Modeling Beat Perception with a Nonlinear Oscillator

Edward W. Large
Institute for Research in Cognitive Science
University of Pennsylvania
3401 Walnut St., Suite 400C
Philadelphia, PA 19102
large@grip.cis.upenn.edu

Abstract

The perception of beat and meter is fundamental to the perception of rhythm, yet modeling this phenomenon has proven a formidable problem. This paper outlines a dynamic model of beat perception in complex, metrically structured rhythms that has been described in detail elsewhere (Large, 1994; Large & Kolen, 1994). A study is described in which pianists performed notated melodies and improvised variations on these same melodies. The performances are analyzed in terms of amount of rubato and rhythmic complexity, and the model's ability to simulate beat perception in these melodies is assessed.

Introduction

The ability to perceive beat and meter is, arguably, the most fundamental perceptual capability underlying our experience of musical rhythm. Simply put, *beat perception* refers to the perception of periodicity within a complex rhythm. When one taps one's foot along with a musical performance, for example, one is physically marking beats corresponding to a perceived periodicity. *Meter perception* can be described in similar terms, as the perception of two or more periodicities that coexist on different rhythmic time scales (Lerdahl & Jackendoff, 1983; Yeston, 1976). Relationships among beats of different levels define regular temporal structures that capture relative time relationships. Such *metrical structures* describe patterns of metrical accents that may explain relative prominence: the perception of strong and weak beats that characterizes the experience of musical rhythm. These musical concepts also have close correlates in theories of linguistic rhythm (e.g. Lerdahl & Jackendoff, 1983; Liberman & Prince, 1977).

Researchers in diverse fields have explored the power of such theories to explain various phenomena in the perception of rhythmically structured acoustic signals. As one might expect, this diversity of interest has led to a wide diversity of proposed models, including context-free grammars (e.g. Lerdahl & Jackendoff, 1983; Longuet-Higgins, 1987), symbolic AI algorithms (e.g. Dannenberg & Mont-Reynaud, 1987; Rosenthal, 1992), statistical approaches (e.g. Brown, 1992; Palmer & Krumhansl, 1990; Vercoe & Puckette, 1985), and connectionist models (e.g. Desain & Honing, 1991; Scarborough, Miller, & Jones, 1992). Although each captures certain aspects of beat and meter in idealized rhythms (i.e. rhythms comprised of precise durations as may be found in a musical score), varying levels of difficulty are encountered when models are confronted with the flexible and complex rhythms that humans naturally produce.

The difficulty of modeling the perception of temporal structure in naturally performed rhythms arises from several sources. One source of difficulty is *rubato*. Performers use rubato, or systematic timing deviation, to communicate musical intentions, and such temporal deviation gives rise to nonstationary rhythmic signals. Another source of difficulty is *rhythmic complexity*, which refers to factors such as the number of different duration values present in a rhythm and the use of syncopation. In short, the periodic components of rhythms that correspond to perceived beats are not truly periodic, and even in ideally timed rhythms there are missing events and extraneous events.

In this paper, beat perception is considered as a pattern of coordination that arises between an internally generated periodic process (a self-sustaining oscillator) and a periodicity within a complex external rhythm (Jones, 1976; Large, 1994). A dynamic model of this process, described in detail elsewhere (Large, 1994; Large & Kolen, 1994), is first outlined. Briefly, the coordination of internal and external periodicities is mapped onto the attractor states of a dynamical system comprising an external (driving) rhythm, and an internal (driven) oscillator (cf. Schoner, 1991; Kelso, DeGuzman, & Holroyd, 1990). Further, the intrinsic dynamics of the internal oscillator are assumed to adapt to the external rhythm, accounting for the robustness of beat perception to systematic timing deviations and rhythmic complexities found in naturally produced rhythms. Next, a study is described in which pianists performed notated melodies and then improvised variations on these same melodies. The performances are analyzed in terms of rubato and rhythmic complexity. Finally, the robustness of coordination between a simulated oscillator and a target periodicity within each rhythm is assessed.

A Dynamic Model of Beat Perception

External Rhythms and Internal Rhythms

The current approach relies upon the notion of a (simple) internal rhythm that responds to one periodicity within a (complex) external rhythm. An external rhythm is represented as a sequence of discrete impulses, $s(t)$, each

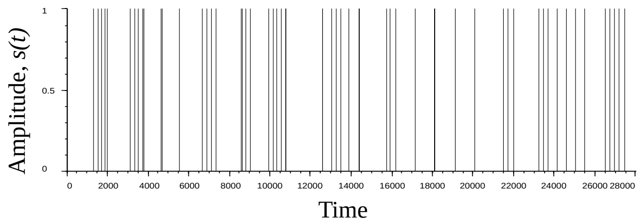

Figure 1: A sequence of discrete impulses, representing the onset of events (notes), drives a nonlinear oscillator.

denoting a single event (e.g. a note onset). Figure 1 shows a such series of impulses, corresponding to note onset times in an improvised melody, collected on a computer-monitored piano. The rhythmic signal serves as a *driver*, and impulses in the signal perturb both the phase and the period of a *driven nonlinear oscillator*, causing changes to the oscillator's behavior.

The internal rhythm is treated as a limit cycle oscillator, a structurally stable dynamical system exhibiting an asymptotically stable limit cycle. The advantage of this approach is that it does not require a great deal of knowledge about the state space or the dynamics of the internal rhythmic process. Rather, the state of the oscillator is reduced to a single variable called *phase*, $\phi(t)$, representing the position of the oscillator around its limit cycle at time t.

Phase is used to model expectations regarding when future events (impulses) are likely to occur. Phase can be defined for $-p/2 \le t < p/2$ as $\phi(t) = t/p$, where p is *period*, the time required to complete one cycle of the oscillation. According to this definition, phase varies from -0.5 to 0.5. The points $\phi(t_x) = 0$ reflect times at which the oscillator maximally "expects" events to occur. When an event occurs at time $t < t_x$, it is said to be early; when it occurs at time $t > t_x$ is it said to be late. Thus, we have the relation:

$$\phi(t) = \frac{t - t_x}{p}, \qquad t_x - \frac{p}{2} \le t < t_x + \frac{p}{2}. \qquad \text{(Eqn 1)}$$

The notion of expectation can be further refined by introducing the notion of a pulse function (Large & Kolen, 1994), describing a "soft" expectancy region around $\phi(t_x) = 0$. A pulse function can be defined as:

$$x(t) = 0.5(1 + \tanh\gamma(\cos 2\pi\phi(t) - 1)) \qquad \text{(Eqn 2)}$$

where the *gain* parameter γ describes the width of the pulse (Large & Kolen, 1994). Figure 2 shows this pulse function for $\gamma = 2$. Pulse amplitude is non-zero for a relatively small portion of the oscillator's cycle, defining a *temporal receptive field*. The temporal receptive field corresponds to a sensitive phase for the oscillator; adaptation to external events (below) occurs only when events fall within this region.

Phase Entrainment and Adaptation of Parameters

Coordination of the internal oscillator with an external periodicity is described as phase entrainment (a form of synchronization), supplemented by adaptation of oscillator parameters. Phase entrainment is modeled using a *phase attractive circle map* (cf. Kelso, DeGuzman, & Holroyd, 1990). The circle map predicts the phase, ϕ_{i+1}, at which the next event will occur, as:

$$\phi_{i+1} = \phi_i + \frac{t_{i+1} - t_i}{p_{i+1}} + \eta_\phi f(\phi_i, \gamma_i), \qquad \text{(Eqn 3)}$$

where ϕ_i is the phase of the oscillator at which the i^{th} impulse occurs, $t_{i+1} - t_i$ captures the sequence of interonset intervals present in the driving rhythm, and $f(\phi_i, \gamma_i)$, is a nonlinear phase coupling term that describes the alteration of phase brought about by the i^{th} input impulse. The coupling strength, η_ϕ, describes the amount of influence that the driver has upon the attentional oscillator.

The primary advantage to modeling beat perception in this way is temporal stability: the ability of such a system to sustain coordinated patterns in a the face of a fluctuating environment, and to reestablish coordination after perturbations (Schoner, 1991). Because of the special complexities of performed musical rhythms, however, adaptation of oscillator parameters is also required. First, because performed musical rhythms are nonstationary (i.e. the period changes in systematic ways), the model oscillator also adapts its intrinsic period:

$$p_{i+1} = p_i + \eta_p h(\phi_i, p_i, \gamma_i). \qquad \text{(Eqn 4)}$$

The coupling function, $h(\phi_i, p_i, \gamma_i)$, describes the adaptations of oscillator period that result from individual input impulses. Adaptation rate η_p (analogous to coupling strength in Equation 3) determines the rate at which oscillator period adapts to changes in the stimulus period.

The model handles rhythmic complexity by defining the functions $f(\phi_i, \gamma_i)$ and $h(\phi_i, p_i, \gamma_i)$ in such a way that phase and period change only when impulses fall within the temporal receptive field. Impulses that fall outside the field do not affect phase and period. This makes the size of the temporal receptive field crucial: if it is too small the oscillator will not robustly handle rubato, yet if it is too large, the oscillator will be led astray by complex rhythms. Thus, γ also adapts to the stimulus, according to the relation:

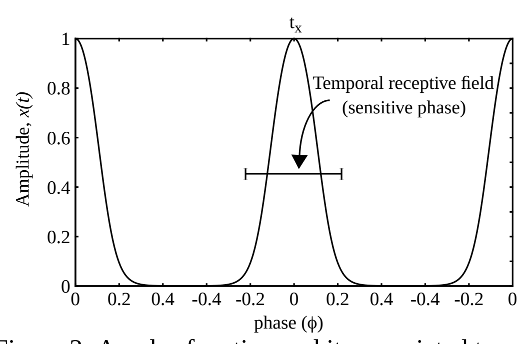

Figure 2: A pulse function and its associated temporal receptive field (sensitive phase).

$$\gamma_{i+1} = (\gamma_i + \eta_\gamma g(\phi_i, \gamma_i)) e^{\eta_\gamma r \frac{(t_{i+1} - t_i)}{p_i}}, \qquad \text{(Eqn 5)}$$

where $g(\phi_i, \gamma_i)$ describes adaptations of gain that result from individual input impulses. γ also decays each cycle, thus if there is no event in the oscillator's current cycle, γ decreases, widening the temporal receptive field. Finally, confidence, c, a measure derived from gain, varies from zero to one, measuring the overall success of the oscillator in finding a periodicity in the input signal. The functions $f(\phi_i, \gamma_i)$, $h(\phi_i, p_i, \gamma_i)$, and $g(\phi_i, \gamma_i)$ are derived in (Large, 1994).

Performances of Notated Melodies and Improvised Variations

The Test Data Set

To test the robustness of the model in response to musical performances, a test set of sixty melodies was collected as follows (for detailed description of the data collection procedure, see Large, Palmer, & Pollack, 1995). Two pianists performed melodies on a computer-monitored Yamaha Disklavier acoustic upright piano. Three children's melodies were chosen as performance material. For each melody, the pianists performed and recorded the melody, as presented in musical notation, five times. With the musical notation remaining in place, the pianists were then asked to play five improvisations. All performances were of a single-line melody only; pianists were instructed not to play harmonic accompaniment.

Next, skilled musicians transcribed the improvisations in standard musical notation. To assess the amount and distribution of rubato among the performances, a measure of timing deviation was calculated for each performance as a coefficient of variation of performed inter-onset intervals (IOIs), based either on the notation (for performances) or the transcriptions (for improvisations). This measure of deviation was then averaged across the five performances of each melody or improvisation by each pianist. An analysis of variance (ANOVA) on mean rubato by performance type (notated melody vs. improvised variation), subject, and tune was conducted. There was a significant main effect of performance type ($F(1,4) = 33.46$, $p < 0.01$), indicating that, on average, more rubato was used in the improvisation of variations than in the performance of the melodies from notation. Mean rubato was 0.05 for notated melodies, and 0.10 for improvisations. There was also a significant interaction between tune and subject ($F(2, 8) = 13.89$, $p < 0.01$). Pianist 1 performed the melodies and improvisations for the first two tunes with little rubato, but for the third tune with high rubato. Pianist 2 performed tune three with little rubato, and performed tunes one and two with relatively high rubato.

Beat Tracking Performance

Next, the oscillator's ability to model beat perception in these melodies was assessed. The response of the oscillator was intended to model the perception of beats at a particular level in a metrical structure (Lerdahl & Jackendoff, 1983). For each performance, the modal inter-onset interval (IOI) category was determined from the score or transcription, and chosen as the target periodicity. For each performance, the unit was initialized such that $\phi = 0$ at the initial onset, and p was set to the initial IOI of the target periodicity. Thus, the oscillator did not have to cope with finding initial phase or period.

A number of statistical measures of performance were collected. Only mean absolute value of relative phase, $\langle|\phi|\rangle$, is reported here because this measure corresponded most closely to intuitive impressions of successful coordination, gleaned by listening to oscillator output. $\langle|\phi|\rangle$ measures performance as the mean absolute phase of the oscillator on which events marking the beat occurred (as determined by the scores and transcription). $\langle|\phi|\rangle = 0$ means perfect performance, while $\langle|\phi|\rangle = 0.5$ is the poorest performance possible (meaning that the unit was $180°$ out of phase with the beat throughout the melody), and $\langle|\phi|\rangle < 0.1$ generally corresponded to a subjective impression of good performance. For each melody, oscillator performance was assessed by comparing $\langle|\phi|\rangle$ with the mean rubato score.

Performances of notated melodies and performances of improvised variations differed qualitatively in level of rhythmic complexity, and differed significantly in the magnitude of timing deviations, so results are discussed separately.

Performances of Notated Melodies. First, the oscillator was exposed to the thirty performances of notated melodies. Performances of notated melodies provided a controlled level of rhythmic complexity. Each melody contained three intended duration categories: sixteenth note, eighth note, and quarter note. Statistics were collected, and an analysis of variance (ANOVA) was conducted with factors tune, subject, and analysis type (mean rubato vs. average absolute phase). The ANOVA showed a main effect of analysis type ($F(1, 4) = 27.73$, $p < 0.01$), with mean rubato = 0.05, and average phase = 0.06. Thus, for these performances oscillator performed slightly worse than mean rubato would predict. This value of $\langle|\phi|\rangle$, however, indicates that on average the was able to achieve robust coordination with target periodicities.

To illustrate the nature of oscillator coordination, Figure 3 gives an example of the oscillator's behavior in response to a performance of *Baa baa black sheep*. Panel A provides a notated version of the melody (transcriptions of improvisations do not include grace notes or other ornaments) and a single row of dots from a metrical structure grid (Lerdahl & Jackendoff, 1983) marking the target events. Notes that are not marked by dots correspond to extraneous events; dots that do not correspond to notes mark times when events are "missing" from the target periodicity. Panel B shows both input and output of the oscillator. The dashed lines show impulses in the signal (marking event onset times). Because of the scale, full output pulses are not shown, rather discrete output pulses (shown as solid lines) are displayed at $t = t_x$. These two lines overlap when a target event is performed at precisely the time predicted by the oscillator, that is, at phase zero, $\phi(t) = 0$, of the driven oscillator. Amplitude of the discrete oscillator pulses corresponds to confidence, c. High amplitude of the discrete pulse corresponds to a small temporal receptive field, low amplitude corresponds to a wide receptive field.

Panel C shows a tempo curve for the performance as a solid line. This curve was derived by extracting the target events from the performance and graphing IOIs for these events. This curve gives the IOIs to which the oscillator should respond. Panel C shows actual observed cycle times of the oscillator using a dotted line. Observed cycle time takes into account not only the intrinsic period, p, of the oscillator, but also phase $\phi(t)$ as it is adjusted in each cycle. Beginning at the initial tempo, the unit effectively calculates a local tempo, and follows performance tempo as the performer speeds up and slows down.

Improvisation of Variations. Next, oscillator performance on the thirty improvised variations was examined. The improvisations provided a more difficult situation than the performances of notated melodies for two reasons. The rhythms of the improvisations were more complex than the rhythms of the melodies, making use of syncopation, and containing up to seven different levels of intended durations according to the transcriptions. Also the improvisations showed significantly greater timing deviation than did the performed melodies. The oscillator was exposed to the melodies, statistics were collected, and an analysis of variance (ANOVA) was conducted with factors tune, subject, and analysis type (mean rubato vs. average absolute phase). The ANOVA showed no main effect of analysis type ($F(1, 4) = 0.005$, $p = 0.947$), with mean rubato = 0.10, and average phase = 0.10. This result shows that for these performances, oscillator performance is on par with mean rubato. The ANOVA also indicated a significant interaction of tune and subject ($F(2, 8) = 4.0$, $p < 0.05$), indicating that the oscillator had more trouble with some performances than with others. Case by case examination revealed that in 20 out of the 30 cases, the oscillator coordinated well with its target periodicity ($\langle|\phi|\rangle < 0.10$). In 10 cases the oscillator had some difficulty. The 10 difficult cases were examined; two of the most difficult are discussed here.

Pianist 1's improvisations on *Mary had a little lamb* were performed in a freely timed blues style. The first improvisation had the highest rubato score (rubato = 0.25), and highest mean phase ($\langle|\phi|\rangle = 0.17$). The oscillator's behavior in this case was representative of its performance on this group of melodies, so it was chosen for further study. The time series corresponding to the performance of the oscillator are shown in Figure 4. The tempo curve indicates the presence of large timing deviations at several points in the melody. Points of particular interest are around t = 3000*ms*, t= 8000*ms*, t= 16000*ms*, and t = 23000*ms*. At these points, γ drops (correspondingly, confidence drops) allowing the oscillator to continue to synchronize with the target in spite of the large deviations.

In spite of these difficulties, however, the figure shows that the oscillator did a respectable job of entraining to its target periodicity in this rhythm. Beats are output at approximately the correct times throughout the piece – the oscillator is not lured away by the many distractor events in this rhythmically complex performance. Another way to see this is to note that the value of average absolute phase ($\langle|\phi|\rangle = 0.17$) is lower than mean rubato (0.25). Additionally, oscillator confidence is high for large sections of the piece; by the oscillator's internal measure its performance

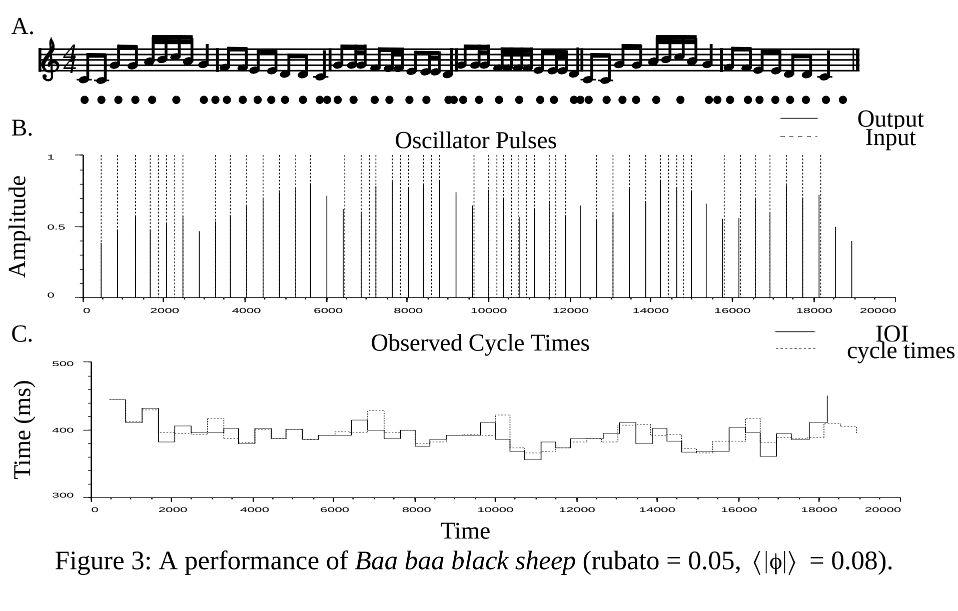

Figure 3: A performance of *Baa baa black sheep* (rubato = 0.05, $\langle|\phi|\rangle = 0.08$).

is good. 9 of the 10 problem cases examined fit the profile of this case. The oscillator had difficulty in certain sections of the performances, but always recovered gracefully, reestablishing coordination after large perturbations.

In 1 of the 10 difficult cases examined, however, the result was poor. Pianist 2's improvisations on *Hush little baby* were the most varied of all the improvisations studied and made heavy use of rubato. The improvisation that proved the most difficult for the model to handle was the third variation. This improvisation made moderately heavy use of rubato and had the highest mean phase, (rubato = 0.16, $\langle|\phi|\rangle = 0.30$). These numbers suggest extreme difficulty in coordination. Figure 5 shows the actual time series corresponding to the performance of the oscillator. Throughout this improvisation the performer makes use of a sort of "jagged" rubato. The tempo curve, shown in panel C, reveals timing deviations that strictly alternate: slower, faster, slower, faster. Panel C also shows the effect of this pattern on observed cycle times. Cycle times are always one step behind the performed durations because changes to the oscillator's phase and period in the current cycle effect oscillator cycle time for the following cycle.

Because timing deviations zigzagged in this fashion, cycle time decreased when performed duration increased, and vice-versa. This rubato pattern occurred in other performances as well, however in this case the amount of rubato was large enough to pose a serious difficulty for the oscillator. The oscillator responds to the correct events, and outputs pulses at more-or-less the correct locations throughout the piece. However, confidence is low throughout as the oscillator attempts to establish a coordinated pattern. Thus large, alternating rubato patterns represent a limiting case for the single oscillator model.

Discussion

Melodies are perhaps the most difficult cases for beat perception models, because they provide fewer reliable cues than accompanied melodies. Bass lines and harmonic accompaniment tend to be more rhythmically consistent, providing additional information. In this study, performances of notated melodies provided a controlled level of rhythmic complexity, while improvised variations provided syncopation and a great variety of duration categories. Both types of performance contained timing deviations, making the task of coordinating with a single periodicity a challenging one. Yet, in 49 out of 60 cases, the oscillator performed robustly by an objective measure ($\langle|\phi|\rangle$).

In 11 cases, difficulties were encountered ($\langle|\phi|\rangle > 0.10$). These difficulties were caused by large temporal deviations, stemming from three sources: heavy use of rubato including 'phase-shifts', actual timing errors on the part of performers, and jagged rubato curves resulting from alternating shortened and lengthened durations (Large, 1994). In 10 of the 11 cases, however, the oscillator was well coordinated for large sections of the melodies, having trouble in some areas but reestablishing coordination after large perturbations.

In the most difficult case, an improvisation that coupled heavy rubato with an alternating tempo profile, the oscillator performed poorly $\langle|\phi|\rangle = 0.30$. Although it did not lose the beat altogether, it was not able to adequately follow the tempo changes, and its internal measure of performance was consistently low. This case illustrates the limits of a single oscillator model. Note, however, that tempo changes of approximately the same magnitude strictly alternate (Figure 5, Panel C). This means that had the oscillator been operating at the next larger periodicity, it would have found almost no rubato at all. Cases such as this would tend to argue for a multiple oscillator model. In a multiple oscillator model different oscillators operate at different time scales (Large & Kolen, 1994). Internal interactions synchronize internal oscillators, and the perception

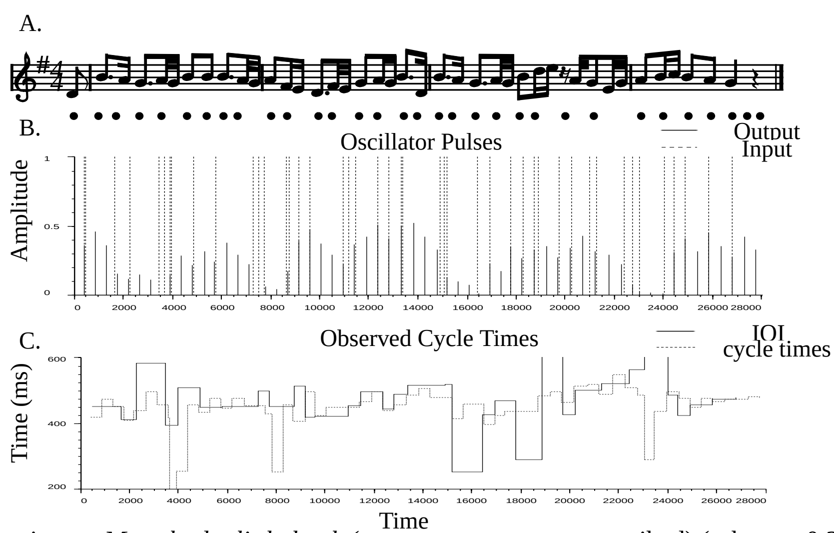

Figure 4: An improvisation on *Mary had a little lamb* (grace notes are not transcribed) (rubato = 0.25, $\langle|\phi|\rangle = 0.17$).

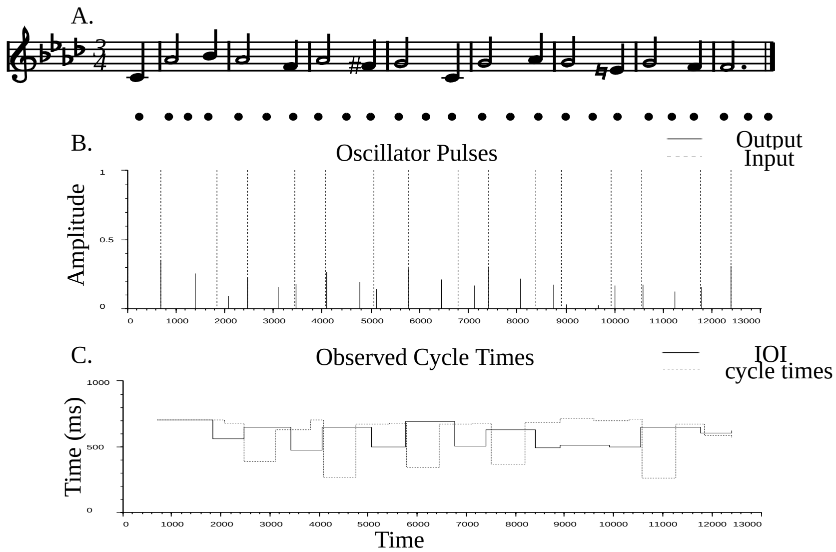

Figure 5: An improvisation on *Hush little baby* (rubato = 0.16, $\langle|\phi|\rangle = 0.30$).

of metrical structure can emerge from the perception of beat on multiple levels. Building networks of interacting oscillators to model the perception of metrical structure is a focus of current research.

Overall, the single oscillator model coordinated remarkably well with complex rhythms given no information other than event onset times. These analyses suggest that nonlinear oscillators, driven with complex, nonstationary rhythms that arise from musical performance, can adequately model the perception of musical beat, and may ultimately be used to model the perception of musical meter as well.

Acknowledgments

The author wishes to thank Caroline Palmer, Brent Stansfield, and three anonymous reviewers for helpful comments on an earlier version of this paper. This work was supported by The Ohio State University Presidential Fellowship.

References

Brown, J. C. (1992). Determination musical meter using the method of autocorrelation. *Journal of the Acoustical Society of America, 91*, 2374-2375.

Dannenberg, R. B. & Mont-Reynaud, B. (1987). *Following an improvisation in real time*. In Proceedings of the 1987 International Computer Music Conference. Computer Music Association.

Desain, P. & Honing, H. (1991). The quantization of musical time: A connectionist approach. In P. M. Todd and D. G. Loy (Eds) *Music and Connectionism* (pp. 150-160). Cambridge: MIT Press.

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

Kelso, J. A. S., deGuzman, G. C., & Holroyd, T. (1990). The self-organized phase attractive dynamics of coordination. In A. Babloyantz (Ed) *Self Organization, Emerging Properties, and Learning* (pp. 41-62). NATO ASI Series B: Physics, Vol. 260.

Large, E. W. (1994). *Dynamic representation of musical structure*. Unpublished Ph.D. Dissertation. The Ohio State University, Columbus, OH.

Large, E. W. & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science, 6 (1)*, 177 - 208.

Large, E. W., Palmer, C., & Pollack, J. B. (1995). Reduced memory representations for music. *Cognitive Science, 19*, 53-96.

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

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

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

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

Rosenthal, D. (1992). Emulation of human rhythm perception. *Computer Music Journal*, *16*, 64-76.

Schoner, G. (1991). Dynamic theory of action-perception patterns: The moving room paradigm. *Biological Cybernetics*, *64*, 455-462.

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* (pp. 427-447). Cambridge: MIT Press.

Vercoe, B. & Puckette, M. (1985). *Synthetic rehearsal: Training the synthetic performer.* In Proceedings of the 1985 International Computer Music Conference (pp. 275-278). Computer Music Association.