CREATING A NESTED MELODIC REPRESENTATION: COMPETITION AND COOPERATION AMONG BOTTOM-UP AND TOP-DOWN GESTALT PRINCIPLES

Jane Singer

The Hebrew University of Jerusalem
The Department of Musicology
jsinger@cc.huji.ac.il

Abstract

A set of principles (based on Gestalt theory) governing how we group notes into meaningful groups has been widely accepted in the literature. Based on these principles, many divergent theories of melodic segmentation and representation have been proposed. However, these theories have not succeeded in achieving a comprehensive and verifiable representation of melody. This is largely due to the fact that multiple competing segmenting factors produce, for any single melody, a large number of possible segmentations and therefore representations. Here a model is proposed, which incorporates widely accepted principles of segmentation. These rules govern three types of factors: (1) changes in proximity (for producing disjunctive segmentation), (2) changes in overall contour and intervallic texture and (3) patterns and periodicity that create parallelism among segments. Because of the nature of the segmentation rules, these same rules establish the attributes of the groups they produce. Based on original research in Singer 2004, principles for establishing preferences among competing rules are formulated in order to create a few preferred representations for approximately 1,000 monophonic folksongs.

1. Introduction

Gestalt rules of grouping, adapted from the field of visual perception, are widely cited in order to explain how the listener groups a stream of notes into meaningful groups. Since different rules deal with different parameters of varying magnitudes (note length, interval size, rests) a number of possible points of segmentation become apparent. Rather than pointing to shortcomings in the Gestalt theory, the competing factors do in fact reflect the ambiguity that exists within the listening experience. In order to build a comprehensive model of melodic segmentation, this ambiguity needs to be represented. On one hand, this eases the demands made on the model, since it is not expected to produce a single “correct” representation. On the other hand, representing such ambiguity is not achieved by simply enumerating alternate representations: each proposed description represents a choice among different organizing principles, adhering to some rules while violating others. Various writings (such as Lerdahl and Jackendoff 1983, Narmour 1990 and Bregman 1990, Tenney 1980, Temperley 2001) largely agree on the major role that proximity, similarity and symmetry play in melodic segmentation and abstraction. However, the application of these rules has remained problematic. Most models either describe only specific factors that influence segmentation or fail to implement the model on more than a few examples chosen specifically for their adherence to the relevant principles.

The proposed model includes both bottom-up rules of proximity and similarity, along with top-down rules that detect periodicity and patterns (intervallic and rhythmic) among groupings.

This paper describes a model that was described in detail in Singer 2004. This model is unique in these aspects:

(1) It includes conditions for conjunctive, as well as disjunctive segmentation (perceptible changes in the melodic surface, without an intervening disjunctive interval).

(2) It integrates top-down processes (repetition and other forms of symmetry) that help reduce the number of possible parses (from among those suggested by bottom-up parses, based mostly on proximity).

(3) It includes a full formalization and is implemented on a large body of real-world data (the Essen Folksong Database, henceforth EFD).

The purpose of this paper is not to detail the individual rules of the model, but rather to explain the logic of different rule types, and to show how they work together to choose a few preferred segmentations.

Other models of disjunctive melodic segmentation have chosen the EFD data as a source for verifying their segmentation algorithms. Bod bases his theory on long-term memory and asserts that these principles have preference over the bottom-up proximity principles, such as those proposed in Temperley and Tenney. Both Temperley and Bod’s claims are supported by quantitative results (that is, their model output matches a high percentage of the EFD segmentations).

All these models make highly valid contributions to the understanding of the level and type of segmentation represented in the EFD. Upon examining the EFD segments, however, it becomes clear that changes in the melodic surfaces suggest intermediate points of segmentation, and that in order to describe the whole segment, it becomes necessary to break it down into subunits (such as upbeat figures, ascending and descending figures, etc.). Consequently, in order to create a representation of melodic abstraction, the melodic instances need to be exhaustively segmented into highly coherent units that can be accurately described. These types of groupings cannot always be accounted for by disjunctive segmentation.

In order to define changes in melodic surface, it becomes necessary to define conjunctive as well as disjunctive segmentation. The proposed model considers changes in melodic movement, such as overall contour and intervallic texture or makeup as segment-inducing factors. Conjunctive segmentation is necessary in order to represent melodies, since these attributes (contour and intervallic texture) can change without being reinforced by disjunctive segmentation. The conditions for perceiving overall contour of a segment were discussed in Singer 2004. Chapter 4 of this thesis conducts a listener experiment to show that under certain circumstances listeners are willing to fully ignore note-to-note movement and choose overall contour as the most salient feature. (Briefly, this is shown to be true by the fact that the listener identifies “most similar” pairs as being those with similar contour, not as those with matching note-to-note patterns, under the defined conditions.) The constraints imposed on the examined material in this listener experiment (the melodies) were incorporated into the model’s definition of single-contour segments.

Chapter 5 of the same thesis discusses the competition between top-down and bottom-up processing. It is shown that within the EFD, segmentation is imposed on the note stream with a preference towards top-down principles. Top-down processing benefits from the wisdom of hindsight: it is able to judge the relative saliency of past events, detect parallelism between groups, and create an optimal segmentation hierarchy. Although top-down parsing probably always takes place (listeners always review past events, while they are taking in the present-sounding note), it revises, rather than nullifies, past formulations. Past and present events remain in competition.

One of the most difficult problems of implementing the Gestalt principles within a model is the fact that too many possible points of segmentation are suggested by the rules. Because a change in proximity according to the bottom-up principles is relative, almost every interval can be a prospective boundary. Although many possible points of segmentation can be eliminated as having little segmenting influence, many competing elements emerge as possible points of segmentation.

The proposed model contends with the multiplicity of contributing factors according to principles gleaned from the EFD, and introduces non-disjunctive segmentation according to changes in contour and intervallic textures (Singer 2004, chapter 4).

The purpose of the present paper is to demonstrate how these multiple rules work together, specifically how some rules help eliminate the multiplicity of parses produced by the proximity principles.

2. Grouping principles

The proposed model is based on two types of principles: top-down and bottom-up. The bottom-up/top-down dichotomy is not a clear one since many top-down principles operate on a level very close to the melodic surface, while many bottom-up processes (such as extended pauses), can reinforce the separation between entire movements. The segmentation principles are described here as intra-group (evaluating a possible note group) and inter-group (evaluating contiguous proposed groups to detect different types of parallelism). The first group is event based, whereas the second seeks out patterns.

2.1. Intra-group rules (“bottom-up”)

Upon hearing a melodic unit (phrase, period or other), the listener may perceive a number of possible groupings. The level of ambiguity may vary from melody to melody, from performer to performer and even from listener to listener. The distinctness of a group is determined by (1) the cohesiveness of the group, as well as (2) the contrast between the attributes of the group (such as contour) with those of the neighboring groups.

The cohesiveness of any single grouping is dependent on a number of factors:

1. The size of the intervals within the proposed group (small intervals).
2. The durations of the notes within the group (short notes).
3. The continuity of overall direction (no peaks).
4. The level of similarity of intervals.

Two important features of melodic description (nos. 3 and 4 above) are **overall contour** and **intervallic texture**; these consist of abstractions of intervallic information that are not linked to a single event, and can only be formulated from all intervals within a segment. If the group is not delimited by a disjunctive interval, there is no actual separation (disjunction), and the saliency of such group borders is low. Because the cognition of contour and texture requires an abstraction based on information from all the intervals within the segment, it constitutes a top-down process, although it operates at a level close to the melodic surface. **Figure 1** demonstrates identical contours, each with different intervallic makeup.

Within the proposed model, figures that adhere to a single contour (with no salient peaks), maintain consistent melodic makeup and have no intervening disjunctions are referred to as **Primitives**. These groupings constitute the most coherent and basic melodic units. Consecutive conjunctive Primitives that maintain a single direction (without intervening peaks) are recombined into **Contours**. Contour pairs that consist of an anacrusis followed by a trochaic are joined into **Constructs**.

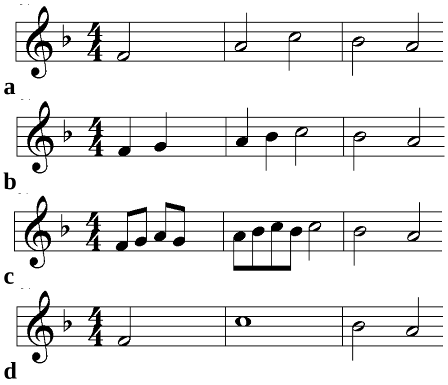


Figure 1: Identical contour, differing intervallic makeup (for the ascending figure): (a) thirds, (b) ascending seconds, (c) zigzag, (d) jump.

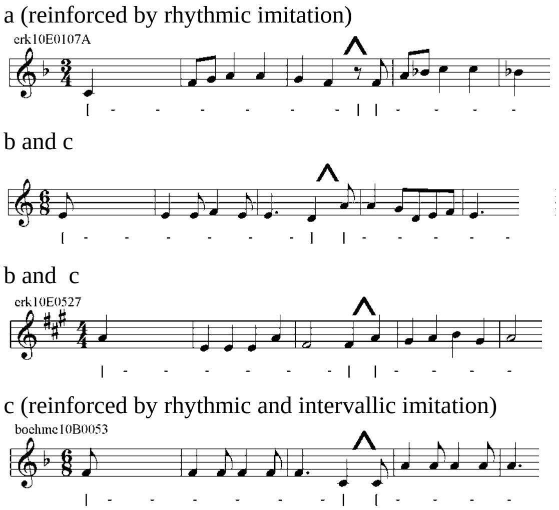
Consecutive Contours can be either conjunctive or disjunctive. A **Disjunction** consists of a specialized interval that is perceived as a separation between segments, rather than belonging to a segment. Within the model, the rules governing the identification of Disjunctions are largely based on Gestalt principles. In addition a number of metric qualities that are typical to disjunctive intervals are considered as factors.[1]

2.2. Inter-group rules

The level of segmentation identified in the EFD can be accounted for by parallelism-identifying principles. At this level, one of the most important principles is that of equal-length segments. It was found (Singer 2004, chapter 5) that within the EFD, of the 10,000 melodies that were checked (the first two segments of each), 9410 of these had segments of equal measure count. Since most segments do not begin on a downbeat (meaning that segments are not made up of whole measures), the number of measures is expressed in downbeat count, rather than beat count. Since evoking this method isolates the measure (referred to as the median measure) where the disjunction probably occurs (and not the exact event), other factors needed to be considered. Proximity (specifically the relative size of the IOI[2]) was one factor that helped to isolate the exact position of the boundary within the median measure, even at this higher level.

A group of principles were gleaned from the EDF that accounted for over 90% of the segmentations examined (9545 of 10395). These

a. The longest note (IOI) in the median measure
b. The largest interval in the median measure
c. The point in the median measure that matches the metric position of the first note of the incipit


a (reinforced by rhythmic imitation)

b and c

b and c

c (reinforced by rhythmic and intervallic imitation)

Figure 2: Some examples of parallelism-creating boundaries.

2.3. Competition and cooperation among the rules

In order to incorporate multiple rules within the model, it becomes necessary to define a system for weighting the principles and factors that is capable of choosing the preferred segmentations.

2.3.1. *The problem of over-segmentation*

Figure 3 demonstrates the recombination of Primitives into Contours (Contours often contain only a single Primitive), and Contours into Constructs. In addition, a top-down segmentation is included (specifically the segmentation represented in the EFD). Here, the number of measures (4) divides evenly into two segments (2+2), and the top-down (EFD) segmentation conforms to rules a, b, and c in 2.2, as well as being reinforced by a repeating rhythmic pattern. The EFD segmentation concurs with the bottom-up parsing, since the last note of the segment is longer than any previous note (reinforcing the proximity principle). Nonetheless, in the case where little or no ambiguity is generated by competing rules, even a single rule can trigger over-segmentation. According to the proximity principle, when applied to the time axis (duration), every long-short note pair can be a possible point of segmentation (Disjunction), and every short-long note pair becomes a possible segment (Primitive). The rules for creating single-direction groups (Contours) help overcome this over-segmentation by preferring longer groups (greater than two notes) of a single Contour. Therefore in Figure 3, we see two alternate interpretations at the Contour level. One inserts a disjunction and upbeat; the other prefers including the entire first measure within a single Contour. The preference for longer groups, together with metric

[1] In Singer 2004 (chapter 5), it is shown that disjunctive intervals show a strong tendency to be intervals that begin on the beat, and a lesser tendency to end off the beat. Therefore, most EDF segments began with upbeat figures.

[2] Inter-onset-interval, the time span between the onset of a note and the onset of the next note.

restrictions imposed on the Primitive (that forbid a Primitive from extending beyond the first note of the next measure[3]), restricts the length and number of Primitives that are included within the preferred set of segmentations.

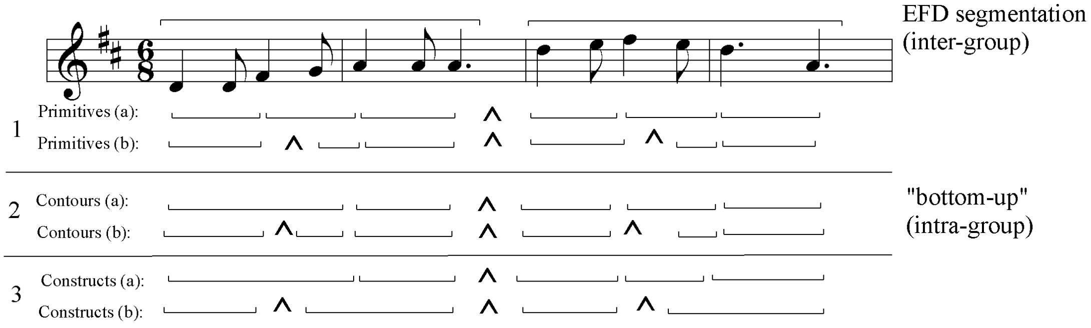
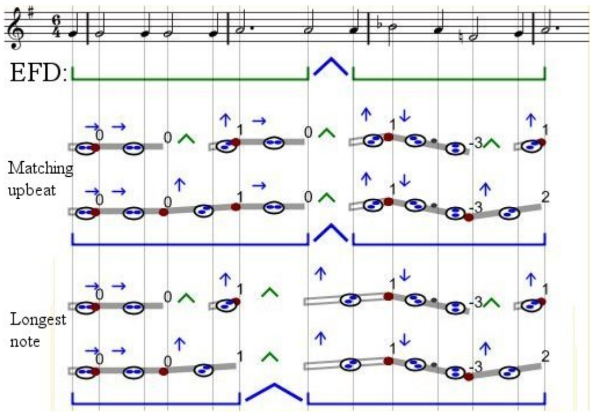


Figure 3: Intra-group (bottom-up) and inter-group (bottom-down) segmentations of a melody.



2.3.2. *The problem of top-down/bottom-up rule competition*

In order to replicate the type of segmentation found within the EFD, it is often necessary to override salient expressions of the proximity rule. The top-down rules often break up groupings that are strongly suggested by long notes and large intervals. The proximity rule, however, is not totally ignored; many points of segmentation are located at the longest note, within the median measure. Figure 4 (an example of the actual model output) depicts a melody that shows a clear parallel division into two parts. As in Figure 3, the two-measure segments are strongly reinforced by the repetition of a rhythmic pattern. Therefore, although the second measure is strongly suggested as the measure of segmentation, the exact point of segmentation is not apparent. The point chosen in the EFD can only be accounted for by the matching upbeat metric positions and durations (upbeat to m. 1 and m. 3). This disregards the first note of m. 2 as a point of segmentation (the longest note in the proposed segment). Within the model output (Figure 4), this second point of segmentation is suggested as an alternative.

3. Conclusions

The proposed model is to identify and represent the most salient features of the EFD melodies. The graphic output of the model (see Figure 4) includes a few (1-3) alternate parsings for all levels of the model. Over 90% of the melodies in the output include a segmentation that matches the original EFD segmentation.

The full set of graphic representations of the model output (along with the input data and XML output) can be found at: http://shum.huji.ac.il/~jsinger/thesis/files.htm.

References

Bod, R. “Memory-based models of melodic analysis: challenging the Gestalt principles,” *Journal of New Music Research* 30, 3 (2001): 27-37

Bregman, A.S. Auditory Scene Analysis: the Preceptual Organization of Sound. (Cambridge, Massachusetts, 1990).

Lerdahl, F. and R. Jackendoff, *A Generative Theory of Tonal Music*. Cambridge, MA, 1983.

Narmour, E., *The Analysis and Cognition of Basic Melodic Structures*. Chicago, 1990.

Schaffrath, H. The Essen Folksong Database , 1990 (including later data encoded by Ewa Dahlig, Damien Sagrillo and David Halperin).

Singer, J. *A Model of Melodic Representation: Towards an Implementation of Music Information Retrieval* (PhD Thesis). The Hebrew University of Jerusalem, 2004.

Temperley, D. *The Cognition of Basic Musical Structures*. Cambridge, MA, 2001.

Tenney, J. and L. Polansky, "Temporal Gestalt Perceptions in Music," *Journal of Music Theory* 24 (2000): 205-241.

[3] Additional metric constraints are applied to the segments at all levels. Briefly, the strong beats, like peaks, are salient events that interfere with the coherency of the groupings at different levels; therefore they induce some degree of segmentation.