12 Defining the Features of Monophonic Vocal Melodies

Jane Singer

Jewish Music Research Center
National Jewish University Library
Givat Ram, Jerusalem
Israel
jsinger@cc.huji.ac.il ; *janes@nomadiq.com*

Abstract

This paper proposes a system of representing melodic abstraction at a near-surface level. This representation consists of a series of note-groups, each described by a number of attributes. The segmentation is generated according to a grammar that defines and ranks possible note groupings, based on the Gestalt principles of proximity, similarity, and good continuation. Each segment is represented by an instance of a segment-type class, along with additional distinguishing attributes. Some initial methods of comparing the melodic segments are suggested.

Supplementary materials for this article are available at *www.ccarh.org/publications/cm/12*.

COMPUTING IN MUSICOLOGY 12 (1999-2000), 185–206.

Our ability to perceive similarities among non-identical melodies is dependent upon many factors. In order to identify certain similar aspects, these must first be defined, isolated, and named for each melodic instance. Only when a satisfactory representation of the individual melody has been achieved can the process of comparison succeed. According to Tversky (1977), the degree and quality of similarity between any two objects is a function of feature-matching. It involves a process of determining common and distinct attributes; the process of determination is called the Contrast Model.

Theories of melodic similarity have recently found practical applications. The creation of large databases of melodic information has been facilitated over the years by (1) the creation of various systems of encoding notated melodies and (2) the increase in storage capacity of computers. Despite these advances in the storage of melodic information, the organization and retrieval of such material remains problematic. Although such systems are able to store general information such as composer, key, and informant, they understand little about the melodic makeup of the entities they store. Various melodic attributes can be unambiguously defined, among them ambitus, length, key signature, time signature, pitches, and durations. However, these attributes are insufficient for identifying similar, but not identical, melodies (such as variants) that typically exist in large song corpora. Various systems have been proposed for recognizing similarity among stored melodies (Cope 1998; Howard 1998; Kornstädt 1998).

In order to retrieve melodies from within a database, a representation must be formulated. Among the many existing systems,[1] some (e.g. those described in Adams 1976) describe overall contour, while other more analytical systems depict exact intervallic movement in terms of limited specific operators that function within an alphabetic context (Deutsch and Feroe 1981). The retrieval of similar melodies according to attributes describing overall contour would produce an unacceptably large number of hits. For example, a set of all the ascending melodies could consist of close to half of the melodies in any given song corpus. The description is far too general. A retrieval system using a highly specific representation (such as that of Deutsch and Feroe), would retrieve few, if any, similar melodies. A third type of system, pitch-reduction (Lerdahl and Jackendoff 1983; Schenker 1956), represents melodies as hierarchies of structural notes. Such systems are both too general and too specific (for the purposes of comparison): the structural notes are not abstracted (they are

[1] Selfridge-Field (1998) discusses a large number of diverse systems of melodic representation, along with their individual strengths and limitations.

represented only by themselves) and the eliminated pitches are not at all represented. Therefore few melodic hierarchies would actually match, and those that did could not be further differentiated nor compared. Clearly a need exists for abstracting melodies in such a way that a reasonable number of melodies could be identified as similar without losing all knowledge of their distinct attributes. Therefore the goal of the system described here is not to achieve a description capable of representing all of the many facets which listeners and scholars could associate with a melody (an impossible task), but rather to extract a number of features capable of distinguishing a group of similar melodies within a corpus. These features must be valid and together should create an overall depiction of the melody.

In attempting to ascertain the features of a musical instance, many problems arise. Should we describe the entire movement, a phrase, a period, a motif or a single note? How can a unit such as a motif be distinguished from its surroundings in an efficient and meaningful way?

12.1 A System for Feature-Based Classification

The system proposed here is capable of isolating various units within a melodic incipit, ranging from a single interval to a larger unit, roughly equivalent to a motif. These units or segments are assigned attributes and can be recombined to form larger, more abstract but still meaningful groupings. In a given melody, the attributes of a unit on one level determine (1) the organization of these units into objects of the next higher level and (2) the comparison of this level with the corresponding level of another melody. The music is limited to vocal pieces, including monophonic folksongs and some art music.[2] The three corpora of Yiddish folksongs used in this project are located in the Vernadsky Central Scientific Library in Kiev, Ukraine.

[2] Vocal melodies tend to be made up of small intervals, with a restricted number of large intervals (Dowling 1978: 352). Vocal art music often adopts instrumental melodic features, including series of large intervals.

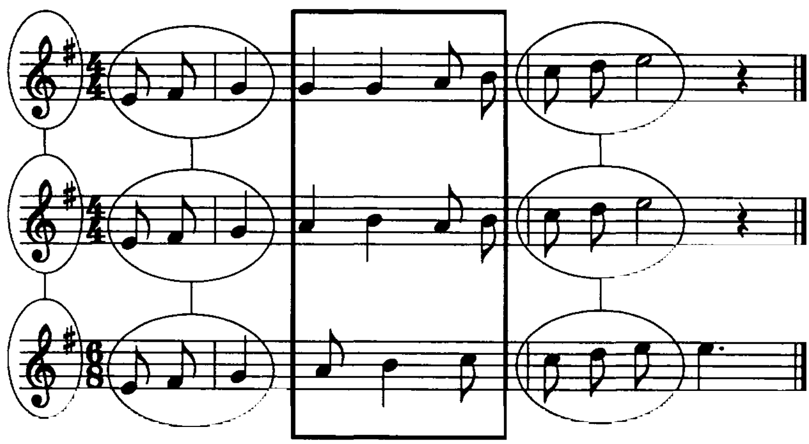
The creation of a successful storage-retrieval system for melodies involves two processes: (1) the creation of an optimal representation and (2) the formulation of an appropriate query strategy. Both of these tasks are enormous; the present paper deals only with the first. Although the current system does not presently deal with the actual design of a full retrieval system, the representation it creates is designed for later processing. The construction of the representation is comparable to the indexing of bibliographic databases, with the individual attributes functioning as keywords.

In the proposed system, the process of creating a representation of a melody involves two stages: (1) the segmentation of the melody and (2) the description of each segment.

12.2 The Melodic Segmentation Process

The quantity and quality of similarity that exists between any two melodies will vary from note to note and from measure to measure (see Figure 1). Therefore, in order to define the attributes that describe this similarity, the melody must be divided or segmented. In a more general context, segmentation, or phrasing, is essential to both performance and active listening. In spite of this need, segmentation can be highly ambiguous.

Figure 1. Varying degrees and types of melodic similarity and dissimilarity. Ovals indicate some very similar areas, the rectangle marks a less similar area.



12.2.1 The Segmentation Method

The literature on melodic segmentation (Deutsch and Feroe 1981; Oura 1991) is largely based on the Gestalt principles of perception, specifically the laws of similarity, proximity, and "good continuation." The application of the proximity rule to pitch-time space[3] has had an overwhelming influence on recommended procedures for the segmentation of melodies, especially at levels close to the melodic surface (see Narmour 1990; Meyer 1973; Meyer 1956; Wertheimer 1938; Lerdahl and Jackendoff (1983): 43 [GPR2[4]]). Specifically, the most convincing groupings consist of notes of shorter duration that form smaller intervals, differentiated from the preceding and subsequent notes by larger intervals and longer notes or rests. Because of the large range of durations and intervals that may occur in a melody, the segmentation process allows for many alternative groupings, based on different applications of the rules. Therefore, although there is a general consensus on the actual rules, their application to a single musical instance remains problematic.

[3] The x-y axes of physical space are often used as analogies for pitch and time. Similarly, musical notation represents pitch and time through the placement of notes in a two-dimensional space.

[4] This and subsequent "GPR" rules refer to the numbering which appears in Lerdahl and Jackendoff (1983: 43).

The present system attempts to supply a practical application of these theoretical rules. An eventual listener study will help adjust the weight given to the individual rules used by the system for choosing a particular segmentation.

12.2.2 Segment Types: The Formation Rules

Here we describe the actual applications of Gestalt principles in the proposed system. *Formation rules* accept or reject the creation of segment-type instances. *Preference rules* (described later) score the possible segments according to factors also based on Gestalt principles.

The segment types are organized into three levels: (1) the primitive, (2) the prefix-contour-suffix, and (3) the construct. Some segment types have a number of subtypes. Each instance of a type has additional distinguishing features. Since Gestalt principles in general have the greatest effect on local groupings (such as motifs), rather than higher structures (such as movements), their influence is greatest at the lowest level, that of the primitive.

12.2.2.1 The Primitive Level

The *primitive* represents the smallest meaningful group of notes and exhibits the highest degree of similarity among the component notes and intervals. Under the proximity and similarity rules, similar small intervals are combined into units of consistent overall direction (good continuation). Large intervals cannot be combined at this level, since this would violate the proximity rule. Therefore a primitive may consist of a single large interval.

Accordingly, the different *interval classes* used at the primitive level vary in their degree of similarity. The classes are:

> IDENTICAL: two intervals are *identical* if both their direction and diatonic value are identical, such as the case of an ascending M2 and ascending m2. The interval classes sharing this level of similarity are: *primas*, *ascending seconds*, *descending seconds*, *ascending thirds*, *descending thirds*, [*ascending skips*], [*descending skips*].
>
> SIMILAR: two intervals are *similar* if they both belong to one of the following categories of interval size, regardless of their direction: *seconds*, (both major and minor), *small intervals* (up to a major third), and *large intervals* (intervals larger than a major third).

Since some intervals belong to more than one category, the similarity rule alone allows for more than one possible grouping (preference rules which deal with different degrees of similarity are given below).

Several additional points pertaining to categories within the primitive level should perhaps be mentioned here. 1) Narmour (1990: 78) applies the similarity rule only to small intervals: when a listener hears a small interval, he expects another interval of a similar size.[5]

[5] "Similarity between two intervals in the same registral direction will be said to exist when the difference between them is not greater than one whole step" (Narmour 1984: 88).

2) The diatonic scale-step value, rather than the absolute pitch value of a note, is used for calculating the value of an interval, in order to eliminate those factors that are modal rather than purely melodic. In doing this, motifs which occur on different degrees of a single mode, or are transposed to different modes, maintain their melodic identity. The present system operates within any 7-step diatonic system.

3) Balzano and Liesch (1982) established interval groups by testing the "confusability factor," that is, which intervals are often mistakenly perceived as being equivalent. According to their study the m2, M2, and m3 belong to such a group. Narmour (1990) groups large and small intervals differently according to the context of the discussion, but for the purposes of continuity ("implication"), he includes the prima, m2, M2, m3, and M3 (Narmour 1990: 76).

4) The "seconds" category is actually a subcategory of the small interval class and represents a higher degree of similarity, although they (seconds) do not represent identical intervals.

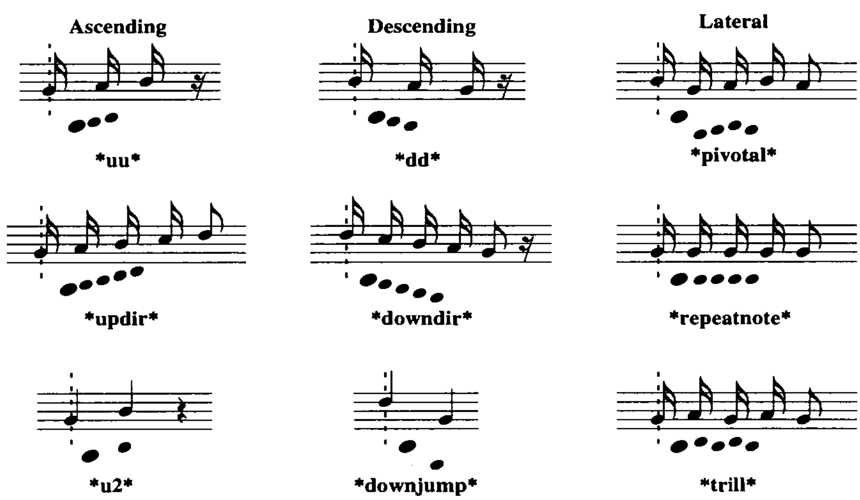
Overall direction on the primitive level is defined as:

- LATERAL: a group of similar, but not identical, small intervals whose ambitus does not exceed a third, or a group of primas.
- ASCENDING: a group of identical ascending small intervals.
- DESCENDING: a group of identical descending small intervals.

These three primitive types are not distinct categories. Admittedly, the differentiating factors chosen here need further investigation. A listener study is presently being carried out to better understand to what degree these groups can be delimited.

The primitive grouping represents the highest level of similarity: both overall direction and intervallic makeup are consistent throughout. Figure 2 depicts some primitive types.

Figure 2. Some primitive types.



KEY TO PRIMITIVE TYPES:

u - ascending second **d - descending second**
u2 - ascending third **downjump - descending large interval**

12.2.2.2 The Prefix-Contour-Suffix Level

The successful primitives are recombined and further qualified at this level. The *prefix* defines the metric quality of some primitives. The *contour* combines primitives of similar direction. The *suffix*, made up of a single primitive, has no constraints at this level.

The groupings at this level are less dependent on Gestalt principles than those on the prefix level. Instead, the different qualities of the prefix, contour, and suffix are largely dependent on metric factors. This metric information is communicated concurrently with the melody itself. All metric melodies are dependent on the ability of the sounding note-pattern to transmit the differentiating qualities of the beats. It is postulated that these constraints have had enormous influence on the development of a large range of musical styles, and specifically that they have restricted greatly the types of rhythmic constructs that are generally found within vocal melodies (Clynes 1997: 204; 1986: 175).

Early psychological experiments (Stetson 1905; Woodrow 1909) confirmed that weak notes will normally be associated with a stronger note that precedes them (Figure 3). However, if the accented note is longer, or a larger interval intervenes, the weak beats are associated with the long note that follows (Figure 4). This segmentation concurs with the proximity rule, yet the metric qualities attached to the resulting segments cannot be accounted for by Gestalt principles. Cooper and Meyer (1960: 10) referred to these two kinds of groupings as *beginning-accented* and *end-accented* groupings.

Figure 3. Beginning-accented groupings.

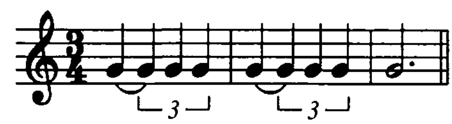


Figure 4. End-accented groupings.

In the present system, the contour, possibly combined with a suffix at the construct level, forms stable units which begin on a strong beat of the measure, like those shown in Figure 3.

The prefix (or anacrusis) pattern is demonstrated by the second and third notes of the triplets in Figure 4, and consists of a few notes of relatively short duration and weak metric stress, ending on a stronger note. The prefix occurs only after a disjunct, usually after a longer note on a stronger beat or after a large interval. Typically these units are decorative in nature, due to their metric position. Their embellishing role is confirmed by the fact that variants of a single melody will often appear with or without these upbeat figures.

THE PREFIX

The *prefix* consists of a single primitive and satisfies the following metric constraints:

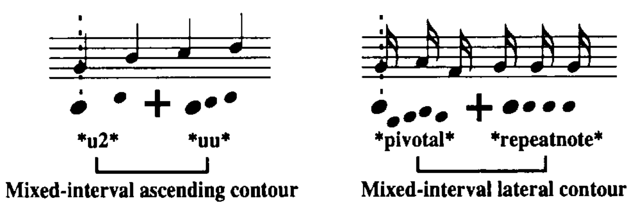
- the first note begins on a beat other than the first beat of the measure; [6]
- the last note of the note prefix falls on a beat which is stronger than all beats spanned.

[6] The most satisfying prefixes begin on a note whose metric position is weaker than that of the second note (see below).

THE CONTOUR

The *contour* groups primitives to form a unit which is consistent in overall direction, but whose intervallic makeup may vary (see Figure 5).

Figure 5. Mixed-interval contours.



The following conditions pertain:

- The first note functions as a structural note of the melody and begins on a beat that is stronger than all beats spanned by the contour.
- A contour can absorb an initial on-beat ornamental figure such as a single second, or a short trill, pivotal figure, or repeated note pattern.
- A contour cannot span the first beat of a measure (although it may end there).

On the contour level, overall direction is inherited from the primitives it combines (a contour can also consist of a single primitive).

The suffix

The suffix consists of a single primitive and is ornamental in nature. A possible suffix is defined only in relation to the main element with which it is to be combined (see below).

12.2.2.3 The Construct Level

Constructs are composed of parts, each of which is qualitatively different yet complementary in its role. Those factors which contributed to discontinuity at a lower level (for instance, the different metric qualities of the prefix and contour) now contribute to the coherence of the constructed whole. Whereas on the lower level the weak-strong and strong-weak relationships between notes were not in themselves connective, the construct level defines the context in which the metric quality of a group contributes to its connective power.

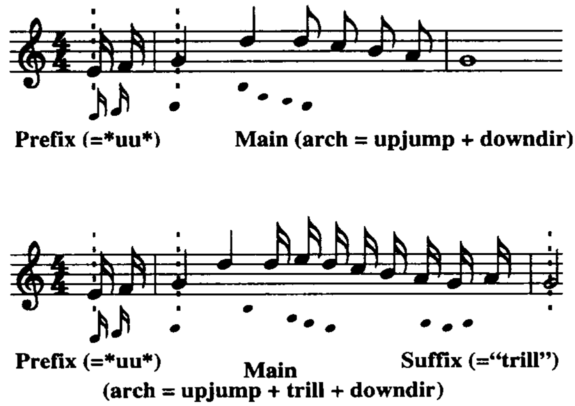
The construct, which imposes further constraints on the prefix, contour, and suffix that it combines, may comprise selected elements of these kinds:

- a *prefix* element, consisting of a single primitive, preceded by a disjunction;
- a *main* element, consisting of a single contour;
- a *suffix* element, composed of a single primitive which could not be attached to the main contour because of the difference in overall direction and which has an ambitus of less than a third.

The strong beat of the main is one of the focal points of the melody, and typically constitutes a structural note. The metric quality of the suffix can be either weak-strong or strong-weak.

This model can appear with components missing. The following sub-groupings of prefix/main/suffix are possible: (1) prefix only, (2) prefix/main, (3) main only, and (4) main/suffix. Figure 6 shows the first construct of two similar melodies.

Figure 6. First construct of two similar melodies.



12.2.3 Preference Rules

In addition to the constraints imposed on a segment by the formation rules, additional factors add to or subtract from its "goodness." Tables 1 and 2 list some of these factors.

Table 1. Intervallic factors that add to a segment's "goodness."

THE "GOOD" COMBINATION	COGNITIVE PRINCIPLE
A pivotal pattern ending on the pivotal note	Good continuation
A trill ending on the note on which it began	Good continuation
A repeating intervallic pattern that is longer than two intervals	Similarity; GPR6
A longer segment	Good continuation; GPR1
A group of the most similar (or identical) intervals[7]	Similarity
An ascending figure, followed by a descending one, together forming an arch	Symmetry
A disjunction consisting of a large interval	Similarity Proximity

[7] For instance a triad, being made up of two identical intervals, is preferable to an alternative segmentation which would split up these two intervals (in the case in which all other segmenting factors are equal).

Table 2. Rhythmic factors that add to a segment's "goodness."

THE "GOOD" GROUPING	COGNITIVE PRINCIPLE
A repeating rhythmic pattern (which is longer than two notes)	Similarity; GPR6
A contour whose last note is longer than all other notes in the contour or A construct whose last note is longer than all other notes in the construct	Proximity Similarity? GPR1
A disjunction which spans a rest	Proximity; GPR1
A disjunction that moves from a longer to shorter note	Proximity
A disjunction whose first note is longer than the preceding note	Proximity
A prefix which is shorter than the main element	Metric constrictions and conventions
A suffix which is shorter than the main element	Metric constrictions and conventions

According to the above formation and preference rules, each interval or segment is evaluated in order to determine (1) if it can be included in the previously established group (or grouped with the previous interval) and (2) how strong a connection it has to the previously established group (or interval). From this information, the system proposes a number of good segmentations.[8]

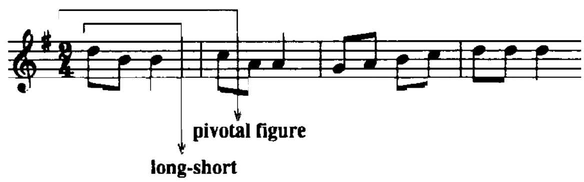
[8] This quality of "continuity" has been referred to as "implication." Narmour (1990: 49) similarly refers to units which can consist of same or identical intervals, same or different registral directions, the level of "implication" varying accordingly.

12.2.4 Resolving Ambiguities of Segmentation

Two types of rules govern segmentation—the formation rules (which define possible combinations) and the preference rules (which rank the possible groupings). Both types of rules usually produce several possible segmentations of a single melody. The segmentation of a melody is clearest when all applicable rules agree on a specific segmentation. However, a melody often arouses more interest when its division is ambiguous.[9] In Figure 5, the first three notes form a segment which ends with a longer note. However, the first four notes form a pivotal figure with close proximity of the third and fourth notes. Both rules add weight to their corresponding segments. The first is preferable because of the overriding importance of the proximity rule when applied to durations.

[9] Western art music, especially instrumental music, thrives on ambiguity of segmentation, including overlapping groups and non-contiguous phrasing. However, the musical material under consideration here, monophonic folksong, is relatively straightforward in this regard.

Figure 7. Contradictory implications of different rules.



12.3 Segment Description and Feature Definition

Once a satisfying segmentation is achieved, the description of the segment is relatively straightforward.

12.3.1 The Primitive, Prefix, and Suffix

The same uniform intervallic texture and direction of the primitive which influenced its cohesiveness as a unit now describes it. The primitives that are further defined as prefixes and suffixes inherit this description.

Additional information, such as length (number of notes/intervals, as well as the sum of the beats spanned), ambitus, metric position, first-note position, last-note position, high-note position, and low-note position are used by the rules to determine their recombination into higher-level units and to create higher-level descriptions. The individual note and interval objects are embedded in the final representation, allowing comparison of the melodic surfaces.

12.3.2 The Contour

The overall shape of a contour is determined by the interval formed by the first and last note ("slope"). Varying degrees of descent and ascent may be more or less convincing, and may be influenced by other factors, such as ambitus. Additional attributes of the contour, such as intervallic texture, length, and note positions, are inherited from the primitive level.

12.3.3 The Construct and Melody

The description of the construct is an aggregate of its components and their descriptions. The *melody* representation consists of a series of constructs and disjuncts. Constructs may occur conjunctly or be separated by a disjunct.

12.3.4 Description of Ambiguous Segmentation Results

Just as various factors contributed to the ambiguity of segmentation, the assigning of features to a segment instance can achieve varying levels of certainty.

> CONTOUR: The overall contour of a segment is established by the slope, or first-to-last interval, and is influenced by the highest-to-lowest interval. Both the ascent and descent are most convincing when they exceed the interval of a third. Therefore, all segments (primitives and contours) that do not exceed a third exhibit a degree of similarity, whether they are classified as ascending, descending, or lateral.

> TEXTURE: A construct can consist of different intervallic textures. These mixed textures, inherited from the primitive level, are named by the more general applicable attribute, such as the mixed interval texture in Figure 5.

12.4 From Theory to Practice: The Computer Application

The computer application of the above rules involves two systems—(1) an object system and (2) a rule system (grammar). The program which implements them, written in Lisp, is adapted from a natural-language parsing program described in Norvig (1992). This adaptation of his program involves the integration of the object system with the grammar.

12.4.1 The Object System

Each real-world note, interval, and segment is represented in the system by an object, with slots representing their attributes. In addition, each object belongs to one or more classes. These classes define the similarity that exists between various intervals and segments. Figure 8 shows the interval of an ascending third (called `*u2*`) and the classes to which it belongs.

Figure 8. The interval of an ascending third (*u2*) and its object classes.

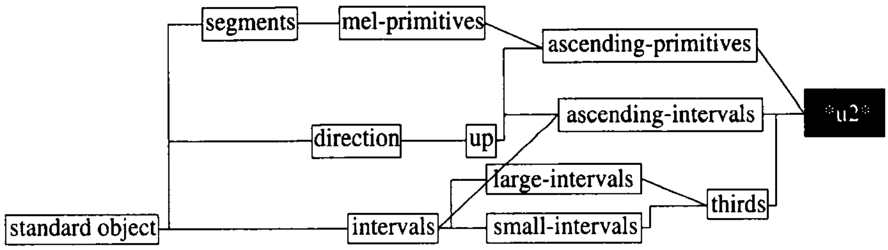
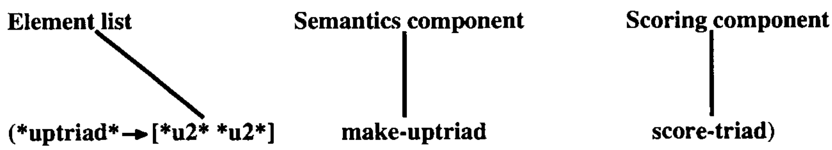


Figure 9. The derivation rule.



12.4.2 The Grammar

The grammar consists of derivation rules, each one consisting of three parts which appear on the right-hand side of the arrow (–>):

- A list of the objects (by their class name) to be combined (element list)
- A rule which either accepts or rejects this combination (semantics component)
- A rule which scores the combination according to its "goodness" (scoring component)

The first two parts of the right-hand side of the rule—the element list and the semantics component—constitute the formation rules. The scoring component contains the preference rules.

The *element list* suggests a possible combination. The semantics component consists of a single function (which in turn can call other functions), which imposes further constraints on the recombination, and may reject it. For example, an element list may contain two small intervals whose combined ambitus must be verified as being less than a third. The scoring component imposes the preference rules and determines the weighting factor of each parse.

12.4.3 How the Rules Work Together with the Object Classes

The encoded melodies of the raw input are initially converted into note-objects. These objects represent the surface elements of the melody: duration, metric stress, and pitch. These are subsequently converted into interval objects, each object containing all the information about the notes it spans. In addition, each interval belongs to a number of classes (seconds, small intervals, large intervals, thirds). These classes are referred to by the rules (see above).

The parser activates the rules by referring to the first object in the input stream (which is initially a series of interval objects) and looks for a rule that can be triggered with this class of object. For instance, the rule in Figure 9 can be triggered by any object which belongs to the `*u2*` class (any ascending third). Once this rule is fired, the parser checks whether the next object in the input stream belongs to the next class named in the element list (in this example, also `*u2*`). The *semantics component* then checks whether these two instances can be combined (for example, whether the resulting ambitus is within the required range), and the scoring component decides how satisfying the combination is.

Since the class system can describe objects in different ways and on different levels, the rules can be very specific or very general. If a combination is successful, the parser looks for a rule to trigger with the new object and continues in a similar manner. Since more than one rule can possibly be applied to any given object, several different segmentations are produced by the parser. These range from the very convincing to the unlikely, sorted according to the score assigned them by the rules.

12.5 Results

The representation produced by the system consists of a list of the best parses, including all the attributes of the embedded objects.

12.5.1 Feature Matching

The representation produced by this system is multi-faceted: it describes many hierarchical levels of each segmentation possibility and provides alternative segmentations. The richness of the description suggests many different approaches to comparison. Those mentioned here are sorting, selection, and multivariate analysis.

As an initial attempt to organize the various attributes, the data were read into an *Access* database table. The following data-elements were exported for the highest-ranking segmentation of each parse:

- melody identification number
- the size of the disjunction between the first and second construct (or nil in the case where the first two constructs are conjunct).

In addition the following information was retrieved for the first two constructs:

PREFIX	MAIN	CONSTRUCT	SUFFIX
ambitus	ambitus	tilt	ambitus
texture	texture		texture
length	length		length
direction	tilt		direction
	shape		
	intervals		

12.5.2 Sorting

The database table was joined with two tables that supplied the sorting values for the *texture* and *shape* fields. The corpus was then sorted according to the following keys (attributes of the first main, listed in the sort order): shape, texture, ambitus, intervals, and length. All fields were sorted in ascending order, with the exception of the ambitus (the SQL statement, along with the data retrieved for this and subsequent queries, appears on the CCARH Web site). Although the data-sort took into consideration only the attributes of the first main, many groups of similar melodies emerged (appeared consecutively). Figures 10 and 11 show two such groups.

These results alone cannot provide a conclusive demonstration of the validity of the data. The approach is clearly limited, since the similarity is represented by a seriation, making it impossible to map two or more features equally. The similarity between any two melodies is a function of the distance between them in the sort order.

Figure 10. One group of similar melodies from the JMF1 corpus.

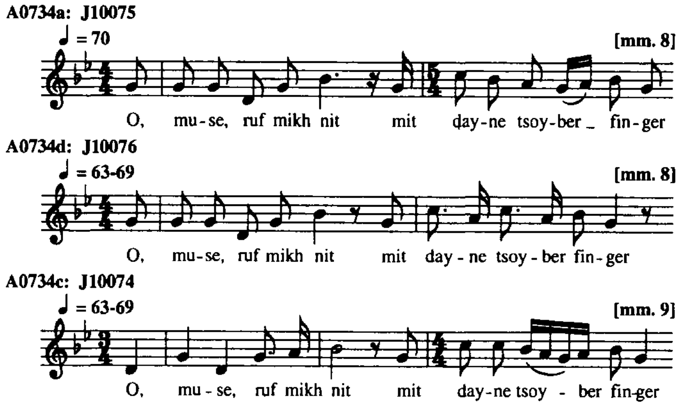
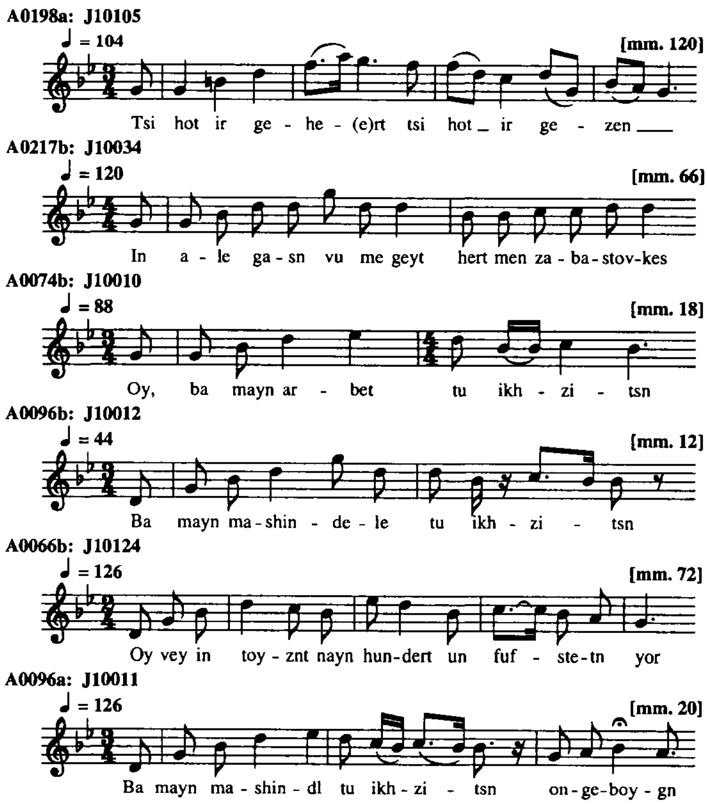


Figure 11. Another group of similar melodies from the JMF1 corpus.



12.5.3 Selecting

The sorting of the corpus according to selected attributes provided an encouraging approach to classifying an entire corpus, but may not be successful in retrieving specific melodies. In order to retrieve a specific group of similar melodies, an SQL SELECT statement was used. In this way groups of melodies that match certain criteria can be retrieved. These attributes may not normally be chosen for defining a general classification scheme. Figure 12 shows the actual parses of the melodies (from outside the Kiev corpus) shown in Figure 1.

Figure 12. The most successful parses of the melodies shown in Figure 1.

These three melodies do not share a common main contour. Furthermore, even though the ambitus of each is less than or equal to a third, a simple joint sort of these melodies with those of the Kiev corpus (according to the ambitus of the main element) did not result in any significantly increased degree of proximity between the first melody and the other two (any melody whose first main element has an ambitus of 0 would be judged more similar to the first than either of the other two).

Since the correlation between all other attributes of the melodies in Figure 12 is very high, an SQL SELECT query based on both the ambitus of the first main (those not exceeding a third) and the attributes of all other similar segments was successful in retrieving these melodies when they were added to those of the Kiev corpus.

12.5.4 Multivariate Analysis

Multivariate analysis consists of a two-dimensional mapping of the similarity between entities and is dependent on a number of weighted variables. A distance matrix lists each attribute of each melody, along with its distance from the corresponding attribute of all other melodies. The distance between any two melodies would be based on any number of relevant variables.

Although the query strategy used here is capable of combining any selection of attributes for the purpose of retrieving similar melodies, the choice of variables is relevant only to that group and must be known *a priori*. For discovering groups of similar melodies, or viewing the distribution of features within a corpus, multivariate analysis is a possible solution.

12.6 Implications for Music Information Retrieval

The eventual goal of the present system is to create a full storage and information retrieval system for melodic data. This process could eventually involve a classification system for organizing entire corpora and an indexing system used for retrieving similar melodies.

In order to evaluate such a system, benchmarks must be set. In the field of information retrieval, there are various methods of evaluating query output (Sparck Jones 1972 and 1975). Selfridge-Field (1998) discusses a wide range of requirements specific to melodic comparison. In the field of information retrieval a successful search is evaluated in terms of recall (number of hits) and level of precision (a ratio of relevant to non-relevant hits). It is important to note that a successful search result does not necessarily consist of 100% relevant hits (Lancaster 1993). This eases greatly the requirements of the system: there are no "wrong" answers.

A successful search is considered to be dependent on the exhaustiveness and specificity of the index terms used to describe the retrieved entities (Sparck Jones 1972). Exhaustiveness refers to the breadth of the description, i.e. how well it describes the totality of the retrieved item. The specificity of index terms refers to the level of precision which the terms attain (i.e. how well the specificity of the description matches the specificity of the item). The overall goal here is to be capable of supplying an overall ("exhaustive") description of a melodic incipit while providing an abstraction which is very close to the surface ("specific").

References

Adams, Charles R. (1976). "Melodic Contour Typology," *Ethnomusicology* 2, 179–215.

Balzano, Gerald J., and Barry W. Liesch (1982). "The Role of Chroma and Scalestep in the Recognition of Musical Intervals in and out of Context," *Psychomusicology* II/2, 3–31.

Clynes, Manfred (1986). "When Time Is Music" in *Rhythm in Psychological, Linguistic and Musical Processes*, ed. James R. Evans, Manfred Clynes (Springfield IL: Charles C. Thomas), pp. 169–224.

Clynes, Manfred (1997). "What Can a Musician Learn about Music Performance from Newly Discovered Microstructure Principles (PM and PAS)?" in *Action and Perception in Rhythm and Music*, ed. Alf Gabrielsson (Stockholm: The Royal Swedish Academy of Music), pp. 201–231.

Cooper, Grover, and Leonard B. Meyer (1960). *The Rhythmic Structure of Music*. Chicago: University of Chicago Press.

Cope, David (1998). "Signatures and Earmarks: Computer Recognition of Patterns in Music," *Melodic Similarity* (*Computing in Musicology* 11), 129–138.

Deutsch, Diana, and J. Feroe (1981). "The Internal Representation of Pitch Sequences in Tonal Music," *Psychological Review* 88, 503–522.

Dowling, W. Jay (1978). "Scale and Contour: Two Components of a Theory of Memory for Melodies," *Psychological Review* 85, 341–354.

Edworthy, Judy (1982). "Pitch and Contour in Music Processing," *Psychomusicology* 1, 45–46.

Howard, John (1998). "Strategies for Sorting Musical Incipits," *Melodic Similarity* (*Computing in Musicology* 11), 101–118.

Kornstädt, Andreas (1998). "*Themefinder*: a Web-based Melodic Search Tool," *Melodic Similarity* (*Computing in Musicology* 11), 231–236.

Lancaster, Frederick Wilfrid (1993). *If You Want to Evaluate Your Library*. Champaign: University of Illinois.

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

Meyer, Leonard B. (1956). *Emotion and Meaning in Music*. Chicago: University of Chicago Press.

Meyer, Leonard B. (1973). *Explaining Music*. Berkeley: University of California Press.

Narmour, Eugene (1984). "Toward an Analytical Symbology: The Melodic, Harmonic and Durational Functions of Implication and Realization" in

Musical Grammars and Computer Analysis, ed. Mario Baroni and Laura Callegari (Florence: Leo S. Olschki Editore), pp. 83–114.

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

Norvig, Peter (1992). *Paradigms of Artificial Intelligence Programming: Case Studies in Common Lisp*. San Francisco: Morgan Kaufmann Publishers.

Oura, Yoko (1991). "Constructing a Representation of a Melody: Transforming Melodic Segments into Reduced Pitch Patterns Operated on by Modifiers," *Music Perception* 9, 251–266.

Schenker, Heinrich (1956). *Der Freie Satz*. Vienna: Universal Edition.

Selfridge-Field, Eleanor (1998). "Conceptual and Representational Issues in Melodic Comparison," *Melodic Similarity* (*Computing in Musicology* 11), 3–64.

Sparck Jones, Karen (1972). "A Statistical Interpretation of Term Specificity and its Application in Retrieval," *Journal of Documentation* 1, 11–21.

Sparck Jones, Karen (1975). "A Performance Yardstick for Test Collections," *Journal of Documentation* 4, 266–272.

Stetson, R. B. (1905). "A Motor Theory of Rhythm and Discrete Succession," *Psychological Review* 12, 250–270.

Tversky, Amos (1977). "Features of Similarity," *Psychological Review* 84:4, 327–352.

Wertheimer, Max (1950). "Laws of Organization in Perceptual Forms" in *A Source Book of Gestalt Psychology*, ed. Willis Davis Ellis (London: The Humanities Press), 71–88.

Woodrow, H. (1909). "A Quantitative Study of Rhythm," *Archives of Psychology* 14, 1–66.