5 Towards a Measure of Cognitive Distance in Melodic Similarity

Eleanor Selfridge-Field

Center for Computer Assisted Research in the Humanities
Stanford University
Stanford, CA 94305-3076, USA
esfield@stanford.edu

Abstract

We propose a still preliminary cognitive-distance metric to rank possible melodic matches. Its primary purpose is to discriminate between literal matches and psychologically valid matches in searches of large symbolically encoded datasets. Cognitive weightings are based selectively on studies in several disciplines related to music. Examples are drawn from both "intentional" searches (those collected manually by musicologists and ethnomusicologists) and those culled in computer "automatic" searches (those collected by computer) of the same melodic prototypes. Two variants of a provisional scoring system are considered.

Music Query: Methods, Models, and User Studies (Computing in Musicology 13).
Published by CCARH and the MIT Press, 2003.

5.1. Melodic Similarity and Feature Prioritization

[1]This paper is an adapted and expanded version of my talk "Accidental and Intentional Melodic Similarity: Towards a Measure of Cognitive Distance," *International Symposium of Music Information Retrieval (ISMIR), Paris, October, 2002.*

The notion of similarity[1] (music excluded) is problematical in ways which arguably make it ill-suited to computer query. For tangible objects with such defining characteristics as shape, color, size, volume, texture, weight, and so forth, it is obvious that within a group of non-identical specimens, the number of combinatorial possibilities for matching one or more properties among all the pairs that can be formed is considerable, and that the feature categories themselves are of an arbitrary degree of complexity. Sound objects too contain an arbitrary number of properties (pitch, pitch fluctuation [ornamentation], duration, articulation, dynamic level, and so forth), many of which can be further refined. The ranking of these properties in relation to the establishment of discrete melodic identity has not been established. It seems fair to say that in similarity studies generally there is no obvious or inherently systematic way to rank objects by degrees, although that is what potential matches expressed as probabilities of an exact match purport to report.

The phrase "melodic similarity" harbors a large array of goals which are divergent and depend on one's purpose. I shall cite four, although many others exist. (1) Musicians, like ordinary listeners, consider melodic similarity to be a concept which enables identification and location of a specific piece of music. Every potential match is correct or incorrect; judging degrees of similarity is usually irrelevant to the goal. (2) Publishers and copyright registrars want to know whether the same "essential" work has been printed or recorded before. The details need not be exactly the same in order for a violation of intellectual-property rights to occur. (3) Musicologists may be seeking to establish clusters of related melodies. These scholars may want to know how musical (or musico-social) identities were formed, preserved, and varied. (4) Music recommendation services, however, want to persuade customers who bought one product that a "similar" product has sufficient likeness to be worthy of their consideration. For them alone a list of probabilities in declining order lies on the direct path to the goal.

It falls to the community of music-query software developers to serve all of these needs. Even at this early date, there is much specialization within the field and little uniformity in the assumed definitions that inhere in these different realms of application. To date there has been little interaction between these developers and researchers in music cognition, where related research has frequently focused on Gestalt principles of melody (pattern completion, extension, expectation) and what is, in a sense, its obverse: segmentation and its tiers of generalization, with occasional consideration given to melodic associations with lyrics, with emotions, and sometimes with reference to specific qualities of memory and attention. Ultimately, music cognition is concerned with minds rather than music, but by offering important clues to the salient principles of melodic perception it identifies possibilities which deserve acknowledgment.

5.2. Towards a Cognitive-Distance Metric

5.2.1. Modeling Melodic Reduction: An Example

The human phenomena we associate with melody (hearing, recognizing, recalling, or performing a melody) exist within a paradoxical universe: the process of realization is necessarily sequential because it happens in time, while the process of "imaging" appears to be iconic and therefore static in time. What gives melodies their theoretical "shape"? We might presume that salient features of a melody, individually perceived and prioritized, play a role in the formation of these configurations.

The purpose of a cognitive-distance metric is to serve all sides of this relationship, to mediate between the experiential reality and the distilled idea. This implies mediation between time and non-time. On the real-time track, we begin by assigning an ordinal value from which the linear position of every sound event can be tracked. To get to the non-time (ideal) track, which retains a sense of time but does not operate in the same temporal space as sounding material, a qualitative weight which takes account of features of local and global importance within the melody is added.

This aim raises the question of note-functions within a melody. On the one hand, any single note in a melody contributes only its relationships, not its pitch content, to the melody as a whole. On the other hand, there are no firm rubrics for determining what the intermediate containers of musical meaning are, despite the fact that numerous models for melodic segmentation have been proposed (Lerdahl and Jackendoff 1983, Halperin 1990, Cambouropoulos 1999, Singer 2003). This chameleonic quality of melody is undoubtedly part of its charm.

Our sense of the simplicity of melody is imposed effortlessly by our minds, for the details of melodic fabric are frustratingly complex (see, *inter alia*, Selfridge-Field 1998). This discloses another paradox: when we try to model the cognitive process that makes melodies seems simple, we discover that the level of discarded detail varies from person to person and can move along the continuum of complexity—simplicity within a single melody. Apropos of this potential for shifting levels of reductions, Figure 5.1 shows a short example from a Bach minuet with three possible reductions.

Figure 5.1. Bach keyboard minuet in Bb Major (right hand on top staff) with three levels of melodic simplification: (a, b) mixed levels of reduction per bar, (c) scale reduction.

In Figure 5.1 Reduction (a) preserves the event-rhythm of Bars 1 and 3 but reduces that of Bar 2. Reduction (b) is an inversion of (a) in that it reduces the activity in Bars 1 and 3 by two levels but that of Bar 2 by only one level. Reduction (c) opts for the least amount of melodic activity required by implied harmonic change. [Obviously, many other reductions are possible.] There is no experimental evidence to suggest that any of these reductions or more or less valid than the others. There are faint indicators that as a collection they represent reductions that could each occur to some individuals. [Some might be more or less likely in different listening contexts, a subject which will not be discussed here.]

That harmonic implication plays some role in the cognitive dynamics of tonal melodies seems to be beyond dispute, even if the details require further investigation. Several studies call attention to the importance of harmonic implication in melodic recognition. Holleran and Jones (1995) suggest its primacy at a general level, while studies by Cuddy (1993) and Dowling (1994) underscore its relevance at a more specific level. Harmonic modeling by Sapp (2004) suggests that divergent harmonic inferences for identical musical passages operate systematically but (vis-à-vis Krumhansl 1990: 99–102) reflect distinctly different tiers of temporal resolution.

Harmonic implication can be seen to play an invisible role in the processes represented by Reductions (a) and (b). Reduction (c) is too shapeless to facilitate the modeling of melodic cognitive processes. A great many melodies of both classical and folk music can be reduced to scales by one means or another. Rhythmic patterning and pitch contour are essential to a fundamental sense of distinct identity.

Adjudicating between Reductions (a) and (b) is a more difficult task. A perfunctory search of the *Themefinder* database (for holdings see Sapp, Liu, and Selfridge-Field 2004), here using *c.* 40,000 incipits of European classical and folk music, culls only seven match-candidates (all of them imperfect) for works in a major key, in 3/4 meter, and initiated by the scale degrees 3–3–4–5. This suggests that Reduction (b) is adequately differentiated for searching purposes. It does not, however, indicate that (b) is more valid cognitively. Neither can one infer that melodies with similar rhythmic patterns could all be reduced as successfully with the (b) template. An implied harmonic change in Bar 2 could would call the reductions into question.

5.2.2 From Modeling to Metrics

Several efforts to create cognitively sensitive search tools for melodic representation and segmentation (Cambouropoulos 1999) have employed elements of the Lerdahl-Jackendoff (1983) hierarchical categorization model for listening. What the reductions shown in Figure 5.1 are intended to demonstrate is that perceptual feature-selection is likely to influence the results and therefore the efficiency of the resulting approaches. If one takes all three reductions and awards one point for the mere retention of each event on each tier, then the scoring shown in Table 5.1 results.

Event	**1**	2	3	4	5	6	**7**	8	9	10	11	12	**13**	14	15	16	17	18
Score	**4**	2	2	2	2	2	**3**	1	3	1	3	1	**4**	2	2	2	2	2

Table 5.1. Simple scoring method for each melodic event based on its presence in each of the reductions of Figure 5.1. Bold type indicates downbeats.

This approach avoids the difficult task of adjudicating between reductions. At the same time, each is retrievable: scores ≥ 4 produce Reduction (c), those ≥ 3 produce Reduction (b), those ≥ 2 Reduction (a).

A ubiquitous property of the Baroque repertory from which Figure 5.1 comes was its "sewing-machine" rhythms, from which buried melodies were to be excavated by the listener. Here the rhythmic figuration is monotonous. The ease with which listeners can construct something more tolerable in their minds is remarkable, but it is undetermined whether simple principles of reduction can be defined. It is premature to speak of a process whereby the "reduction" of a melody of this kind actually amounts to varying the weights of each item in its repetitive background to make the melody more definable.

While the scores of Table 5.1 contain no metrical information *per se*, they imply it. Here too there is something of a circular argument, for the coincidence of harmonic change precipitates the necessity to introduce notes which will be rendered essential to the melodic definition by most reductive approaches.

What can we learn from this example? One point is that metrical and harmonic information are indispensable elements of melodic identity. Whether they are captured implicitly or explicitly, some account must be made of them.

This simplistic approach to measurement attempts to get at the conundrum posed by Cuddy (1993), who asked [I paraphrase], "Why does a single note matter so much in one melody and so little in another?" One way of modeling a search designed to capture the notes that "matter so much" is to address the fact that some notes in a melody play multiple roles, which makes their presence essential, while others are simply present as "filler" and are hardly missed if absent. In the context of a given melody, some notes might be more salient than others because they complete a melodic leap which is out of character with the context of the surroundings, are associated with a rhythmic deviation (such as syncopation), or effect some other attention-getting feature (a change in articulation or dynamics, the presence of ornamentation, etc.). Narmour (1990) has investigated these kinds of deviations extensively in fleshing out his implication-realization model of melody. No attempt is made here to deal with this rich literature, nor with the variable weightings of particular features which are implied by different performances, nor to deal with the much-discussed question of phrase-boundary ambiguities. It is acknowledged, however, that some of the properties that may be captured in the proposed metric can be strengthened or weakened in particular performances and highlighted (or not) in refinements to this kind of approach. Further experimental investigation of all of these excluded areas is encouraged.

5.2.2. General Approach

Many authors have offered models of hierarchical weights in pursuit of a general model of melodic perception and cognition. Elements of this hypothetical approach can be found, *inter alia*, in Jones's concept of Joint Accent Structure (Ralson 1991, Jones 1993, Jones 1997), in the well-known Lerdahl/Jackendoff Generalized Theory of Tonal Music (GTTM 1983), and its extensions and refinements by Cambouropoulos for segmentation (1998) and by Temperley for music cognition (2001). Because actual implementation would require an immensely tedious set of procedures for multiple evaluations of every note in a set of melodies before search algorithms could be tested, some simpler means for representing some sense of this kind of searching are used for experimentation here.

The scheme proposed here is vastly oversimplified from a cognitive perspective, but it has the practical virtues that it can be implemented using existing data structures and algorithms and that it can therefore be tested on large databases of melodic material. The database we have used for testing is *Themefinder* (*www.themefinder.org*), on which see Sapp, Liu, and Selfridge-Field (2004).

The proposal as described here comes with several caveats:

1. **Metrical characteristics.** The criteria and methods for weighting are subject to "tuning" for specific repertories. Computer analysis of diverse repertories suggests wide variance in feature-frequency from one repertory to the next. For example, vocal music (particularly popular music) has a more restrained tessitura than instrumental music, so that octave differentiation may be more important for searching the latter than it is for the former. The attribute profiles of unmetered music (e.g., chant repertories, some Asian repertories of the present, French keyboard music of the seventeenth century) are obviously unsuited to this kind of evaluation. So too are improvised repertories which are characterized by syncopation (e.g., jazz) and those in which the use of melody is clandestine (e.g., chorale melodies in elaborate chorale preludes). As we know from the legacy of the mensural music of the fifteenth and sixteenth centuries, the arithmetic of triple and some compound meters is necessarily different from that of duple meter.

2. **Modal characteristics.** The behavior of melodies which come from works in minor keys is necessarily more complex than that for works in major keys, since ascending and descending versions of the scale are variable. This produces an expanded array of possible intervallic sizes and tonal qualities.

5.3. Feature Selection and Prioritization

Melodies have an uncalculated number of properties. We do not really know how many of them are essential to recognition. It seems likely that the number of relevant ones and, potentially, the relative weightings among them in the establishment of melodic identity may vary from one repertory to another. This possibility militates against one-size-fits-all solutions to melodic-search algorithms.

String-matching in melodic searches has thus far tended to mean iterative pitch-matching, irrespective of the labels used to describe them. While the recognizability of meter and accent (that is, beat organization) may play a strong role in song recognition and recall, temporal aspects of music are rarely considered in string-matching techniques. The most prevalent elaboration of one-dimensional string searching in musical applications has been edit distance,[2] but when applied to music it represents a categorically different set of problems from those considered here. Foreground/background distinctions, which bring the need of high relief to the representation of a linear process, are irrelevant to typing. Sophisticated varieties of edit distance have been introduced in several fields (primarily scientific) in recent years, but the results of edit-distance procedures in music retrieval are often unconvincing.

[2]The original goal of the Levenshtein edit-distance metric (1966) was automatic error-checking of a performance task (text input, in Russian) in the era of key-punches and paper tapes.

In art music, melodic "foreground" may be perceptible at different levels of temporal resolution, at variable lengths, and in multiple transformations. Melodic identity is considered to be singular even when successive presentations exhibit changes in order (inversion, transposition), key (modulation), and completeness (motivic development). Melodic manipulation has been one of the chief playgrounds of both the composers of European art music and the performers of the improvisational art music of the Middle East and the Indian subcontinent. By conferring a wide range of attributes—from internal coherence to human memorability to cultural significance—on music itself, melodic identity contributes significantly to the notorious elusiveness of "musical meaning."

Musical features which confound string searches come largely from the rhythmic domain. They include ties, rests, syncopation and, most particularly, accentuation, since it is not directly represented in musical notation or in representations of it. Experimental research in music cognition does not provide much guidance on the role that rhythmic and metrical features play in melodic recognition and recall. What are the expectancy thresholds for rhythmic deviation? Here no answers are forthcoming, despite a number of important results reported recently for the perception of rhythm generally and its role in individualizing musical performance.

5.4. Human Criteria for Melodic "Matches": A Case Study

The original purpose of the talk which underlies this study (ISMIR, Paris, 2002) was to adduce from human comparisons of melodies some general principles and then to contrast the results with those from computer searches for approximate matches. In the writing up, the characterization of human search criteria (to produce "intentional" matches) became unmanageable. It seems that some "monothematic" collections of melodies favor examples that satisfy individual criteria, while others are motivated by confirmation of social definitions of "belongingness." For this reason, the present examination considers a collection which I regard as "best of breed" in the first category, an expert system devised (without explicit criteria) by one individual.

Of several highly detailed longitudinal studies of tune preservation and dissemination in the early modern period, one by the noted Italian musicologist, organist, and harpsichordist Luigi Ferdinando Tagliavini (1994) affords a particularly fertile field for consideration. It provides 68 examples (of instantiations of one "tune"). Tagliavini has enjoyed a long career in professional performance, so his perceptual faculties might be considered to be somewhat superior to those of musically well-trained undergraduates. His sources come exclusively from printed sources of the sixteenth through eighteenth centuries. The prints are from various parts of Europe.

It is fundamental to the use of highly subjective expert systems to understand the expert's goal. Tagliavini's purpose is to show melodic persistence across centuries, language contexts, and titles. This has been a frequent goal in the study of mainly oral traditions (for example, of the preservation and dissemination of religious chant and folk songs). It has also received attention in the studies of the development of polyphonic music of the fifteenth and sixteenth centuries (often again with an emphasis on embedded chants). Tagliavini's focus heads off in a different direction, for it emphasizes printed music in collections assembled for diverse purposes and a range of instrumentations. His examples show many variations of meter, mode, ornamentation, and, less frequently, accentuation and harmonization. It is transcribed for many different instruments, and a few examples are transcribed from recordings. The theme he tracks is one which he identifies as the *ballo di Mantova* ("the dance from Mantua"), a generic title which in sixteenth- and seventeenth-century instances lacks lyrics but occurs under many titles. Like printed music itself, the audience for the *ballo* would have been Italian nobles. Over the next two centuries the *ballo* cropped up in France, Belgium, the Netherlands and elsewhere. Apart from record transcriptions, Tagliavini does not include nineteenth- or twentieth-century examples, but the theme of Smetana's "Ma Vlast" ("My Fatherland") and of Israel's current national anthem "Hatikvah" ("The Hope," adopted in 1948) would both qualify. In a way, this diffusion only tells us an obvious truth: music traveled in the heads of musicians as well as on paper.

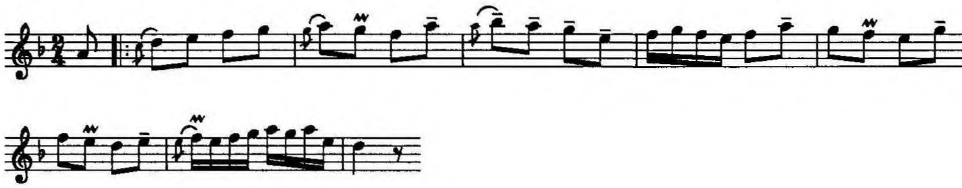
Tagliavini's cache of examples raises two central questions about melodic identity. First, where are the outer limits of melodic recognition? In such a big sea of examples, which one should be considered the basic instantiation of the melody (i.e., the target) in a computer model of the selection process? A music historian would normally proceed as Tagliavini has and would select the earliest printed example as the target. Since the aim here is to compare the results with those produced in a computer search, a target most amenable to computer searching was selected instead. It is a texted instantiation called "Fuggi da questo cielo" and was printed in Florence in *c.* 1625. Its advantage in a computer model is simply that the melody is set syllabically, so that nothing considered at the time to be fundamental should be missing and, conversely (on account of the date and genre), nothing extraneous should be present. "Fuggi" and eleven of Tagliavini's variants are shown in Figures 5.2 (a–k).

Target: "Fuggi, fuggi, fuggi da questo cielo"—Florentine monody (*c.* 1625) attributed to Gherardo Pedrali as shown:

5.2a. Instrumental transcription—Gasparo Zanetti (1645).

5.2b and 5.2c. Rhythmic variants on transcription—(b) keyboard transcription attributed to Gio. Battista Ferrini (1661) and (c) keyboard transcription attributed to Gaetano Greco (*c.* 1715).

5.2d. Ornamental variant—transcribed from a twentieth-century violin performance (with guitar accompaniment) by Melchiade Benni.



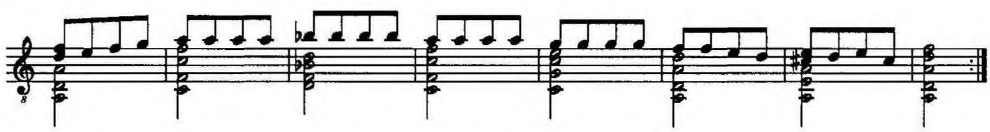
5.2e. Harmonic variant—"Cecilia," from a Dutch anthology of 1700.

5.2f and 5.2g. Metric variants—"Cecilia" in (f) a Belgian variant transcribed *c.* 1850 by Jan Frans Willems and (g) as given in the Dutch collection *Evangelische leeuwerck* (1682).

5.2h. Rhythmic replacement of melodic tones—Nicolas Saboly (*c.* 1670).

5.2i. Rhythmic simplification (with retexting).

5.2j. Reharmonization—arrangement for Spanish guitar by G. B. Granata, 1646 (reharmonization).



5.2k. Reharmonization with chordal arpeggiation (self-accompaniment, embedded notes, etc.)—anonymous transcription for lute or guitar, early 18th century.

Figures 5.2a–k. The target (the ballo di Mantova) and eleven match candidates provided in a musicological study by Luigi Ferdinando Tagliavini.

Tagliavini's selection is not so much representative of musicological studies of tune families (Selfridge-Field 2005) as it is of an ear steeped in classical music, particularly the keyboard repertory of the sixteenth through the eighteenth century. This is a repertory in which arrangements and transcriptions, some of significant complexity, are prevalent.

5.5. Computer-Produced Match Candidates

Before applying our provisional cognitive-distance metric to Examples 5.2a–k, we ran the first five notes of the *ballo* through *Themefinder* to see what match candidates would be produced by a linear search. Linear searches in *Themefinder* operate on pitch only at five levels on a continuum from generality (gross contour) to specificity (note and inflection names). The intermediate levels of search are refined contour (steps, skips, direction), scale-degree profile, and intervallic profile.

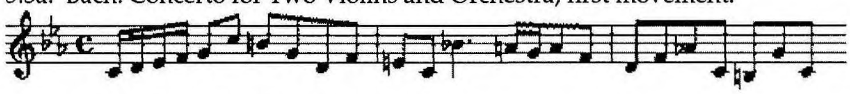
Wildcard searches are permitted. All levels can be filtered by meter and mode. Incipits in the *Themefinder* database come from three fundamentally different repertories: classical instrumental music, monophonic folk songs, and Renaissance polyphony. They are therefore somewhat different in character from many of the repertories represented in Tagliavini's collection (in it keyboard, lute, and string ensemble music make up a substantial portion). The component databases of *Themefinder* are searchable separately or together. The subset of data used included *c.* 12,000 classical incipits and *c.* 8,600 folk songs.

In a first-six-pitches search of *Themefinder*'s classical section at the (intermediate) scale-degree level, 22 matches were returned when no filters were applied, 6 when the search was limited to works in a minor mode, 4 when the search was limited to 4/4 meter, and none when both filters were in place. In other words, there was no candidate that might be termed "excellent" in its likelihood of matching the target. The pitches of repeated notes were ignored. (Because of the ambiguity that the treatment of repeated notes introduces into melodic profiles [e.g., should we substitute the sixth discrete pitch for the sixth event, which employs the same pitch as the fifth event?], we opted to exclude the sixth event in the melody.) A first-five-note search resulted in much higher rates of identification: 116 matches with no filters, 43 matches in the minor mode, 26 in 4/4, and 8 with both filters in effect. The most promising candidates from each search are shown in Figures 5.3a–b (5-note), and 5.3c–d (6-note).

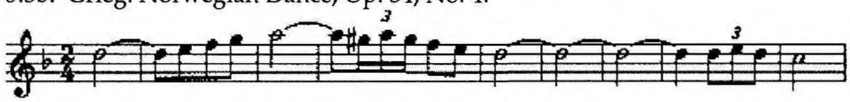
In searches of the folk-music database, the statistics on machine matches for the 6-note search were 20 matches with no filters, 1 with a mode filter, 4 with a meter filter, and none with both filters. The results for a 5-note target were 45 without filters, 6 with mode filter, 12 with meter filter, and none with both.

Other statistical studies have demonstrated that the folk songs in *Themefinder* have a much greater concentration in regular *n*/4 and *n*/8 meters than do the classical incipits, where *n*/2 and *n*/16 meters are also encountered and where *n*'s that are prime numbers proliferate in nineteenth- and early twentieth-century items. In both the classical and the folk-song portions of the database, examples in major mode are far more numerous than those in minor modes. This confers obvious biases when filters are used.

5.3a. Bach: Concerto for Two Violins and Orchestra, first movement.

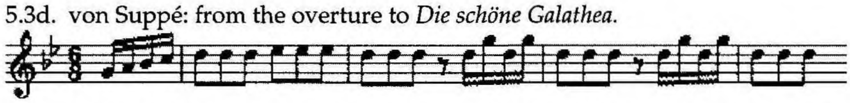


5.3b. Grieg: Norwegian Dance, Op. 34, No. 4.



5.3c. Dutch folk song: "Die Schlangenkoechin Isabelle, mijn dochterken."

5.3d. von Suppé: from the overture to *Die schöne Galathea.*



Figures 5.3a–d. Accidental matches in Themefinder for the first six distinct pitches of the ballo di Mantova (repeated pitches ignored).

The positive effect of greater string length comes as no surprise, since studies by both computer methodologists (Selfridge-Field 1991, 1993; Uitdenbogerd 2003) and psychologists (Deliège, 2001) show that a more articulate target (suggested by greater length) produces more appropriate results than a less articulate one.

What is instructive in this group of match candidates is that the musical features which are most conspicuous in the examples which have been automatically culled are distinctly different from those collected by analogue means in Figure 5.2a–k. Excluding 5.3c, which resembles 5.2g in its metrical orientation (both come from Dutch folk songs), the other candidates all seem much more remote, in cognitive terms, from the target for a single reason: the relative accentual weights of the pitches that ostensibly suggest a match are privileged in the human collection and ignored in the machine collection. This difference is reinforced by the divergent harmonic implications of each set.

The cognitive-distance metric that we propose is tuned especially to this set of examples. That is, it attempts to model the privileges tacitly granted in Figures 5.2a–k over those of Figure 5.3a–d. This is hardly a scientific approach, but the opportunity to compare results for a fully automatic process and a fully human one where both provide ground for extensive investigation is actually quite rare.

5.6. Parameters and Scoring

The scoring system presented here is based on two goals—one perceptual, one practical. Whatever happens in a melody only attracts attention if it is conspicuous. Any downbeat is more likely to be noticed than any upbeat, although in future work it could be instructive to explore possible trade-offs between accent and other attention-attracters, such as intervallic leaps, changes of direction, or rhythmic deviations from an established pattern.

For easy computation of results, the provisional scoring system is a ten-point scheme. The first part assesses pitch in relation to metrical and accentual information. Because our patience with false matches has been sorely tried by the large caches of poor candidates that are retrieved from large datasets, we set fairly strenuous thresholds.

1. Basic Pitch-Accent Structure			Range = 0-4
A. If meter matches target			Max = 1.00
	and	If subunit (e.g., quarter note) is the same	Score = 1.00
	or	If subunit is different (e.g., 4/8 vs. 2/4)	Score = 0.50
Else			Score = 0.00
B. Percentage of matched pitches on primary beats*			Max = 2.00
If matching number of scale degrees = 100%			Score = 2.00
	or	If matching number of scale degrees => 90%	Score = 1.33
	or	If matched number of notes/unit => 80%	Score = 0.67
Else			Score = 0.00
C. Percentage of matched pitches on secondary beats			Max = 1.00
If matching number of scale degrees = 100%			Score = 1.00
	or	If matching number of scale degrees => 90%	Score = 0.67
	or	If matched number of notes/unit => 80%	Score = 0.33
Else			Score = 0.00

Table 5.2. Pitch-accent scoring.

The second part of the metric assesses harmonic conformance. This assessment will appear to be quite arbitrary, given that the database contains only monophonic material. The auditory context that a user brings to melodic incipits will, of course, make potential matches seem more or less appropriate. Yet it is clear from a close look at the *ballo di Mantova* candidates that some pairings with the target stand a low probability, at face value, of perceptual recognition because they are implicitly so distant harmonically.

Example 5.2k is a "bonus" example intended to show how a melody can be conceptually close to a model while appearing to be quite distant in a literal item-per-item search. It represents the ultimate challenge for a cognitive-distance metric, since the ideal metric would rank such a candidate highly. This is a self-accompanied melody, so the harmonic fabric is more nearly literal than implied. Because the harmonic "fill" falls off the beat, however, it creates a very different silhouette from source (2j).

Examples 5.2j and 5.2k raise a separate issue, too. It might be typified as a "real-life" problem which is not normally discussed in musical literature. The issue is that, as shown by Tagliavini, these examples (which are variants of one another) have an implied harmonization (the harmonization that the outline of the melody suggests) which is flawed (in comparison with familiar renditions). This kind of flaw is very commonly found in popular guitar books and lead sheets that rely on a very limited menu of chords in order to give beginners access to the repertory in some fashion. Practitioners will want to consider how best to accommodate these kinds of effects. For the sake of this theoretical discussion, we reharmonized. That is, we corrected the harmonization of four of the eight accented events present so that instead of the sequence i-III-VI-III-vii-i-V-I, which suggests a modulation to the major mode, the sequence was changed to i-i-iv-i-iv-i-V-i, which retains the minor mode except for the inevitably sharped third in the dominant at the cadence. (The harmonizations as given in 5.2j and 5.2k are not those in common use today, with the exception that Smetana's "Moldau" theme is harmonically similar, while being melodically slightly different.)

In the proposed metric, both examples would score well if the harmony is "corrected" (that is if the searcher were actually imaging in the i-i-iv-i-iv-i-V-i) but poorly if it were left unaltered. Example 5.2k has the potential for a much higher score than 5.2j because of its greater level of activity at the quarter-note level; its melodic deviations from the target fall on unaccented beats. The thresholds are slightly more liberal in harmonic-assessment sections because domino effects introduce a greater degree of interpretive uncertainty. One false step will push all subsequent events into a lower-scoring category, although in human cognition the damage to recognition may be much less dramatic. Obviously greater refinement is needed to deal with this possibility.

II. Basic Harmonic-Accent Structure	Range = 0–6
A. Mode of work (major, minor, other)	Max = 1.00
If modes match	Score = 1.00
Else	Score = 0.00
B. Percentage of matched chords on downbeat**	Max = 2.50
If unambiguous matches on primary beats => 90%	Score = 2.50
or If unambiguous matches on primary beats => 80%	Score = 2.00
or If unambiguous matches on primary beat => 70%	Score = 1.50
Else	Score = 0.00
C. Percentage of matched chords on secondary beats**	Max = 2.00
If unambiguous matches => 90%	Score = 2.00
or If unambiguous matches => 80%	Score = 1.50
or If unambiguous matches => 70%	Score = 1.00
Else	Score = 0.00
D. Percentage of matched chords on tertiary beats	Max = 0.50
If unambiguous matches => 90%	Score = 0.50
Else	Score = 0.0

Table 5.3. Harmonic accent scoring.

The method for computing the score of beat-sensitive items (5.1b, c; 5.2b, c, d) addresses each individual beat indicated in the numeral of the time signature as shown in Table 5.4.

Meter	**Primary**	**Secondary**	**Tertiary**
2 / 4	1	2	
3 / 4	1	2, 3	
4 / 4	1	3	2, 4
6 / 8	1	4	2, 3, 5, 6
12 / 8	1	7	4, 10

Table 5.4. Beat values for various meters.

As currently developed it is focused only on tonal works in regular meters. It is not suitable for music with persistent syncopation, such as much of the jazz repertory, without adaptation.

The rather stringent thresholds for any score to be achieved may be surprising to some, but in practical situations involving large quantities of musical data, it is essential to have fairly rigorous cutoffs. The difficult question is precisely how to retain what may be psychologically valid even when it may be quite remote on a note-per-note basis. For use with very large datasets such as the *Themefinder* database, we would set the thresholds considerably higher than they are here. At the same time it is acknowledged that ideas of what does or does not match an intended target may vary from user to user. We believe that the final selection of a match should therefore be determined by the user.

5.7. Tabulation and Comparison

In a first pass, the melodic/accentual and harmonic/accentual scores were to be added. Points were predicated on the notion of keeping the numbers simple. The scores obtained in this way are shown in Table 5.5.

Table 5.5. Scores for examples in Figure 5.2a–k (human matches), using addition as a basis for ranking.

Example	Pitch-Accent score		Harmony-Accent score		Total score (additive)	
	Raw	Ranked	Raw	Ranked	Raw	Ranked
2a	3.67	2	5.5	3	9.17	2
2b	3.67	2	5.0	4	8.67	3
2c	2.67	6	6.0	1	8.67	3
2d	1.17	9	4.5	5	6.67	8
2e	2.67	6	4.0	9	6.67	8
2f	2.33	8	4.5	5	6.83	7
2g	1.00	10	2.0	11	3.00	11
2h	3.50	4	4.5	5	8.00	6
2i	4.00	1	4.5	5	8.50	5
2j	1.00	10	4.0	9	5.00	10
2k	3.33	5	6.0	1	9.33	1

These results were then compared with scores for the machine matches from *Themefinder* (Table 5.6). (This comparison is flawed by the fact that there were so many fewer candidates.) The computed results are notably poorer, so much so that ranks were not assessed in relation to those of Table 5.5.

Example	Pitch-Accent score	Harmony-Accent score	Total score (additive)
3a	0.50	1.00	1.50
3b	1.33	1.00	2.33
3c	1.50	3.00	4.50
3d	0.83	1.00	1.83

Table 5.6. Scores for incipits shown in Figures 5.3a–d ("accidental" matches).

Since the "accidental" matches were not originally ranked, a comparison of scores by both approaches to it only serves to point out that their multiplicative scores are very low indeed. While that dramatizes their distance from the "intentional" matches, the additive scores make the categorical separation equally noticeable. Thus this approach could be of value in cases in which little is known about the underlying data or in which the quantity of potential matches is likely to be large.

5.8. Discussion

The most common way in which perceptually different melodies can appear (in sorted symbolic code) to be the same is when pitch alone is used as the yardstick, unless the level of pitch-definition is relatively precise (Sapp et al. 2004).

In relation to the literature on melodic cognition, the numerous studies of Mari Riess Jones (e.g., 1991, 1997) verify the important role of accentual information in human melodic assessment. Jones's concept of JAC is difficult to implement automatically because of its reliance on specific pitch-content in situations where, from our point of view, fuzzy searching may be required in the human-computer interface. Yet the idea of the coupling of accentual information with pitch and harmony separately (but without assessing "rhythm" or "accent" directly) is offered as a practical adaptation of this approach. Halperin's finding on the large degree of human tolerance for modal variability when coupled with the preservation of other musical features (1998) accounts for the low penalty in our metric for modal digression. The unpredictable relationship between detail and generalization in melodic comparison is highlighted in Cuddy (1993), wherein the variable effect of a single note on the recognizability of melody is confronted. We return to this study to underscore the author's attribution of this unpredictability to that fact that no parameters have been established for any aspects of melody other than accent.

The cognitive-distance measure proposed here does not go far beyond accent and it does not come with any claim for universal suitability. Its immediate purpose is to initiate more focused discussion of what those other attributes may be and how they should be tuned under different circumstances. Until these are better defined, search algorithms cannot dramatically improve what might be termed "cognitive satisfaction" levels. We do not know, for example, what the prevalent human ranking order might be for rhythmic substitutions. To what extent might the preferred rankings be accent-dependent? Are these rankings coincident with, stronger than, or weaker than other known attention-getters (melodic leaps, changes in melodic direction, accentual displacement, etc.)? Since the actual cognitive validity and relative cognitive weights of some of the factors included are still undetermined, we offer this material primarily as a basis for further discussion and experimental verification or refutation.

Acknowledgments

The *Themefinder* project was initiated by David Huron in 1996 for the purpose of capturing user-query information. The original user interface was developed by Andreas Kornstädt. Further development and management have been overseen by Craig Stuart Sapp for most of the intervening years. Most of the development work has been supported by the Center for Computer Assisted Research in the Humanities at Stanford University.

References

Cambouropoulos, Emilios (1998). "Towards a General Computational Theory of Musical Structure." Ph. D. thesis, University of Edinburgh.

Campouropoulos (2001). "Melodic Cue Abstraction, Similarity, and Category Formation: A Formal Model," *Music Perception* 18/3 (2001): 347–370.

Cuddy, Lola (1993). "Melody Comprehension and Tonal Structure" in *Psychology and Music: The Understanding of Melody and Rhythm*, ed. Thomas Tighe and W. Jay Dowling (Hillsdale, NJ: Lawrence Earlbaum Associates, Inc., 1993), pp. 19—38.

Dowling, W. Jay (1994).

Deliège, Irène (1997). "Cue Abstraction in the Representation of Musical Form" in *Perception and Cognition of Music*, ed. Irène Deliège and John Sloboda (Hove, UK: Psychology Press), 387–412.

Halperin, David (1990). "A Segmentation Algorithm and its Application to Medieval Monophonic Music," *Quantitative Linguistics*," 43, 107–119.

Holleran, Susan, Mari Riess Jones, and David Butler (1995). "Perceiving Implied Harmony: The Influence of Melodic and Harmonic Context," *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21/3 (1995), 737–753.

Jones, Mari Riess (1993). "Dynamics of Musical Patterns: How Do Melody and Rhythm Fit Together?" in Thomas J. Tighe and W. Jay Dowling (eds.), *Psychology and Music: The Understanding of Melody and Rhythm* (Hillsdale, NJ: Earlham Affiliates, 1993), 67–92.

Jones, Mari Riess, and Peter Q. Pfordresher (1997). "Tracking Musical Patterns Using Joint Accent Structure," *Canadian Journal of Experimental Psychology* 51/4 (1997), 271–291.

Krumhansl, Carol L. (1990). *Cognitive Foundations of Musical Pitch.* New York: Oxford University Press.

Lerdahl, Fred, and Ray Jackendoff (1983). *A Generative Theory of Tonal Music.* Cambridge: The MIT Press.

Levenshtein, V. I. (1966). "Binary Codes Capable of Correcting Deletions, Insertions, and Reversals," *Soviet Physics Doklady*, 10/8, 707–710.

Muellensiefen, Daniel, and Klaus Frieler (2003-04). "Cognitive Adequacy in the Measurement of Melodic Similarity: Algorithmic vs. Human Judgments," *Music Query: Methods, Models, and User Studies (Computing in Musicology 13)*, 145–176.

Narmour, Eugene (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model.* Chicago: University of Chicago Press.

Ralson, Jacqueline T., and Mari Riess Jones (1991). "Some Influences of Accent Structure on Melody Recognition," *Memory and Cognition* 19/1, 8–20.

Sapp, Craig Stuart (2004). "Computational Harmonic Analysis of Western Tonal Music." Ph. D. thesis, Stanford University.

Sapp, Craig Stuart, Yi-Wen Liu, and Eleanor Selfridge-Field (2004). "Search-Effectiveness Measures for Symbolic Music Queries in Very Large Databases," *Proceedings of the Fifth International Symposium on Music Information Retrieval (ISMIR)*, Barcelona.

Singer, Jane (2003). "A Model of Melodic Representation: Towards an Implementation of Music Information Retrieval." Ph. D. thesis, The Hebrew University of Jerusalem.

Selfridge-Field, Eleanor (1991). "Issues of Representation in Melodic Comparison," *Computers in Music Research*, II (Queen's University, Belfast, 1991).

Selfridge-Field, Eleanor (1993). "Music Analysis by Computer: Approaches and Issues" in *Music Processing* (*Computer Music and Digital Audio Series*, 9; gen. ed. John Strawn), ed. Goffredo Haus. Madison, WI: A-R Editions, Inc., and Oxford: Clarendon Press, 1993, pp. 3–24.

Tagliavini, Luigi Ferdinando (1994). " 'Il ballo di Mantova', ovvero, 'Fuggi, fuggi da questo cielo', ovvero, 'Cecilia', ovvero..." in *Max Lutolf zum 60. Geburtstag: Festschrift* (Basel: Wiese), 135–175.

Temperley, David (2001). *The Cognition of Basic Musical Structures.* Cambridge, MA: The MIT Press.

Uitdenbogerd, Alexandra, and Justin Zobel (1999). "Matching Algorithms for Large Music Databases: A Technical Report," Melbourne: Dept. of Computer Science, RMIT University.