Title
PDP Models for Meter Perception

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 12(0)

Author
Jones, Jacqueline A.

Publication Date
1990

Peer reviewed

PDP Models for Meter Perception

Don L. Scarborough, Benjamin O. Miller,

and

Jacqueline A. Jones

Brooklyn College of the City University of New York

Brooklyn College, Brooklyn, N.Y. 11210

Bitnet: dosbc@cunyvm, bmiller@bklyn, jajbc@cunyvm

Abstract

A basic problem in music perception is how a listener develops a hierarchical representation of the metric structure of music of the sort proposed in the generative theory of Lerdahl & Jackendoff (1983). This paper describes work on a constraint satisfaction approach to the perception of the metric structure of music in which many independent "agents" respond to particular events in the music, and where a representation of the metric structure emerges as a result of distributed local interactions between the agents. This approach has been implemented in two PDP simulation models that instantiate the constraints in different ways. The goal of this work is to develop psychologically and physiologically plausible models of meter perception.

Introduction

Many of the achievements in artificial intelligence have been based on rule-based symbolic approaches. On the other hand, there has been widespread interest in parallel distributed processing (PDP) systems which often use a non-symbolic (or a distributed symbol) constraint- satisfaction approach toward solving problems (Marr, 1982). In an earlier report (Miller, Scarborough & Jones, 1988) we described the **Beats** model which is a rule-based approach that builds on earlier work by Longuet- Higgins & Lee (1982). In this paper we describe a PDP approach to the simulation of musical rhythm perception.

Metric Analysis

Meter, in music, is a perceived pulse that marks off equal temporal intervals in the music. These pulses tend to be grouped perceptually, with the first of each group heard as accented; within groups, pulses appear to alternate between weak and strong in a regular way that reflects a hierarchical organization, as illustrated below. In the notation introduced by Lerdahl & Jackendoff (1983) in their **Generative Theory of Tonal Music** (henceforth, **GTTM**), 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 and might represent, say, time intervals corresponding to eighth notes. The second row of pulses, corresponding to every other pulse of the first row, groups the pulses in the first row by twos and therefore represents a quarter note pulse rate. The third row represents half notes. Moments with dots at more than one level (e.g. time points 1, 3, 5 and 9) represent perceived stresses relative to others, and in this example, 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**, a level in the hierarchy that corresponds to a moderate foot- tapping tempo. It is important to note that such a hierarchical structure is heard even when listening to a sequence of equally spaced, equally intense tones (Fraisse, 1982).

	Time	1	2	3	4	5	6	7	8	9	10	etc.
	1	.	.	.	.	.	.	.	.	.	.	
Metric Level	2	.		.				.		.		
	3	.				.				.		

A metric structure provides the framework for the emergence of rhythm. For example, syncopated rhythms occur when perceived musical accents are heard as occurring at relatively unstressed times in the metric structure. Thus, musical events are heard within the context of the metric structure. However, a listener must use the same musical events to discover the metric structure in the first place. We are interested in how a listener induces a metric hierarchy while listening to the music.

We have developed a PDP model of the process by which listeners determine the meter of a piece of music. A goal of the model is to generate output that conforms to a well-formed metric hierarchy (Lerdahl & Jackendoff, 1983) under conditions where a human listener would perceive such a hierarchical structure. The simulation model makes a single left-to-right pass through the piece and constructs candidate metric levels.

A Constraint-Satisfaction Approach to Meter Perception

In the BEATS rule-based model of meter perception (Miller et al. 1988), metric levels are generated and selected on the basis of rules that embody expectations about meter, such as well-formedness. BEATS can be said to make decisions. Our PDP model, **BeatNet**, performs more or less the same task not by making decisions but by satisfying a set of constraints. Decision making in rule-based BEATS requires a single agent, or inference engine, with knowledge about states of affairs and knowledge of rules to follow with respect to those states. A distributed constraint-satisfaction approach, by contrast, has no single agent, no centralized knowledge of states, and no rules. Rather, in such a network there are many individual agents or processes; each process acts on the basis of local constraints and knows nothing about the state of other processes but can interact with them. Through this interaction, processes affect one another's behavior in a way that is simple at the level of the processes but complex for the network taken as a whole.

The BeatNet model is based on the idea that every level in a metric hierarchy will correspond to the duration and phase of a single note or small group of adjacent notes somewhere in the piece. This suggests that by determining the durations of notes and note groups we will have a set of candidate metric levels that can be used to construct the correct metric hierarchy. This is in fact the approach we have taken in the BeatNet model. However, many durations in a piece will not correspond to a proper metric level. Therefore BeatNet's task is to identify which durations found in the score belong in the meter.

BeatNet can be thought of as a network of very low frequency oscillators whose periods correspond to the onset-to-onset intervals that occur in a piece. BeatNet makes a single, left-to-right pass though a piece without backtracking. It ignores all information other than onset-onset intervals. As BeatNet moves through the score, the time between one note onset and another leads to the excitation of an oscillator, or metronome, a process that embodies this periodicity. A metronome is defined by where it starts, i.e. its phase, and the aggregate duration of the note(s) it spans, i.e. its period. As a piece is heard, a subset of the oscillators is excited by the time intervals that occur. The activation of a metronome oscillator, then, depends on the detection of a specific time interval or period in the stimulus. If a particular period is detected with more than one phase, each period/phase combination activates a corresponding oscillator with the corresponding period and phase. Oscillators not set in motion are generally inactive.

Simply activating the metronome oscillators that correspond to onset-onset intervals does not lead to a coherent metrical structure. For example, a series of eighth notes will activate metronomes corresponding to an eighth note, a quarter note, a dotted-quarter note, a half note, and so on at various phases. However a metric analysis that conforms to the Lerdahl & Jackendoff rules will only allow a subset of the metronomes that form a well structured hierarchy. This requires that metronomes that appear to be inconsistent with the metric structure be inhibited, while metronomes that are consistent should be further strengthened. To do this, we let metronomes interact via excitation and inhibition in ways that represent constraints on the network that are likely to lead to a coherent metric structure.

In BeatNet, each metronome has a variable strength, or activation level, which is modified in four ways. A metronome's strength decays steadily over time, and is also lowered from time to time by inhibitory inputs from other metronomes. Activation is increased by excitatory inputs from other metronomes. Finally, a metronome's activation increases if it predicts a note onset, i.e. it ticks concurrently with a note onset. Strength is indirectly affected by two other features of BeatNet. First, each metronome has a threshold. If its strength falls below threshold, a metronome can neither inhibit nor excite another metronome, but it continues to tick and can still receive excitation and inhibition. Second, the notion of the tactus, described earlier, suggests that metronomes with moderate periods should have more influence in the network than those with very large or small periods. Accordingly, several parameters are scaled so as to increase the influence of metronomes close to the tactus.

We have created two versions of this BeatNet metronome model. In the first model, called the **Broadcast** model, we have tried to build in very few of the required constraints into the network structure. That is, we configured the network based on a few simple principles that only indirectly embody the constraints that a well formed hierarchy must have. The issue was whether such a simple architecture might nonetheless produce metric analyses that were well formed in the Lerdahl & Jackendoff sense. The second model, called the **Resonance** model, makes much stronger assumptions about the nature of the interactions between metronomes.

The Broadcast Model

In the Broadcast model, we tried to avoid assuming specific patterns of hardwired interaction among the metronomes. Instead, interaction among metronomes occurs only in response to signals that are broadcast throughout the system in response to particular events. These broadcast signals originate from both external and internal events. An externally generated signal is broadcast whenever a note onset occurs. The effect of this signal is that all metronomes that time out or tick at that instant are activated. We also assume that metronomes differ in sensitivity. That is, we assume that, as with pitch perception, we are most responsive or sensitive to events of a particular period, and that this peak sensitivity is the basis for the tactus, the metric level that is perceived as most salient, and which is generally around 2 Hz (Fraisse, 1947-48). The second source of broadcast signals comes from the internal ticking of the metronomes themselves. A tick of a metronome is, functionally, much like a note onset in that it excites other metronomes that tick at the same time. The strength of this signal depends on how many metronomes tick at the same time. However, in the case of these metronome ticks, we assume that the metronomes are organized into a linear structure like a basilar membrane, and that the effect of one metronome on another decreases with the distance between the two. Close neighbors are excited or inhibited strongly, while more distant metronomes are hardly affected. Distance, for purposes of calculating the effect of one metronome on another, is defined in terms of the ratio of the periods of the two metronomes.

The idea that metronomes that tick at the same time strengthen each other is a weak way to implement the constraint that in a well formed hierarchy, a metronome with a period T should be in phase with other metronomes whose periods are related to T by simple integer ratios, e.g. 2T or 3T, and 1/2T or 1/3T. Metronomes that are in phase with each other are more likely to tick together than out-of-phase metronomes. In addition, the tick of a metronome broadcasts an inhibitory signal to all metronomes with shorter periods that do <u>not</u> tick at that time. This implements the constraint that in a well formed metric hierarchy, a pulse at one metric level should coincide with pulses at all the lower metric levels. The asymmetry between excitation and inhibition can be seen in the nature of a well-formed metric hierarchy: a beat at a given level need not be a beat at higher levels, but it must be a beat at all lower levels (Lerdahl &Jackendoff, 1983). The Broadcast model's pattern of inhibition and excitation is such that metronomes that together constitute a well-formed hierarchy will tend to strengthen one another and inhibit outsiders. Finally, as mentioned above, we assume that metronomes tend to run down. That is, without sustaining input events, the activation of a metronome will decay.

The interactions can be summarized as follows:

metronome	ticks												
1	.	.	.	.	.	.	.	.	.	.	.	.	
2	.		.		.		.		.		.		.
3		.		.		.		.		.		.	
4	.				.				.				.
	a	b	c	b	a	b	c	b	a	b	c	b	a

At the 'a' steps metronomes 1, 2 and 4 excite one another and metronome 3 is inhibited by 4, which has a larger period, and by 2, which has the same period and a different phase. At the 'b' steps metronomes 1 and 3 excite one another while 3 inhibits 2, and at the 'c' steps there is only excitation between metronomes 1 and 2. Nothing inhibits metronome 4, because it is the largest, or metronome 1, because it ticks together with all higher-level metronomes.

Figure 1 illustrates metronome interactions. Each curve represents a single metronome. The ordinate represents the activation level of a metronome. The abcissa intersects the ordinate at an activation level of 0. The horizontal dashed line above the abcissa represents the threshold that a metronome must reach before it can affect other metronomes. The abcissa represents time, and the vertical dashed lines mark note onsets. The very first note onset occurs at the ordinate. Figure 1a (left panel) illustrates the emerging activation pattern that occurs in various metronomes when the

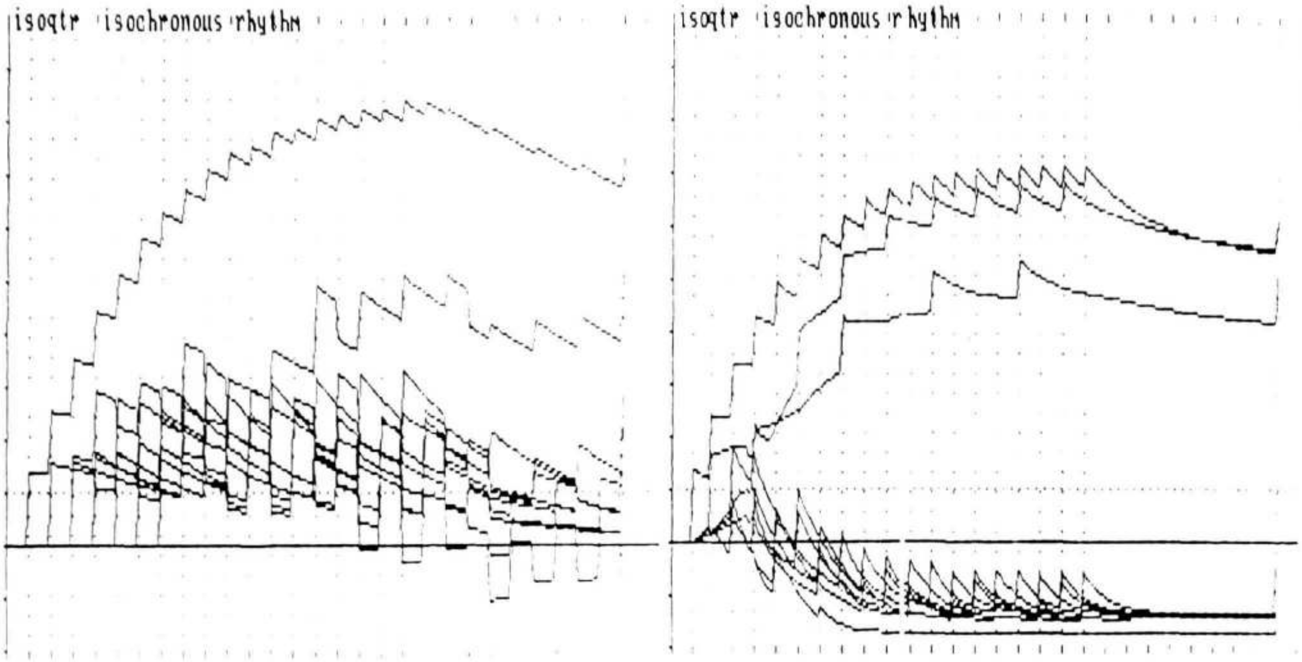

Fig. 1, Quarter Notes: a) Broadcast; b) Resonance

Broadcast network hears an isochronous sequence of 20 quarter notes followed by two whole rests. The jaggedness of the activation patterns occurs because of changes in activation that occur in response to external and internal signals. In between these signals, activations decrease as a result of decay. Though this pattern appears to be rather chaotic, a quarter note metronome (top curve) and a half note metronome (second highest curve) clearly emerge from the noise. Further, these two metronomes are in phase with each other. Less clearly evident is a whole note metronome. It too is in phase with the two more active metronomes. However, it is in competition with other whole note metronomes that differ in phase. In addition, there are metronomes with a period of three quarter notes at various phases. No clear winner emerges from this competition. However, this is not necessarily a failing of the model for two reasons. First, with an isochronous sequence, there is no clear basis for inducing one well- formed metric structure over another. Second, the lack of clarity at higher levels of the metric structure matches the idea that the tactus is an intermediate level of the metric structure that is most strongly perceived.

Fig. 2, London Bridge is Falling Down

Figure 2 shows the note duration sequence for the song, London Bridge is Falling Down, which is in 4/4 time, with the first dotted quarter note beginning on the down beat. The response of the Broadcast model to this piece is shown in Figure 3a (left panel). The top line represents a quarter note metronome, and the next two lines down are an eighth note and a half note metronome, respectively. These three metronomes form a proper metric hierarchy. No whole note metronome emerges clearly, but the next strongest metronome is, in fact, a whole note metronome with a phase that agrees with the other levels of the metric structure.

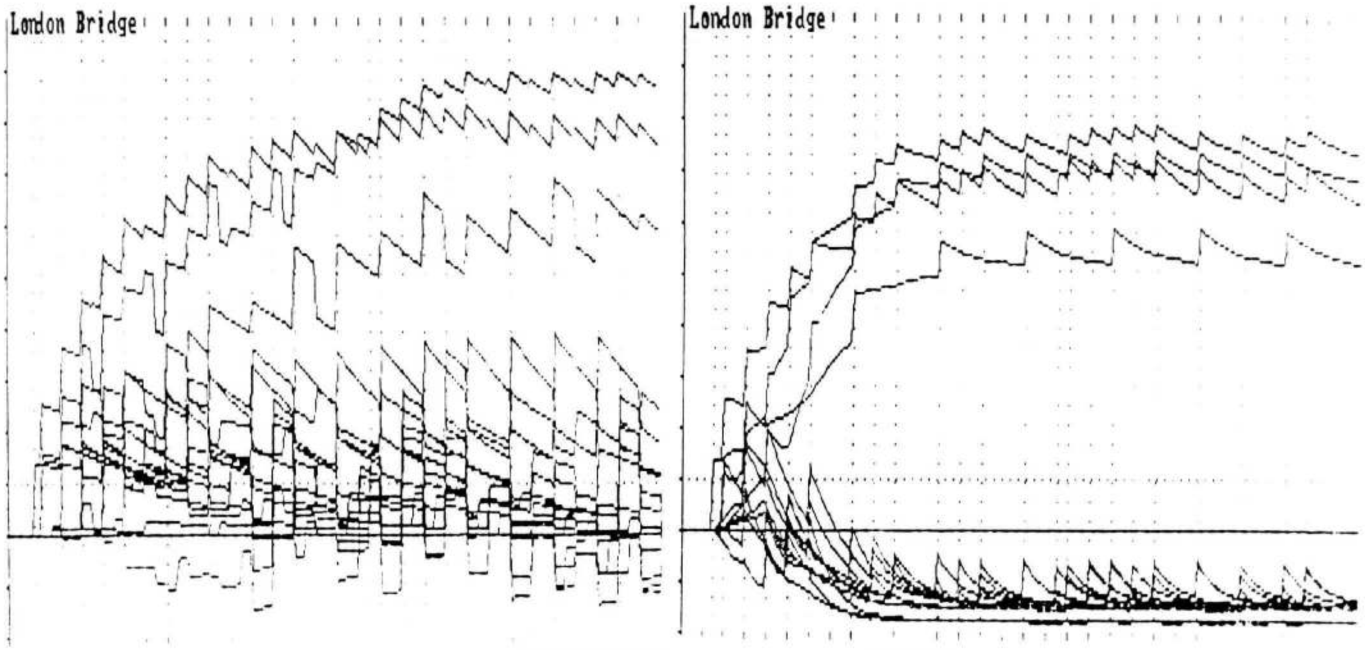

Fig. 3, London Bridge: a) Broadcast; b) Resonance

All in all, the Broadcast model seems to do a fairly good job of inducing appropriate metric structures on the basis of some fairly simple assumptions about how the metronomes might interact. Further, the interaction pattern does not appear to depend critically on the choice of parameters that determine the magnitude of the interaction effects within fairly wide limits. The important point is that the principles for metronome interaction do not explicitly embody the concept of a metric hierarchy.

The Resonance Model

In the Resonance model, as with the Broadcast model, we assume that a metronome is excited by the occurrence of note-onset intervals that match its period and phase. However, the Resonance model makes more specific assumptions about the architecture. In particular, this model assumes that excitatory and inhibitory interactions continue to occur in the interval between ticks and note onsets. For this to occur, we have to assume that each pair of metronomes is connected via an excitatory or an inhibitory connection depending on whether or not that metronome pair is in phase, and on whether their periods form a simple integer ratio. This assumption leads to a pattern of interactions that produces much quicker sorting out of "bad" metronomes. In addition, depending on the strength of the interactions, the Resonance model literally does resonate in that a particular subset of the metronomes can form a pattern of interactions in which the strength of the internal signals alone is sufficient to hold the metronomes in a stable pattern of activation. This may not be an unreasonable characteristic of the model because it means that the same structure that recognizes a metric structure can also produce a metric structure, just as a listener, after hearing a few bars of a piece, can anticipate the rhythmic pattern thereafter.

Figure 1b (right panel) illustrates the emerging metric structure with the Resonance Model in response to the isochronous quarter note sequence. Here, an unambiguous metric structure emerges consisting of a quarter note, a half note, and a whole note metronome all with the appropriate phase relationship. All other candidate metric levels become quickly inhibited. Figure 3b (right panel) shows the response of the Resonance model to London Bridge. Here too a complete metric hierarchy emerges. The top four lines represent a quarter note, an eighth note, a half note and a whole note level. Again, other candidate metronome levels are strongly inhibited within a relatively few notes.

The Resonance model induces metric structures much more clearly than does the Broadcast model. However, before we can entertain the Resonance model as a plausible description of how people might recognize metric structures, we need to have an account of how the specific pattern of excitatory and inhibitory weights between metronomes that is involved in this model might arise. At this point, we do not have a clear answer to this issue.

Conclusion

The BeatNet models succeed in producing reasonable metric analyses for many pieces. In other cases, the output does not coincide with the interpretation specified by the score. However, it is possible that a human listener would make the same errors if deprived of all information other than time intervals. While BeatNet has some intuitively appealing properties, it is not without problems. A difficulty with BeatNet that is inherent in network models is the virtual impossibility of predicting the model's output and the difficulty in understanding the relation of this output to the system parameters. However, in general we have tried to constrain the system parameters based on a priori judgments about what was reasonable. On the other hand, while a rule-based system such as BEATS (Miller et al. 1988) has few parameters and is very predictable -- on the whole it can produce only one analysis -- BeatNet can produce different output as the values of its initial state and parameters are changed. Exploring the various configurations over a number of parameters is a daunting task, but we are encouraged by the fact that a configuration that works well with one score tends to work well with other scores.

Acknowledgments

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

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.

Marr, D. (1982). **Vision.** San Francisco: W.H. Freeman.

Miller, B., Scarborough, D. & Jones, J. (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.