Models of Metrical Structure in Music

Edward W. Large
Department of Computer and Information Science
The Ohio State University
Columbus, OH 43210
large@cis.ohio-state.edu

Abstract

Recent models of metrical structure in music rely upon notions of oscillation and synchronization. Such *resonance* models treat the perception of metrical structure as a dynamic process in which the temporal organization of musical events synchronizes, or entrains, a listener's internal processing mechanisms. The entrainment of a network of oscillators to an afferent rhythmic pattern models the perception of metrical structure. In this paper, I compare one resonance model with several previously proposed models of meter perception. Although the resonance model is consistent with previous models in a number of ways, mathematical analysis reveals properties that are quite distinct from properties of the previously proposed models.

Introduction

The perception of rhythm in music involves, among other things, the perception of *metrical structure*. Metrical structure describes an important part of musical phenomenology, the sense of alternating strong and weak beats that accompanies the experience of listening to music (Lerdahl & Jackendoff, 1983). It is related to the metrical structure of language (Lerdahl & Jackendoff, 1983), however, in music metrical structure refers to a distinctive form of temporal organization. Psychologically, metrical structure may be viewed as a perceptual framework that affects human perception, attention, and memory for complex, temporally structured event sequences (Jones & Boltz, 1989; Palmer & Krumhansl, 1990; Povel & Essens, 1985).

Theories of metrical structure attempt to describe the perceived temporal organization of rhythmic patterns. Generative models posit rules for describing legal metrical structures (e.g. Lerdahl & Jackendoff, 1983; Longuet-Higgins & Lee, 1978). The perception of rhythm then involves parsing a rhythmic pattern to retrieve its metrical structure (Jackendoff, 1992; Longuet-Higgins & Lee, 1982). Clock models have been proposed (e.g. Essens & Povel, 1985; Scarborough, Miller, & J. Jones, 1992) that explain meter perception as the process of aligning the ticks of an internal clock with the onset of events in an acoustic signal.

Dynamic attending models (Jones, 1976; Jones & Boltz, 1989) propose that rhythm perception is a dynamic process in which the temporal organization of rhythms synchronizes, or entrains, a listener's attention. Connectionist accounts of the perception of metrical structure have recently been proposed that are more or less consistent with this view (Large & Kolen, in press; McAuley, 1994; Page, 1994). In this paper, I focus on one such model (Large & Kolen, in press). According to this model oscillatory units phase- and frequency- lock to periodic components of a rhythmic pattern, and the self-organizing response of a network of oscillators embodies a dynamic perception of metrical structure. I compare this model with previous accounts of metrical structure in music. I argue that resonance models of musical meter provide advantages over previous approaches for describing the perception of temporal structure in music.

Models of Metrical Structure

Models of metrical structure must account for several aspects of the perception of temporal organization in rhythmic patterns. First, a model must specify *structural constraints*, describing which temporal structures correspond to legal metrical structures. This is not enough, however, because the relationship of rhythmic patterns to legal structures may in principle be many-to-many. Thus, a model must also specify a set of *metrical preferences* (Lerdahl & Jackendoff, 1983), to predict which structure a listener would perceive for any given pattern. Finally, the durational patterns that musicians perform never correspond precisely to the durations called for in a musical score, rather they contain intentional and unintentional timing variability (e.g. Sloboda, 1983). This is a problem for most models of metrical structure, because the models refer to nominal note durations. Therefore, a perceptual model must also explain how listeners deal with *performance timing*.

Generative Models

Generative models of metrical structure handle structural constraints by proposing grammars that describe the organization of events in time. For example, Lerdahl and Jackendoff (1983) have proposed a generative theory in which the metrical structure of a piece is transcribed as a grid (see Figure 1) similar to linguistic metrical structure grids (cf. Liberman & Prince, 1977). Each row of dots represents a level of beats, and the relative spacing between dots of adjacent levels describes the relationship between

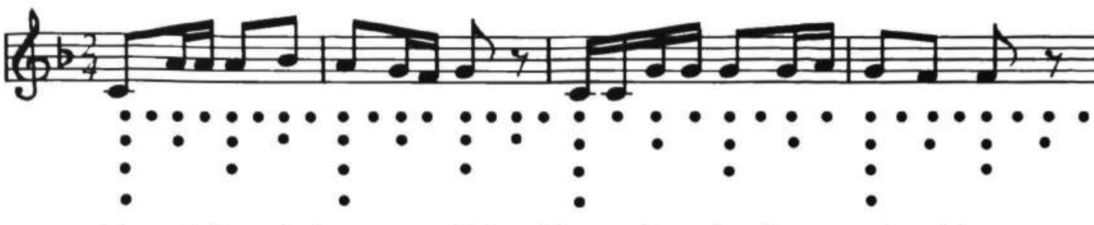

Figure 1: A metrical structure grid describing an alternation of strong and weak beats.

the beat periods of adjacent levels. Lerdahl and Jackendoff's (1983) metrical grids explicitly capture the perception of regularly recurring strong and weak beats. Points where the beats of many levels coincide are strong beats; points where few beats coincide are weak beats. Lerdahl and Jackendoff (1983) give a set of *well-formedness rules* to describe legal metrical structure hierarchies. They give a set of *preference rules* to describe which metrical structure a listener would perceive for any given rhythmic pattern. They do not, however, deal with the issue of performance timing.

Longuet-Higgins (1978) has proposed that the meter (i.e. the time signature) of a melody may be regarded as a generative grammar, and the rhythm of a melody may be described as a tree structure generated by the grammar. Terminals and non-terminals correspond to periods of time, and productions describe temporal nestings corresponding to integer ratios (2:1, 3:1, 4:1, etc.). The "task" of rhythm perception is to infer from a rhythmic pattern a metrical grammar and a parse tree. Figure 2 shows a simple melody, and the parse tree generated by a metrical grammar that describes binary temporal nestings (corresponding to the 2/4 time signature). Longuet-Higgins and Lee (1982) handle metrical preferences as processing constraints imposed by the demands of real-time parsing. Longuet-Higgins (1987) deals with performance timing by proposing that the length of a beat period may be adjusted throughout the performance as the performer speeds up or slows down. According to this model a parser uses a static tolerance window, within which it will treat any onset as being "on the beat". Incoming durations are assigned to one of a few duration categories (i.e. quarter note, eighth note, and so forth), so that they can be properly dealt with by the grammar.

Clock Models

Models of meter perception sometimes invoke the notion of a clock that attempts to align its output ticks with event onsets in the acoustic signal. For example, Essens and Povel (1985) model the perception of meter as the induction of a "best clock" according to which rhythmic input is coded and stored in memory. Input to their system consists of a pattern of inter-onset durations with associated accent information (strong/weak). The system hypothesizes clocks of every different period and phase that could be implied by the pattern. Each clock is then ranked according to counter-evidence, which includes the number of ticks that would coincide with weakly accented events, and the number of ticks that would coincide with no event onset. In this way, the pattern of accent in the sequence determines the beat period, phase, and "strength" of the induced clock, implementing metrical preferences. The model does not deal with timing variability.

The BeatNet model (Scarborough, Miller, & J. Jones, 1992) was designed to implement Lerdahl and Jackendoff's (1983) generative theory. BeatNet is based on a connectionist parallel constraint satisfaction paradigm. The BeatNet network is a one-dimensional array of idealized oscillators of different beat-periods that operate to align their discrete output ticks with event onsets, producing a metrical grid of the style proposed by Lerdahl and Jackendoff (1983). A metrical structure emerges from local interactions between units, rather than from the global effect of rule-based analysis. The interactions are handcrafted to implement Lerdahl and Jackendoff's well-formedness and preference constraints. The model does not deal with performance timing.

Resonance Models

Resonance models of the perception of temporal organization posit the existence of one of more oscillators that entrain to rhythmic patterns (Large & Kolen, in press;

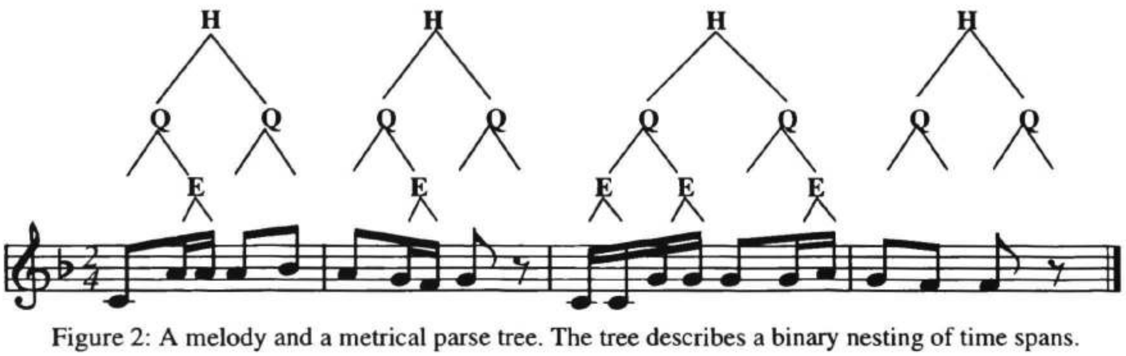

Figure 2: A melody and a metrical parse tree. The tree describes a binary nesting of time spans.

McAuley, 1994; Page, 1994). Resonance models are consistent with dynamic attending models (Jones, 1976; Jones & Boltz, 1989) in that they treat rhythm perception as a dynamic process in which the temporal organization of external musical events entrains a listener's attention. For specificity, I describe one resonance model in detail (Large & Kolen, in press).

Large and Kolen (in press) have proposed an approach to modeling the perception of metrical structure that is based on the mathematics of coupled oscillation. A rhythmic pattern serves as a driving signal. Through non-linear coupling, the rhythm perturbs both the phase and the intrinsic period of individual oscillators, modeling musical beat. A system of oscillators entrains to the periodic components of a rhythmic signal at different time-scales, and to the outputs of one another. Thus metrical structure is modeled as the collective consequence of mutual entrainment among many constituent processes.

The rhythm is represented as a series of discrete pulses, $s(t)$, corresponding to the onset of individual musical events (e.g. notes) such that $s(t) = 1$ at the onset of an event, and 0 at other times. Each oscillator adjusts both its phase and period so that during stimulation the unit's output pulses become phase- and frequency-locked to a driving rhythm. The output of the oscillator is given by:

$$o(t) = 1 + \tanh\left(\gamma\left(\cos\frac{2\pi}{p}(t-t_0) - 1\right)\right), \qquad (1)$$

where t is time, p is the period of the oscillation, $(t-t_0)(mod)p$ is the phase, and γ is the output gain. Each output pulse (see Figure 3) instantiates a temporal receptive field for the oscillatory unit – a window of time during which the unit "expects" to see a stimulus pulse. The width of the receptive field can be adjusted by changing the unit's output gain, γ. The unit responds to stimulus pulses that occur within this field by adjusting its phase and period, according to the following rules:

$$\Delta t_0 = -\eta_1 s(t) p \operatorname{sech}^2 \gamma\left(\cos\frac{2\pi}{p}(t-t_0) - 1\right) \sin\frac{2\pi}{p}(t-t_0), \qquad (2)$$

$$\Delta p = -\eta_2 s(t) p \operatorname{sech}^2 \gamma\left(\cos\frac{2\pi}{p}(t-t_0) - 1\right) \sin\frac{2\pi}{p}(t-t_0). \qquad (3)$$

Figure 3 shows three such units responding to a piano performance of a short melody (from Figure 1). The oscillators interacted with one another through discretized output pulses. Each oscillator was also allowed to adjust its own output gain, γ. Figure 3 shows a stable metrical interpretation of the input rhythm emerge, with strong and weak beats observable as points in time when the output pulses of several oscillators occur simultaneously. The

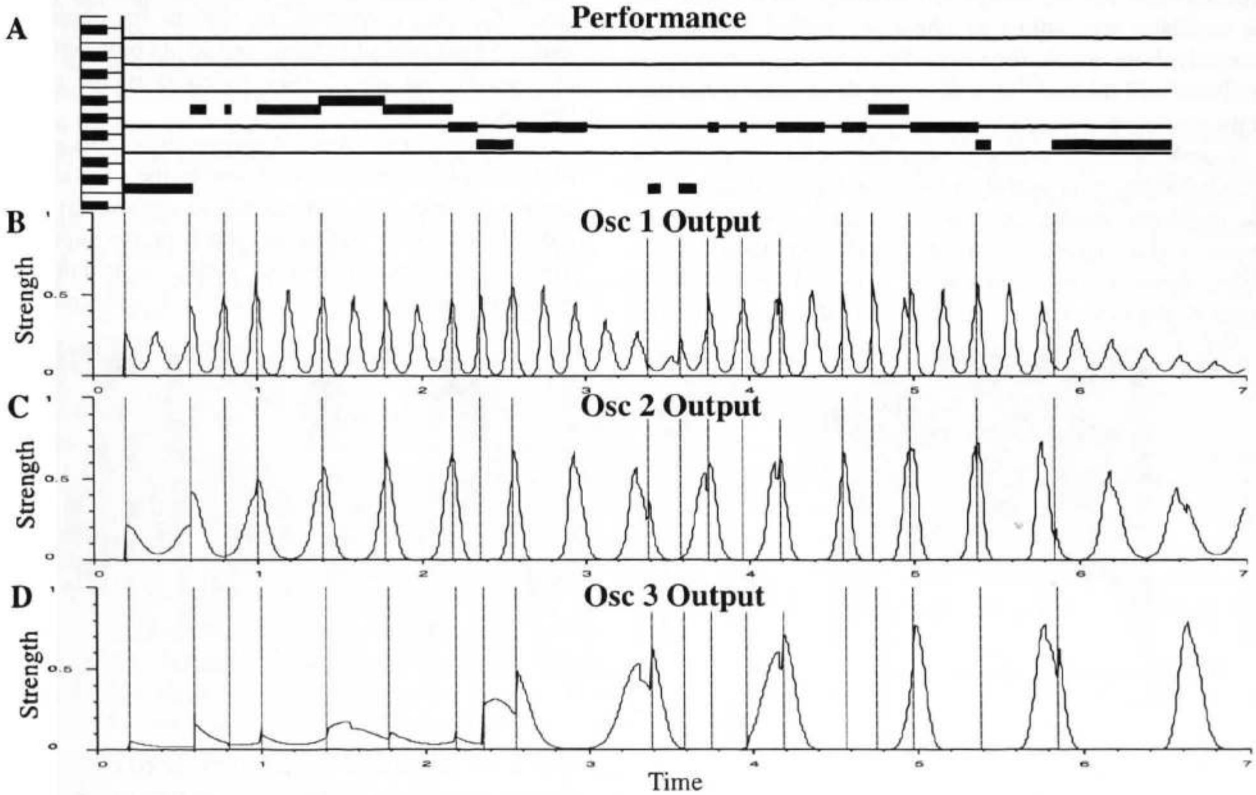

Figure 3: The response of a network of three oscillators to a performance of the melody of Figures 1 and 2. A) Piano-roll notation of the performance, B - D) Output of the oscillators. Vertical lines show event onset times.

three oscillators are correctly responding to the metrical structure at the sixteenth, eighth, and quarter note levels. This example shows the network responding to the metrical structure of a rhythmic pattern in the presence of systematic performance timing deviations.

Comparison of Models

It is informative to compare the above resonance model with the previous models we have described above, because there are important similarities and important differences. For instance, the output pulses of each oscillator are analogous to the beats of a metrical structure grid, or the ticks of a clock model. Points in time when the output pulses of several oscillators occur simultaneously, may be thought of as strong beats. In this section I present a comparison of this approach with previous models of metrical structure. First, I discuss the issue of how the resonance model deals with performance timing deviations. Next, I address how the model implements structural constraints. I do not address the implementation of metrical preferences in this paper.

Performance Timing

In the resonance model, an output pulse fills the role of a beat. However, output pulses differ from beats, or the ticks of an idealized clock, in that each output pulse has a width. Thus an event that occurs a little earlier or later than expected may still be interpreted an being "on the beat" and the oscillator will adjust its phase and period accordingly. Precisely how much deviation from timing regularity an oscillator will tolerate depends on the three basic parameters of the model, η_1, η_2, and γ.

To understand the effect of these parameters on an oscillator's behavior, it is useful to examine regime diagrams for the oscillator model. A regime diagram summarizes the response of a single oscillator to periodic stimulation, a simplified form of input that can be expected either from an external rhythm or from another oscillator in the network (Large & Kolen, in press). In the following regime diagrams, I assume that q is the period of the input (driver) and p is the intrinsic period of the responding (driven) oscillator. The ratio, q/p, is plotted along the x-axis. η_1 varies along the y-axis, and individual diagrams are given for different values of γ.

The regime diagrams show parameter regions that result in stable phase-locked states, called Arnol'd tongues (Schroeder, 1991). Figure 4 shows the effects of phase-locking alone (Eq. 2). For example, in an uncoupled system (η_1 = 0), if q/p = 0.5 then the driver fires twice for each time the driven oscillator fires. In a phase-coupled system (Eq. 2, $\eta_1 > 0$), even when q/p is different from 0.5, the system may still lock in a 1:2 relationship (the center "tongue" in the diagrams of Figure 4), because each time the driven oscillator fires, its phase is perturbed slightly by the coupling to the driver. Each diagram identifies parameter regions that result in phase-locked ratios, q/p such that $p \leq 8$. Darker regions reflect faster convergence on stable phase-locks.

The oscillator also adjusts its intrinsic period (Eq. 3). This is important for modeling musical beat because when the driving signal is temporarily removed, the oscillator continues at the driver's period, "expecting" the driver's eventual return. Figures 5A-C show the effects of adding the delta rule for period (Eq. 3). For easy comparison with Figure 4, η_2 was fixed at a small positive value, and again η_1 varied along the y-axis. These entrainment regions are larger than the corresponding regions for phase-locking alone. Adjustment of intrinsic period not only acts as a sort of memory, but also causes widening of the resonance tongues.

The width of each Arnol'd tongue reflects the sensitivity of the coupled system to deviations in the q/p ratio. In the current context, this corresponds to the sensitivity of an individual unit to timing deviation in the input signal. Thus, these Arnol'd tongues predict sensitivity of the model to timing deviation in musical performance. The

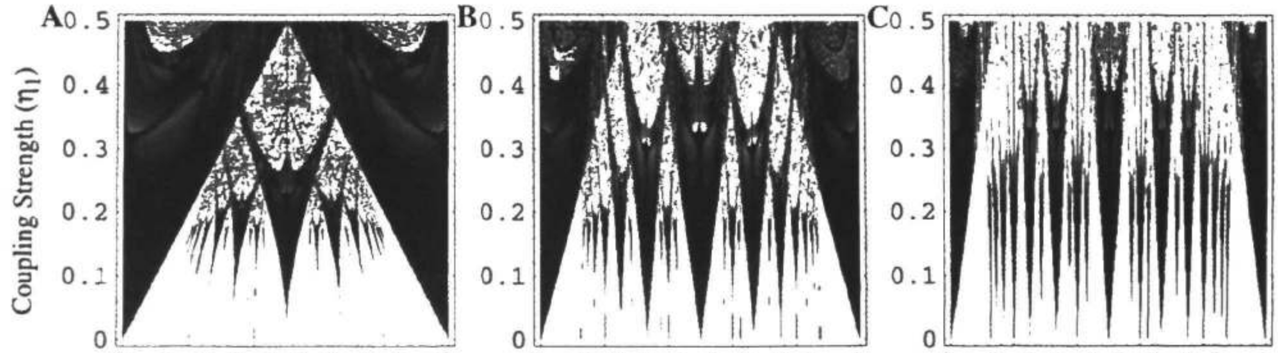

Figure 4: Regime diagrams describe stable phase-locking regions as a function of driver/driven period ratios and coupling strength. A) $\gamma = 0$, B) $\gamma = 2$, C) $\gamma = 8$. Each diagram shows the following regions (left to right): 0:1, 1:8, 1:7, 1:6, 1:5, 1:4, 2:7, 1:3, 3:8, 2:5, 7:3, 1:2, 4:7, 3:5, 5:8, 2:3, 5:7, 3:4, 4:5, 5:6, 6:7, 7:8, 1:1.

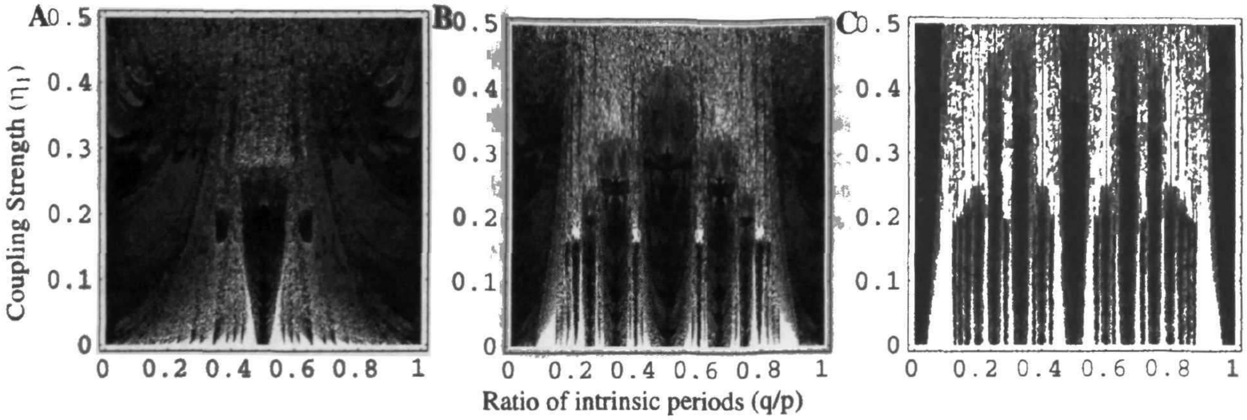

Figure 5: Phase-locking regions as a function of driver/driven period ratios and coupling strength with frequency-locking enabled. A) $\gamma = 0$, B) $\gamma = 2$, C) $\gamma = 8$. Each diagram shows the following regions (left to right): 0:1, 1:8, 1:7, 1:6, 1:5, 1:4, 2:7, 1:3, 3:8, 2:5, 7:3, 1:2, 4:7, 3:5, 5:8, 2:3, 5:7, 3:4, 4:5, 5:6, 6:7, 7:8, 1:1.

response depends on the ratio of the intrinsic periods of the input and the responding oscillator, as well as on the values of η_1, η_2, and γ. Adjustment of intrinsic period enhances the stability of the oscillator's response in the presence of timing deviations

Structural Constraints

As external input brings the response periods of two oscillators in a network close to some rational ratio, interaction between the oscillators will tend to keep the two oscillators locked at that ratio. Thus, assuming that the interaction is one-way (i.e. that one oscillator's discretized output forms the input to a second oscillator) the regime diagrams of Figures 4 and 5 describe allowable relationships between the outputs of the two oscillators much as the rules of a generative model describe allowable ratios between adjacent levels of beats.

The stability of the locking behavior is determined by sensitivity to deviation in the ratio of oscillator periods. Deviation in this ratio may be caused, for example, by the oscillators' response to the external input. Figures 4 and 5 show that, in general, 1:1 phase-locks are more stable than 1:2 phase-locks, which are more stable than 2:3 phase locks and so forth. Thus, the Arnol'd tongues summarize the structural constraints of the model. Two oscillators can lock in any rational ratio, with the stability of the lock determined (more or less) by the identity of the ratio.

This is an important departure from the previous models we have described. Each of the previous models allow only integer ratio relationships between adjacent levels of beats. This restriction of legal metrical structures to strictly nested hierarchies limits the scope of such theories, however. Only some music can be described in this way. Much non-Western music, as well as contemporary Western Art music, Jazz, and popular music make use of *dissonant* rhythmic structures (Yeston, 1976), known as *polyrhythms*. A polyrhythmic relationship between two levels of beats is a relationship of beat-periods such that N beats at one level occupy the same amount of time as M beats at the next level. Polyrhythmic relationships are described using *rational* ratios (3:2, 4:3, 5:4, and so forth) (Yeston, 1976).

The resonance model handles polyrhythmic structure naturally. Rather constructing rules that describe allowable beat period ratios, polyrhythmic relationships arise given interaction between oscillators in the network. Further, the stability of polyrhythmic structures is limited by the stability of the corresponding phase-lock, preventing extravagant claims regarding the perception of polyrhythmic structures in music.

Conclusions

Resonance models of metrical structure posit oscillators that synchronize periodic output pulses with afferent rhythmic patterns. The response of a network of oscillators models the perception of metrical structure as a process of self-organization. The model I have described here (Large & Kolen, in press) is consistent with results in human rhythm perception, which show that temporal pattern structure affects human abilities to perceive, remember, and reproduce rhythmic sequences (e.g. Deutsch, 1980; Essens & Povel, 1985; Jones, Boltz, & Kidd, 1982). In addition, the dynamic approach to metrical structure perception is consistent with dynamic theories of motor coordination. The literature on motor coordination reveals a number of activities, including rhythmic hand movements and cascade juggling, to be consistent with mathematical laws governing coupled oscillations (see Beek et al., 1992 for a review of recent models).

The output of each oscillator models the experience of musical beat, and the coincident output of several oscillators models metrical accent, consistent with previous models of metrical structure perception. However, analysis of the behavior of a single oscillatory unit in response to periodic stimulation reveals complex dynamics, and important differences compared with previous models. Regime diagrams describe the content of resonance models regarding the well-formedness of metrical structures, and show how resonance models handle polyrhythmic structure. In addition, resonance models account for the perception of meter in the face of timing variability. Thus resonance models may provide the basis for more comprehensive accounts of the perception of metrical structure.

Acknowledgments

This research was supported by the Ohio State University Presidential Fellowship to Edward W. Large. The author would like to thank B. Chandrasekaren, Ray Jackendoff, Mari Jones, and John Kolen for enlightening discussion of the topics presented in this paper.

References

Beek, P. J., Peper, C. E. & van Wieringen, P. C. W. (1992) Frequency locking, frequency modulation, and bifurcations in dynamic movement systems. In G.E. Stelmach and J. Requin (Eds.) *Tutorials in motor behavior II.* Elsevier Science Publishers B. V.

Deutsch, D. (1980) The processing of structured and unstructured tonal sequences. *Perception and Psychophysics*, 28, 381-389.

Essens, P. J. & Povel, D. (1985) Metrical and nonmetrical representation of temporal patterns. *Perception and Psychophysics*, 37, 1-7.

Jackendoff, R. (1992) Musical processing and musical affect. In M. R. Jones and S. Holleran (Eds.) *Cognitive bases of musical communication.* Washington: American Psychological Association.

Jones, M. R. (1976) Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, 83, 323-335.

Jones, M. R. & Boltz, M. (1989) Dynamic Attending and Responses to Time. *Psychological Review*, 96, 459-491.

Jones, M. R., Boltz, M. & Kidd, G. (1982) Controlled attending as a function of melodic and temporal context. *Journal of Experimental Psychology: Human Perception & Performance*, 7, 211-218.

Large, E. W. & Kolen, J. F. (in press) Resonance and the perception of musical meter. *Connection Science.*

Lerdahl, E. & Jackendoff, R. (1983) *A generative theory of tonal music.* Cambridge: MIT Press.

Liberman, M. & Prince, A. (1977) On stress and linguistic rhythm. *Linguistic Inquiry, 8*, 249-336.

Longuet-Higgins, H. C. (1978) The perception of music. *Interdisciplinary Science Review*, 3, 148-156.

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

Longuet-Higgins, H. C. (1987) *Mental Processes.* Cambridge: MIT Press.

McAuley, J. D. (1994) Finding metrical structure in time. In M. C. Mozer, P. Smolensky, D. S. Touretsky, J. L. Elman & A. S. Weigend (Eds.) *Proceedings of the 1993 Connectionist Models Summer School.* Hillsdale, NJ: Erlbaum Associates.

Page, M. P. A. (1994) Modelling aspects of music perception using self-organizing neural networks. Unpublished doctoral dissertation, University of Exeter.

Palmer, C. & Krumhansl, C.L. (1990) Mental representations of musical meter. *Journal of Experimental Psychology: Human Perception & Performance*, 16, 728-741.

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

Scarborough, D. L., Miller, P. & Jones, J. A. (1992) On the perception of meter. In M. Balaban, K. Ebcioglu, & O. Laske (Eds.) *Understanding Music with AI: Perspectives in Music Cognition*, 427-447. Cambridge: MIT Press.

Schroeder, M. (1991) *Fractals, Chaos, Power Laws.* New York: W. H. Freeman and Company.

Schulze, H. H. (1989) The perception of temporal deviations in isochronic patterns. *Perception & Psychophysics*, 45, 291-296.

Sloboda, J.A. (1983) The communication of musical meter in piano performance. *Quarterly Journal of Experimental Psychology*, 35, 377-396.

Yeston, M. (1976) *The stratification of musical rhythm.* New Haven: Yale University Press.