underlies it offer apparently contradictory statements. Selfridge-Field (1998, p. 31): "Recent studies in musical perception suggest that durational values may outweigh pitch values in facilitating melodic recognition." On the other hand, Downie (1999, p. 15) remarks that "Psychoacoustic research has shown the [pitch] contour, or shape, of a melody to be its most memorable feature." In any case, it is evident that the pitch contour of a melody is by no means its *only* memorable feature.

One obvious question is what is the relative weight of information carried by each of our four parameters in a given style of music. Curiously, there does not appear to be any published work on this question, [2] but for the music we are focusing on, mainstream Western music in general, reasonable figures might be pitch 50%, rhythm 40%, timbre and dynamics 10%. Note that pitch occurs in both the horizontal (melodic) and vertical (harmonic) dimensions, and rhythm is not just strings of durations: it also involves accent patterns resulting from the meter (essentially, time signature). [3]

2.5. Pitch matching and realistic databases

In any case, it is clear that a great deal of the information in music is not in pitch, and certainly not in horizontal (melodic) pitch. Yet almost all music-IR work to date has focused primarily on pitch matching, and in the horizontal dimension alone – and that work has enjoyed a fair amount of success (cf. Downie, 1999). (One of the very few papers to focus on rhythm matching is Chen & Chen, 1998.) However, almost all music-IR work has also focused exclusively on monophonic music, and has been tested with moderate-sized databases (10,000 documents or so) of music that is relatively simple (often folksongs) as well as monophonic. For comparison, it is estimated that

---

[2] Boltz (1999) considers the relative cognitive effects in memorizing melodies of pitch and rhythm, and includes some discussion of style-related factors.

[3] A caveat here. Aside from questions of what the figures should be, citing any relative-weight figures makes it sound as if the factors are independent and can be combined linearly. In reality, these factors are clearly not independent. We might have to make the assumption of independence to make building a music-IR system a tractable problem, but we should always bear in mind that this is an oversimplification.

the music holdings of the Library of Congress amount to over 10,000,000 items, including over 6,000,000 pieces of sheet music and tens of thousands, perhaps hundreds of thousands, of scores of operas and other major works (K. LaVine, personal communication, May 2, 2000). As for polyphony, a symphony by Mozart might at times employ 12 voices; Stravinsky's *Le Sacre du Printemps* uses a maximum of about 38. Popular music is generally simpler than this, while most movie and TV music is probably in the same range as symphonic music. Will melodic pitch alone be adequate for large databases and complex music? Some evidence of the need to consider other information follows.

2.6. Salience

Salience in music is tremendously dependent on factors like dynamics (loudness) and thickness of texture. In fact, in works for large ensembles like the symphony orchestra, a substantial fraction of the "melodies" played by individual instruments are completely indistinguishable in the overall effect. This can lead to what appear to be excellent "matches" for queries that are actually of little or no interest.

2.7. Duration patterns and rhythm

Selfridge-Field (1998) gives several examples of ridiculous matches based on pitch alone (pp. 27, 32). The main cause in all cases is ignoring rhythm (though in some cases ignoring melodic direction is also a factor).

There are many melodies in which most interest is rhythmic. Extreme cases include those which begin with distinctive rhythms but with many repetitions of the same pitch, e.g., Beethoven's Symphony no. 7, III, main theme (12 repetitions); Bartók's Piano Sonata, II (20); and Jobim's One-Note Samba (no fewer than 30).

2.8. Confounds

Melodic-pitch-based music IR systems generally try to match either contours, or actual profiles of successive pitch intervals. But Selfridge-Field (1998, p. 30) comments that "three elements...can confound both contour and intervallic-profile comparisons. These are *rests*, *repeated notes*, and *grace notes* [italics ours]. Researchers focused on contours often argue that all three disrupt the 'flow' of the line." Other confounding elements include such "ornaments" as turns and *trills* (Fig. 8).

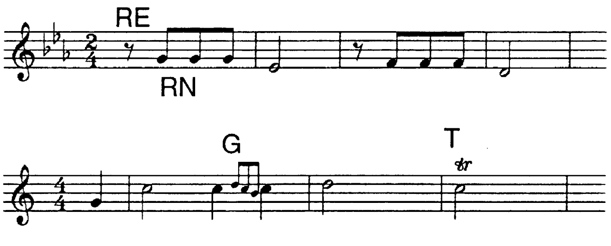

Fig. 8. RE = rest, RN = repeated notes, G = grace notes, T = trill.

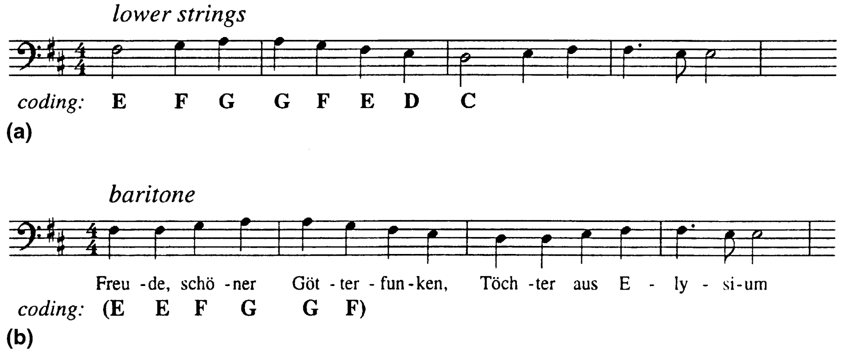
lower strings
coding: E F G G F E D C
(a)
baritone
Freu -de, schö -ner Göt - ter - fun - ken, Töch -ter aus E - ly - si-um
coding: (E E F G G F)
(b)

Fig. 9. Beethoven: Ode to Joy.

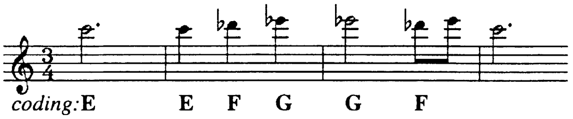

Fig. 10. Dvorak: The Wood Dove.

In many styles of music, these elements are common enough to be a serious complication. Our Appendix A: Melodic Confounds gives statistics on Barlow and Morgenstern's "classical" themes, as well as statistics on tunes in a "fake book" [4] and a hymnal. Approximately one-third of Barlow and Morgenstern's themes contain rests, and fully two-thirds of our sample of the fake book contain them.

Here is a real-life example that illustrates all three of the above-mentioned problems: salience, rhythm, and confounds. One of the current authors looked in Barlow and Morgenstern's index for the main theme of the last movement of Beethoven's Ninth Symphony, the famous "Ode to Joy" (shown in Fig. 9(a), with their index entry: letter names for the notes of the melody transposed to the key of C). The index contains an entry that matches the first six notes, but it is a little-known piece by Dvorak, The Wood Dove, Op. 110 (Fig. 10), that sounds hardly at all like the Beethoven. The main cause of the false positive is that the index ignores rhythm. The false negative is more interesting. Most instances of the theme in the Beethoven work, especially the more salient ones, involve trivial melodic ornamentation, specifically the "repeated notes" confound: subdivision of the first note (Fig. 9(b)). The latter version was, in fact, the one the current author searched for, while the former version, which occurs first in the Symphony, is the one Barlow and Morgenstern chose: four pages separate the entries for the two versions in the index!

At least 42% of Barlow and Morgenstern's themes contain repeated notes. Their claim of completeness (in the epigraph at the beginning of this paper) may be literally correct, but this incident shows that – at least with manual lookup – the 10,000 index entries are not sufficient to support retrieval in all reasonable cases. Mongeau and Sankoff (1990) discuss both our situation,

---

[4] This is a collection of popular song melodies with chord symbols so that musicians who do not know a given tune can "fake it".

which they call *fragmentation*, and the inverse situation, combining repeated notes into a single note, which they refer to as *consolidation.* The essence of the problem in either case is disagreement between the query and the score over the number of instances of a note.

2.9. Cross-voice matching

It is tempting to assume that one can search in polyphonic music for matches to a query one voice at a time, but in a great many cases, this will not be workable. For one thing, music in time-stamped event form generally does not have complete voicing information, and music in audio form has none at all (see Fig. 2). (Uitdenbogerd & Zobel, 1998, reports work on algorithmic treatment of MIDI for music-IR purposes.) Even when complete voicing information is available – usually where the database is in notation form – matching across voices will sometimes be necessary. An example is Mozart's Variations for piano, K. 265, on "Ah, vous dirais-je, Maman": the theme, otherwise known as "Twinkle, Twinkle, Little Star", is shown in Fig. 11(a). In Variations 2 (Fig. 11(b)), 4, and 9, the melody starts in one voice, then, after four notes – not enough for a reliable match – moves to another. We know of no prior work on cross-voice matching. But intuition suggests (and our preliminary research supports it) that cross-voice matching will be a disaster for precision if only melodic pitch is considered, because it is likely to find many spurious "matches" with totally different rhythm, buried in inner voices or accompaniment. Many of these problems would be alleviated by considering dynamics or timbre: for example, Wessel's effect discussed above strongly suggests that cross-voice matches are better evidence of a document's relevance when similar timbres are involved than when the timbres are very different.

2.10. Polyphonic queries

In searches "by example", which almost any user might want to do, queries will generally be polyphonic, just like the music that is sought.

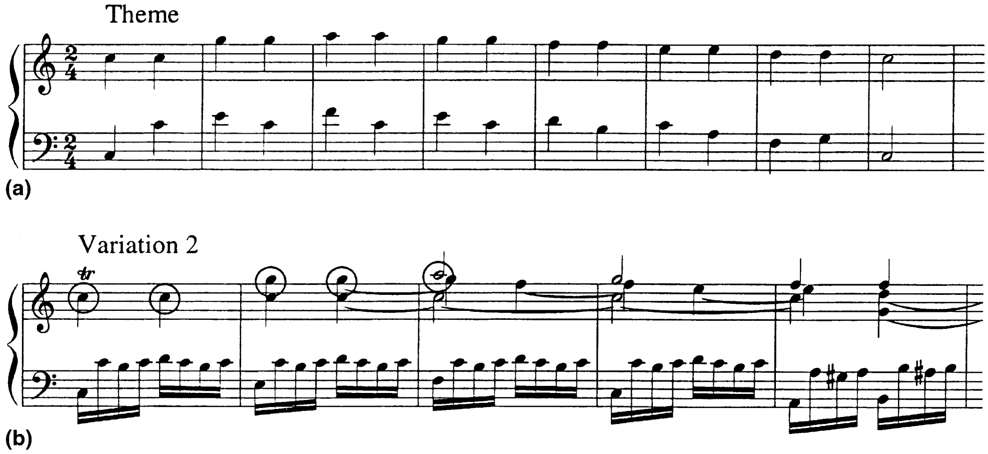

Fig. 11. Mozart: Variations on "Ah, vous dirais-je, Maman".

A more specialized case applies only to musically trained users: music scholars and students, jazz musicians, etc. Such users will sometimes want to find instances of chords and chord progressions: of course, such queries are inherently polyphonic and require considering more than just melodic pitch. One of the present authors (Byrd) has been interested for years in finding examples of the final cadential progression of the Chopin Ballade in F-minor, Op. 52, that have the same soprano line as Chopin's.

2.11. Calling the question

We can now return to our question: will melodic pitch alone be adequate for large databases and complex music? It seems very likely that it can be answered in the negative: melodic pitch will not be adequate for anywhere near all users and situations, and even melodic and harmonic pitch together will often fail for searching larger databases and/or more complex music. The obvious way to improve results is to match on duration patterns as well as pitch (Smith, McNab, & Witten, 1998). Note that duration matching can and probably should be as flexible as pitch matching: in our own research, we have implemented matching on "duration contour" in a way that is exactly analogous to pitch contour (Byrd, 2001). In some situations, it should also help to match relative loudness and/or timbre. And even if matching on loudness is not explicitly required, loudness and thickness of texture should probably be considered as affecting salience, and used to adjust ranking of search results. Dovey and Crawford (1999) discuss several factors they feel should be considered in relevance ranking, including salience.[5]

2.12. Special cases and sidestepping the issues

It might be argued that, in one particular case, we can completely sidestep all of these issues and rely on techniques that are not even specific to music. That case is where both the query and the database are in audio form, and the query is an actual performance of the exact music desired (perhaps from a CD in the user's possession). But audio signals contain so much extraneous information related to room acoustics and microphone placement as well as to fine details of performance that, even in this case, the problem may be intractable. Foote (2000) suggests otherwise, and his ARTHUR system showed good results with orchestral music; but he tested it with "an extremely modest corpus" and cautions that his approach may not scale well.

A situation that seems clearly to be manageable with pure audio techniques is the even more specific case of identifying different recordings, or different versions of one recording, of a single performance: this has been attacked, and with considerable success, by Gibson (1999). Gibson comments that his system "assumes that the [query] sample is no more than a rerecording of the original."

---

[5] Notice Dovey and Crawford's assumption of best-match rather than exact-match retrieval. The advantage of best match – that the user can look as far down the result list as they want and thereby choose the tradeoff between recall and precision they want – appears at least as important for music as it is for text. But this is necessarily speculative: as we have said, work on music-IR evaluation has hardly begun.

3. Causes: Why is music IR hard?

3.1. Segmentation and units of meaning

In a recent paper, one of the present authors wrote: "The distinction between concepts and words underlies all the difficulties of text retrieval. To satisfy the vast majority of information needs, what is important is concepts, but – until they can truly understand natural language – all computers can deal with is words." (Byrd & Podorozhny, 2000, p. 4) To put it differently, in text, there are many ways to say the same thing, and users cannot possibly be aware of all the ways when they formulate their queries. Therefore, it is important for an IR system to conflate variants of the same word. It is also important to conflate different but (in the context of the user's information need, and in a statistical sense) synonymous words. If a system does not do both, recall will suffer. [6] (This is admittedly oversimplified: very often a concept is represented not by a word but by a noun phrase or something even more complex. But that does not affect our point.)

Thus, a basic requirement of text IR is conflating units of meaning, normally words. On the other hand, the conflation must be done judiciously or precision will suffer. Note that this principle holds regardless of the retrieval model, be it exact-match or best-match, and regardless of whether term matching or language modelling is used.

Essentially the same principle applies to music in any of our three representations. In music as in text, there are many ways to "say" the same thing (see the list of "Objective" matching problems in Section 4), and again, a user cannot be aware of all. But it is not clear that music *has* units of meaning: a music "word list", i.e., a dictionary of musical symbol sequences without definitions, is very difficult to imagine, and a music dictionary *with* definitions is even harder to imagine. There is simply no predictable association of musical entities with meanings. [7] And even if music has "words", in many cases, experts will not agree on where the boundaries are.

Segmenting English into words is relatively easy: a rather good first-approximation method is just to look for white space or punctuation marks. In Chinese, among other languages, words have no explicit delimiters, so segmentation is much more difficult; nonetheless, experts generally agree on where word boundaries are (D. Moser, personal communication, July, 1999), and algorithmic solutions have been reasonably successful (Ponte & Croft, 1996). In music, however, experts do not generally agree on segmentation except in unusually clear-cut cases – barlines are entirely useless for this, and rests are of limited help, even when they occur – and automatic segmentation even of monophonic music (e.g., Cambouropoulos, 1998) is at an early stage. It is clear that segmentation in music is vastly more difficult than in Chinese. [8]

---

[6] Blair and Maron (1985) make a similar argument very effectively. The main difference is that they assume exact-match evaluation, and this and other questionable assumptions lead them to far too sweeping conclusions. But, for example, their description of a concerted attempt to find all references in a large database to a certain concept is extremely thought-provoking.

[7] A few musical techniques do have conventional associations with emotional states: the use of the minor mode to express "sadness", for example. But such associations are, notwithstanding Cooke (1959), notoriously unreliable and inconsistent.

[8] Byrd (1984,pp. 49–55) compares the difficulty of formatting in Chinese, mathematics, and music notation, and argues that music is the most difficult. The situation with respect to segmentation exactly parallels that for formatting, and for similar reasons.

In fact, it can be argued that *overlapping* segments (perhaps "motives" or "phrases") are common, even within voices. Of course, music in event format may not have complete voicing information, while music in audio form will have no voicing information at all, and one cannot even begin to look for boundaries within a voice without knowing what events are in the voice. Selfridge-Field's confounds aggravate the situation further: they mean that conflating even fragments that are obviously closely related, the way stemming and case folding in text conflate closely related strings, requires considerably more sophistication than with text.

Overlapping segments are certainly common in musical texture as a whole, and the problem is far worse when polyphonic music is taken into account. By the very nature of the independence of voices in polyphony, it is always possible for phrases or motives to overlap in different voices; in fact, the technique of counterpoint to a large extent depends on this. For example, the only remotely clear divisions in the first page and a half of J.S. Bach's "St. Anne" Fugue, BWV 552, are at measure 21 and possibly measure 11 (marked "A" and "B" in Fig. 12).

When full voicing information is explicitly present, the problem might be sidestepped by treating each voice as an independent monophonic string, but in most cases of music in event format, and all cases of audio recordings, it will be extremely difficult to disentangle these overlappings.

3.2. Polyphony

Downie (1999) speculates that "polyphony will prove to be the most intractable problem [in music IR]." We would put it a bit differently, namely that polyphony will prove to be the *source* of the most intractable problems.

Polyphonic – that, is, most – music involves simultaneous independent voices, something like characters in a play. Ordinarily, of course, only one character in a play is active (speaks) at a time, and when more than one does speak at a time, the (temporal) relationship between them is defined in the simplest possible way. Exceptions are such 20th century works as Caryl Churchill's *Top Girls* (1982) (Fig. 13, from Act 1, Scene 1; font changes added for clarity). However, most music is much more complex than this: see Fig. 12, from J. S. Bach's "St. Anne" Fugue. An obvious reason is that complex parallelism is greatly facilitated by sophisticated rhythmic notation, which text lacks: Churchill's notation of asterisks and slashes is adequate for her purposes but very limited.

We have already pointed out the necessity of cross-voice matching in unvoiced polyphonic music. Of course, without the multiple voices polyphony involves, the problems of cross-voice matching would not exist. A less obvious consequence of multiple voices is the issue of salience. Salience is essentially ignored by all text-IR systems we are aware of. [9] But without some consideration of at least the audibility of likely matches in their context in a polyphonic score, the risk of being overwhelmed by false matches is quite serious. Early experiments have been made with a simple cross-voice musical-matching algorithm suitable for unvoiced polyphonic scores, using an extract from the first movement of Beethoven's *Eroica* Symphony (Dovey,

---

[9] An obvious analogue in text IR might be to take into account a formatted document's typography, so that for example text styled as "bold", "emphasis", "strong", or "italic" is assigned a higher weight than plain text.

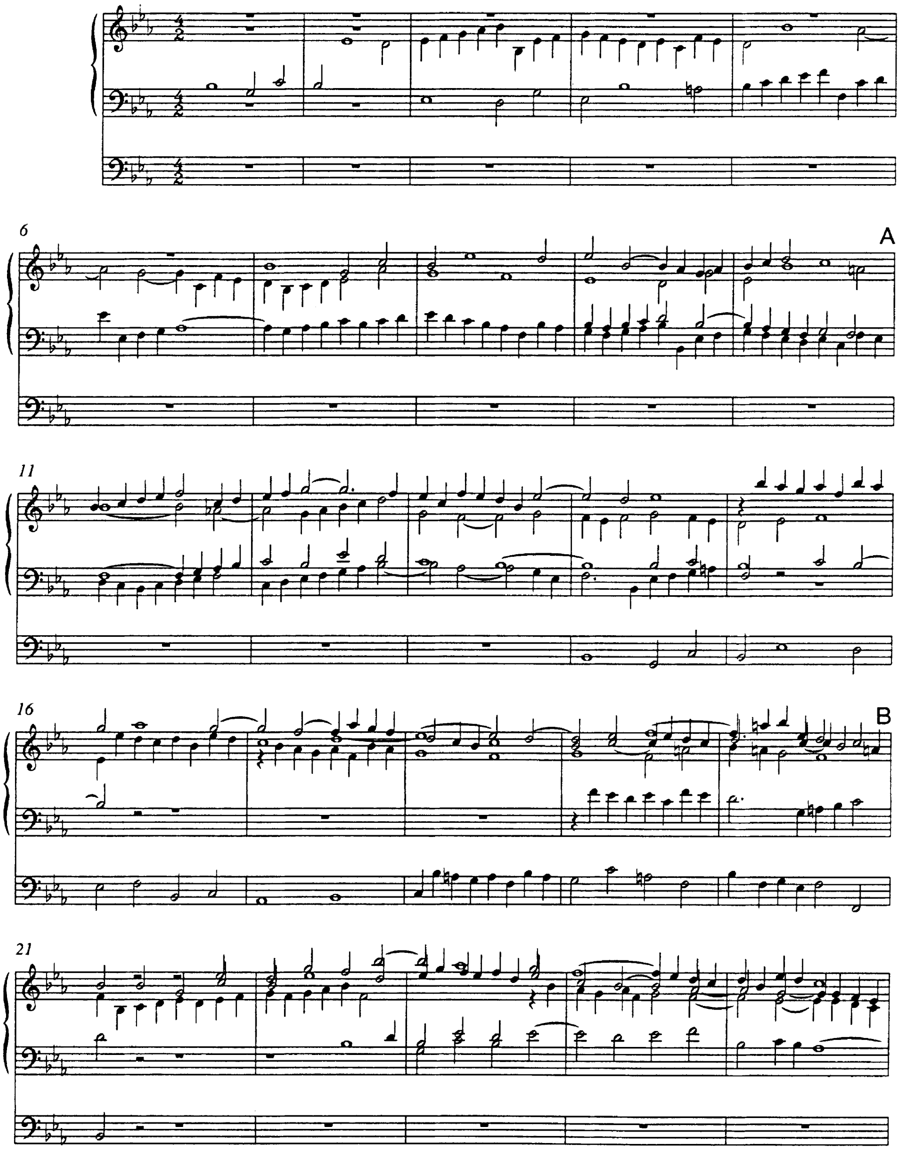

Fig. 12. Bach: “St. Anne” Fugue, BWV 552.

1999). It was found that a very recognizable woodwind phrase (Fig. 14(a)) which appeared audibly only once in the extract occurred 92 times buried within a passage of repeated chords that happened to contain the nine notes in the correct sequence (Fig. 14(b))! This already seems disastrous in terms of precision, but consider that this is a case of exact matching of the note sequence; allowing common musical transformations would have damaged precision to an even greater extent. (In this case the “real” match has the woodwind phrase as the highest sounding note throughout; this fact certainly contributes to its salience, but it often happens that the

Text of the play:

MARLENE. What I fancy is a rare steak. Gret?

ISABELLA. *I am of course a member of the / Church of England.**

GRET. Potatoes.

MARLENE. *I haven't been to church for years. / I like Christmas carols.

ISABELLA. *Good works matter more than church attendance.*

Performance (time goes from left to right):

M: What I fancy is a rare steak. Gret? I haven't...

I: *I am of course a member of the Church of England.*

G: Potatoes.

Fig. 13. Churchill: *Top Girls*, Act 1, Scene 1.

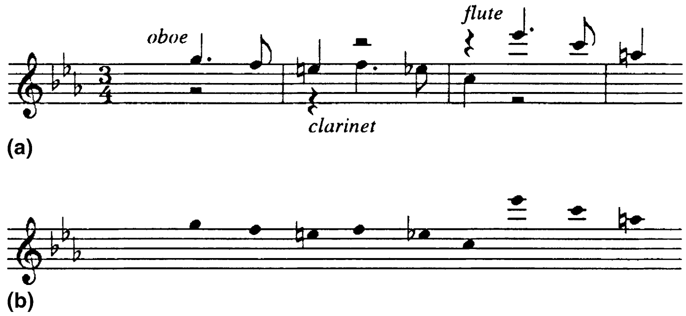

Fig. 14. Beethoven: Symphony no. 3, I, woodwinds.

highest notes are not very salient. Even piccolos, the highest instruments in the orchestra, sometimes play accompaniment.)

3.3. *Efficiency*

With music as with text, acceptable efficiency requires an approach other than sequential searching (this applies to all three representations of music). On a useful-size collection, indexing via inverted lists – the standard solution – is undoubtedly thousands of times faster.

In monophonic music, matching on one of our four parameters at a time, indexing is not too hard. In fact, Downie (1999) adapted a standard text-IR system to music, using n-grams as words and ignoring the units-of-meaning question; the results with a database of 10,000 folksongs were quite good. But, as we have observed, 10,000 monophonic songs are not a lot of music, and polyphony makes things much more difficult, especially for matching on more than one parameter at a time (pitch and rhythm being the obvious combination). A recent paper (Lee & Chen, 2000) compares several approaches to indexing monophonic music; at least one seems adequate for demanding situations in terms of both scalability and flexibility, but it is not at all clear how to adapt this work to polyphonic music.

It is important to bear in mind that inverted lists are not the only way, and may not be the best way, to avoid the efficiency disaster of sequential searching. For example, *signatures* have been studied for text IR and found to be inferior to inverted lists in nearly all real-world situations (Witten, Moffat, & Bell, 1999, pp. 143–145); but the tradeoffs for music IR might be very different.

3.4. Recognizing notes in audio

The fundamental problem of audio music recognition (AMR) is simply separating and recognizing the notes (obviously, this applies to the audio representation only). Castan (2000) discusses stand-alone AMR systems, which nearly always output MIDI files; he comments "There is no such thing as a good conversion from audio to MIDI. And not at all with a single mouse click." He concentrates on programs that are actually available, most of them commercial; among those he lists are no less than four that claim to handle polyphony. For research on AMR, see Sterian, Simoni, and Wakefield (1999), Martin and Scheirer (1997), and Walmsley (1999).

Difficulties of AMR include "masking", which leads to notes being missed, and the fact that every musical note consists of many partials, which leads to non-existent notes being found; these difficulties increase very rapidly with the number of notes actually present simultaneously. The Web site for one commercial system comments that music-recognition systems "work with an exactitude [sic] of 70–80% but only for single-voice melody. For polyphonic music the exactitude is even lower. The variety of musical timbres, harmonic constructions and transitions is so great that, for example, there will be no computational capabilities of all computers in the world to recognize [the] musical score of a symphonic orchestra." (AKoff, 2000)

Notice that for query input, monophonic AMR is quite helpful, e.g., to let users hum or whistle queries, and several existing music-IR systems – for example, the early system of Ghias, Logan, Chamberlin, and Smith (1995) and the recent MELDEX (Bainbridge et al., 1999) – support audio queries. For databases, monophonic AMR will rarely be helpful.

3.5. User interfaces

The general topic of user interfaces for music IR deserves an entire paper of its own. We simply note that good user interfaces for music are extremely challenging to develop, even for the apparently routine task of musical score editing and printing (Byrd, 1984, 1994), and very few of these problems can be disregarded for music notation-format query interfaces and result displays. For audio or MIDI, the problems are easier in some ways, but harder in others: if a system cannot show content in a result list graphically, it may take a user a very long time to choose among, say, 100 proposed matches.

4. Symptoms: problems matching musical data

4.1. Query "quality control"

Search queries in a music-IR system might be constructed using a variety of input method. These may include: direct manual coding; translation from score-notation files; MIDI-keyboard performances; manual editing within a graphical or textual search dialog; or even whistling, humming, or singing into a microphone. The important thing is that each input method is subject to its own characteristic errors. Assuming the user is competent to use the method, these errors might be caused by imperfect specification of a query (possibly due to over-simplification or to "false memory") or by its incorrect interpretation by the search program. An MIDI keyboard cannot distinguish between enharmonic pitch spellings; with audio input, a user's performance may be inaccurate in pitch or rhythm, and the pitch-tracking system may not handle such errors correctly.

4.2. Database "quality control"

Similar comments apply to the musical databases being searched (Huron, 1988). There is, typically, very little "quality control" of publicly available musical data, and, again, there are characteristic forms of error arising from a wide range of musical ambiguities. A piece of music saved as an MIDI file may contain unexpected extra data, such as the explicit realization of trills and other ornaments which would simply be represented by signs in score notation. There may be errors which have escaped an editing or data-checking process (a particularly insidious kind of error is one that fits the harmonic or melodic context even though it is clearly wrong; such an error is very hard to spot in aural monitoring).

On the other hand, the encoding of the musical data may be perfectly accurate, but from a source that differs in some respect from the user's expectations. On a trivial level, a piece familiar to the user from a recording in one key may be encoded from an edition in another; at a more subtle level, certain performance-related characteristics which are the subject of the performer's personal choice (e.g. the complex of time-based performance choices classed under the headings of rubato and articulation, or chord-spreading) may be encoded in performance-based data in a manner that conflicts with the user's expectations based on the appearance of a printed score. Furthermore, by their very nature, "performances" of a musical work (in any style or genre) are *inherently* diverse and divergent from their model: the number of possible ways of performing any one work is enormous.

Assuming that an identical musical score is being used, performance A of a given work may take longer overall than performance B, yet some segments of A may be done *faster* than in B; groups of notes (chords) that are sounded simultaneously in A may appear in close succession (spread) in B; partially specified items in the score (such as grace notes, or ornaments like trills) may be interpreted differently in the two scores, with the result that any two performances of a given score will probably contain different numbers of sounding notes.

All the examples given here are within the bounds of "accurate" performance of the music: neither is less "correct" than the other.

4.3. Implications and a catalog of problems

In our discussion of Section 3.1, we commented that "in music as in text, there are many ways to 'say' the same thing." The identities of musical entities are stubbornly resistant to certain types of transformation. Simple examples include mutation (roughly, changing from minor to major or vice-versa); diatonic “transposition” (really scale-degree shifts); “tonal answers” to fugue subjects (where repetitions of the subject have pitch intervals distorted to stay within the scale); and varying the number of repetitions of a note. More complex examples include a myriad ways of ornamenting melodies. This is analogous to the problem of conflating various ways of expressing the same concept in text: through variants of the same words, synonymous words and phrases. These considerations mean that searching for exact matches is of no more use – and quite possibly less – in music than in text IR.

Appendix B contains a first attempt at a catalog of the problems.

4.4. Prospects for solutions

Huron (1988) gives a list of and a thoughtful discussion of “error categories” for music databases that applies to our type 9 and, to a lesser extent, to all of our “subjective” types.

All of the problems we have listed are common now. But how good are the prospects for solving them, one way or another? Objective problems are inherent in music, so they will certainly remain common. Subjective problems and mistakes by user result from human nature, so they also will remain common. Outright mistakes from conversion are common now in optical music recognition (OMR), and much more in AMR, systems. As technology improves, they may become less common in OMR. But we must assume they will remain common in AMR, at least for many years to come: one expert commented that AMR is “orders of magnitude more difficult” than OMR (C. Raphael, personal communication, September 1999).

To sum up, we can expect most, if not all, of these problems to be with us for the foreseeable future.

5. Conclusions

In a paper like this, summarizing the challenges of a significant new area of technology, the only “conclusions” we can offer are suggestions for future research.

5.1. User-interface issues

In recent years, text-IR researchers have tried to leverage user-interface techniques first applied in database systems to overcome the difficulty of achieving high precision and high recall simultaneously; results are very promising. The idea is summarized in Shneiderman’s “Visual information seeking mantra”: “Overview first, zoom and filter, then details on demand” (North, Shneiderman, & Plaisant, 1996). For music-IR, a list of scores might be presented with user control over relative ranking according to the criteria, preferably using Shneiderman’s (1994) dynamic-queries techniques, e.g., with sliders controlling relative weights and the display reacting interactively. (It would be better to use real dynamic queries instead of just dynamic ordering of the results of a static query, but that would also impose much greater computation and data-transfer demands.)

5.2. Units of meaning revisited

Even on the level of individual instances of a musical motif or theme within a work, repeated occurrences are rarely identical; musical entities are recognizable even when they objectively differ quite significantly. *If a musical entity is recognizable, it is likely to be the subject of a search query.* Therefore, more attention needs to be paid to the work of music psychologists and researchers in music cognition, especially into musical recognition and memory.

It is generally recognized that partial and approximate matching is a sine qua non for successful music IR: see Crawford, Iliopoulos, and Raman (1998) and Smith et al. (1998), and Section 4, above. Specialized string-matching techniques, such as those sometimes used in text IR to recognize words unusually or incorrectly spelt, have been successfully applied to monophonic music IR (see, e.g., Downie, 1999), but – as usual – the problem is much more difficult for polyphony.

5.3. Scale and performance

As we have seen, with sequential searching, musical-similarity matches in useful-size polyphonic databases are likely to be unacceptably slow. Obviously, we need to develop polyphonic indexing (or signature-based) methods; research like Lee and Chen (2000) is just beginning to show how this might be done.

5.4. Relevance and music

It is not at all clear that the standard IR evaluation model is valid for music. “Information”, the explicit goal of conventional IR, has an unquestioned correspondence (albeit complex and ill-defined) with the concepts expressed in words in a query. The notion of “relevance”, on which standard IR strategies depend, is bound up with the relations between concepts in a way that has little or no parallel in music. The question of whether relevance is the proper goal even for text IR has received much attention in recent years: see for example the discussion of “topicality” vs. “utility” in Blair (1996).

Acknowledgements

Besides commenting on drafts of this paper, as participants in the OMRAS project, Matthew Dovey and Jeremy Pickens made many contributions to the thinking that underlies the paper. Alexandra Uitdenbogerd and the anonymous referees also made helpful comments.

Appendix A. Melodic confounds

The term “melodic confounds” is due to Selfridge-Field (1998).

In the statistics below, rests are counted only if internal (not at the very beginning or end). Repeated notes in the “musical” sense exclude cases like appogiaturas reiterated across the barline, or where there are intervening rests.

1. Barlow and Morgenstern’s (1948) *Dictionary of Musical Themes* contains incipits of a few measures each for about 10,000 themes of classical-tradition instrumental pieces. We checked 400 themes (all of pages with numbers ending with 00, 20, 50, and 70).
2. The anonymous The Real Vocal Book (a “fake book”, undoubtedly crammed with blatant copyright violations; undated but c. 1980) contains melodies and chord symbols for about 225 complete pop songs. Starting with number 1, we considered every fifth song. As a rough analog of incipits, we scanned the first two systems, but ignoring pickups ending the first ending; then the first two systems of the bridge/chorus, if any, for a maximum of two themes per song. The 45 songs we considered contain 81 themes. There appear to be no grace notes, trills, or turns in the entire volume.
3. Worship in Song: A Friends Hymnal (1996) contains 335 hymns. Starting with number 5, we considered every fifth hymn: 67 in all. As an analog of incipits, in each, we scanned the first two systems. There appear to be no grace notes, trills, or turns in the entire volume.

Here are percentages of each sample containing each type of confound. Values of 37% and above are in *italics*; other values are no greater than 15%.

	B&M Dictionary	Real vocal book	Friends hymnal
Repeated notes (musical sense)	*42%*	*46%*	*75%*
Repeated notes (other)	11%	7.5%	12%
Rests	*37%*	*48%*	13%
Grace notes	15%	0	0
Trills and turns	7.5%	0	0

Appendix B. Preliminary catalog of problems

The list below is a first attempt at categorizing the problems of music IR. “M” means the problem applies to monophonic music; “P” means it applies to polyphonic music. (Notice however that, in view of the music-perception phenomena we discussed earlier, even this distinction is not clear-cut: a single monophonic line is sometimes heard as polyphonic.)

B.1. Objective

M, P: 1. Replacement (note-for-note). Cases include tonal answer (Bach: “St. Anne” (Fig. 12) and very many other fugues); mutation; diatonic transposition.

M: 2a. Melodic ornamentation, simple: subdividing a note into repeated notes (Mongeau & Sankoff, 1990, call this “fragmentation”). Example: Beethoven: Symphony no. 9, IV, main

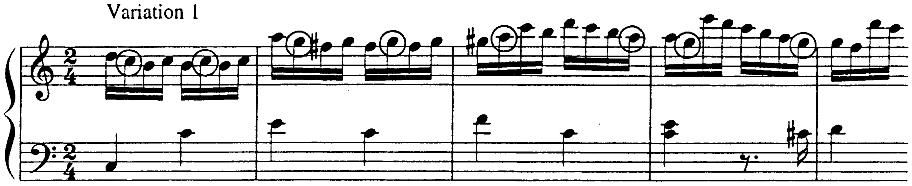

Fig. 15. Mozart: variations on "Ah, vous dirais-je, Maman".

theme (the "Ode to Joy") appears first beginning with a half note followed by two quarters; but in most subsequent appearances, including those that are most salient, the first note is subdivided into two (Fig. 9).

M: 2b. Melodic ornamentation, complex: insertion. Example: Mozart's Variations K.265, Variation 1 (Fig. 15; this is essentially the "submerged" melody of Selfridge-Field, 1998).

M: 3a. Melodic simplification, simple: combining repeated notes into a single note (Mongeau & Sankoff, 1990, call this "consolidation").

M: 3b. Melodic simplification, complex: deletion. As compositional devices, 3a and 3b are inverses to 2a and 2b, but as far as music-IR is concerned, they are effectively identical. If the query is the version of the "Ode to Joy" theme that starts with a half note and the document includes only the repeated-quarter-notes version, the problem is fragmentation; if the query is the repeated-quarter-notes version and the document includes only the half-note version, it is consolidation.

P: 4. Melody crossing voices. Example: in Mozart's Variations K.265, Variation 2 (Fig. 11). (Selfridge-Field, 1998, calls this "roving"; Crawford et al., 1998, calls it "distributed matching".)

M, P: 5. "Linear" transformations of the entire query. Cases include transposition and time scaling (i.e., augmentation/diminution).

B.2. Subjective: errors of perception, cognition and memory

See the discussion of "Human performance in melody recall" in McNab et al. (1996), as well as general discussion of perceptual, cognitive, and memory errors in Uitdenbogerd and Zobel (1998).

M: 6. The version remembered is simplified.

P: 7. The version remembered mixes voices, typically melody and accompaniment.

M, P: 8. The version remembered is incorrect in some other way.

B.3. Other: discrepancies in input

M, P: 9. Outright mistakes in query and/or database. These may be automatic (from the conversion process) or manual (by the user).

M, P: 10. Performance-related issues. These are affected by style/genre (e.g., "swing" in jazz, the similar "notes inégales" in baroque music, chord-spreading and -breaking in piano, guitar, and flute music).

References

AKoff Sound Labs. (2000). What is music recognition? Retrieved January 31, 2001, from the World Wide Web: http://www.akoff.com/about.html.

Bainbridge, D., Nevill-Manning, C., Witten, I., Smith, L., & McNab, R. (1999). Towards a digital library of popular music. In *Proceedings of digital libraries '99 conference*. New York: Association for Computing Machinery.

Barlow, H., & Morgenstern, S. (1948). *A dictionary of musical themes*. New York: Crown.

Blair, D., & Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document-retrieval system. *Communications of the ACM*, *28*(3), 289–299.

Blair, D. (1996). STAIRS redux: Thoughts on the STAIRS evaluation, ten years after. *Journal of the American Society for Information Science*, *47*, 4–22.

Boltz, M. (1999). The processing of melodic and temporal information: Independent or unified dimensions? *Journal of New Music Research*, *28*(1), 67–79.

Byrd, D. (1984). *Music notation by computer.* Doctoral dissertation, Indiana University, Ann Arbor, Michigan, UMI order no. 8506091.

Byrd, D. (1994). Music notation software and intelligence. *Computer Music Journal*, *18*(1), 17–20.

Byrd, D., & Podorozhny, R. (2000). *Adding Boolean-quality control to best-match searching via an improved user interface*. Technical Report IR-210. Computer Science Department, University of Massachusetts, Amherst.

Byrd, D. (2001). *Music-notation searching and digital libraries*. Technical Report IR-220. Computer Science Department, University of Massachusetts, Amherst. Also accepted by Joint Conference on Digital Libraries (JCDL, 2001).

Cambouropoulos, E. (1998). Musical parallelism and melodic segmentation. In *Proceedings of the XII Colloquio di Informatica Musicale, Crorizia, Italy*.

Castan, G. (2000). Converting audio to MIDI. Retrieved January 31, 2001, from the World Wide Web: http://www.s-line.de/homepages/gerd_castan/compmus/audio2midie.html.

Chen, A.L.P., & Chen, J.C.C. (1998). Query by rhythm: An approach for song retrieval in music databases. In *Proceedings of the Institute of Electrical and Electronic Engineers eighth international workshop on research issues in data engineering: Continuous-media databases and applications (RIDE)* (pp. 139–146).

Churchill, C. (1982). *Top girls*. London: Methuen.

Cooke, D. (1959). *The language of music*. Oxford, UK: Oxford University Press.

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

Deutsch, D. (1972). Octave generalization and tune recognition. *Perception and Psychophysics*, *11*(6), 411–412.

Dovey, M. (1999). A matrix based algorithm for locating polyphonic phrases within a polyphonic musical piece. In *Proceedings of AISB '99 symposium on artificial intelligence and musical creativity*. Edinburgh, Scotland: Society for the Study of Artificial Intelligence and Simulation of Behaviour.

Dovey, M., & Crawford, T. (1999). Heuristic models of relevance ranking in searching polyphonic music. In *Proceedings of Diderot forum on mathematics and music* (pp. 111–123). Vienna, Austria.

Downie, J. S. (1999). *Evaluating a simple approach to music information retrieval: Conceiving melodic n-grams as text*. Doctoral dissertation, University of Western Ontario.

Downie, J.S., & Nelson, M. (2000). Evaluation of a simple and effective music information retrieval system. In *Proceedings of ACM SIGIR conference on research and development in information retrieval*. New York: Association for Computing Machinery.

Foote, J. (2000). ARTHUR: Retrieving orchestral music by long-term structure. Read at the first international symposium on music information retrieval. Retrieved January 31, 2001, from the World Wide Web: http://ciir.cs.umass.edu/music2000.

Ghias, A., Logan, J., Chamberlin, D., & Smith, B. C. (1995). Query by humming: Musical information retrieval in an audio database. In *Proceedings of ACM international conference on multimedia*. Retrieved February 20, 2001, from the World Wide Web: http://www.acm.org/pubs/articles/proceedings/multimedia/217279/p231-ghias/p231-ghias.html.

Gibson, D. (1999). *Name that clip: Music retrieval using audio clips*. Presentation at SIGIR 1999 Workshop on Music Information Retrieval. Abstract retrieved January 31, 2001, from the World Wide Web: http://www.cs.berkeley.edu/~dag/NameThatClip.

HARMONICA (1999). Accompanying action on music information in libraries: HARMONICA. retrieved January 31, 2001, from the World Wide Web: http://www.svb.nl/project/harmonica/harmonica.html.

Huron, D. (1988). Error categories, detection, and reduction in a musical database. *Computers and the Humanities*, *22*, 253–264.

ISMIR (2000). MUSIC IR 2000: International symposium on music information retrieval. Retrieved February 28, 2001, from the World Wide Web: http://ciir.cs.umass.edu/music2000.

Kassler, M. (1966). Toward musical information retrieval. *Perspectives of New Music*, *4*(2), 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.

Lee, W., & Chen, A.L.P. (2000). Efficient multi-feature index structures for music data retrieval. In *Proceedings of SPIE conference on storage and retrieval for image and video databases* (pp. 177–188).

Lemström, K., Laine, P., & Perttu, S. (1999). Using relative interval slope in music information retrieval. In *Proceedings of the 1999 international computer music conference* (pp. 317–320).

Martin, K. D., & Scheirer, E. D. (1997). Automatic transcription of simple polyphonic music: Integrating musical knowledge. Retrieved March 5, 2001, from the World Wide Web: http://sound.media.mit.edu/Papers/kdm-smpc97.txt.

McAdams, S., & Bregman, A. (1979). Hearing musical streams. In Roads, C., & Strawn, J. (Eds.), *Foundations of computer music* (pp. 658–698). Cambridge, MA: MIT Press (1985). (reprint).

McNab, R., Smith, S., Witten, I., Henderson, C., & Cunningham, S.J. (1996). Towards the digital music library: tune retrieval from acoustic input. In *Proceedings of digital libraries '96 conference*. New York: Association for Computing Machinery.

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

North, C., Shneiderman, B., & Plaisant, C. (1996). User controlled overviews of an image library: A case study of the visible human. *Proceedings of digital libraries '96 conference*. New York: Association for Computing Machinery.

OMRAS (2000). Online music recognition and searching. Retrieved January 31, 2001, from the World Wide Web: http://www.omras.org.

Pierce, J.R. (1992). *The science of musical sound* (revised ed.). New York: W.H. Freeman.

Ponte, J., & Croft, W.B. (1996). *Useg: A retargetable word segmentation procedure for information retrieval.* Technical Report IR-75. Computer Science Department, University of Massachusetts, Amherst.

The Real Vocal Book (n.d.). Title page lists as publisher "Real Vocal Book Press".

Selfridge-Field, E. (1998). Conceptual and representational issues in melodic comparison. In W. Hewlett and E. Selfridge-Field (Eds.), Melodic similarity: Concepts, procedures, and applications (Computing in Musicology 11, pp. 3–64). Cambridge, MA: MIT Press.

Shneiderman, B. (1994). Dynamic queries for visual information seeking. *IEEE Software*, *11*(6), 70–77.

Smith, L. A., McNab, R. J., & Witten, I. H. (1998). Sequence-based melodic comparison: A dynamic programming approach. In Hewlett & Selfridge-Field.

Sparck Jones, K., & Willett, P. (Eds.). (1997). *Readings in information retrieval*. San Francisco: Morgan Kaufmann.

Sterian, A., Simoni, M. H., & Wakefield, G. H. (1999). Model-based musical transcription. In *Proceedings of the 1999 international computer music conference* (Beijing, China). Retrieved January 31, 2001, from the World Wide Web: http://musen.engin.umich.edu/papers/transcription.pdf.

Tseng, Y.-H. (1999). Content-based retrieval for music collections. In *Proceedings of ACM SIGIR conference on research and development in information retrieval* (pp. 176–182). New York: Association for Computing Machinery.

Uitdenbogerd, A.L., & Zobel, J. (1998). Manipulation of music for melody matching. In *Proceedings of ACM international conference on multimedia* (pp. 235–240). New York: Association for Computing Machinery.

Uitdenbogerd, A.L., Chattaraj, A., & Zobel, J. (in press). Music information retrieval: past, present and future. In D. Byrd, J. S. Downie, & T. Crawford (Eds.), *Current research in music information retrieval.* Boston: Kluwer Academic Publishers.

Walmsley, P. (1999). Bayesian graphical models for polyphonic pitch tracking. In *Proceedings of diderot forum on mathematics and music*, Vienna, Austria. Retrieved January 31, 2001, from the World Wide Web: http://www-sigproc.eng.cam.ac.uk/~pjw42/ftp/didlt.pdf.

Wiseman, N., Rusbridge, C., & Griffin, S. (1999). The Joint NSF/JISC International Digital Libraries Initiative. *D-Lib Magazine* 5(6). Retrieved January 31, 2001, from the World Wide Web: http://www.dlib.org.

Witten, I., Moffat, A., & Bell, T. (1999). *Managing gigabytes* (2nd ed.). San Francisco: Morgan Kaufmann.

Worship in song: A friends hymnal (1996). Philadelphia: Friends General Conference.