Discovering Grouping Structure in Music

Jacqueline A. Jones, Benjamin O. Miller, and Don L. Scarborough
Department of Computer and Information Science and Department of Psychology
Brooklyn College, City University of New York
jajbc@cunyvm, bombc@cunyvm, dosbc@cunyvm

Abstract

GTSIM, a computer simulation of Lerdahl and Jackendoff's (1983) A Generative Theory of Tonal Music, is a model of human cognition of musical rhythm. GTSIM performs left-to-right, single-pass processing on a symbolic representation of information taken from musical scores. A rule-based component analyzes the grouping structure, which is the division of a piece of music into units like phrases and the combination of these phrases into motives, themes, and the like. The resulting analysis often diverges from the analysis we would produce using our musical intuition; we explore some of the reasons for this. In particular, GTSIM needs to have an algorithm for determining parallel structures in music. We consider alphabet encoding (Deutsch and Feroe, 1981) and discrimination nets (Feigenbaum and Simon, 1984) as algorithms for parallelism.

Introduction

We have been developing a computer simulation of Lerdahl and Jackendoff's (1983) A Generative Theory of Tonal Music (henceforth GTTM) as a model of human cognition of musical rhythm. Our computer simulation, called GTSIM (Jones, Miller & Scarborough, 1988) is a rule-based model with a neural network component. The simulation performs left-to-right single-pass processing on a symbolic representation of information taken from musical scores. Three aspects of music are analyzed: a rule-based component determines metric structure (Miller, Scarborough & Jones, 1988), a neural network determines the tonality, or perceived key, at any point in the score (Scarborough, Miller & Jones, 1989), and another rule-based component determines some aspects of the grouping structure.

We have recently integrated several modules of our model. Now that the modules have been integrated, we are beginning to construct algorithms for their interaction. In particular, we are trying to use strong beats in the metric structure to help find the correct grouping analysis. Grouping analysis is the process by which we divide a piece of music into units like phrases, and then combine these phrases into motives, themes, and the like. While the integrated analysis provides an approximation of the lowest level of grouping boundaries (phrase boundaries) in many cases, the cases for which it fails raise questions about the theory. Recognition of parallelism in music, not yet implemented in our model, seems to be an essential component for producing correct grouping analyses.

Background—GTTM

Lerdahl & Jackendoff's GTTM partitions rhythm into two independent hierarchical components: metric structure and grouping structure. Metric analysis yields a hierarchical representation of metric structure which conforms to

traditional intuitions about meter and accent. The hierarchy represents the strength of the beat at evenly spaced times in the music. Stressed notes (strong beats in the music) correspond to the highest levels of the metric hierarchy. Grouping analysis yields another hierarchy, reflecting intuitions about musical phrases, motives, themes, etc. Grouping preference rules (GPRS) tell us where to find group boundaries, while grouping well-formedness rules tell us how to construct a legal grouping hierarchy from the first level groups. Grouping and metric analysis in GTTM are largely independent, and each can, to a large degree, be carried out without the other. While GTTM's analysis tries to find the best fit between meter and grouping, one cannot be inferred from the other.

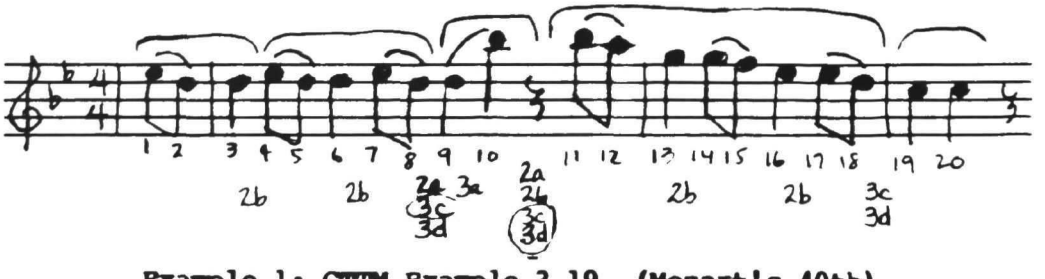
Our Model—GTSIM

We have attempted to devise a model of the process by which a human listens to and understands music. To this end, we have devised a system which processes music from beginning to end, without backtracking. Backtracking - going back and making a second pass through the music once one has heard the entire piece - is not a reasonable model of how humans process music. All our algorithms are constrained by the limits of human memory.

Application of the GTTM Grouping Rules

The grouping module of GTSIM identifies potential grouping boundaries in the score, based on proximity of note onsets or offsets and on significant differences in such attributes as pitch, duration, and articulation (defined in Lerdahl & Jackendoff's grouping preference rules (GPRs) 2 and 3, and their subrules). It places a marker between two notes if there is an application of a rule at that point. The transition point thus marked is a candidate for being an actual group boundary. Our module has successfully marked the rule applications which Lerdahl & Jackendoff find in their own examples. However, we also tend to find spurious candidate boundaries as a result of rigorous application of the rules.

Example 1 shows our initial grouping analysis of the melody at the beginning of Mozart's 40th Symphony, Lerdahl & Jackendoff's Example 3.19. The rule applications which we find and that they do not are circled. In all examples, rule applications are shown below the score, and the groups determined by algorithm INTEG2, described below, are marked above the score.



Example 1: GTTM Example 3.19 (Mozart's 40th)

The three extra boundaries come about through rigorous interpretation of the slur-rest rule, the articulation rule, and the duration rule. We interpret the 8-9 and 10-11 transitions as being boundaries between notes of different articulation, since the notes are not slurred together, and since notes 9 and

11 must be articulated at their onset just like any other non-slurred note. The duration rule applies at 10-11 as well, since notes 9 and 10 are the same length, as are notes 11 and 12, while notes 10 and 11 are of different lengths.

Choosing First Level Boundaries

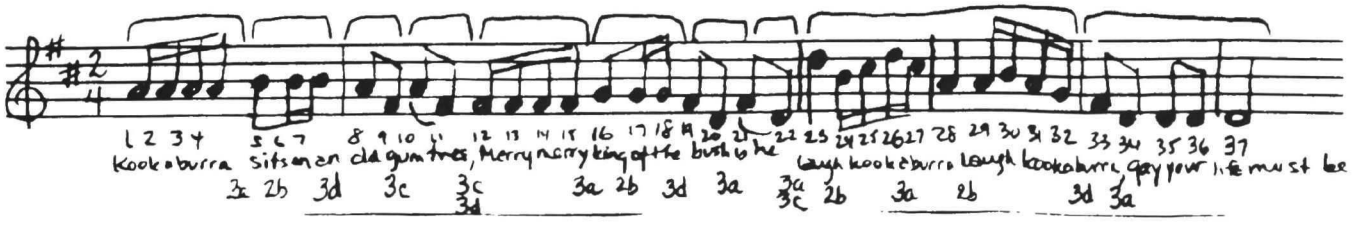
Marking all the rule applications is only the first step in creating the grouping hierarchy. Once the candidate boundaries have been identified by the rule applications, we must decide which of the candidate boundaries are the actual boundaries between groups (phrases). These boundaries divide the piece into the groups that constitute the first level of the grouping hierarchy. Next, we must begin to combine these groups into ever larger units--groups of groups--limited in principle only by the length of the piece of music itself.

Not all of the transitions marked as candidate boundaries can be actual group boundaries. In Example 1, selecting the boundaries at transitions 8-9, 9-10, and 10-11 would violate the GTTM rule that no group should consist of one note. Furthermore, a grouping structure of this sort would violate our musical intuition of how this piece is grouped. In other pieces, there are candidate boundaries between 7 (or more) notes in a row. "Greensleeves," for example, has candidate boundaries at almost every transition (see discussion below). It is not possible for each of these transitions to mark a new phrase in the music.

Our initial attempt to select first level group boundaries used a simple counting algorithm. The algorithm counts the candidate boundaries at each point; the candidate boundaries with more rule applications are selected as actual boundaries; those with the most rule applications are considered larger-level boundaries. The algorithm works well for some pieces; for "Row, Row, Row Your Boat," it nicely divides the piece into four lines with a major break after line 2; for other pieces, it produces no groups at all (some pieces have no transitions with more than one rule application).

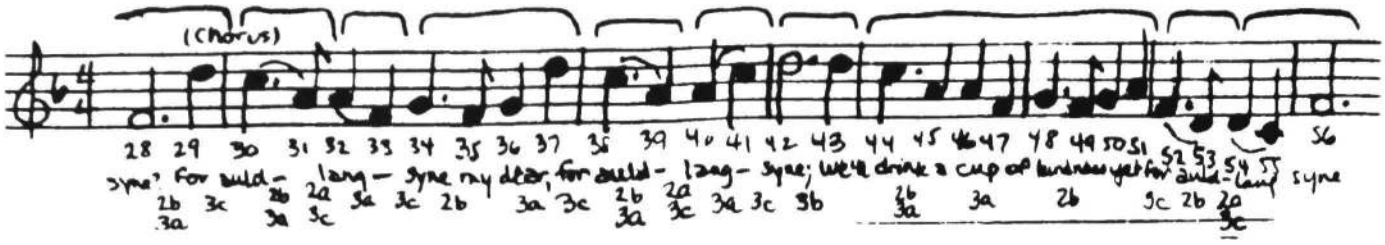
Two more recent algorithms, INTEG1 and INTEG2, are based on the integrated metric and grouping analysis. One of GTTM's metric preference rules says to prefer a metric analysis in which the first note in a group falls on a strong beat. INTEG1 looks at candidate boundaries with more than one GTTM rule application and checks to see whether the note that would begin the group so delineated falls on a strong beat in the metric hierarchy. INTEG1 also produces uneven results; first, many pieces have no transitions with more than one rule application; and second, sometimes a single application of GPR 2 (which identifies rests, among other things) outweighs many applications of GPR 3.

INTEG2 provides better results. All notes which seem to begin groups, based on the fact that they follow one or more rule applications, are checked to see whether they are at a higher metric level than the two adjacent notes. If they are, they are considered to be first level group boundaries. Example 2 illustrates a successful analysis.



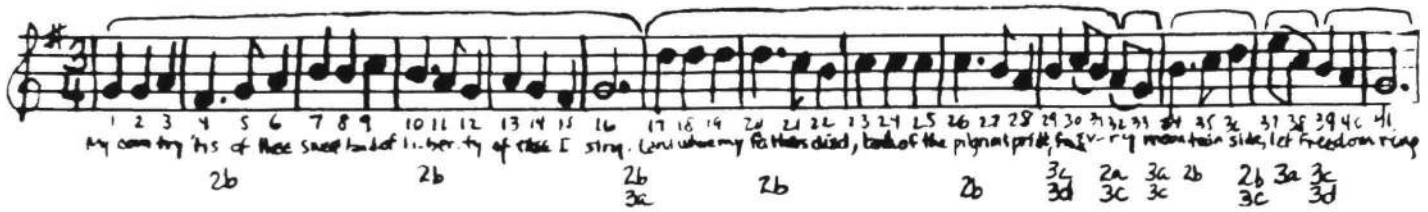
Example 2: Kookaburra

INTEG2 ensures that we will capture only groups which begin on primary or secondary strong beats in a bar; phrases that begin on upbeats, such as those in "Farmer in the Dell" and "Auld Lang Syne" (Example 3, transitions 28-29, 36-37, 42-43, and 50-51) will be ignored.



Example 3: Auld Lang Syne (chorus)

The groups which are established by INTEG2 are not always those which we would pick by looking at the score or by listening to the music. Often the groups are of irregular size throughout a piece; one section will be completely undivided, while another section will be overly divided into many small groups.



Example 4: America

Partly this reflects the fact that we have not yet implemented all the grouping preference rules; one of the significant markers of phrasing in music is parallelism, yet we have not yet attempted to implement GTTM's parallelism rule (see discussion below).

Higher Level Boundaries

Eventually, all the first level groups must be combined into a single well-formed hierarchy of groups. This will be carried out by imposing the higher level rules on the evidence accumulated by the lower level rules. We have attempted to do this only with the first simple counting algorithm.

Discussion

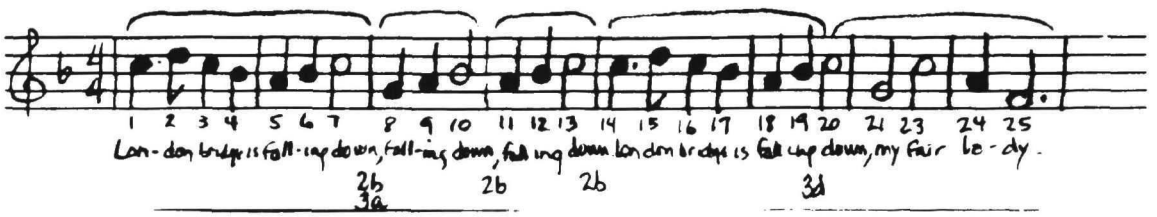
Our attempt to develop a computational model of this formal theory has brought us to a clearer understanding of the limitations of the theory. We also have more awareness of the difficulties in modelling the theory. There are problems at each of the levels of analysis.

Rule Application Level

One problem at this first level is that strict interpretation of the rules causes GTSIM to find excessive numbers of rule applications. Automated application of the rules finds duration differences, as in Example 1, transition 10-11, which Lerdahl & Jackendoff ignore, seemingly because it crosses a rest, an issue unaddressed by GTTM. In other pieces, like "Auld Lang Syne" (Example 3),

"We Three Kings," and "America" (Example 4), there are sections where almost every note has one or more rule applications. Music which alternates long and short notes, particularly music with dotted rhythms, tends to produce many extra attack-point rule applications (GPR 2b). Music that has melodic pitch skips will produce extra pitch rule applications (GPR3a), and music that alternates between slur and standard articulation will have extra articulation (GPR 3c) and slur-rest (GPR 2a) rule applications. Most of these rule applications ought to be ignored but, once produced, must be processed.

A second problem is that other clearly heard boundaries are excluded by the GTTM rules. For example, in "London Bridge," transition 20-21 is not an application of the pitch rule, although a listener feels that it ought to be, because the change in pitch from notes 20 to 21 is the same as the change in pitch from notes 21 to 22. If note 22 were even a half-step lower, 20-21 would be marked as a pitch boundary. The rule correctly rejects the 21-22 transition for the same reason (it is the same pitch difference as 20-21), but this still does not seem satisfactory.



Example 5: London Bridge

A third problem is that some of the folksongs we have used as sample scores do not have clear grouping boundaries to the eye or ear. We intuitively assume that the music divides at commas, or at the ends of lines, but there is often no evidence other than linguistic for finding group boundaries at these points.

On closer inspection, the music sometimes does provide evidence of grouping at such points. Often the music contains parallel sequences which correspond to the linguistic divisions. Two sequences can be said to be parallel if they instantiate the same pattern. A melodic figure and its verbatim repeat are certainly parallel sequences, but parallelism is not necessarily limited to identity. Two musical sequences can be considered parallel if they are similar in rhythm, pitch contour, or internal grouping, or if one embellishes or simplifies the other without changing its essential melodic or rhythmic characteristics. GTTM has a parallelism rule, which says that parallel segments should be combined into parallel parts of higher level groups. However, it does not specify how to pick out the parallel groups and mark them in such a way that they can later be combined into larger groups.

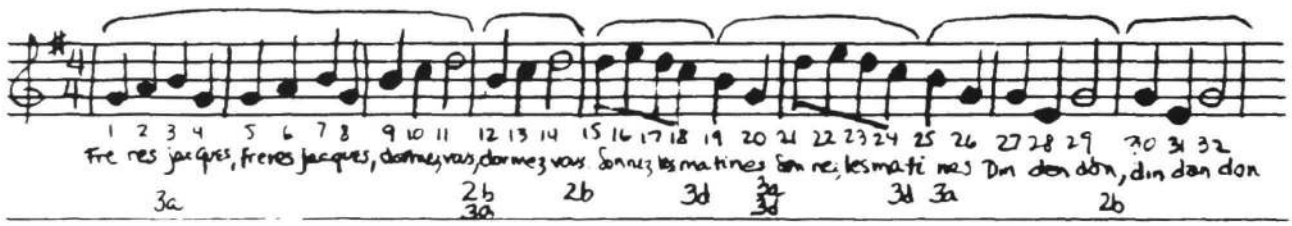
Application of a parallelism rule would find other candidate boundaries, and often these would coincide with the linguistic boundaries. Examples of this can be seen in the first four bars of "America" (Example 4), where there are only two candidate boundaries, neither of which reflects our perception of the actual grouping. However, the first six notes are clearly melodically and metrically parallel to the second six notes; a parallelism rule would find boundaries at transitions 6-7 and 12-13. An algorithm which discovered this parallelism would greatly enhance the analysis. Similarly, the first eight notes of "Freres Jacques" (Example 6) form two parallel groups that are not delineated in any way by the existing rules.

Selecting Group Boundaries Level

Selecting the appropriate first level boundaries is not as simple in a computational model as in a formal theory. The formal model explains the concept, but doesn't need to worry about reasonable implementation of that concept. Thus Lerdahl & Jackendoff's discussion of the selection process involves backtracking, weighting of rules, and determination of parallelism, all of which are computational problems.

Backtracking--applying rules in retrospect, after hearing the entire piece--is not a reasonable model of how humans process music. In their explanation of Example 1, Lerdahl & Jackendoff reject the boundary at transition 9-10 by noting that the boundary at transition 10-11 divides the example in half, and is therefore correct; since there cannot be a group consisting of one note, the 9-10 boundary is incorrect. As a listening model, this assumes that one can listen ahead to the end of the section, determine where the section ends, and then backtrack and decide that the 10-11 boundary marks its division into two large groups. This much backtracking, which requires storing 20 notes in memory, seems unlikely as a model of human performance.

Another problem is selection of first level group boundaries when there is a conflict. Even one rule application, of the right kind, can indicate a true boundary; the conjunction of many rule applications is also strong evidence. However, without some weighting, this decision process cannot be automated. Lerdahl & Jackendoff suggest using a system of weights applied to the rules, where GPR 2 outweighs GPR 3, except when GPR 3 measures a change in dynamics. They do not specify the weighting further. Thus they apply the rules in a rather ad hoc fashion: in Example 1, two rule applications at transition 8-9 are ignored in favor of two at transition 10-11. In other of our own examples, a transition with one rule application clearly (to the listener) outweighs a transition with two or more rule applications. In "Freres Jacques" (Example 6), the boundary that divides the piece in half (transition 14-15) has only one rule application (GPR 2b), while the less important transition 11-12 has two rule applications. At other times, that same single rule application must be ignored, as in "America" (Example 4).



Example 6: Freres Jacques

GTTM's Intensification rule (GPR 4), says that a larger-level boundary may be placed where the effects picked out by GPRs 2 and 3 are relatively more pronounced. This also suggests the need for weighting the effects discovered at the rule applications. Although we have rejected unchanging weights, we plan to incorporate dynamically determined weighting into our next algorithm, noting Deliege's (1987) experiments with weighting GTTM's grouping rules.

A final problem is parallelism. Parallelism is represented in GTTM as a higher level rule, that is, a rule by which to form larger groups from smaller groups (GPR 6). Even if there are other rule applications at the ends of groups, it is the perception that one set of notes is parallel to another set

of notes that enables us to select the correct grouping boundaries. Without parallelism, grouping analyses are improperly segmented. By finding out what we can do without parallelism, we have discovered just how potent a psychological argument parallelism is in determining grouping structure.

Recognition of parallelism, however, is a difficult pattern recognition problem. There are several possible approaches. One algorithm will use a modified discrimination net, modeled after EPAM-III, a model of recognition and learning devised by Simon and Feigenbaum (Feigenbaum & Simon, 1984). EPAM-III was developed to learn to recognize strings of symbols such as a sequence of phonetic features, or a letter sequence.

Another model which we expect to use as a guide to recognizing parallel structure is alphabet encoding. Any sequence can be described in terms of an alphabet that contains all the elements that occur in that sequence and a set of operators that describe transitions between elements and groups of elements. We can encode the same sequence differently by using a different alphabet or set of operators. Alphabet-based coding allows us to represent the hierarchical structure of a sequence; in this way, sequences are reduced to coded chunks that are easier to remember and match with other chunks.

Deutsch and Feroe (1981) assume that pitch sequences are stored internally as hierarchical networks, and use alphabet encoding to represent a hierarchy of nested patterns and subpatterns. The idea that music is represented by such structures accounts nicely for the fact that recognition of melodies is not affected by transposition or, within limits, by tempo changes. At a more detailed level, the complexity of a formula can predict how accurately the corresponding musical sequence is perceived (Jones, Maser & Kidd, 1978). While their model is a strong intuitive representation of the concept, it is not obvious how to implement it within the constraints of our model.

Conclusion

Our computer model of Lerdahl & Jackendoff's GTTM calculates preliminary grouping structure. However, without weighting the grouping rules, and without adding a component which recognizes parallelism, an important psychological factor in determining grouping, we will not get accurate results. Since GTTM does not address these issues, we must develop our own algorithms.

Acknowledgements

This research was funded in part by PSC-CUNY grants to Jones and Scarborough, and by an NSF graduate fellowship to Miller.

References

Deliege, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. Music Perception, 4, 325-360.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. Psychological Review, 88, 503-522.

Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. Cognitive Science, 8, 305-336.

Jones, J. A., Miller, B. O., & Scarborough, D. L. (1988). A rule-based expert system for music perception. Behavior Research Methods, Instruments and Computers, 20(2), 255-262.

Jones, M. R., Maser, D. J., & Kidd, G. R. (1978). Rate and structure in memory for auditory patterns. Memory and Cognition, 6, 246-258.

Lerdahl, F., & Jackendoff, R. (1983). A Generative Theory of Tonal Music. Cambridge, MA: MIT Press.

Miller, B. O., Scarborough, D. L., & Jones, J. A. (1988). A model of meter perception in music. Proceedings of the Tenth Annual Conference of the Cognitive Science Society (pp. 717-723). Hillsdale, NJ: Lawrence Erlbaum Associates.

Scarborough, D. L., Jones, J. A., & Miller, B. O. (1988). An expert system for music perception. Proceedings of the First Workshop on Artificial Intelligence and Music, AAAI-88 (pp. 9-19). Menlo Park, CA: American Association for Artificial Intelligence.

Scarborough, D. L., Jones, J. A., & Miller, B. O. (1989). Modelling music cognition: An expert system. Proceedings of The Arts & Technology II: A Symposium (pp. 132-146). New London, CT: Center for Electronic and Digital Sound at Connecticut College.