A Model of Meter Perception in Music

Benjamin O. Miller, Don L. Scarborough and Jacqueline A. Jones
Brooklyn College of the City University of New York

ABSTRACT

A fundamental problem in music cognition is the question of how the listener extracts the music's temporal organization. We describe a model, implemented as a computer simulation, that constructs a hierarchical representation of metric structure that conforms to the requirements of Lerdahl & Jackendoff's (1983) generative theory. The model integrates bottom-up processing of score data with top-down processes that generate predictions of temporal structure, and with rules of organization that correspond to musical intuition. Several examples of the program's output are used to illustrate these processes.

INTRODUCTION

People perceive patterns in temporal events even in the absence of any physical cues. For example, we hear a sequence of identical equally spaced tones as being grouped by twos or possibly threes (Fraisse, 1982), with the first tone of each group accented. When physical cues, such as accents or pauses, are present, perceptual grouping is that much more robust. A result of such perceptual processes is meter. A meter specifies a perceived pulse that marks off equal intervals in time. Pulses tend to be grouped, with the first of each group heard as accented. Within groups, strong and weak pulses alternate in a way that reflects hierarchical organization, as illustrated below. In the dot notation introduced by Lerdahl & Jackendoff (1983), the numbers at the top represent successive equally spaced points in time, and the dots below are pulses. The first row of dots shows a pulse at each successive point in time. Pulses with dots at more than one level are perceived as stressed relative to others, and they hierarchically organize the pulses into pairs, pairs of pairs, and so on. Of these levels, the most perceptually salient is what we intuitively call the **beat**, or what Lerdahl & Jackendoff call the **tactus**.

```
          Pulse  1  2  3  4  5  6  7  8  9  10  etc.
              1  .  .  .  .  .  .  .  .  .  .
Metric level  2  .     .     .     .     .
              3  .           .           .
```

Rhythm emerges from the interaction of metric structure with auditory events. For example, syncopated rhythms occur when perceived musical accents are heard as occurring at unaccented positions in the metric structure. Thus, musical events are heard within a framework established by the metric structure. On the other hand, the music itself must also guide the listener in establishing a metric framework for the interpretation of the music. We are interested in how a listener discovers a metric structure.

We have been developing a model of meter and rhythm perception based on Lerdahl & Jackendoff's (1983) *Generative Theory of Tonal Music*, or GTTM. GTTM attempts to formalize the intuitions of a human listener regarding classical Western tonal music. The theory does this by means of four stages of analysis, each embodied in two sets of rules. Well-formedness rules (WFRs) are analogous to grammatical rules, in that they specify legal structures within a stage, and preference rules (PRs) correspond to laws of perceptual organization. We are modeling, as a computer simulation, the first stage, metric analysis, which yields a hierarchical representation of metric structure that conforms to traditional intuitions about meter and accent.

In previous work, there have been five very different computer simulation models of meter and rhythm perception. Differences among the models have to do with: the method of analysis--that is, note-by-note or all-at-once (i.e. the entire score is available in constructing the analysis); the type of data used (i.e. time intervals only or time and pitch); their ability to deal with anacrucis (initial unaccented notes), syncopation, and ambiguity; and the nature of their output. All simulations are limited to single voice input.

Simon's (1968) all-at-once LISTENER program groups the note durations it uses as input into units of equal duration. From this it extracts repeating rhythmic phrases that may span several measures. LISTENER does not distinguish between meter and rhythm. Longuet-Higgins & Steedman (1971), developed a note-by-note parser that adopts the first note-value as the basic metric unit and adjusts it based on successive note values. It cannot handle passages of notes of equal duration. A later program by Steedman (1977), using the output of the Longuet-Higgins & Steedman program, makes a second pass through the score, considering not only note-values but also melodic repetition, assuming that size and separation of a figure and its repeat reflect the metric structure.

A fourth program (Longuet-Higgins & Lee, 1982) returns to a time-only orientation as well as a note-by-note approach. The program takes a list of note values (onset to onset) as input, and uses four productions to generate a metric unit and the location of bar lines as output. This program is a major improvement over earlier programs. It handles syncopation and anacrucis, and it is also hierarchical, identifying more than one metric level as it progressively expands the metric unit. However, since it analyzes until the metric unit reaches a maximum size of one whole note and then stops, it cannot detect changes in meter in the middle of a piece, nor does it produce an analysis of the whole piece.

Grid theory (Povel, 1984) is a different approach to time-only all-at-once analysis. A grid is analogous to a metronome that ticks at a uniform rate. If a metronome ticks while music plays, we can talk about the fit between the two in terms of how many ticks coincide with note onsets (**hits**; the more the better) and how many note onsets are not accompanied by ticks (**misses**; the fewer the better). For a given piece, the fit of a grid is a function of that grid's unit (inter-tick interval) and phase (where the ticks start relative to the music). The levels of a metric hierarchy in GTTM can be thought of as grids with the same phase and different units. Povel's insight provides a way, in principle, of quantifying the notion of the tactus, an aspect of GTTM theory that is of great psychological importance. We suggest that the tactus is that level (grid) in the metric hierarchy that maximizes the ratio of positive evidence (hits) to negative evidence (misses).

Each of the above models produces as output some aspect of the metric or rhythmic structure: LISTENER extracts rhythmic groups or phrases; the programs by Longuet-Higgins, Lee, and Steedman extract the time signature, with some additional information about the grouping structure within each measure; Povel's program identifies a single metric unit as the beat or tactus. None of these simulations provides a complete analysis of the metric hierarchy. Furthermore, only Longuet-Higgins & Lee provide a psychologically plausible model of how listeners discover metric structure. The present work addresses these shortcomings.

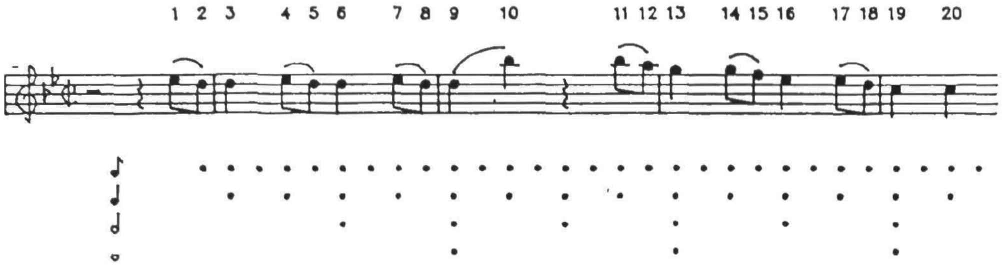
Where Povel's program selects, out of a large set, a single best-fitting grid, our model generates a family of grids representing a metric hierarchy that satisfies the GTTM metric well-formedness rules. In addition, this must be done as the program "listens" to the score rather than all-at-once. To this end, we have adapted the production rules developed by Longuet-Higgins & Lee (1982). We have added to these rules by providing: a way to generate the entire hierarchy; criteria for excluding levels generated by the rules but not acceptable in the context of GTTM; and a means of generating those levels not generated by Longuet-Higgins & Lee's rules but required by GTTM.

THE MODEL

In overview, our model has three types of processes: 1) Bottom-up processes that note time intervals between successive note onsets as they occur; 2) Top-down processes that take these intervals and combinations thereof and use them to predict the time of future events, with different intervals leading to different predictions. These predicted event times correspond to Povel's grids; and 3) Processes that evaluate the various predicted metric units or "grids" for consistency with a well-structured hierarchy as specified by GTTM.

Below is our analysis of Mozart's Symphony no. 40 (first violin part). It is comparable to GTTM's analysis (Lerdahl & Jackendoff, 1983, p. 23), with two differences. First, for reasons to be described, our analysis does not include the larger metric levels. Second, none of the metric levels in our analysis begins on the first note of the symphony, because doing so is not psychologically plausible. Once the program has identified metric levels, it generates expectations at each level, like a listener who has "got the beat." Getting the beat, however, takes some time. Once this is done, the program generates a metric structure that conforms to the well-formedness rules (WFRs): there is a dot at every note onset (WFR 1); if there is a dot at level L there is a dot at L-1 (WFR 2); dots at level L group dots at L-1 either by twos or by threes (WFR 3); temporal intervals (represented by distance) between dots at any level are uniform (WFR 4). Note also that strong beats (those with dots at higher levels) coincide with onsets of longer notes. This agrees with Mozart's bar lines and with preference rule 5.

The program begins by placing a marker, T1, at the onset of the first note. When it reaches the beginning of the second note, it places another marker, T2, and establishes the T2 - T1 interval as the current metric unit (MU). The hypothesis that this interval is a unit in the piece's metric structure predicts that another onset will occur one metric unit further on, so the program projects a third marker, T3, one MU into the future, i.e. at T2 + MU. The hypothesis is made explicit in a **metronome**, a process that is used to generate a level in the metric hierarchy. At this point the MU is an eighth-note, so an eighth-note metronome is made. When this metronome "ticks," it places a dot in the growing metric hierarchy and then sets itself to tick again one MU later. The metronome counts its dots, and whether a dot coincides with a note onset (a hit).



Mozart's Symphony No. 40, First Movement (1st Violin part)

This completes processing at the onset of the second note. The next location is always the lesser of two distances: the distance to the next note onset and the distance to the next tick of the lowest level metronome. The distance to the sole metronome's next tick is the same as the distance to the next eighth-note, so the program moves to this location.

The program finds T3 at this new location, supporting the hypothesis that the eighth-note is a metric interval. On the strength of this, the program hypothesizes a higher-level temporal grouping at double the current (eighth-note) metric unit. There are two reasons for doubling rather than tripling the current MU to generate the next higher level. First, ratios of two are far more common. Second, two is the most common grouping in the subjective organization of identical, isochronous tones (Fraisse, 1982), and in spontaneous tapping (Fraisse, 1947-1948). The doubling operation (Longuet-Higgins & Lee's "*Conflate*" procedure) holds T1 fixed and moves T2 to where T3 is (i.e. at T1 + 2MU), recalculates the MU (T2 - T1) and projects T3 to T2 + MU. The new quarter-note MU generates a new metronome which is set to tick at T1 + MU, i.e. T2. There are now two metronomes, with different units. The program examines each in turn to see if it is due to tick at the current location. If it is, the metronome produces a dot, counts it, and scores a hit if there is a note onset at the current location. If a metronome is not due to tick but there is a note onset, the metronome records a miss, an onset it did not predict. In our example, both metronomes are due to tick at the current location (the third note onset), so now the metric hierarchy contains two dots at the eighth-note level and one at the quarter-note level, all hits.

The distance to the next note onset is a quarter-note, while the unit of the lowest metronome is an eighth-note, so the next location is halfway between the onset of note 3 and the onset of note 4. The program examines each metronome. Only the eighth-note metronome is set to tick here, and it produces a dot, but there is no onset here, so it is neither a hit nor a miss.

T3 occurs at the onset of the fourth note, so we might expect the program again to double the MU and create a new metronome. However, something more important has happened. At the onset of note 4, the program reaches the end of the note 3 quarter-note, and it recognizes that this note is longer than any it has heard before. Since longer notes usually initiate higher-level metric groupings (Povel & Essens, 1985), and since it is still early in the piece, the program retrospectively interprets the first two notes as upbeats to the third note. Accordingly, an "*Update*" process moves T1 forward so as to make the onset of note 3 the anchor point of the metronomes, and projects T2 and T3 at one and two MUs, respectively, from T1. The MU is not changed, and no new metronome is created. Note that the *Conflation* that seemed warranted at this location is no longer possible, since T3 has been projected to a point we have not yet reached. The program now examines the metronomes: both are due to tick, and since there is an onset here both dots are hits.

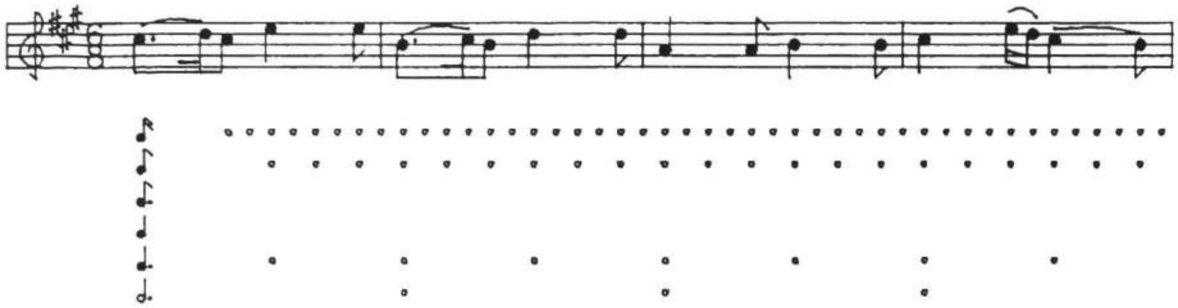
At the onset of note 5 the eighth-note metronome yields a hit and the quarter-note a miss. At the onset of note 6, we have reached T3. The program therefore *Conflates*, doubling the quarter-note MU to a half-note, and makes a third metronome. This metronome is set to tick at T1 + MU, which happens to be where we are. Note that if T1 had not been Updated before this *Conflation*, the first dot at the half-note level would have been at the onset of note 4, which would have been inappropriate. Instead, the half-note metronome produces dots such that every other dot coincides with a bar line in the score. At note 9 the program reaches T3 once again, where another conflation yields a MU of a whole-note and a fourth metronome.

We do not allow enlargement of the MU beyond one whole-note, so our analysis does not produce GTTM's two-measure, four-measure and higher metric levels. The point of this limitation is that higher levels are: a) less perceptually salient (they are far above the tactus); and b) are better understood as defining phrasal boundaries than metric units (Longuet-Higgins & Lee, 1982). Indeed, GTTM allows for metric discontinuities at high levels and for violation of WFR 4 (which normally requires dots at a given level to be equally spaced in time) in such cases. We have not yet attempted to incorporate such intuitions in our program.

The rest of the analysis generates the full metric hierarchy and counts the hits and misses at each of its levels. In addition, a procedure, *Slide*, which we have added to those of Longuet-Higgins & Lee, slides all of the T-markers forward 1 MU whenever the program catches up with T3. We assume that the listener continues to predict future onsets on the basis of the largest experienced metric unit, and uses this projection to detect metric changes. When a metric change occurs, the framework of T-markers may be reset at the point of change. We have not yet implemented these intuitions, but the problem of meter change will be an important test of our model.

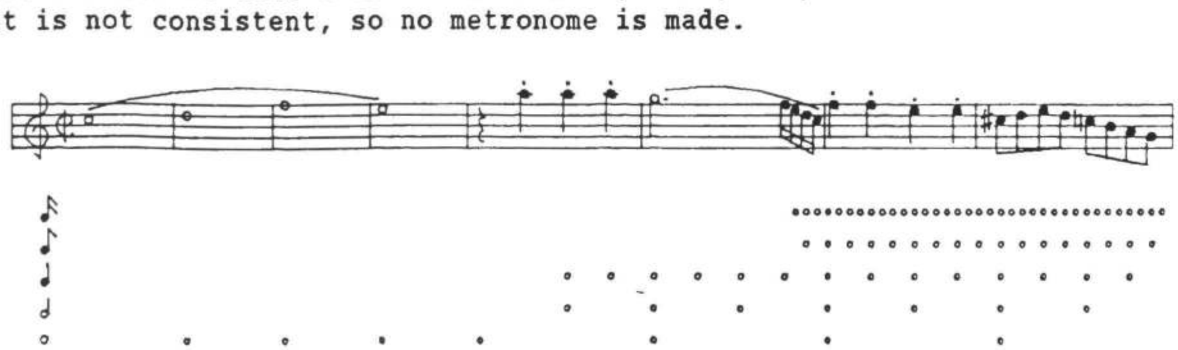
Our analysis of Mozart's Piano Sonata K.331, shown below, illustrates another important operation. When the program gets to note 2 it puts T2 there and declares the MU to be a dotted eighth-note. Before reaching T3 the program realizes that note 3 is longer than the note beginning at T2, i.e. note 2. Again, on the assumption that longer notes should initiate higher level metric units, the MU is now *Stretched* to a quarter-note by moving T2 to the onset of note 3. Intuitively, the purpose of the *Stretch* procedure is to handle dotted notes. A dotted note is usually followed by a complementary note which, added to the dotted note, yields a duration that fits the metric hierarchy at a higher level than either of the notes alone.

The program has so far created sixteenth-, dotted eighth- and quarter-note metronomes. By the end of note 4, it again finds that its hypothesis (that there will be an onset at T3) disconfirmed, since note 4 surrounds T3. This suggests that the MU that generated that hypothesis is incorrect, and that the MU should be enlarged. The most conservative enlargement is made by moving T2 to the onset of note 4, yielding a MU of a dotted quarter-note. Both applications of the *Stretch* procedure illustrate a common principle: whenever the metric unit is enlarged by some means other than doubling or tripling it is assumed that the old MU is incorrect and should therefore be eliminated (by marking it as rejected). This indicates that it is not part of the listener's metric representation.



Mozart's Piano Sonata, K. 331, First Movement

In some pieces the listener may hear a note whose duration is not represented in the metric hierarchy. In such cases we assume that the listener creates a metronome to represent that duration if it is consistent with the current MU. By consistent we mean that the duration in question is an integral multiple or divisor of the MU. Consider the beginning of the last movement of Mozart's Symphony no. 41 ("Jupiter") below. At the outset the listener establishes a MU of a whole-note, and eventually hears the first of the quarter-notes. Longuet-Higgins & Lee's rules do not provide any way to generate smaller metric levels, but we assume that the music itself may directly dictate levels in such cases. Since the first quarter-note is consistent with the MU, the program creates a quarter-note metronome. A few notes later we hear a dotted half-note; this, too, is smaller than the MU, but it is not consistent, so no metronome is made.



Mozart's Symphony No. 41, Last Movement

The "Jupiter" example also illustrates a third way that metric levels are created. For the first four bars the metric hierarchy consists of a single level. When the quarter-note metronome is created, its unit is consistent with the current MU, but the metric hierarchy now violates WFR 3, which limits the ratio between adjacent levels to three or less. The solution is to interpolate a third level that satisfies WFR 3 with respect to both the quarter-note level and the whole-note level, i.e. a half-note level. After the dotted half, we hear a sixteenth-note, and it is again necessary to create a new metronome, interpolated at the eighth-note level.

CONCLUSION

Our program yields a psychologically and musically plausible metric analysis of a wide variety of scores. Using note-by-note processing, it produces a metric hierarchy that conforms to the GTTM rules, and it identifies one of the metric levels as the **tactus** or beat. There are many other scores, however, that it cannot correctly analyze. This is a limitation of duration-only analysis: rhythm is not the sole carrier of information about metric structure. The extent to which our approach succeeds reflects the redundancy between rhythmic, melodic and harmonic dimensions in most music. The important question, therefore, is whether those scores that our program cannot analyze are also difficult for human listeners when only the rhythm is presented. Answering this question will be an important test of our model.

ACKNOWLEDGMENTS

This work was supported by an NSF Pre-Doctoral Fellowship to Miller and a PSC-CUNY Faculty Research Award to Jones.

REFERENCES

Fraisse, P. (1947-48). Mouvements rythmiques et arythmiques. *L'Annee Psychologique, 47-48,* 11-21.

Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 149-180). New York: Academic Press.

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

Longuet-Higgins, H.C. & Lee, C.S. (1982). The perception of musical rhythms. *Perception, 11,* 115-128.

Longuet-Higgins, H.C. & Steedman, M.J. (1971). On interpreting Bach. In D. Michie and B. Meltzer (Eds.), *Machine Intelligence 6.* Edinburgh: Edinburgh University Press.

Povel, D.-J. (1984). A theoretical framework for rhythm perception. *Psychological Research, 45,* 315-337.

Povel, D.-J. & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2,* 411-440.

Simon, H.A. (1968). Perception du pattern musical par AUDITEUR. *Sciences de l'Art, V-2,* 28-34.

Steedman, M.J. (1977). The perception of musical rhythm and metre. *Perception, 6,* 555-569.