

Dynamics and Embodiment in Beat Induction

Douglas Eck, Michael Gasser and Robert Port
{deck,gasser,port}@cs.indiana.edu
Cognitive Science Program
Indiana University

1 Introduction

We are interested in modeling an aspect of human rhythm perception and production called beat induction. Roughly, *beat induction* consists of finding the downbeats in a metrical signal. The most common example of beat induction in human performance is foot tapping to music, and one way to state our goal is that we want to build feet which can tap to the radio as well as people do. However, we are interested in more than the end-state of adult rhythmical behavior; we also focus on the path that people take in perfecting this skill.

We build *embodied models*, ones with actual robotic components that interact with and constrain computational components. This commitment to a physical model of beat induction might seem wrong-headed. After all, beat induction is a perceptual phenomenon in adults, not necessarily involving motor control at all. But we believe that the interactions between body and brain in developing infants and toddlers cannot be ruled out as an important factor in beat induction learning. Infants exhibit rhythmical motor movements from birth: the tongue movements that make up the suckling reflex recur rhythmically with rates ranging from 600ms to 1200ms (Fraisse, 1982), and these rhythmical movements occur in concert with a perceptual preference for periodic sounds in roughly the same rate range (Demany, McKenzie, & Vurpillot, 1977).

In short, we see beat induction as an excellent micro domain for studying the interactions between brain, body and environment. A beat induction mechanism could plausibly be induced in the body as physical oscillation. Of course, one could argue interminably about whether the body is leading the brain or vice versa, and this is not our desire. (“Which came first, the action or the perception?”) We wish to avoid this trap by starting with an explicit body and an explicit yperceiving device from the start. In this way we can explore the interactions and mutual constraining that arises as these two systems interact without making a commitment to the relative importance of either. We do not wish to make Todd’s argument (Todd, 1985) that the body is required for adult beat induction. In fact it seems obvious to us that the oscillatory trunk movement in Todd’s model is not necessary for downbeat prediction. However, in the arena of developing rhythmical abilities, it is clear that the brain and body are wed from birth.

Our model includes a network of oscillators which interacts with auditory signals and a proposed mass spring arm which is driven by these oscillators to beat rhythmically. With this

framework we are able to model many aspects of human rhythm perception and production, including not only beat induction but also rhythmical pattern learning and metrical preference learning. (See (Gasser, Eck, & Port, 1999) for more on learning metrical preferences with similar networks.)

This chapter is devoted to contrasting our approach to other common approaches. We will start by examining a formal approach to rhythm cognition, offer a dynamical approach as a better way to capture some inherent properties of real performed rhythms, and end by arguing that embodiment constrains a dynamical system in a way which makes it easier to model the development of rhythmical abilities in people.

2 Formal approaches to rhythm

Rhythmical music is organized hierarchically such that the timing of events in the music is described with some success using integer ratios. This is borne out in sheet music notation, where the duration of a quarter note in 4/4 music is related to the duration of a bar by an exact 1:4 ratio. These relationships are not limited to notes and bars; in sheet music, all of the levels of the metrical hierarchy from 64th-notes to multi-bar phrases are related by perfect integer ratios.

Formalists such as Lerdahl and Jackendoff hold that the information found in sheet music is all that is necessary for a mostly-complete model of music cognition (Lerdahl & Jackendoff, 1983-1984). That is, along the dimension of time, all we need are perfect integer ratios. This downplays the importance of music performance, where messy non-integer ratio timing is always found. In fact, Lerdahl and Jackendoff make this point directly when they write, "It appears crucial to characterize mental structures before asking how they are computed over time (1983,p230)."

For formalists like Lerdahl and Jackendoff, representations come in the form of symbol strings. In general, these symbol strings are manipulated by a parser using rules taken from a grammar. Such a mechanism operates in lock-step time, processing each symbol discretely as if crossing a creek by jumping from stone to stone with no sense of what happens in between. This quality makes symbolic systems natural for working with representations having serial order (like notes in sheet music), but unwieldy for working with representations depending on continuous timing properties, like genuine performed music.

It is interesting that the very property which makes *formal systems* good for building digital computers undermines their effectiveness in working with performed music. When we do general computation we want our computers to be completely decoupled from real world time. This ensures that when we upgrade to a faster processor in our home PC, our spreadsheet software won't suddenly come up with different answers. It will simply come up with the same answers faster. But when we do want to couple with events in real world time, this same discreteness makes symbolic systems unworkable.

2.1 Reasons why the formal approach fails

A defender of the formal approach could argue that the real-time aspects of music are exceedingly minor and not worth modeling. To counter this argument we outline four particularly

important ways in which real performed music diverges from the perfect integer ratios found in sheet music. We believe that all four of these phenomena are important not only to beat induction but to rhythm cognition in general. Because they reveal our sensitivity to the real temporal organization of music, they go further than sheet music in telling a story about how humans are rhythmical.

2.1.1 Performance noise

When human subjects are instructed to tap periodically as if imitating a metronome, they cannot perform perfectly and end up producing output which is to some extent noisy (Collier & Wright, 1995; Wing, Church, & Gentner, 1989). Due at least in part to motor control limitations, subjects consistently produce beats that are a little too early or late as compared to a target metronome. In terms of performance, there are no perfect integer ratios. However, when listeners are asked to judge these noisy metronomes, they report them as being free of error (Clarke, 1987). That is, the auditory perceptual system is robust to this kind of noise, extracting the same sense of periodicity from signals with varying levels of temporal noise. With human perception, people make sense of noisy signals by extracting periodicity from them.

2.1.2 Expressive timing

Even if one controls for noise, a quarter note is still not a perfect integer ratio because musicians purposefully perturb the timing to accentuate different parts of the music (Repp, 1992). Musicians slow down and speed up in a way which helps bring alive the music they play. For example in twelve-bar blues the ending of a twelve-bar phrase (one layer in the metrical hierarchy) coincides with a deceleration in tempo. This kind of deviation from pure periodicity is preferred by listeners (Clarke, 1989), and so cannot be factored out as noise. When expressive timing information is removed from a performance, it is often described as being wooden or artificial.

2.1.3 Tempo invariance

From performance to performance, the tempo of a piece changes, sometimes drastically. Yet it is easy to observe that we make reliable downbeat predictions across a wide range of tempos (Port, Cummins, & McAuley, 1995). Sheet music notation may account for this by having a preferred performance rate in the form of a metronome marking, but this does little to account for the ranges of valid performing rates or how those rates are connected to preferred tempos.

2.1.4 Preferred tempo

Nor does sheet music account for a listener's preferred tempo or tactus. After considerable acceleration, a listener will change the metrical interpretation of a piece (violating tempo invariance) in an attempt to keep the rate of the downbeat assignment in a comfortable

range. There is evidence that this tactus is centered between 500ms and 600ms inter-onset interval (Fraisse, 1982) which is consistent with this reassignment of downbeats.

With these four phenomena we outline ways in which real time is an important aspect of rhythm perception and production. We believe that this sensitivity to real time is deeply embedded in our ability to perceive and perform music, and we reject the formalist claim that the appropriate cognitive structures can be found via representations which ignore the rich timing structure found in performed music.

3 From formal to dynamic

Having pointed out some problems with a symbolic approach to rhythm perception and production, it is now our task to offer an alternative framework for laying out theories and building models. We have taken great pains to show that the key problem with formal models is that they compute in lock-step time across discrete symbols. We offer an alternative framework of dynamical systems to address this problem.

We believe that cognitive systems are dynamical systems (Gelder & Port, 1995). A *dynamical system* is one where a rule of evolution determines the trajectory of a point through some state space. (For more on dynamical systems, see (Beek, this volume)). In the domain of beat induction, the state space might be defined in terms of phase and frequency, and the rule of evolution might define a trajectory that passes through the same region repeatedly, each time registering a beat. Systems like this are well-modeled using tools from dynamical systems theory.¹ This set of mathematical tools has proved to be useful in modeling all sorts of systems which evolve over time, from the aerodynamics of an airplane wing in flight to fluctuations in zoo-plankton populations. For the beat induction example, our modeling tool would presumably be taken from the class of dynamical systems with trajectories which exhibit limit-cycle behavior. These mathematical tools are of course called oscillators and are popular mechanisms for modeling beat induction (McAuley, 1995; Large & Kolen, 1994). 1994).

Of particular relevance for rhythm cognition is the way dynamical systems handle time. The crucial observation is that a dynamical system is defined at all points in time. The temporal resolution of a dynamical system is limited only by one's computational resources when running a simulation. Thus when one runs a dynamical system on a digital computer, the simulation can come arbitrarily close to continuous time presuming one wants to wait long enough. If one needs to do online processing in real time and so cannot afford to wait, the computing device of course becomes a limiting factor.

A dynamical system has no discrete points where computation must occur. The system is defined not only at the stepping stones in the creek but at all points in between as well. This makes it relatively easy to build simple systems which are sensitive to temporal processes. To be sure, we do not know how far we can get with dynamical systems in explaining human cognition, but we believe that many cognitive phenomena which have previously been attributed to rule following can be explained more simply and with greater predictive power using simple dynamical systems. As an example of a simple dynamical we present a network of oscillators which synchronizes with periodic signals.

4 Dynamical systems which synchronize

4.1 Adaptive oscillators

There are exist dynamical models which achieve beat induction by modifying their frequencies and phases to match the characteristics of some input signal. These adaptive oscillators (McAuley, 1995; Large & Kolen, 1994) are limit cycle oscillators which use a nonlinear phase-dependent function to match the phase and frequency of the oscillator to that of the input signal. fullciteAmcauley:1995 used a particular adaptive oscillator to model human sensitivity to relative timing in auditory signals. Large and Kolen use a slightly different adaptive oscillator as the underlying mechanism for a theory of attention (Jones & Boltz, 1989).

4.2 Stability and synchronization in networks of relaxation oscillators

We have used similar oscillators as the basic computational units in a network which learns metrical preferences (Gasser et al., 1999). Though we were successful in our goal to learn patterns, we found that adaptive oscillators of this type were unstable when coupled together in the large numbers necessary to support network learning. In looking for a more stable, fast synchronizing oscillator we recently explored the synchronizing properties of coupled networks of relaxation oscillators, as are found in models of neural action potential (Hodgkin & Huxley, 1952; Fitzhugh, 1961; Nagumo, Arimoto, & Yoshizawa, 1962; Morris & Lecar, 1981).

(Somers & Kopell, 1993, 1995) showed that networks of Morris-Lecar neural oscillators self-synchronize faster than do simple phase-pulling oscillators when coupled together via excitatory connections. Huber (1998) reported less-optimal but similar results using a simpler coupling function and a simpler relaxation oscillator called a Fitzhugh-Nagumo oscillator (Fitzhugh, 1961; Nagumo et al., 1962).

We believed that this fast coupling behavior was suggestive of an ability to entrain with simple rhythms. However, self-synchronization as described in the reports above does not involve an external input signal. We wished to know how such a system would respond to metrically-organized signals and began by testing the response of a network of Fitzhugh-Nagumo neurons using the simple case of an isochronous sequence of pulses. We found that these systems readily synchronize their firing with external metronomes (see figure 1) over a wide range of oscillator frequencies. For a mathematical description of the network, see (Huber, 1998).

This is a marked departure from the behavior of simple sinusoidal oscillators coupled using phase pulling, which were unable to synchronize reliably with the input (see figure 2).

Though these systems do phase and frequency lock with the periodic input, a closer look at their behavior reveals that they are not in fact very sensitive to periodicity. That is, they don't respond much differently to periodic signals than they do to aperiodic signals. This point is revealed by adding temporal noise to the signal, making it less periodic. Under a wide range of conditions, the network in general synchronizes almost as tightly with these

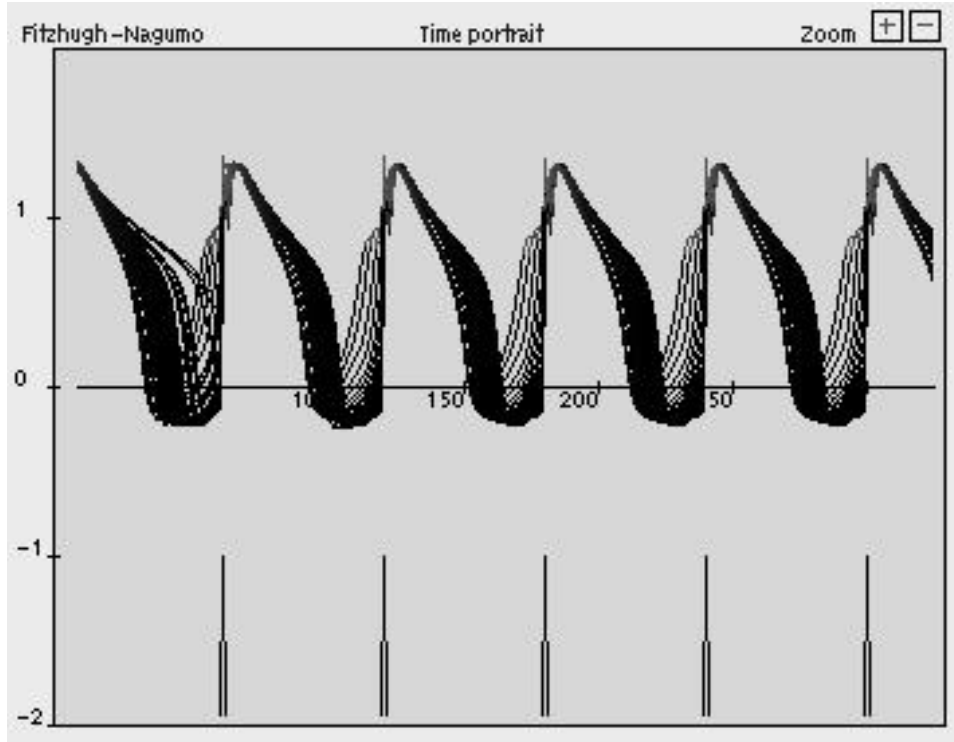


Figure 1: A network of 40 fully-connected Fitzhugh-Nagumo neurons with resting periods from 60 to 180 synchronizing with a period 60 pulse. In this example all connection weights are set to +1.

noisy signals as it does with periodic signals (see figure 3).

Also, the network is incapable of sustaining synchronization for even a short time when the input signal is removed. As soon as the input signal is turned off, the network begins to return to its resting state, losing whatever phase and frequency locking it had with the input. If the system had something akin to momentum it would continue moving in a similar fashion and gradually rather than catastrophically desynchronize.

4.3 Meter through fractured synchrony

We continued this exploration with a network of Fitzhugh-Nagumo oscillators like that of Huber. However, instead of using non-varying input for self-synchronization, we introduced a continuous-time metrical pattern. It came as no great surprise that a simple network with all excitatory connections in fact did phase lock and frequency lock with the input signal. After all, the input signal was sufficiently strong to force the oscillators to fire with each beat, causing them to align. What was more surprising was the behavior of the network using inhibitory connections. When the network was organized in a ring topology with nearest neighbors inhibiting one another, it was able to couple with a signal organized in 3/4 meter such that most energy was placed on the downbeat (see figure 4).

That is, the oscillators did not fire en masse each time a pulse occurred, but rather in repelling one another via the inhibitory connections, they distributed the firing out in such

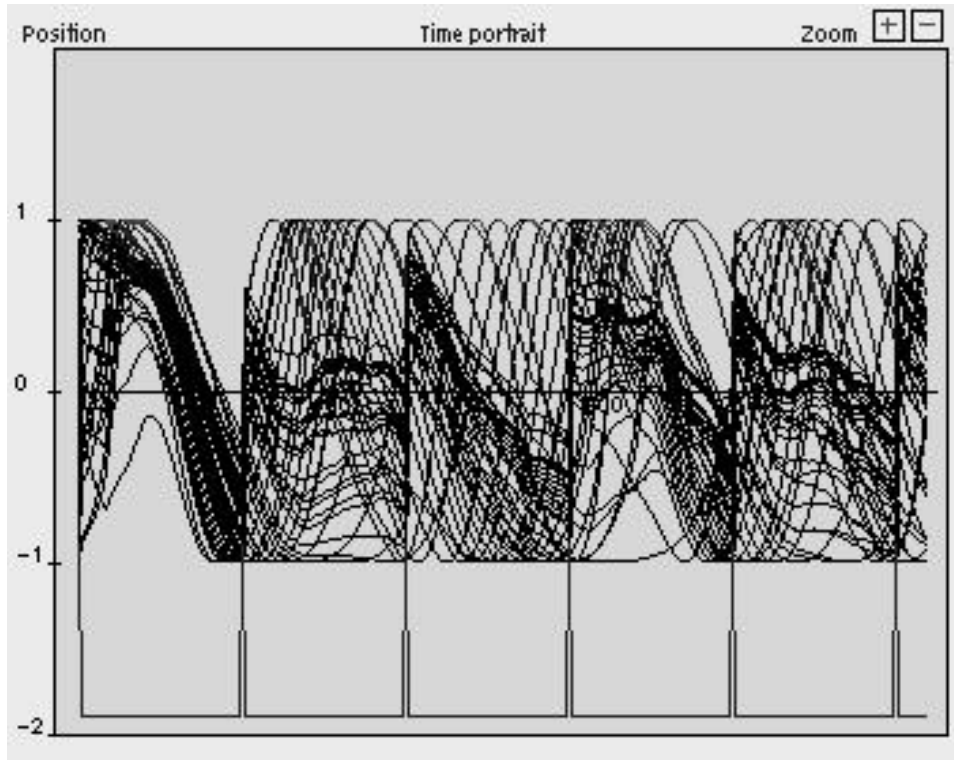


Figure 2: A network of 40 fully-connected phase pulling oscillators with resting periods from 60 to 180 synchronizing with a period 60 pulse. In this example all connection weights are set to +1.

a way that it reflected the meter of the pattern. This behavior mirrors fractured synchrony, a phenomenon claimed by Somers and Kopell (1995) to be unique to networks of nonlinear coupled oscillators.

4.4 Towards entrainment

We believe that these networks of relaxation oscillators show great promise as devices for synchronizing with metrical patterns. They synchronize very quickly to periodic patterns and, via fractured synchrony, respond to some of the metrical structure present in a pattern. That is, at least for a very simple pattern of weak and strong beats, the oscillators phase lock to all of the beats with more oscillators tending to phase lock with stronger beats. Also, they settle into stable limit when coupled together in large numbers.

What is lacking is a suitable energy source to keep the network synchronized in the absence of the input signal. The adaptive oscillators mentioned above (McAuley, 1995; Large & Kolen, 1994) are equipped with such an energy source in the form of a sinusoidal forcing function which is modified to allow phase resetting and frequency modification. Our relaxation oscillator networks receive only a trickle of constant energy input to keep them moving and are otherwise wholly dependent on the input signal. We resist adding a forcing function to our networks because it would change their dynamics so much that we could no longer rely on the findings of Somers and Kopell (1995) which guarantee stability in these

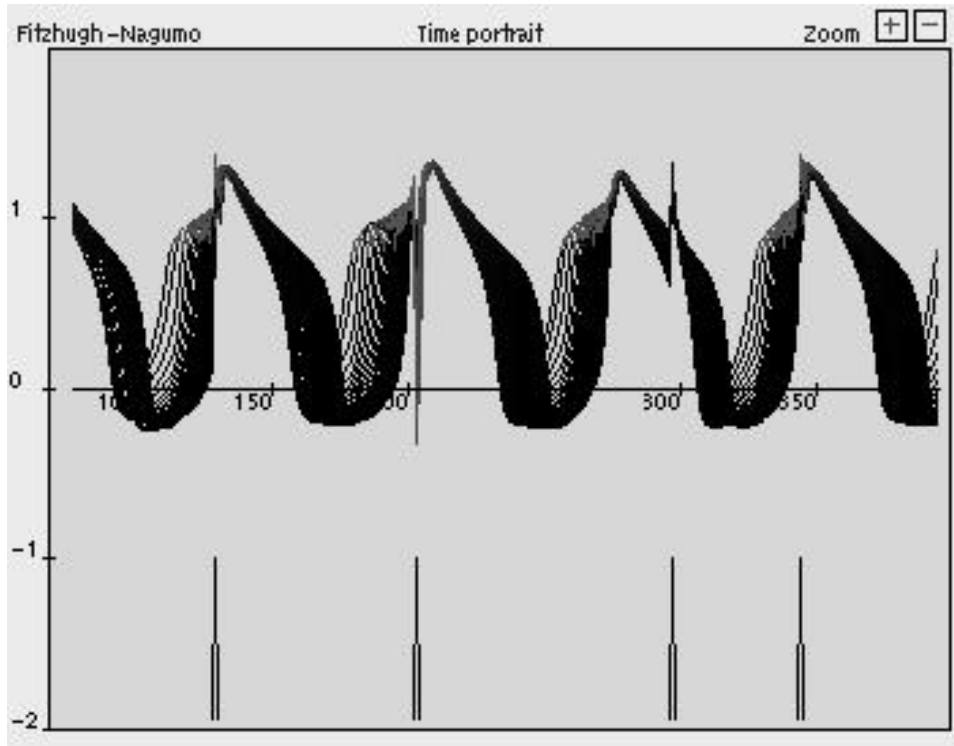


Figure 3: This network is the same as the one in figure 1. This time the input signal is a metronome plus some amount of random noise. Note that synchronization is still quite strong despite the noise in the signal.

networks over a large range of conditions. Hence we are left looking for a mechanism which can take advantage of the fast synchronizing properties of the relaxation oscillator network but which has the ability to continue moving in the absence of input. Our solution is to use a mass spring oscillator in the form of a robotic arm.

This last section of the chapter introduces some ideas central to embodiment and discusses the role of a robotic arm in our beat induction mechanism. In short, the arm is built to achieve two things: one, it is to provide the energy source mentioned above and, two, it is to constrain the relaxation oscillator network to prefer periodicities and hence to couple less readily with non-metrical stimuli.

5 From dynamic to embodied

If there is any lesson to take from robotics research in the last decade, it is this: do not underestimate the richness of a real environment. For an overview see [fullciteAclark:1997](#). If these roboticists were to have an ad campaign to further their cause, the slogan would surely be “Just Build It.” For building a physical model is the best way to couple with the physical environment.

Of course if simple mass spring dynamics were all we wanted, we could use a forced damped mass spring as derived from Hooke’s Law and Newton’s Second Law of motion:

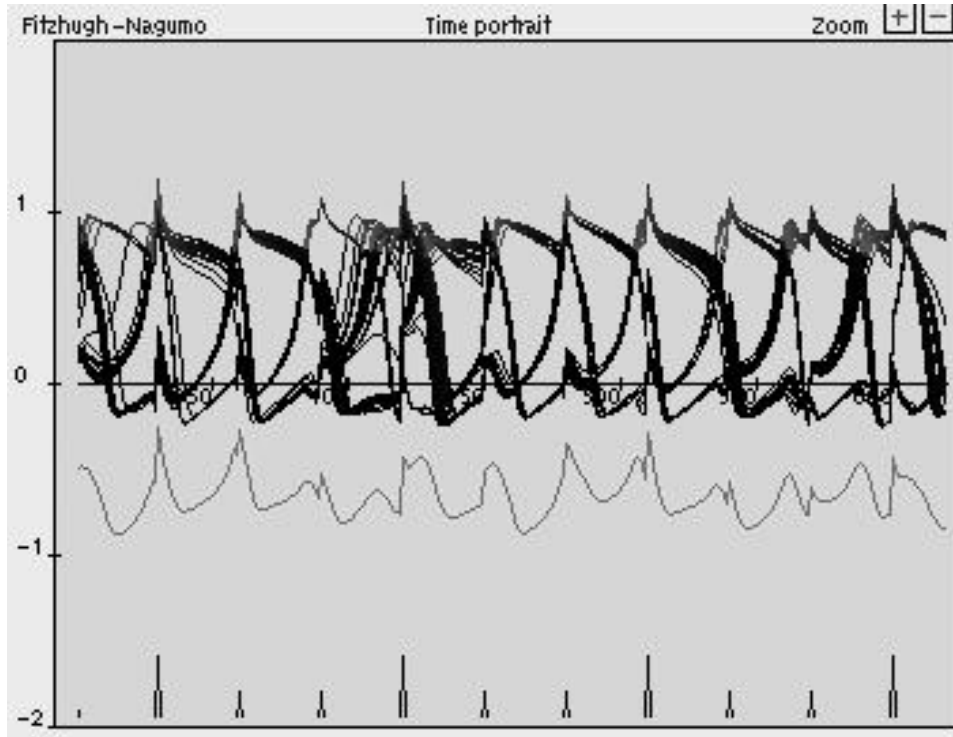


Figure 4: This network of 40 oscillators ranging in period from 95 to 105 exhibits fractured synchrony. All connection weights are set to -1. Note that despite some instability there are three stable groups of oscillators tracking the three elements of the period 30 input pattern consisting of two weak beats followed by a strong beat. The line in the middle is the sum of the fast (x-axis) variable for all the oscillators.

$$m\ddot{x} + 2\zeta\dot{x} + kx = \text{input} \quad (1)$$

This simple spring could then be driven by the oscillators using the network's summed output as a driving signal. And in fact as part of our simulations we do exactly this. But simulating even the simplest perturbation of such a system requires thought. How, for example, would we model someone knocking the arm or tightening the springs or changing the mass? To be sure, each of these actions can be modeled if sufficient care is taken. However, we do not want the responsibility for simulating such an environment when a real mass spring working in a real environment would do a better job.

This approach is not new in robotics. An excellent example of this idea at work is seen in Randal Beer's robot cockroaches (Beer & Chiel, 1993; Beer & Gallagher, 1992). These tiny mechanisms use their legs as an active part of computing their gait. Beer achieves this by distributing the computation to all six legs by equipping each limb with a few artificial neurons. These neurons get position information from the leg via a proprioceptive connection and use that leg position to help drive the computation. The results are impressive. If for example a leg is wounded, other legs compensate by modifying gait. Our point in bringing up this example is that Beer has gotten some computational mileage out of simple

interactions between the legs and the terrain. If he had simulated the robots, he would have been responsible for simulating the terrain, including leg entrapments, adding perhaps more complexity to his modeling framework than was added by his use of real robots.

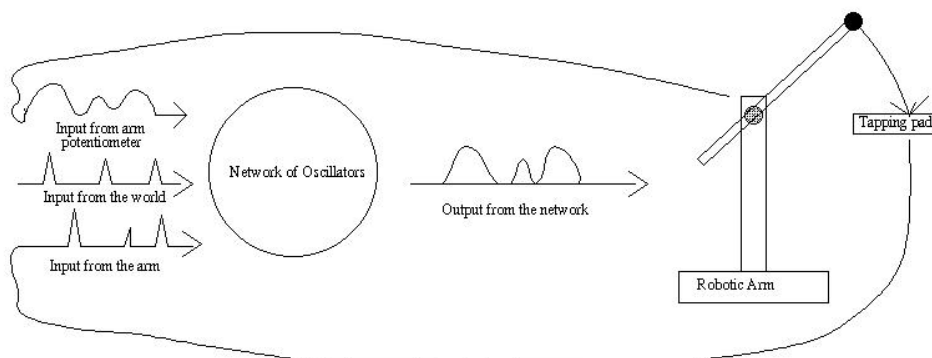


Figure 5: An idealized diagram of our mass-spring robotic arm model.

Our robotic arm though simple in construction is capable of exhibiting complex mass-spring behavior. (See figure 5) The robot consists of a weighted baton mounted on an axle. This axle is coupled by springs to two servomotor-driven rods. By sending appropriate signals to the motors, the rods retract and extend, changing the tension on the spring and pivoting the arm on its axle. A potentiometer is mounted at the axle, allowing us to monitor arm angle as a continuous signal. Via its two-motor configuration, the arm is capable of simulating the antagonistic muscles in the human arm. Note that the arm looks more like a pendulum turned on its side than it does a mass spring. In fact, mass spring dynamics are equivalent to pendulum dynamics when the pendulum is stiff and the swings of the pendulum are not of high magnitude. We use the term "mass-spring" because it focuses on the slow-to-change "massy" properties of these oscillators more than does the term "pendulum".

We outlined above some aspects of beat induction that are ignored by symbolic models. We return to these and outline ways in which they can be addressed by a system having both the inertia of a mass-spring dynamics and the fast synchrony of a relaxation oscillator network.

5.1 Performance noise

Due to its momentum, a mass spring system can resist the tiny random changes in trajectory caused by performance noise. Heavier mass springs are proportionally better at this filtering than light mass springs. However, this resistance to noise also makes mass springs relatively slow to synchronize. Our design is to take advantage of the fast synchronization properties of Fitzhugh-Nagumo oscillators to quickly couple with an input signal. We can then use the output from these oscillators to drive the mass spring system into a stable entrainment pattern.

5.2 Expressive timing

Expressive timing requires sensitivity to events which diverge from rigid periodicity. We note that the network of relaxation oscillators can couple and synchronize successfully with expressively-timed input signals. This in itself is no great feat since these networks can synchronize successfully with most signals. One of our future goals is to study methodically the ways in which the mass-spring arm handles output from the oscillator network as the network is being driven by expressively-timed pieces of music.

5.3 Preferred tempo and Tempo Invariance

Fitzhugh-Nagumo oscillators, like van der Pol oscillators, increase in frequency as signal amplitude increases. This makes it unlikely that such neural oscillators will naturally exhibit a strong tactus. Mass spring systems on the other hand have resonant frequencies correlated to their mass and stiffness. This causes them to exhibit a preferred tempo that can be adjusted by modifying the mass and or spring stiffness of the system. In this way the system can be tuned to (a) prefer a base tempo which a function of the mass of the arm and (b) couple with patterns having frequencies near that base tempo.

6 Future work and conclusions

We intend to model different aspects of rhythm perception and production with particular focus on the development of rhythmical abilities like beat induction. Like Beer's cockroach robots, our mass spring arm is able to send proprioceptive information back to the network of oscillators, allowing us to modulate the extent to which the network of oscillators has access to arm position information. This gives us a way to explore how the body and brain might work together via proprioception to solve the problem of rhythm. One area of particular interest involves network learning: in what ways does the arm constrain the kinds of patterns that can be learned? Our intuition is that the network will show a preference for rhythmical signals in proportion to the amount of proprioceptive information available.

We have outlined a dynamical, embodied approach to beat induction that focuses on the interactions between body, brain and environment. We focused on two aspects of the task, synchronization and prediction, and offered as a solution a system which incorporates simulated neurons and a mass spring arm. We believe that this dual approach has great potential in modeling beat induction and perhaps even rhythm cognition in general.

References

- Beer, R., & Chiel, H. (1993). Simulations of cockroach locomotion and escape. In R. Beer (Ed.), *Biological neural networks in invertebrate neuroethology and robotics*. New York: Academic Press.
- Beer, R., & Gallagher, J. (1992). Evolving dynamic neural networks for adaptive behavior. *Adaptive Behavior*, 1, 91–122.

- Clarke, E. F. (1987). Categorical rhythm perception: An ecological perspective. In A. Gabriellsson (Ed.), *Action and perception in rhythm and music*. Stockholm: Royal Swedish Academy of Music.
- Clarke, E. F. (1989). The perception of expressive timing in music. *Psychological Research*, *51*, 2–9.
- Collier, G. L., & Wright, C. E. (1995). Temporal rescaling of simple and complex ratios in rhythmic tapping. *JEP-HPP*, *21*(3), 602–627.
- Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature*, *266*, 718–719.
- Fitzhugh, R. (1961). Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, *1*, 455–466.
- Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York: Academic Press, Inc.
- Gasser, M., Eck, D., & Port, R. (1999). Meter as mechanism: A neural network that learns metrical patterns. *Connection Science*, *1*.
- Gelder, T. van, & Port, R. F. (1995). It's about time: An overview of the dynamical approach to cognition. In R. F. Port & T. van Gelder (Eds.), *Mind as motion* (pp. 1–44). Cambridge, Massachusetts: The MIT Press.
- Hodgkin, A., & Huxley, A. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, *117*, 500–544.
- Huber, D. E. (1998). The development of synchrony between oscillating neurons. In *Proceedings of the 20th annual conference of the cognitive science society*.
- Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*, 459–491.
- Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science*, *6*, 177–208.
- Lerdahl, F., & Jackendoff, R. (1983-1984). An overview of hierarchical structure in music. *Music Perception*, *1*(2), 229–252.
- McAuley, J. D. (1995). *On the perception of time as phase: Toward an adaptive-oscillator model of rhythm*. Unpublished doctoral dissertation, Indiana University, Bloomington, IN.
- Morris, C., & Lecar, H. (1981). Voltage oscillations in the barnacle giant muscle fiber. *Biophysical Journal*, *35*, 193–213.
- Nagumo, J., Arimoto, S., & Yoshizawa, S. (1962). An active pulse transmission line simulating nerve axon. *Proceeding IRE*, *50*, 2061–2070.

- Port, R. F., Cummins, F., & McAuley, J. D. (1995). Naive time, temporal patterns, and human audition. In R. F. Port & T. van Gelder (Eds.), *Mind as motion* (pp. 340–371). Cambridge, Massachusetts: The MIT Press.
- Repp, B. H. (1992). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition*, *44*, 241–281.
- Somers, D., & Kopell, N. (1993). Rapid synchronization through fast threshold modulation. *Biological Cybernetics*, *68*, 393–407.
- Somers, D., & Kopell, N. (1995). Waves and synchrony in networks of oscillators of relaxation and non-relaxation type. *Physica D*, *89*, 169–183.
- Todd, N. (1985). A model of expressive timing in tonal music. *Music Perception*, *3*, 33–58.
- Wing, A. M., Church, R. M., & Gentner, D. R. (1989). Variability in the timing of responses during repetitive tapping with alternate hands. *Psychological Research*, *51*, 28–37.