ELSEVIER

Human Movement Science 19 (2000) 527–566

HUMAN MOVEMENT SCIENCE

www.elsevier.com/locate/humov

On synchronizing movements to music

Edward W. Large *

Center for Complex Systems and Brain Sciences, Florida Atlantic University, 777 Glades Road,
P.O. Box 3091, Boca Raton, FL 33431-0991, USA

Abstract

How do people synchronize movement patterns with music? Most likely, when people listen to a musical rhythm, they perceive a *beat* and a *metrical structure* in the rhythm, and these perceived patterns enable coordination with the music. Here, a model of meter perception is proposed in which a musical stimulus provides input to a pattern-forming dynamical system. Under rhythmic stimulation, the system undergoes bifurcations that correspond to the birth of self-sustained oscillations and the formation of temporally structured patterns of oscillations. The resulting patterns dynamically embody the perception of beat and meter, and they are stable in the sense that they can persist in the in the face of rhythmic conflict. The performance of the model is compared with the results of a recent beat induction study (J. Snyder, C.L. Krumhansl, Tapping to ragtime: Cues to pulse-finding. Music Perception, 2000 (under review)) in which musicians tapped along with musical rhythms. The network closely matched human performance for natural musical signals and showed a similar pattern of breakdowns as the input degraded. The theoretical implications of these findings are considered. © 2000 Elsevier Science B.V. All rights reserved.

PsycINFO classification: 2326; 2340

Keywords: Auditory perception; Music perception; Cognitive processes; Rhythm; Perceptual motor coordination; Finger tapping; Task complexity; Time perception; Dynamical systems

---

* Tel.: +1-561-297-0106; fax: +1-561-297-3634.
E-mail address: large@walt.ccs.fau.edu (E.W. Large).

PII: S0167-9457(00)00026-9

1. Introduction

Perhaps the most natural response to music, one that most of us observe or take part in every day, is to synchronize our bodily movements with musical rhythms. We clap our hands, snap our fingers, stomp our feet, sway, dance or in some other way coordinate our movements with the temporal structure that we perceive in musical events. The organic, effortless, and often completely spontaneous nature of musical synchronization suggests a deep relationship between movement and music. Yet it also masks a phenomenon of extraordinary complexity. How do we synchronize our movements to music?

Many investigators have studied the ability to synchronize motor behavior with auditory sequences using simple isochronous and/or tempo modulated tone sequences as stimuli and a number of models have been proposed to explain the behavioral findings, including patterns of variability, compensation for temporal stimulus perturbations, and a general anticipation bias (e.g., Chen, Ding & Kelso, 1997; Hary & Moore, 1987; Kelso, DelColle & Schöner, 1990; Mates, Radil & Pöppel, 1992; Pressing & Jolley-Rogers, 1997; Semjen, Vorberg & Schulze, 1998; Vorberg & Wing, 1996). In some ways, these studies are relevant to the question at hand, since the simplest example form of musical coordination is to tap our fingers, more or less isochronously, with the beat of the music. Yet, even relatively simple music is far more elaborate than the monotonous ticks of an auditory metronome; one question that such studies do not address is that of stimulus complexity. Thus, while there is a great deal of data available on synchronization with simple acoustic signals, relatively little is known about how people accomplish the similar feat of synchronization with more complex auditory stimuli such as music.

How do people tap along with music? Most likely, when we listen to music we hear a beat – a psychological, more-or-less isochronous, pulse train – that provides the stimulus for synchronization. Although this proposal is intuitively appealing, the subtlety of the rhythmic and tonal materials of most music makes this suggestion more difficult to assess than it might appear at first blush. Except in the case of rhythmically simple music, it is rarely clear from a straightforward analysis of the acoustic signal where "the beat" is located, even though it may be quite obvious upon auditory presentation of the musical material. This has led investigators in the field of music perception to propose various theoretical and computational approaches to beat perception (Desain & Honing, 1991; Brown, 1993; Essens & Povel, 1985; Gasser, Eck & Port, 1999; Goto & Muraoka, 1998; Large & Kolen, 1994;

Large & Jones, 1999; Large & Palmer, 2000; Longuet-Higgins & Lee, 1982; Parncutt, 1994; Scarborough, Miller & Jones, 1992; Scheirer, 1998; Steedman, 1978; Todd, Lee & O'Boyle, 1999; Toiviainen, 1998; Dannenberg & Mont-Reynaud, 1987; Van Noorden & Moelants, 1999; Vercoe & Puckette, 1985; Vos, Van Dijk & Schomaker, 1994). These models, known variously as beat induction, pulse finding, or beat tracking models, attempt to predict the period and phase of the psychological pulse series (the beat) of a complex acoustical stimulus on a moment-to-moment basis. While each has shed some light upon the issues at hand, none to date has been entirely successful in capturing the robustness of human beat perception.

One problem for models of beat induction is that there is still relatively little empirical data available. However, recently a few investigators have studied tapping to music (Jones & Pfordresher, 1997; Repp, 1999a,b,c; Scheirer, 1998; Snyder & Krumhansl, 2000; Van Noorden & Moelants, 1999) while several others have studied the effects of stimulus complexity in tapping and time perception (e.g. Deutsch, 1983; Jones & Yee, 1997; Large & Jones, 1999; Peper, Beek & Van Wieringen, 1995; Povel & Essens, 1985; Pressing, 1998, 1999). Thus, some preliminary findings have been established. People can find and track the beat of music and other rhythmically rich patterns. People exhibit preferences for temporal structures that embody simple time ratios (e.g. 2:1, 3:1, 3:2). Non-temporal cues such as pitch, loudness and harmony, interact with temporal cues in beat finding. People demonstrate preferred absolute tempi around an inter-beat-interval (IBI) of approximately 600 ms, in both perception and production.

Building upon earlier theoretical work (Jones, 1976; Large & Kolen, 1994; Large & Jones, 1999), I present an approach to beat induction that takes into account the empirical results discussed above. The basic premise of this approach is that the stability properties of a specific type of non-linear dynamical system make it appropriate for modeling beat perception in music. I propose a model of beat induction in which a musical stimulus provides input to a *pattern-forming* dynamical system. Under rhythmic stimulation, the system undergoes bifurcations that correspond to the birth of self-sustained oscillations and the formation of temporally structured patterns of oscillations. The resulting patterns dynamically embody the perception of musical beats on several time scales, patterns known in music theory as meter, or metrical structure (e.g. Cooper & Meyer, 1960; Hasty, 1997; Lerdahl & Jackendoff, 1983; Yeston, 1976). These patterns are stable yet flexible: They can persist in the absence of input and in the face of conflict, yet they can also reorganize given potent indication of a new temporal structure. Thus, this approach may be capable of capturing the robustness of human beat perception. As a preliminary test of the model, I compare its performance with the results of a synchronization study (Snyder & Krumhansl, 2000) that was explicitly designed to test beat induction in musical signals.

2. Background

Snyder and Krumhansl (2000) investigated the nature of mental pulse-finding using a synchronization tapping task. They presented participants with excerpts of ragtime piano pieces, and asked them to tap the most comfortable pulse of each excerpt on a piano keyboard.[1] Because their stimuli were relatively complex, and the tapping data provided a good index of human beat induction performance, these pieces were chosen as an initial test of the model of meter perception developed below. A four bar section of one of the Snyder and Krumhansl stimuli will be used to illustrate the basic concepts of rhythm, beat, and meter. The model itself is described in detail in the next section.

The stimuli consisted of eight ragtime piano pieces; an excerpt of one of these pieces (Lily Queen) is notated in Fig. 1. The music notation indicates right-hand (RH) and left-hand (LH) parts, written on separate staves. The pitch of each note is indicated by vertical height of the note head, while the sequential order of notes is indicated by horizontal position; simultaneously sounded notes (chords) are stacked vertically. The relative timing of events is indicated by the style of the note stems and beams. A stemmed note without a flag or horizontal beam is a quarter-note, and its duration is equal to one beat period, a stemmed note with a single beam is an eighth-note, equal to 1/2 of a beat period, and stemmed notes with double beams are sixteenth-notes, equal to 1/4 of a beat period. The sixteenth note rests in this excerpt denote silences (in the RH part) for 1/4 of a beat period. The beat period itself is a relative temporal measure. Absolute timing information is provided by the metronome marking (here, MM = 102 beats per minute). In conjunction with the relative timing information provided by the notation, it indicates absolute durations of 588.2, 294.1, and 147.1 ms, for quarter, eighth, and sixteenth notes, respectively. The bar lines (vertical lines through both staves) denote grouping of events into durations equal to two beat periods (1176.5

---

[1] Each tap triggered a percussive hi-hat sound.

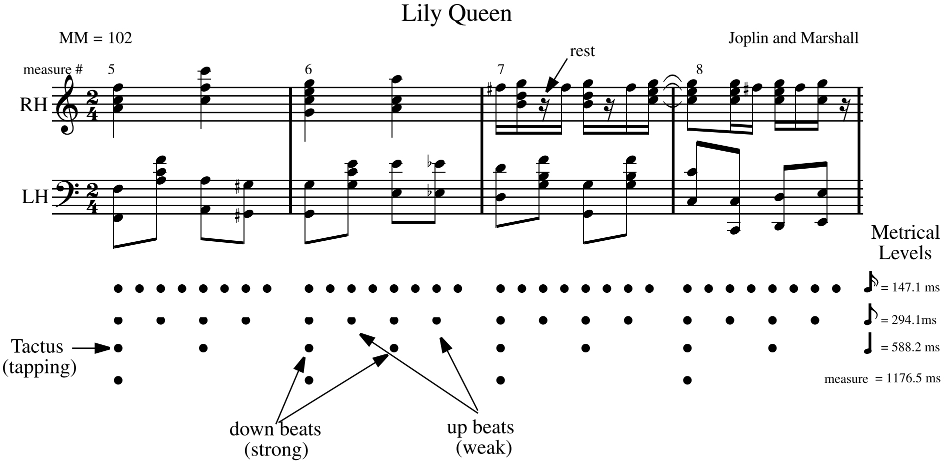

Fig. 1. Four bars from one of the ragtime stimuli used by Snyder and Krumhansl (2000).

ms); these groups are called measures, or bars. Finally, note that in a human performance of this piece, neither the absolute nor relative durations would be precisely reproduced. Rather, musical performances exhibit both systematic and random deviations from metronomic regularity (see Palmer, 1997).

Rhythm refers to the organization of events in time, more specifically the *pattern of timing and accent* associated with a music sequence, as shown in Fig. 1. By accent I mean the phenomenal accent (cf. Lerdahl & Jackendoff, 1983) associated with each sounded event. Phenomenal accent arises through a complex combination of many variables, including pitch, loudness, timbre, and duration, and is thus difficult to define in a straightforward way. One important factor in phenomenal accent, however, is acoustic stress (i.e. intensity accent). So for example, the more notes are sounded simultaneously the stronger the perceived accent is likely to be, if everything else is equal.

In psychological terms, a *beat* is a series of perceived pulses marking subjectively equal units in time. According to Cooper and Meyer (1960):

> Though generally established and supported by objective stimuli (sounds), the sense of pulse may exist subjectively. A sense of pulses, once established, tends to be continued in the mind and musculature of the listener, even though the sound has stopped. For instance, objective pulses may cease or may fail for a time to coincide with the

> previously established pulse series. When this occurs, the human need for the security of an actual stimulus or for simplicity of response generally makes such passages point toward the re-establishment of objective pulses or to a return to pulse coincidence (p. 3).

In other words, perceived beat is an inference from the acoustic stimulus (Lerdahl & Jackendoff, 1983), and functions as an expectation for when events are likely to occur in the future (cf. Large & Kolen, 1994; Large & Jones, 1999). A perception of beat generally arises in response to an actual periodicity present in the musical signal; however, it is stable in the sense that once perceived it may continue even when the periodicity is interrupted in some way. Perceived beats are transcribed in Fig. 1 using horizontal rows of dots beneath the musical score (cf. Lerdahl & Jackendoff, 1983); in this example four levels of beats are shown. The different levels of beats correspond to the timing of measures, quarter-notes, eighth-notes, and sixteenth notes. Finally, when multiple beat levels are perceived one level is generally more salient than the others. This level is called the *tactus*; it is defined by the points at which one would tap along with the music, and the tactus period is generally notated using quarter notes. [2]

Metrical structure, or simply *meter*, refers to the temporal pattern that is created by the simultaneous perception of beats at several different time scales, as shown in Fig. 1. In the figure, the horizontal spacing between beats denotes the beat period, while vertical alignment denotes relative phasing among beats of different levels. Metrical structure describes one of the most important components of rhythmic experience: the feeling of regularly recurring strong beats called *metrical accent* (Lerdahl & Jackendoff, 1983). Metrical accents arise at temporal locations where the beats of many levels come into phase. Points where many pulses coincide are felt as strong; points where few beats coincide are felt as weak. Strong beats are sometimes referred to as down-beats, and weak beats as up-beats. Like beat, meter can be thought of as an inference from the musical signal. It is a perceived temporal structure that is related, but not equivalent to the patterns of timing and accent (rhythm) of the music. Meter functions as a complex form of temporal expectation: Events are more strongly expected at strong beats, and less expected on weak beats (cf. Desain, 1992).

---

[2] The tactus is the basic beat or pulse of the music, in everyday language we sometimes call it *the* beat, even though several beat levels may be perceived simultaneously.

To summarize, a rhythm is a temporal pattern that can give rise to perceptions of beat and meter. Beat is a psychological pulse that functions as a stable expectation: Once established, it tends to continue "in the mind and musculature of the listener," even when the rhythm stops or temporarily comes into conflict with the pulse series. Likewise, a metrical pattern, once inferred, persists in the face of rhythmic conflict, and gives way only to strong or consistent contradiction. But how is a sense of beat established? How do metrical patterns form? And how do they adapt when the music demands?

3. Modeling meter perception

The current approach treats meter perception as an active, anticipatory mode of rhythm perception, and is based upon three key principles. The first is that the perception of musical beat is most appropriately modeled as an active, self-sustained oscillation. This may be conceived as a direct mathematicization of Cooper and Meyer's description, that a sense of beat "once established, tends to be continued in the mind and musculature of the listener, even though...objective pulses may cease or may fail for a time to coincide with the previously established pulse series." Second, the perception of musical meter can be modeled as a network of oscillators that are coupled to one another. Such a network should give rise to a dynamic metrical percept whose structure reflects both the temporal structure of the input as well as internal dynamic constraints. Finally, the oscillators of the network, when driven with a complex external rhythm, should entrain to different periodicities within the temporal pattern. The entrainment of oscillators at multiple time scales provides the listener with a framework that shapes expectations about future events (Large & Kolen, 1994; Large & Jones, 1999).

Self-sustained oscillation arises in a variety of physical systems from electrical circuits to ecology and neurobiology. Such systems can often be modeled accurately using non-linear differential equations, where the state variables and dynamics that capture the nature of the oscillation vary depending upon the physical system under study (see e.g. Hirsh & Smale, 1974; Murray, 1989). The current study, however, focuses on the development of a mathematical framework for capturing musical behavior. Thus, it would be desirable to work at a phenomenological level, a level of description that can capture our observations about perception and action directly, without worrying too much about the details of the neural system that gives rise to the hypothesized oscillations. Fortunately, mathematics provides us with a suitable tool, called *normal form* dynamical systems. A normal form dynamical system abstracts away from the details of particular physical models, in an attempt to succinctly capture the basic properties that are shared among a family of more complex differential equations. Normal form dynamical systems are, in a rather specific mathematical sense, the simplest systems that capture the basic behavior under study (Arrowsmith & Place, 1990). Within the normal form framework, self-sustained oscillations are described by a normal form for the Hopf bifurcation, which when written in the form introduced below, is sometimes called a *Hopf oscillator*.

In this section, a model of meter perception is developed in four steps. First, the Hopf oscillator is introduced and its basic behavior described. Second, a network is defined in which Hopf oscillators compete with one another to describe the temporal structure of the afferent rhythm. Third, a driving term is added that (1) causes each oscillator to synchronize its phase with an incoming musical signal, and (2) adds energy to those oscillators whose activity correlates well with the rhythmic input signal. Finally, a noise term is added to model variability in human behavior. This results in a relatively compact description that gives rise to metrical patterns. In the next section I present some results from stimulation with realistic musical input signals.

3.1. Mathematical model

3.1.1. Hopf oscillators

The first order of business is to describe how to model a stable internal pulse, or beat. As described above, the simplest mathematical model for this job is the normal form for the Hopf bifurcation (Arrowsmith & Place, 1990), which can be written as a differential equation in the complex variable $z(t)$.

$$\dot{z} = \alpha z + \mathrm{i}\omega z - z|z|^2. \tag{1}$$

Here the dot operator denotes the differentiation with respect to time (e.g. $\dot{z} = \mathrm{d}z/\mathrm{d}t$), α is an energy parameter, and ω is the eigenfrequency of the oscillator. [3] The behavior of this system is best understood by transforming to

---

[3] Where possible, I will refer to oscillator periods rather than eigenfrequencies. The relationship of period to eigenfrequency is $p = 2\pi/\omega$. The term frequency will be reserved to refer to frequencies within the audible range, i.e. those frequencies which contribute to the perception of pitch. The term period will be used to refer to the rate of slower periodic events, i.e. those periodicities that correspond to the perception of musical beat.

polar coordinates using the identity $z(t) = r(t)\exp \mathrm{i}\phi(t)$, and separating real and imaginary parts. The transformation yields a system of two differential equations, describing the time evolution of amplitude and phase, respectively.

$$\dot{r} = r(\alpha - r^2),$$

$$\dot{\phi} = \omega.$$

This transformation reveals that the amplitude and phase of the Hopf oscillator are independent and phase is free-running. Thus, by setting $\dot{r}$ to zero, one can find the stable states of the system. Such an analysis reveals two behaviors; the value of the energy parameter determines which behavior is observed, as shown in the Fig. 2(A). When α is less than zero (energy is being removed from the system), the system has a stable fixed point with $r = 0$, and the system behaves as a damped oscillator. For parameter values (α) greater than zero (energy is being added into the system), a stable limit cycle develops with amplitude $r = \pm\sqrt{\alpha}$ and the system generates a sustained oscillation. [4] Sustained oscillation is the basic model of the internal pulse, or musical beat.

3.1.2. Network interaction

The perception of meter can be modeled by constructing a network in which every oscillator has a different period, and the periods span the temporal range corresponding to beat perception (see Section 4.1). How should the oscillators in such a network interact with one another? One simple approach is to assume that they must compete for activation through mutual inhibition. Those oscillations that are the most consonant with the input (i.e. have higher α values, described below) should tend to deactivate those that do a poorer job of correlating with the incoming rhythm. Thus, in response to a rhythmic input signal, only a few oscillators should remain active, those that best reflect the temporal structure of the rhythm.

Such behavior can be captured with the addition of an interaction term to Eq. (1):

$$\dot{z}_n = \alpha_n z_n + \mathrm{i}\omega_n z_n - z_n|z_n|^2 - \sum_{m \neq n} \gamma_{mn} z_n |z_m|^2, \qquad (2)$$

---

[4] The fact that r can be positive or negative at the amplitude/phase level is an artifact of the transformation to polar coordinates. Negative and positive values of r can be considered equivalent, because the amplitude of the oscillation in the underlying system is the same in either case.

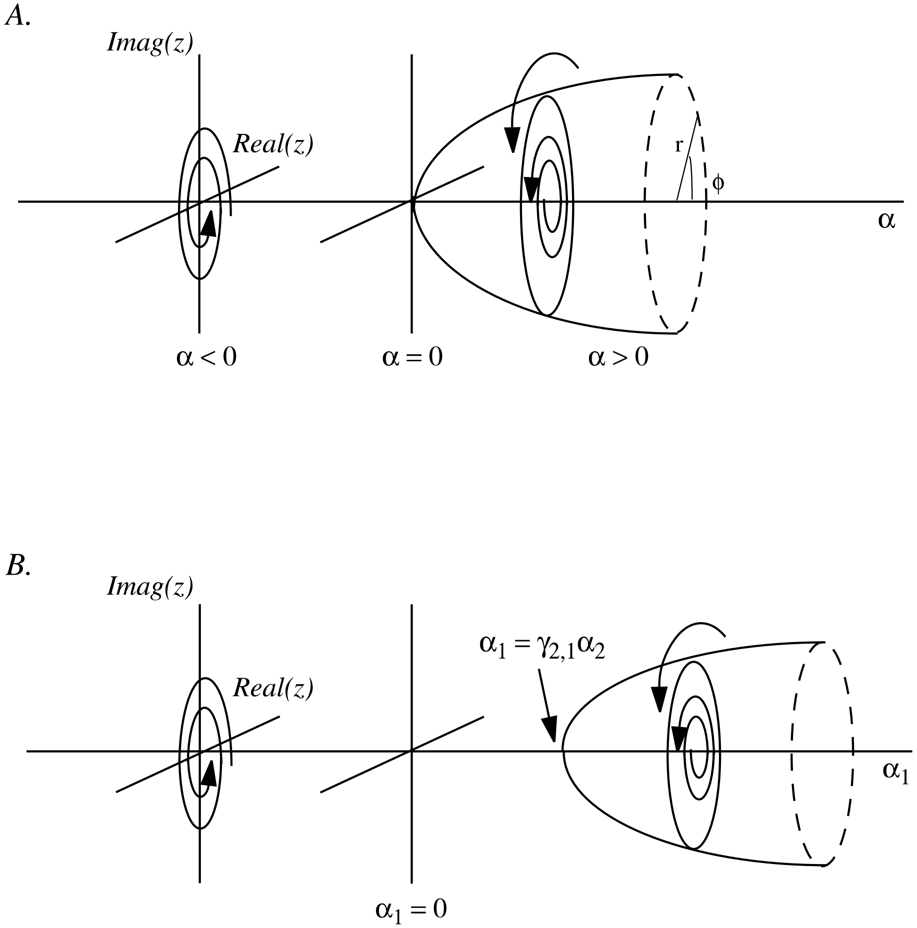

Fig. 2. Schematic overview of a Hopf oscillator. (A) As the energy parameter, α crosses zero a self-sustained oscillation arises. Zero is the bifurcation point. (B) Within the context of a 2 oscillator network (Eq. (2)) the bifurcation point moves to a point that is determined relative to the other oscillator's amplitude.

where γ is an $N \times N$ matrix of inhibition parameters, $\gamma_{mn} \geqslant 0$, and N is the number of oscillators in the network. As above, transformation to polar coordinates is informative.

$$\dot{r}_n = r_n(\alpha_n - r_n^2) - \sum_{m \neq n} \gamma_{mn} r_n r_m^2,$$

$$\dot{\phi}_n = \omega_n.$$

The transformation shows that in Eq. (2) amplitude and phase are independent and phase is free running, as in Eq. (1). Thus, inhibition works on the amplitude dimension only and network interaction can be understood simply, by restricting further the analysis to the amplitude dimension. A local stability analysis (see e.g. Perko, 1991) was performed on this system by first finding the fixed points, or equilibria, of this system, and then linearizing about each fixed point to determine the conditions (relative parameter values) under which each fixed point is stable.

Table 1 summarizes the stability analysis for the case of a two-oscillator network, revealing several interesting features of the system. First, as with individual oscillators (State 1), both oscillators are deactivated when both energy parameters are less than zero. Next (States 2 and 3), a single oscillator is activated when the product of its energy parameter and its inhibition parameter exceeds the energy parameter of the second oscillator. In this case, the limit cycle of the active oscillator has amplitude $\sqrt{\alpha_n}$. Finally, when both energy parameters are high enough (State 4), both oscillators are activated, each with slightly lower amplitude than it would achieve if activated alone (illustrated in Fig. 2(B)). A simple rule of thumb for this system is that when one oscillator is active, the second oscillator's energy must exceed a certain proportion of the active oscillator's energy in order for it to activate; the required proportion is given by the strength of the inhibition parameter. Thus, the inhibition terms in the γ matrix specify relative activation thresholds.

There are two additional points to note about the interaction between oscillators. First, the stability criteria of Table 1 are not all mutually exclusive. It is possible that States 2 and 3 can be stable simultaneously. In this bistable situation either oscillator can be active depending on the past

Table 1
Fixed point and stability/existence criteria for amplitudes in a two-oscillator network described by Eq. (2)

State	Fixed point	Stability/existence
1	$r_1 = 0,\ r_2 = 0$	$\alpha_1 < 0,\ \alpha_2 < 0$
2	$r_1 = \pm\sqrt{\alpha_1},\ r_2 = 0$	$\alpha_1\gamma_{2,1} > \alpha_2$
3	$r_1 = 0,\ r_2 = \pm\sqrt{\alpha_2}$	$\alpha_2\gamma_{1,2} > \alpha_1$
4	$r_1 = \pm\sqrt{\dfrac{\alpha_1 - \alpha_2\gamma_{1,2}}{1 - \gamma_{1,2}\gamma_{2,1}}}$ $r_2 = \pm\sqrt{\dfrac{\alpha_2 - \alpha_1\gamma_{1,2}}{1 - \gamma_{1,2}\gamma_{2,1}}}$	$\alpha_1 > \alpha_2\gamma_{1,2},\ \alpha_2 > \alpha_1\gamma_{2,1}$

history of the system; this phenomenon is known as hysteresis. The implication for beat perception is that once an oscillator becomes active, it remains active unless (1) it is actively inhibited by a competing oscillator, or (2) its energy parameter drops below zero. Second, the analysis of the two oscillator system generalizes to any number of oscillator as follows. For an N-oscillator system, 2^N patterns of oscillator activation can be stabilized: Any combination of units can be activated given the right parameters (see Goldenstein, Large & Metaxis, 1999). When one oscillator is active $r_n = \sqrt{\alpha_n}$, and amplitudes degrade gracefully as more oscillators are activated.

3.1.3. External input

In this third section, I consider how an afferent stimulus (the musical rhythm) drives the oscillator network. To do this, the energy parameter α is replaced with a multiplicative term involving the input signal $s(t)$,

$$\dot{z}_n = \eta_n s(t) V_n(z_n)|z_n| + \mathrm{i}\omega_n z_n - z_n|z_n|^2 - \sum_{m \neq n} \gamma_{mn} z_n |z_m|^2, \tag{3}$$

where η_n is a coupling strength that determines magnitude of the input signal's effect on each oscillator. $V_n(z_n) = \exp(\kappa_n(\mathrm{Re}(z_n)/|z_n|))/(\exp\kappa_n)$ is a transfer function (Fig. 3(A)) that mediates the effect of input to each oscillator as described by Large and Kolen (1994) (see below), where κ_n is fixed for each oscillator based on period. [5] In terms of amplitude and phase,

$$\dot{r}_n = r_n(\eta_n s(t) V_n(\phi_n) \cos\phi_n - r_n^2) - \sum_{m \neq n} \gamma_{mn} r_n r_m^2, \tag{3a}$$

$$\dot{\phi}_n = \omega_n - \eta_n s(t) V_n(\phi_n) \sin\phi_n. \tag{3b}$$

Thus, the external stimulus, $s(t)$ affects both the amplitude and the phase of each oscillator in the network. In the amplitude dimension, the energy pa-

---

[5] In previous work (Large & Jones, 1999), the concentration parameter κ was allowed to adapt. Here, for simplicity is fixed for each oscillator increasing monotonically with oscillator period.

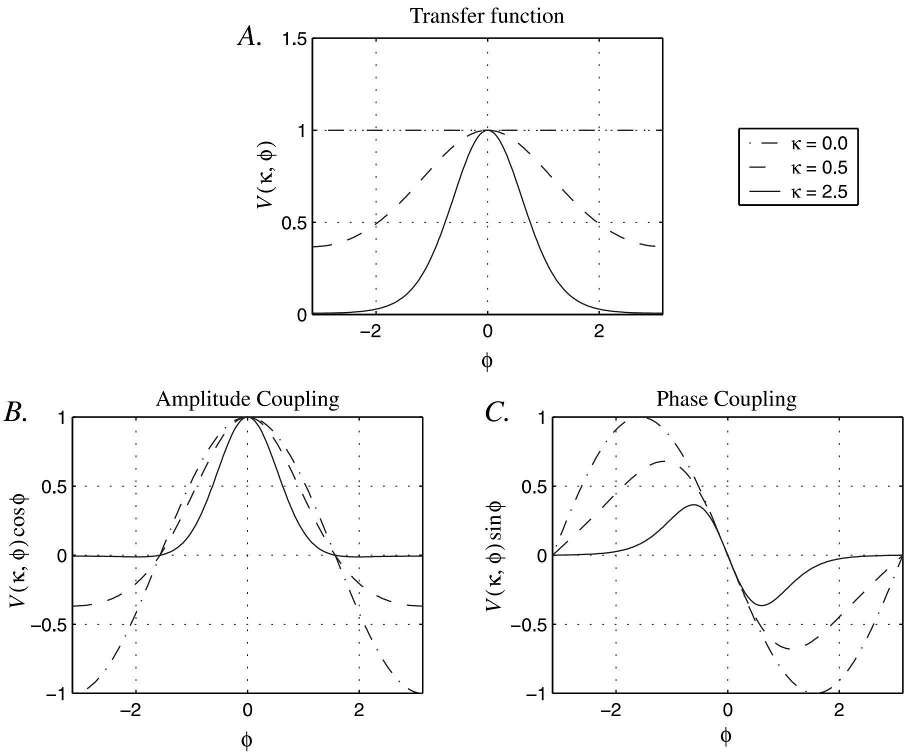

Fig. 3. (A) Transfer function, (B) amplitude coupling function, (C) phase coupling function.

rameter is determined by the product of the input stimulus with a pulse function ($V_n(\phi_n)\cos\phi_n$; Fig. 3(B)) that embodies the oscillator's expectations about when events should occur. Intuitively, oscillator energy is determined by a non-linear correlation of the signal with the oscillator's expectations. The better the oscillator does at predicting the input signal, the higher its amplitude becomes. In the phase dimension, the effect of the coupling term is to cause the oscillator to entrain to the input stimulus (Large & Kolen, 1994; see Fig. 3(C)). When an input pulse is early ($\phi < 0$) or late ($\phi > 0$) phase adapts, adjusting temporal expectations to better predict the input signal.

In both the amplitude and phase dimensions, the transfer function works as a "temporal receptive field" (Large & Kolen, 1994), a temporal gate that limits the sensitive phase of the oscillator. Input affects the oscillator differently depending upon the phase at which it is received.[6] This is depicted in Fig. 3, where the transfer function is shown alone (Panel A), and as it mediates the effect of input in the amplitude dimension (Panel B) and the phase dimension (Panel C). For $\kappa = 0$ all input events affect the oscillator, for $\kappa = 2.5$ the oscillator concentrates its response on those events that coincide with its temporal expectations, ignoring events that occur away from phase zero. This provides an important form of temporal stability in the phase dimension. In a numerical investigation of a closely related system, Large and Kolen (1994) demonstrated that the resonances in such a system exhibit a generalized Arnol'd Tongue structure: The oscillator can stably entrain to input whose period forms a simple ratio (e.g. 1:1, 2:1, 3:1, 3:2) with its own period; the simpler the period ratio, the more stable the entrainment. This is critical, because complex rhythmic inputs contain many different periodicities, thus the oscillator entrains appropriately when the input contains any periodicity that is consistent with its predictions.

3.1.4. Variability

Finally, to simulate variability in human behavior, noise is added to the system.

$$\dot{z}_n = \eta_n s(t) V_n(z_n)|z_n| + \mathrm{i}\omega_n z_n - z_n|z_n|^2 - \sum_{m\neq n} \gamma_{mn} z_n |z_m|^2 + \sqrt{Q}\zeta_n(t). \tag{4}$$

Here $\zeta_n(t)$ denotes a Gaussian white noise source with zero mean and unit variance, and Q is a variance parameter, used to set the noise level. Transformation to polar coordinates reveals

$$\dot{r}_n = r_n(\eta_n s(t) V_n(\phi_n)\cos\phi_n - r_n^2) - \sum_{m\neq n} \gamma_{mn} r_n r_m^2 + \sqrt{Q}\zeta_n(t)\cos\phi_n, \tag{4a}$$

$$\dot{\phi}_n = \omega_n + (-\eta_n s(t) V_n(\phi_n)\sin\phi_n) - (\sqrt{Q}\zeta_n(t)/r_n)\sin\phi_n, \tag{4b}$$

thus, the additional term contributes noise in both amplitude and phase dimensions.

---

[6] In polar coordinates it becomes evident that the transfer function is an amplitude-normalized pulse derived from a von Mises distribution over ϕ with concentration parameter $\kappa_n : V_n(\phi_n) = (\exp(\kappa_n \cos\phi_n))/(\exp\kappa_n)$. Large and Jones (1999) used the von Mises distribution to model performance in a time discrimination task, conceiving of it as the concentration of attentional energy about expected time points generated by an attentional oscillation. This transfer function could also be thought of as modeling the proportion of cells in an underlying neural population that are spiking at any given point during one cycle of the oscillation (see e.g. Recanzone, Merzenich & Schreiner (1992), with implications for a tactile temporal discrimination task).

In summary, the behavioral model consists of a network of Hopf oscillators, driven by a musical stimulus. Under rhythmic stimulation, this pattern-forming dynamical system undergoes bifurcations that correspond to the birth of self-sustained oscillations and the formation of temporally structured patterns of oscillations, modeling the perception of temporal structure. In principle such metrical pattern formation could take place in the absence of overt physical movement; however, synchronization of movement to music depends upon, and perhaps interacts with, the formation of such patterns.

The model is intended to be general enough to capture meter perception for a wide range of input signals; the goal of the remainder of this paper is to test its basic assumptions. Thus, model parameters are chosen based on general considerations, and are not fit to data in the sense of being explicitly adjusted to precisely reproduce observed results (with one exception, see below). In other words, an attempt is made to test the theoretical assumptions without overfitting parameters that likely could be adjusted to reproduce certain specific patterns of results quite accurately. The remainder of this paper discusses such a preliminary test of this model, comparing its meter induction performance with the results of a recent study of synchronization to music (Snyder & Krumhansl, 2000). The study was designed to assess mental pulse-finding, and several of its measures are especially appropriate for gauging model performance. However, because the model is not one of synchronized tapping per se – among other things, it does not explicitly include a model of movement (cf. Wing & Kristofferson, 1973) – some comparisons will have to be interpreted with caution. Furthermore, whereas the model captures the perception of metrical structure, the experiments measured tapping at only a single metrical level. Thus, certain aspects of model performance cannot be directly assessed using these data. Nevertheless, tapping is a direct way to measure many key variables that relate directly to meter perception, and for these, Snyder and Krumhansl have proposed benchmark measures for model performance.

4. Methods

4.1. Network parameters

There are three parameters that determine the behavior of the oscillator network, namely the vector of eigenfrequencies ω, the matrix of inhibition strengths γ, and the vector of input coupling strengths η. The range of oscillator periods (i.e. eigenfrequencies), determines the tempi at which beats can be perceived. The matrix of inhibition strengths will determine what oscillators can be activated simultaneously, and thus constrain the metrical patterns can be represented within the network. Finally, the input coupling strengths determine how strongly the input drives each oscillator.

A number of estimates have been made of the effective range of rhythm perception (see e.g. Fraisse, 1982). For the simulation reported below, the periods of the oscillators were chosen to span the range from approximately 100 to 1500 ms, roughly corresponding to the range over which people accurately perceive changes in the tempo of auditory sequences (Drake & Botte, 1993). Within this range, oscillator periods were even spaced on a logarithmic scale, i.e. according to an equal-tempered strategy. The complete network consisted of four octaves of oscillators, with a largest period of 1500 ms, and 24 oscillators per octave This provided a tempo resolution (Weber fraction) of 2.93% roughly comparable to the best tempo resolution found by Drake and Botte (1993).

A variety of studies have demonstrated a preference for small integer ratios in the perception and production of complex rhythms (see e.g. Collier & Wright, 1995; Fraisse, 1982; Jones & Yee, 1997; Povel & Essens, 1985), with a preference for ratios of 2:1 and sometimes 3:1. To take these findings into account, an interaction matrix was designed so that oscillators whose periods possess 2:1 and 3:1 integer ratios (and multiples of these) compete only weakly with one another, while those at intervening ratios compete strongly. Fig. 4(A) shows the competition kernel: one row of the interaction matrix, describing the inhibition produced by the oscillator with period p on the other oscillators in the network. The interaction kernel is symmetric about each oscillator's period, minima occur at harmonic and subharmonic multiples of 2 and 3, and competition decreases as with the square of the period ratio.

Finally, coupling strength, η, determines the amount of energy added to the oscillator, and the amount of phase adaptation that takes place, in response to each input impulse. This has two important effects on network performance. First, it governs the overall speed with which the network will settle into a stable metrical interpretation, called *relaxation time.* Second, the network definition (see Eq. (3)) admits a different coupling strength for each oscillator, allowing independent control (i.e. separate from the interaction matrix, γ) over the nature of the patterns that arise in the network. This

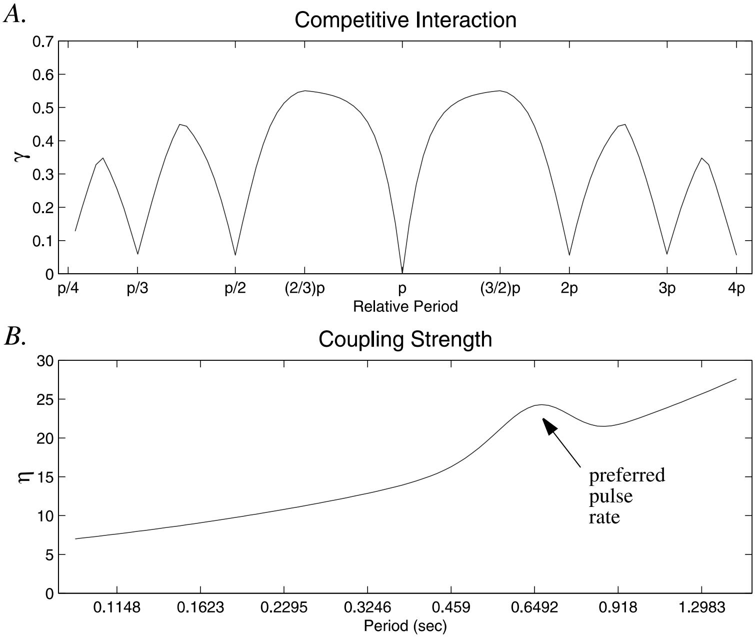

Fig. 4. (A) The interaction kernel used in the simulations reported in this paper. It is symmetric about each oscillator's period, minima occur at harmonic and subharmonic multiples of 2 and 3, and competition decreases as with the square of the period ratio. (B) Coupling strength, as a function of oscillator period.

makes it possible to bias the relative amplitudes that active oscillators can achieve, an ability that has both practical and theoretical significance. As a practical matter, slower oscillators require higher coupling strengths, since they naturally respond to input more slowly. Theoretically this is important because it has been reported that people show preferred tapping rates (see e.g. Fraisse, 1982). Most studies indicate preferred rates centered around 600 ms, but perceptual studies indicate that preferred tempo also adapts to reflect the mean tempo experienced within an experimental session (Jones & McAuley, 2000; Woodrow, 1951). The coupling strengths shown in Fig. 4(B) reflect these three considerations. For base coupling strength, which determined overall network relaxation time, a value of 7 was chosen to provide a reasonable relaxation time, after several preliminary runs of the model on musical input. Second, coupling strength increased with the square of the ratio of oscillator period to smallest period in the network. Finally, a small bias was added, so that the network would exhibit a preferred pulse rate. For simplicity, the preferred rate was chosen to be equal to the mean IBI for the Snyder and Krumhansl stimuli, 638 ms; it was not allowed to adapt. [7]

Although the number of individual model parameters is potentially large (N oscillator periods, N coupling strengths, N^2 inhibition strengths), in practice a much smaller number of free parameters determines network structure. Three parameters determine the range (minimum and maximum) and spacing of oscillator periods. Because relative coupling strengths are fixed, one parameter suffices to set the overall level of external driving force; coupling bias (for preferred pulse rate) is specified by three parameters (height, width, and position). Finally, inhibition parameters are all based on the same symmetric interaction kernel ($N - 1$ distinct values), but by fixing the functional form of the kernel, the number of parameters is reduced to one, specifying the overall level of competition, plus two specifying the locations of the minima of the interaction kernel. Thus, a total of 10 free parameters describes a general model of meter perception. In addition to these parameters, system noise was set at, $Q = 0.0025$. The parameters were chosen as described above and fixed for the subsequent simulations; they were not explicitly adjusted to maximize goodness of fit to the synchronization data described below.

In summary, the parameters of the network were selected based upon general considerations. Oscillator periods were chosen to span the reported range of human rhythm perception. Coupling strengths were biased so that the network would exhibit a preferred pulse rate comparable to the preferred tapping rate observed in humans, providing a connection with the experimental data which used tapping as an index of pulse-finding ability. Finally, an inhibition matrix was defined to favor temporal structures based upon harmonic and subharmonic multiples of 2 and 3, allowing the formation of metrical patterns appropriate to a broad range of musical rhythms. Only duple ratios occurred in Snyder and Krumhansl's stimuli; the inclusion of triple ratios makes the model more general and provides a test of the network's ability to distinguish between duple and triple structures quickly and reliably. With these parameters, the network is capable of representing a large number of possible metrical patterns, it is not specifically tailored to any particular task or data set. The metrical patterns that form in the network

---

[7] The meter perception model performs similarly if coupling bias is omitted, however in such a case the choice of tapping level (determined by the first metrical level to activate) will reflect the spectral structure of the input more faithfully. Furthermore, the location of the bias peak can be moved within the range of reported preferred tempi without strongly affecting the results reported below.

arise out of a combination of the spectral properties of the input and the internal constraints of the network. Thus, the simulations described below test general properties of this dynamic approach to meter perception.

4.2. Computational methods

All simulations were written in Matlab and run on a Sun Workstation (Solaris 4.4.1; Matlab 5.3). The differential equations describing the behavior of the oscillator network (Eqs. (4a) and (4b)), were solved numerically using Euler (fixed time constant) integration, with time constant equal to the sample rate of the input (approx. 10 ms). The network was initialized on each run with small random amplitudes and phases. The output of the simulation yielded a time course of amplitude and phase values, which were converted to pulse times for comparison with the data of Snyder and Krumhansl (2000) using a simple procedure. The amplitude of each oscillator was averaged over the length of the performance, and any oscillator whose activation exceeded an adaptive threshold (7.5% of total amplitude) was considered to have been active during the performance. Active oscillators were grouped, and the activation time of the group was defined as the first point in time when the amplitude of one of the oscillators crossed the activation threshold. The first group to become active was considered to be the tactus (tapping) level; tactus pulses were defined as the pulses of the strongest (highest amplitude) oscillator in the group. [8] Pulse times were recorded when the phase of each oscillator crossed zero. Typically 1–4 oscillator groups were activated in each excerpt, instantiating a metrical structure (see, e.g., Fig. 7(C), below).

4.3. The stimuli

In a recent study, Snyder and Krumhansl (2000) focused primarily on two issues surrounding the perception of beat and meter. The first was that of syncopation. Syncopation can be operationalized as the lack of events (or sometimes the occurrence of unstressed events) on strong beats, accompanied by the placement of stressed events on weak beats. Referring back to Fig. 1, the right-hand part (RH) of measures 5 and 6 are not syncopated because events occur only on strong beats. Examples of syncopation occur in the right

---

[8] The simple strategy was viable because the tempo of the excerpt did not change. In more complex situations, a more sophisticated strategy for reporting beats would be required.

hand part of measures 7 and 8, however. An unstressed event occurs on the strongest beat of measure 7, followed by a stressed event on a very weak beat (the second sixteenth note) and no event on the next, stronger beat. Also, the last event of measure 7 is stressed, with no event on the following downbeat – the strongest beat of measure 8. Syncopation amounts to a violation of the temporal expectations embodied in the perceived metrical structure. However, if structural expectations arise from a process that attempts to synchronize strong beats with points of musical stress, then strongly syncopated music could confuse this process. For example, in the RH part of this excerpt, stressed events (the chords) initially occur at a rate of one per beat, but in measure 7, due to the syncopation, the rate increases to one per three sixteenth notes (3/4 beat). In the full excerpt this rhythmic pattern is disambiguated by the characteristic duple pattern in the LH part. However, listening to the RH part in isolation, one might hear this (correctly) as a syncopation against the original meter, or (incorrectly) as a change from duple to triple meter. Thus, understanding the perception of syncopation is critical to understanding the dynamics of meter perception.

The second key issue involves the role of non-temporal factors, such as bass lines, melody lines, and harmony in influencing beat perception. For example, in ragtime piano the left-hand part usually consists of bass notes (played in octaves) on the down-beats, and chords played using higher pitches on the up-beats. In the current example, only measure 7 strictly follows this pattern, but according to Snyder and Krumhansl's analysis, their stimuli showed this pattern with relative consistency. This type of cue could be exploited by subjects in resolving the issue of where strong beats are located. Thus, listeners might adopt a strategy of aligning strong beats with bass notes, lessening any potential confusion that arises from the use of syncopation in this music. Other types of cues, such as the harmonic changes, and melodic patterns could also be exploited in determining the perception of metrical structure. Therefore, understanding whether and how pitch information influences meter perception is crucial to theoretical modeling of this process.

In Snyder and Krumhansl's study, the musical excerpts were generated by computer so that the timing of the individual events was metronomically precise (to within ±1 ms). The pieces were played to participants via MIDI [9]

---

[9] Musical Instrument Digital Interface, a communications protocol for controlling electronic musical instruments.

using a digitally sampled piano timbre. To control different types of pulse-finding cues that were available, four versions of each piece were generated. First, full versions (both LH and RH parts) and right-hand only (RH) versions were created. The elimination of the left-hand part creates a version of the music that is more syncopated, and potentially more confusing, than the full version because fewer events occurred on the down-beats. Next, based upon each of these two versions, two additional stimuli were created from which non-temporal cues were eliminated. To create the monotonic versions, all pitches were changed to middle C, and where more than one note was sounded simultaneously, additional notes were discarded. Fig. 5 shows the four MIDI versions for measures 5–8 of Lily Queen using piano-roll notation. Here, pitch is displayed on the vertical axis, using MIDI note number (MIDI note 60 corresponds to middle C, $f_0 = 262$ Hz), and time is displayed on the horizontal axis. Here, the temporal and non-temporal features of the stimulus become apparent. For example, time is represented more explicitly, so it is clear that note durations in the musical score correspond to inter-onset intervals (IOIs) rather that onset-to-offset durations.

Next, the MIDI stimuli was played on a Kawai 950 digital piano, and the acoustic signal was passed through a simple auditory model. The auditory model approximated processing in the earliest stages of the auditory system: frequency filtering by the cochlea, mechanical to neural transduction at the level of the primary afferents, and onset responses, which can be observed as early as the cochlear nucleus (see Langner, 1992, for a review). This may be thought of as the minimal neural signal that is available to the beat induction mechanism, making several channels of event onset information available as input to the model of beat induction. This preprocessing step was similar to that used in previous models of rhythm perception (Scheirer, 1998; Todd, 1994).

First, the acoustic signal was digitally recorded on a Macintosh G3 computer at a sample rate of 1,1025 Hz, with a 16 bit sample resolution. Then the digitally recorded signals were passed through the auditory model. A Patterson–Holdsworth auditory filter bank was used (Matlab implementation by Slaney, 1998) to separate the data into 20 frequency channels, spanning the range from 32 Hz to one-half the sample frequency (approximately the frequency range of the piano keyboard). The data in each frequency channel was then full-wave rectified. Each of the 20 frequency channels was decimated by a factor of 110 to a sample rate of approximately 100 Hz ($1,1025/110 = 100.23$ Hz) to speed further calculation. Next, an onset detection procedure was applied to each channel of data. An amplitude

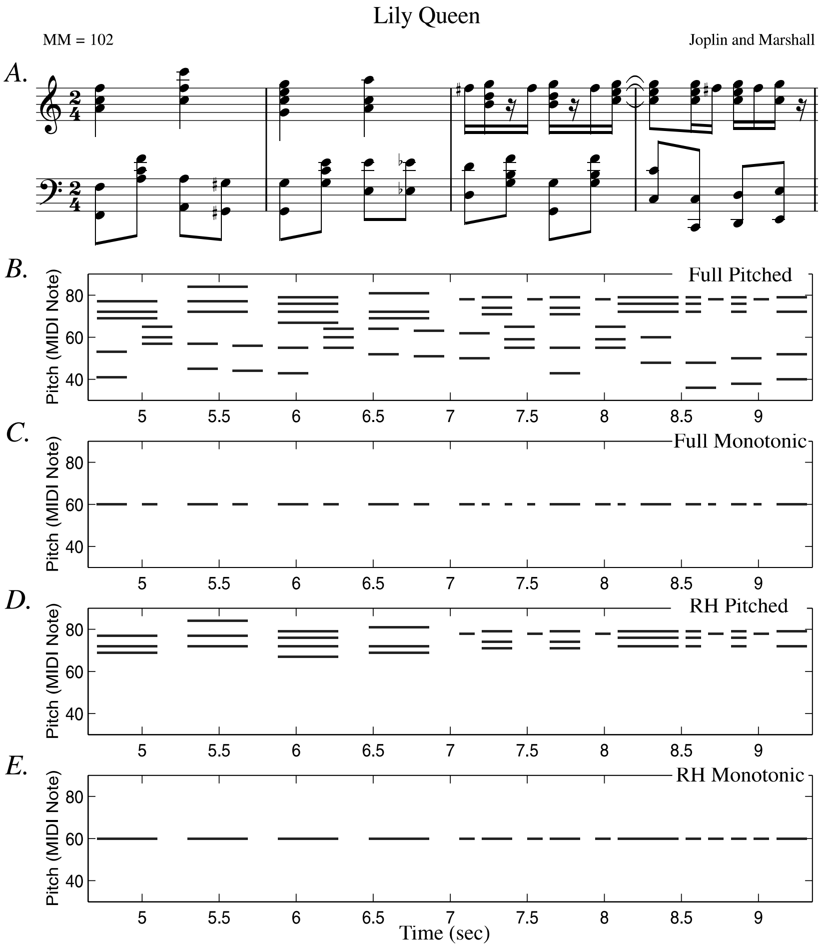

Fig. 5. The four experimental manipulations: (A) the original score, (B) the full pitched version, (C) the full monotonic version, (D) the right-hand version, (E) the right-hand monotonic version.

envelope and its first derivative were calculated in each frequency channel using Gaussian FIR filters. Finally, the first derivative signals were half-wave rectified and amplitude-normalized according to overall signal energy, yielding a primitive representation of event onsets. The result of this process is shown in Fig. 6 for each of the four stimulus conditions.

To summarize, Snyder and Krumhansl's stimuli were constructed to control cues to beat induction in complex musical signals. They controlled

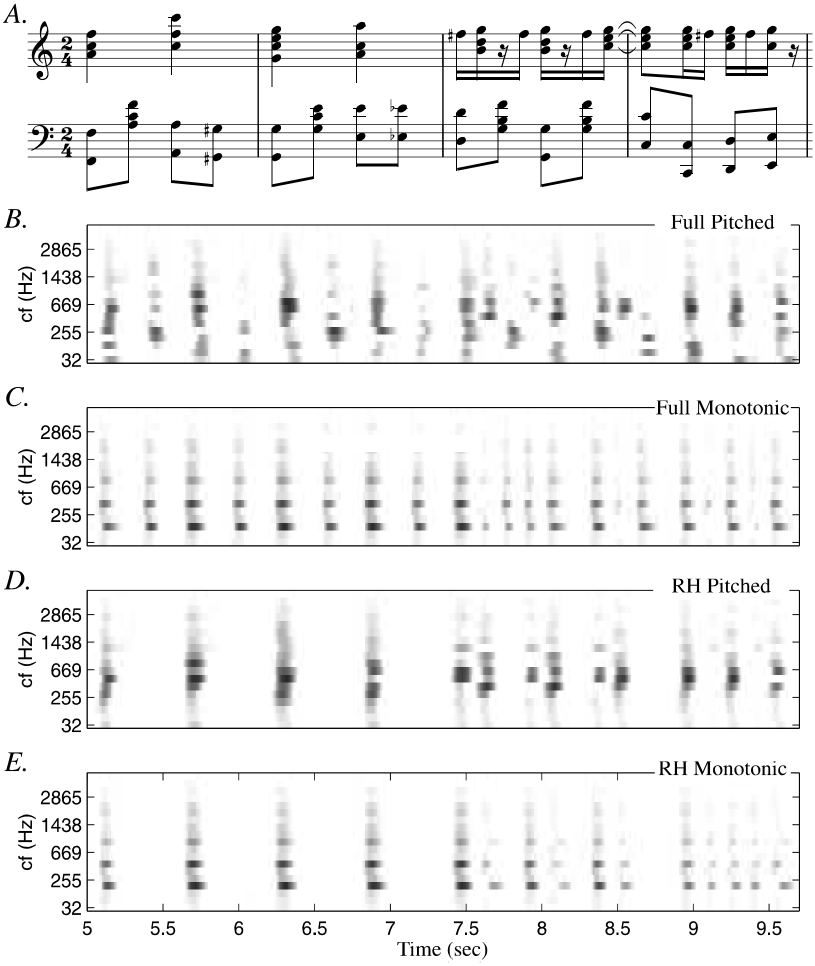

Fig. 6. Input to the oscillator network, produced by a simple auditory model that simulated cochlear filtering, mechanical to neural transduction, and onset detection. (A) The original score, and the auditory input for (B), the full pitched version; (B) the full monotonic version; (C) the right-hand version; (D) the right-hand monotonic version.

level of syncopation and non-temporal information within a realistic musical context. The input to the model consisted of digital audio recordings of the ragtime piano excerpts. The data were preprocessed using a simple auditory model to simulate neurally plausible input signals that may be available to human beat induction. In the next section I test the performance of the model using Snyder and Krumhansl's stimuli, and compare the model's output to the performance of participants in their synchronization task.

5. Results and discussion

Eight pieces used in the Snyder and Krumhansl experiments served as input to the model. These pieces were presented in four versions each – *full pitched, full monotonic, RH pitched*, and RH *monotonic* – yielding 32 separate stimuli. As described above, the different versions controlled amount of syncopation and pitch information available in the input. One of the more interesting experimental findings was that synchronization tapping performance was largely independent of pitch information for these stimuli. To investigate this finding, two different versions of the model were run, Simulation 1, which did not use frequency information, and Simulation 2, which made limited use of frequency information.

In Simulation 1 each oscillator received equal input from all auditory frequency channels. This was accomplished by summing onsets over all frequency bands, and using the summed input signal to drive all oscillators. Thus, the only effect of auditory filtering was to ensure that onsets in high-frequency bands were not masked by energy in the lower part of the spectrum. In general all onsets were identifiable in this input signal, however high-frequency onsets tended to have lower amplitude.

In Simulation 2 each oscillator received a slightly different input signal, depending upon the frequency channel in which each onset arrived. This was accomplished by taking a weighted sum of the onset signals, where the weighting differed depending on the period of the oscillator. Slower oscillators (larger periods) received a greater proportion of their input from lower frequency channels, while faster oscillators received more input from the higher frequency bands. The strategy implemented a simple (perhaps learned) heuristic (cf. Lerdahl & Jackendoff, 1983), that in Western music bass onsets tend to mark strong metrical locations. The effect of this manipulation was that for the slower oscillators, onset rates tended to be slower; often onsets from the higher registers did not appear at all in these input signals. Conversely, for faster oscillators, onsets tended to come at a faster pace; these input signals combined information from events that occurred in the higher registers with the onsets of harmonics corresponding to events in the lower registers. In brief, Simulation 1 did not take advantage of frequency information, while Simulation 2 made limited, heuristic use of the available frequency information.

5.1. Initial observations

Fig. 7 shows the result of one run of the model for the initial portion of Lily Queen. Panel A shows the piano roll notation for the first nine measures and Panel B shows a trace of oscillator activity ($\mathrm{Re}(z) = r\cos\phi$) for the three most highly activated oscillators in the network (Simulation 1), corresponding to the eighth-note (0.398 seconds), quarter-note (0.595 seconds), and measure (1.191 seconds) metrical levels. The metrical structure is shown as pulses (**X**'s) between Panels A and B, derived from the phase of the three oscillators shown in Panel B (described in Section 4). The tactus level pulses, which will be compared with participant tap times below, are the pulses of the

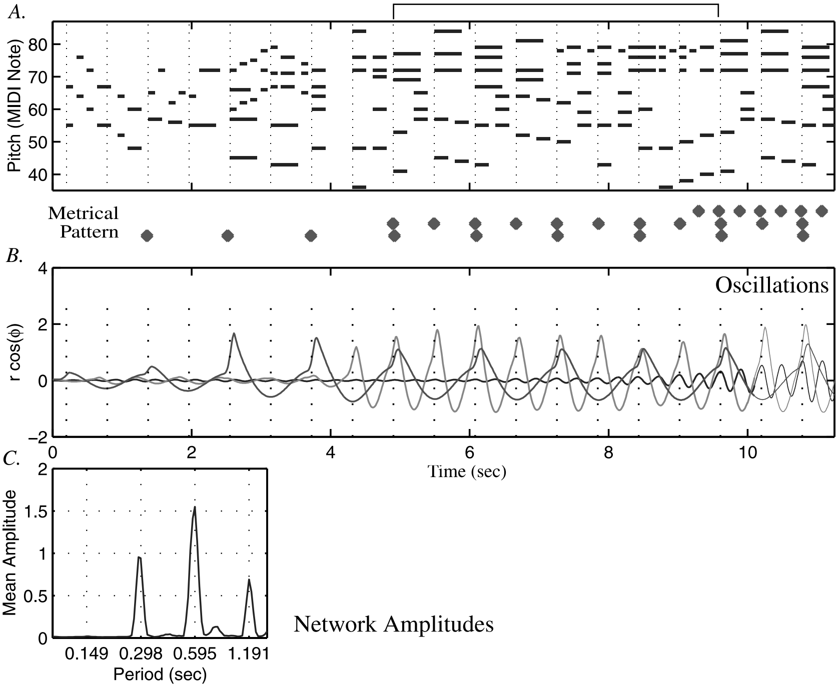

Fig. 7. The results of one run of Simulation 1 on the first nine bars of Lily Queen. (A) Piano roll notation of the stimulus. The bracketed portion corresponds to the fours notated in Fig. 1. (B) The oscillations of the three most highly active oscillators. (C) Mean amplitude for each oscillator in the network over the length of the performance.

first oscillator to activate. Here the measure level is identified as the tactus, however, more often the model identified the quarter note level (at 600 ms) as the tactus level (see below).

Several features of model performance are immediately apparent in the figure. First, the oscillators activate early, entrain to the beat of the music (marked by vertical grid lines on Panels A and B), and quickly attain stable amplitudes. Panel C summarizes network performance (average amplitude) for the entire excerpt. Three main peaks in network amplitude are evident, corresponding to the three oscillations shown in Panel B. An additional, a smaller peak is evidence of an oscillator group that was active at the beginning of the piece, but lost the competition within the first few seconds. This is one of many different metrical patterns that can be stabilized within this network. The spectral properties of the afferent rhythm in conjunction with the inhibitory network interactions gave rise to this three-leveled, duple metrical structure.

Next, each model was run twice on each stimulus excerpt. For most stimulus examples, the model performed comparably between the two (stochastic) runs of the same model, however, for a small number of pieces the results differed qualitatively between the two runs. To determine how effectively the networks represented the metrical structure of the experimental stimuli, a histogram was calculated, showing the period of the highest amplitude oscillator in each active group relative to the IBI of the corresponding stimulus pattern (p/IBI). These results are shown in Fig. 8, for the two runs

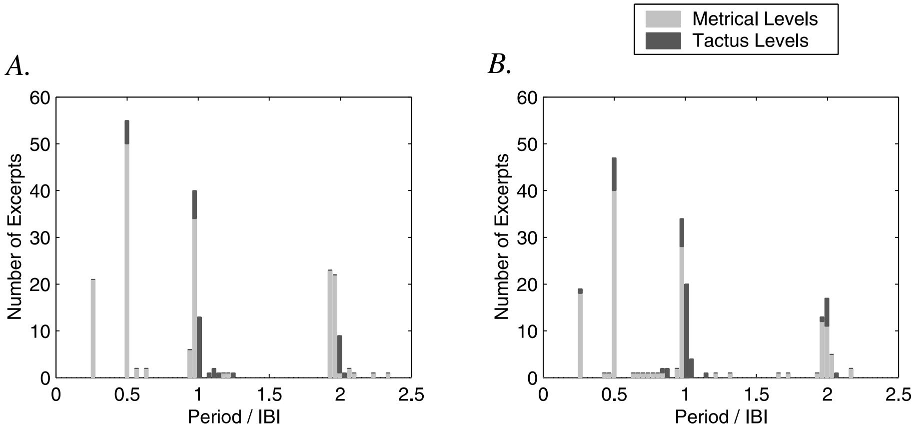

Fig. 8. Metrical structure histograms for (A) Simulation 1, and (B) Simulation 2.

of Simulation 1 (Panel A) and the two runs of Simulation 2 (Panel B). Each histogram represents performance on 64 excerpts (two runs on each of the 32 stimulus excerpts). All activated metrical levels are shown. Several interesting features of network performance are evident in the histograms. First, there are four distinct peaks corresponding, from left to right, with sixteenth note, eighth-note, quarter-note, and measure levels of the metrical structures. Both networks correctly identified eighth and quarter note metrical levels on almost every stimulus excerpt. Next most often the networks activated oscillators at the measure level, and on slightly less than half of the examples, the networks identified sixteenth note levels. This observation is not directly comparable with any experimental performance measure, since participants tapped only one level of beats (the tactus).

The histograms also illustrate which metrical levels were identified as the tactus, shown as darker bars. Most often the quarter note level ($p/\mathrm{IBI} = 1$) was chosen, however sometimes other metrical levels were identified as tactus (see Table 2). This finding is compatible with the performance of participants in the Snyder and Krumhansl study, who also tapped most often at the quarter-note level, but sometimes also tapped at the eighth note level or the measure levels. Participants in both experiments tapped at the quarter note level more often than the model (also shown in Table 2). The histogram also reveals several outliers from the main metrical peaks. In these cases, either the networks found the wrong metrical structure, or sometimes oscillators in addition to those that instantiated the correct metrical structure were active for a significant proportion of the performance. In some cases (eight excerpts in Simulation 1, nine excerpts in Simulation 2) the tactus period differed significantly (by at least 4%) from the true IBI or related metrical period. In these cases relative phase often exhibited occasional phase wrapping, a phenomenon called relative coordination (von Holst, 1937); this was observed mainly in the RH versions.

At this level of analysis, the performance of the two simulations is similar, with both networks correctly identifying the metrical structure for most

Table 2
Metrical levels chosen as tactus (tapping level) by the two simulations compared with the two Snyder and Krumhansl (2000) experiments

Tapping level	Simulation 1 (%)	Simulation 2 (%)	Experiment 1 (%)	Experiment 2 (%)
Sixteenth	0	2	0	0
Eighth	8	11	9	3
Quarter	78	75	88	97
Half	14	12	3	0

excerpts. Both simulations also showed greater variability in choosing a tactus level than the musician participants. Yet there are several aspects of network performance that are not captured in the metrical level histograms. These histograms do not show relaxation time, and they do not show whether the oscillators were properly entrained at their respective metrical levels. In the next section, I will examine the phase, and several other aspects of the behavior of the tactus level oscillators, as network performance is compared in detail with the tapping performance of participants in Snyder and Krumhansl (2000, Experiment 2).

5.2. Comparison with tapping data

5.2.1. Dependent measures

To examine network performance in greater detail, I compared network performance with the tapping data of Snyder and Krumhansl (2000, Experiment 2). To do this, pulse times were derived for the tactus level oscillator, as described in Section 4. Next, canonical beats times were determined for every metrical level by aligning the events of the MIDI score with the peaks in the onset signal produced by the auditory model. There was some variance in onset peak times, and this was dealt with by aligning the MIDI signal to minimize the mean discrepancy. [10] This procedure also compensated for the auditory filter delays by effectively delaying the MIDI signal. Ideal beat times relative to the onset signal were then extrapolated from the MIDI score.

Several dependent measures of model performance were defined to match, as exactly as possible, the dependent measures used in the tapping study (Snyder, personal communication). First, four modes of pulse alignment were defined: down-beat, up-beat, neither, and aperiodic. *Down-beat* and *up-beat* mode were counted when two or more pulses in a row occurred within 100 ms of a down-beat or up-beat, respectively. Two or more consecutive pulses that corresponded to some other metrical position were counted as *neither. Aperiodic* pulses included single deviant pulses and pulses with an inconsistent or drifting phase relationship to the beat. The proportion of time spent in each mode was tallied. Changes from one mode to another were called switches. Switches did not include changes to or from aperiodic pulses. Interpulse interval (IPI) was calculated as the average time between tactus level

---

[10] MIDI note onsets were aligned with peaks in the onset signal after summing across frequency bands.

pulses, excluding IPIs 100 ms or greater than the stimulus IBI, for consistency with Snyder and Krumhansl's measure. [11]

The remaining measures were expressed as proportions of the IPI for each trial. Beats-to-start-tapping (BST) was the time from the first note of the piece to when the model began outputting pulses, divided by the IBI. The coefficient of variability (CV) was defined as the standard deviation (SD) of IPI, divided by IPI, excluding pulses greater than 100 ms from the IBI. The coefficient of delay (CDel) was the average time between a pulse and the closest down-beat note or up-beat note, divided by IPI, where positive values denote pulses that occur after the beat. The coefficient of deviation (CDev) was the average absolute value of time between a pulse and the closest down-beat or up-beat note divided by IPI. For both CDel and CDev, only pulses within 100 ms of a down-beat or an up-beat note were considered.

5.3. Model performance

BST, an index of relaxation time, is shown in Fig. 9 for the musicians (Panel A), Simulation 1 (Panel B), and Simulation 2 (Panel C). Musicians began tapping after three or four beats, and they tended to start tapping soonest for the full-pitched versions, although none of the observed differences were statistically significant. Simulation 2 is superior to Simulation 1, which begins tapping sooner than the musicians by about one beat. Simulation 2 does a fairly good job of capturing mean BST over the four stimulus versions, and the overall magnitude of differences between the means of the different stimulus conditions is comparable to the Experimental data. Thus, the model's relaxation time squares reasonably well with human performance, with Simulation 2 providing a better match to the data than Simulation 1.

Tapping mode, shown in Fig. 10, measures whether pulses occurred on the down-beat (in phase with the ideal beat) the up-beat (antiphase), neither (at a some other consistent phase) or aperiodic (any other relationship). Musicians tapped quite reliably in the down-beat mode for the full-pitched versions, but less often as the stimulus information degraded. They tapped more often on the down-beat for *pitched* than *monotonic* versions, and more for the *full* than

---

[11] Cases in which the model's tactus IPI differed by more than 4% from the IBI (or relevant metrical ratio) were eliminated from these calculations because the calculation of one or more dependent measures broke down, or because phase-wrapping caused an erroneously high number of mode switches to be counted. These totaled eight cases for Simulation 1 and nine cases for Simulation 2.

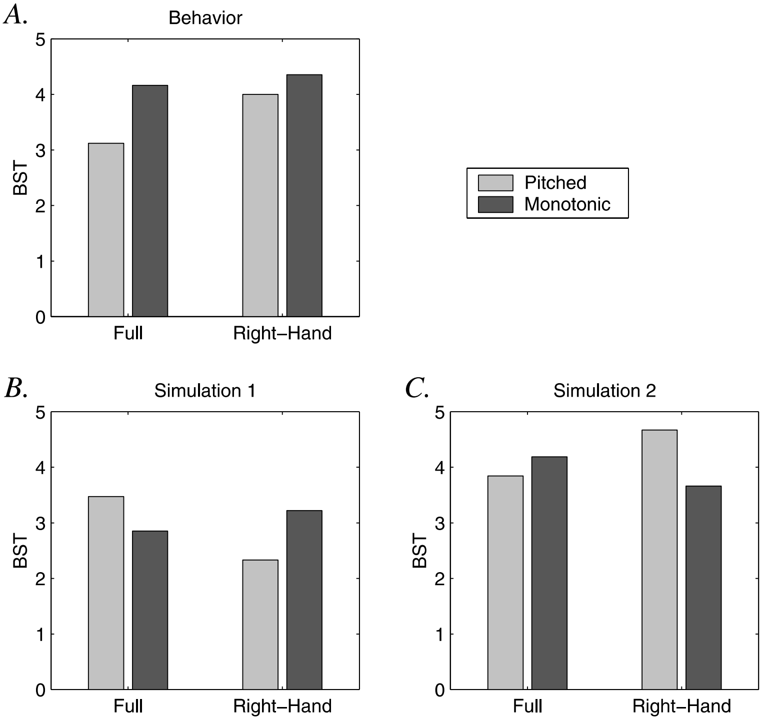

Fig. 9. BST for (A) musicians, (B) Simulation 1, and (C) Simulation 2.

RH versions. Tapping in the neither and aperiodic modes increased with syncopation (more for RH than full versions). Overall, syncopation disrupted performance more than the lack of pitch information. Both simulations match this general pattern of results relatively well. More pulses were produced in down-beat mode for the full-pitched versions, with performance deteriorating for the degraded versions. Pulses were produced more often on the down-beat for pitched than monotonic versions, and more for the full than RH versions; also, pulses were produced more often in neither mode for RH than full versions. The absolute proportion of down beat taps matched musicians' performance closely for Simulation 2, significantly outperforming Simulation 1. The performance of Simulation 2 deteriorated too much for the RH versions, however (less than 60% down-beat taps, greater than 20% neither taps). In addition, there were too many up-beat taps and too few aperiodic taps in the stimulus degraded versions. Overall however, Simulation 2 provided a good match to the tapping mode data, and performed more like musicians than Simulation 1.

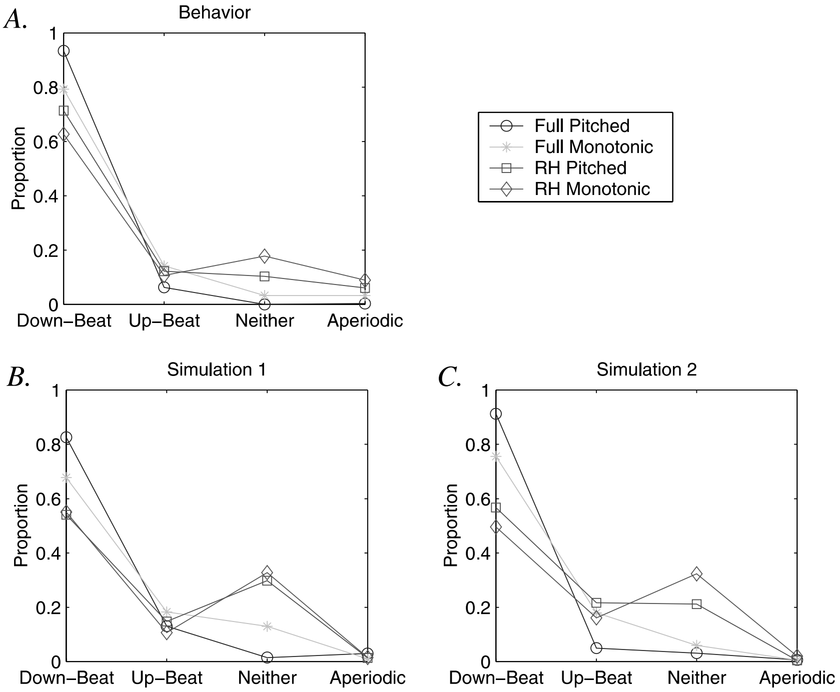

Fig. 10. Tapping mode for (A) musicians, (B) Simulation 1, and (C) Simulation 2.

Switches per excerpt (Fig. 11) served as a key index of phase stability, measuring the ability of the model to persist in its initial interpretation of strong beat locations. The number of switches between tapping modes was low for musicians (less than one in every condition) and stimulus degradation produced a systematic breakdown in phase stability. More switches were observed in RH than full versions, but the difference between pitched and monotonic versions was not significant. Again syncopation was more disruptive to synchronization tapping than lack of pitch information. Both simulations predict this qualitative pattern of results, producing more switches in RH than full versions. In the full versions, both produce relatively few switches overall, less than one per excerpt. However, neither simulation performs as stably as musicians in absolute terms, with both producing about three times as many switches. A follow-up investigation indicated that for both simulations a large proportion of switches occurred in the first few pulses, suggesting that the model may not have been fully relaxed in the phase dimension.

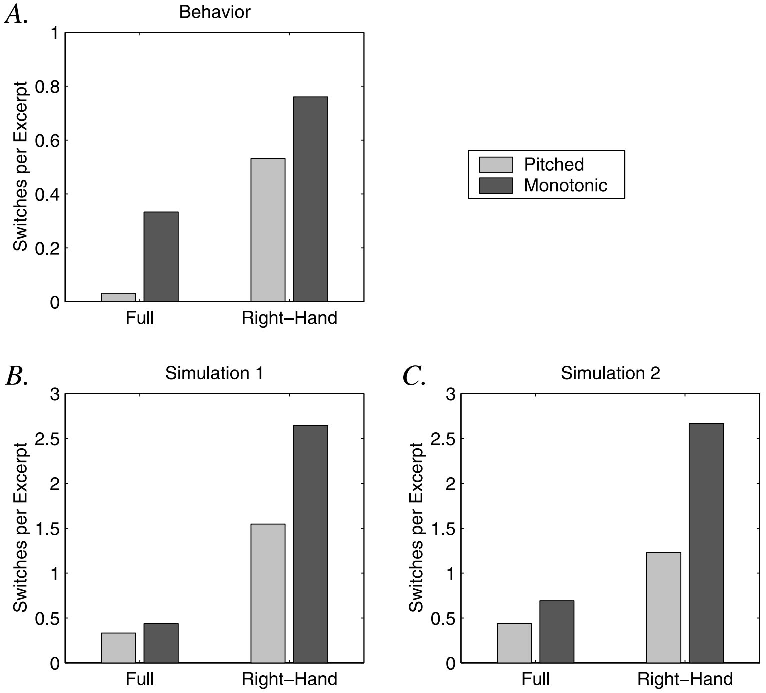

Fig. 11. Switches per excerpt for (A) musicians, (B) Simulation 1, and (C) Simulation 2.

The variability measures (Fig. 12), served as additional indexes of phase stability. CV (Fig. 12) was lower overall for both simulations than for the musicians, which is interpretable under the assumption that some variability arises in the (unmodeled) motor system (cf. Wing & Kristofferson, 1973). Both simulations appear to predict the observed increase of CV for the more syncopated RH versions, although Simulation 2 outperforms Simulation 1 in predicting the lack of effect for pitched versus monotonic versions. The remaining two measures, CDel and CDev, gauged overall phase offsets. Both models diverge from human performance in predicting large phase lags in the full-pitched, with Simulation 2 performing more poorly than Simulation 1 in this regard. Interestingly, aside from this large discrepancy, Simulation 2 predicts CDel in the other conditions with a fair degree of accuracy. It is perhaps unwise to interpret these final two comparisons in much detail, however, because these measures lie beyond the limit of comparability between this model and the current data. The model is one of beat perception, and does not explicitly include a model of tap production. People may tap

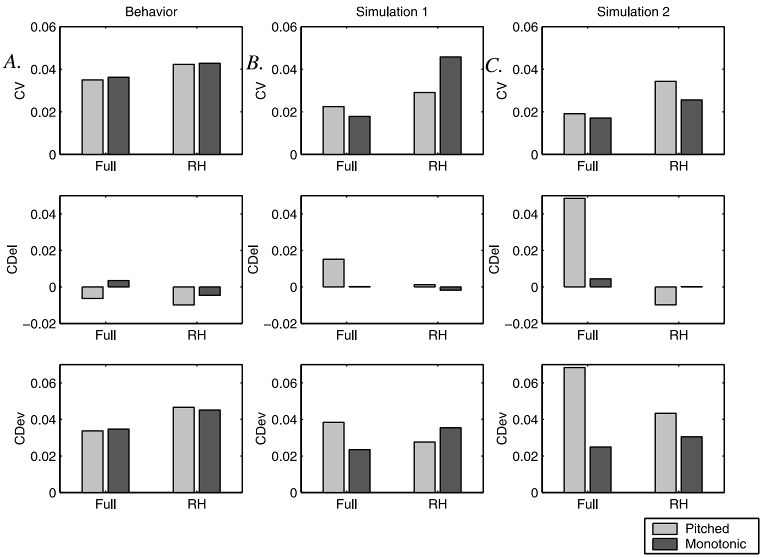

Fig. 12. Variability for (A) musicians, (B) Simulation 1, and (C) Simulation 2.

ahead of the beat, for example, in order to synchronize perceptual feedback from the taps with the perceived beat (see e.g. Semjen et al., 1998), and such considerations are not taken into account in this formulation of the model. Overall, however, the level of variability observed in both simulations is consistent with human performance.

In summary, the pattern forming network identified metrical structures in ragtime piano pieces, and stimulus-degraded versions of the pieces. The model typically identified 3 or 4 metrical levels in each piece, ranging from the lowest level present in these pieces (sixteenth notes) through the measure level. Metrical structure was correctly identified in most cases; however, the network's choice of tactus (tapping) level was more variable than humans'. Both simulations identified metrical structures incorrectly in some other cases; these were almost always stimulus degraded versions. In predicting the tapping data of Snyder and Krumhansl (2000), Simulation 2 (which used frequency information) consistently outperformed Simulation 1 (which did not use frequency information). Simulation 2 matched human performance well for natural music (full-pitched versions), and also predicted patterns of breakdown that were observed in human performance as information was removed from the stimulus. Most importantly, Simulation 2 correctly predicted that syncopation should be more disruptive to pulse-finding than lack of pitch information for these stimuli. This is somewhat surprising given that Simulation 2 is the one that uses frequency information; however, Simulation 1 also predicted the main features of the data. Model parameters could almost certainly be tweaked to provide a closer match to this particular pattern of results, however, this preliminary assessment provides a strong indication that the principles upon which the model is based capture key features of rhythm perception for complex, temporally structured stimuli such as music.

6. Conclusions

The pattern-forming dynamical system described here gave rise to metrically structured patterns of oscillation when stimulated with rhythmic input. It provided reasonable predictions of human performance for natural music, and network performance deteriorated similarly to human performance as stimulus information degraded. For both the model and for the musicians, increased syncopation was more disruptive to synchronization than lack of pitch information. Thus, this preliminary test provides initial support for the model, and also suggests promising areas for future research.

Interpreted in isolation, the empirical results appear to indicate that pitch information is not important in pulse finding, at least not for these stimuli. However, the modeling results contradict this interpretation: The simulation that used pitch information matched human performance more closely than the simulation that did not use pitch information. This apparently contradictory finding leads to a number of insights, however. First, even in the empirical study, pitch information did have some effect (Snyder & Krumhansl, 2000); significantly, musicians tapped less often on the down-beat for monotonic versions of the music. This indicates that pitch information is an important cue to phase, a finding that was reinforced by the modeling results. Pitch information conferred the biggest advantage to the model in phase preference (Fig. 10), assessed here as tapping mode. This interpretation squares with previous work regarding the role of non-temporal information in pulse-finding. Music theorists have suggested that a wide range of non-temporal cues interact in the perception of musical meter (e.g. Lerdahl & Jackendoff, 1983), and the influence of non-temporal factors in tapping behavior has been established (e.g. Repp, 1999b; Jones & Pfordresher, 1997). Thus, it seems likely that pitch information influences the perception of rhythm when it is available, but that people are able to adapt to a lack of pitch information to some extent. Importantly, pitch information can reduce phase ambiguity. Thus, tapping behavior can be significantly disrupted for highly syncopated music in the absence of pitch information. This suggests that one important direction for future research regards the role of pitch information in determining phase preference. More detailed empirical data on how and when pitch information influences rhythmic perception and performance is necessary.

Other factors are also likely to prove significant, especially when it comes to phase preferences for highly syncopated music (e.g., the RH versions) and phase stability, assessed here as switching rate. The most noteworthy of these is phase coherence, which is an important theoretical aspect of meter perception (Lerdahl & Jackendoff, 1983), and has been identified as a key factor in modeling time perception (Large & Jones, 1999; Large & Palmer, 2000). In the current model formulation the phases of the various oscillations are not always coherent, because no phase interaction was defined among the oscillators in the pattern-forming network. Phase interactions have been omitted at this modeling stage for two reasons. First, it is important to understand the phase preferences that arise solely in response to the external signal. In addition, the formal analysis of network behavior is more straightforward in the absence of such interactions. A significant next step in modeling at the behavioral level will be to define internal coupling such that oscillators entrain to one another, preserving metrical relationships. In this scenario, the network would have to settle on a phase interpretation of the rhythm as a whole, and this may be expected to improve predictions regarding strongly syncopated music, and switching of phase interpretations.

As described in the introduction, alternative models of beat induction and tracking have been proposed. While it is not possible to compare our model of each of these in detail, it is informative to make the broad distinction between non-linear and linear models (e.g. Scheirer, 1998; Todd et al., 1999) of beat induction. Linear models use banks of band-pass or comb filters in place of the non-linear oscillator network that is described here. Todd's model also includes a control theoretic component as an explicit model of the role of the motor system in beat perception. Although no linear model has been described that is able to recover an entire metrical structure, these models have had some success at beat finding; and a direct comparison would be informative at this level. In a bank of linear filters, all filters are always active to some extent, and the job of deciding which filter is currently the most highly active becomes a tricky business. This can lead to instability in identification of both beat period and phase, resulting in high rates of interpretation switching (Scheirer, 1998). By contrast, the advantages of the dynamical system approach follow precisely from stability properties. Thus, such pattern forming networks should be expected to display certain benefits over linear models, such as the ability to quickly find a metrical interpretation that remains viable for some time. As demonstrated here, such abilities are similar to those observed in human performance.

Other dynamical models of temporal tracking and/or motor synchronization (deGuzman & Kelso, 1991; Gasser et al., 1999; Large & Kolen, 1994; Large & Jones, 1999; Large & Palmer, 2000; McAuley, 1995; Pressing, 1999; Toiviainen, 1998) have been formulated as discrete-time maps rather than continuous-time differential equations. In fact, the current model has evolved from a generalization of the well-studied sine circle map, first proposed by Large and Kolen (1994) to model entrainment to complex, temporally modulated musical performances. Circle maps arise as mathematical simplifications of self-sustained oscillations, and the phase dynamics of Large and Kolen's original model is precisely the discrete version of Eq. (3b). More recently, sophisticated discrete-time models have been described that assume a small system of harmonically related oscillations with internal phase couplings defined to embody the temporal structure of particular complex sequences (Large & Jones, 1999; Large & Palmer, 2000). Closely related formalizations model complex forms of synchronization of motor sequences to auditory patterns (deGuzman & Kelso, 1991; Pressing, 1999) using harmonically structured phase coupling terms in place of multiple internal oscillations. The limitation of such of approaches is the requirement of specifying the network topology or the coupling parameters based upon an a priori knowledge of the sequences involved. One approach to this problem is to simply assume that the relevant parameters can be stored in memory and retrieved at will (e.g. Pressing, 1999). The current model deals with the problem more directly, by proposing a network that automatically activates the appropriate internal oscillations to reflect the temporal structure of novel sequences.

Finally, although the materials chosen to test the network controlled for two key aspects of stimulus complexity, there are other equally important issues that models of meter perception must address. The most significant of these is tempo modulation: Musical performances exhibit systematic deviations from metronomic regularity. Temporal fluctuations in performance are not noise, they are important cues for structural and emotional content (Large & Palmer, 2000). Earlier modeling work has focussed primarily on this issue (Large & Kolen, 1994; Large & Jones, 1999; Large & Palmer, 2000), however, the current proposal differs from earlier models in the way that tempo tracking is conceived. Rather than explicitly adapting the period of internal oscillations, tempo tracking is handled at the network level. Peaks in oscillator amplitude (Fig. 7(C)) should move when performance tempo changes significantly. Evaluation of this approach on tempo modulated musical performances is an issue for future research.

So, how do people synchronize to music? The answer proposed here is that as people listen to musical rhythms, a stable multi-periodicity pattern arises psychologically, serving as a dynamic embodiment of the temporal structure of the rhythm. In the simplest case, a single periodicity is used to guide tapping along with rhythms. In other situations, more complex metrical patterns may be engaged for synchronization of intricate movements such as in dance. Ultimately, this form of musical behavior speaks to the deep interdependence of action and perception. However, in understanding this connection, complexity of the real-world stimuli should not be underestimated. As the current analysis reveals, many important aspects of coordinated movement may linked directly to the perception of stimulus structure.

Acknowledgements

This research was supported in part by National Science Foundation grant SBR-9809446 awarded to the author. I would like to thank Joel Snyder and Carol Krumhansl for kindly providing me with their stimulus materials and data. I am also indebted to Tjeerd Dijkstra, Armin Fuchs, and Viktor Jirsa for insightful discussions, comments, and suggestions regarding this approach. Thanks also to Philip Fink, Mari Jones, Justin London, Jeff Pressing, Bruno Repp, Joel Snyder, and Michael Stauffer for commenting on an earlier draft of this manuscript. Finally, thanks to Dagmar Sternad and the Sante Fe Institute for arranging and funding the "Debates in Dynamics" workshop. Correspondence should be directed to large@walt.ccs.fau.edu, or to Edward Large, Center for Complex Systems and Brain Sciences, Florida Atlantic University, 777 Glades Road, Box 3091, Boca Raton, FL 3341-0991, USA.

References

Arrowsmith, D. K., & Place, C. M. (1990). *An introduction to dynamical systems*. Cambridge: Cambridge University Press.

Brown, J. C. (1993). Determination of the meter of musical scores by autocorrelation. *Journal of the Acoustical Society of America*, *94*, 1953–1957.

Chen, Y., Ding, M., & Kelso, J. A. S. (1997). Long memory processes (1/fa Type) in human coordination. *Physical Review Letters*, *79* (22), 4501–4504.

Collier, G. L., & Wright, C. E. (1995). Temporal rescaling of simple and complex ratios in rhythmic tapping. *Journal of Experimental Psychology: Human Perception and Performance*, *21* (q), 602–627.

Cooper, G., & Meyer, L. B. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.

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.

deGuzman, G. C., & Kelso, J. A. S. (1991). Multifrequency behavioral patterns and the phase attractive circle map. *Biological Cybernetics*, *64*, 485–495.

Desain, P. (1992). A (de)composable theory of rhythm perception. *Music Perception*, *9*, 101–116.

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

Deutsch, D. (1983). The generation of two isochronous sequences in parallel. *Perception and Psychophysics*, *34*, 331–337.

Drake, C., & Botte, M. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception and Psychophysics*, *54*, 277–286.

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

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

Gasser, M., Eck, D., & Port, R. (1999). Meter as mechanism: A neural network that learns metrical patterns. *Connection Science*, *11*, 187–215.

Goldenstein, S., Large, E. W., & Metaxis, D. M. (1999). Nonlinear dynamical systems approach to behavior modeling. *The Visual Computer*, *15*, 349–364.

Goto, M., & Muraoka, Y. (1998). Music understanding at the beat level: Real-time beat tracking for audio signals. In D. Rosenthal, & H. Okuno (Eds.), *Readings in Computational Auditory Scene Analysis*. Mahwah, NJ: Erlbaum.

Hary, D., & Moore, G. P. (1987). Synchronizing human movement with an external clock source. *Biological Cybernetics*, *56*, 305–311.

Hasty, C. F. (1997). *Meter as rhythm*. New York: Oxford University Press.

Hirsh, M. W., & Smale, S. (1974). *Differential equations, dynamical systems, and linear algebra*. San Diego, CA: Academic Press.

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., & McAuley, J. D. (2000). Categorical time judgements in extended temporal contexts. Journal of Experimental Psychology: Human Perception and Performance (under review).

Jones, M. R., & Pfordresher, P. Q. (1997). Tracking musical patterns using joint accent structure. *Canadian Journal of Experimental Psychology*, *51* (4), 271–290.

Jones, M. R., & Yee, W. (1997). Sensitivity to time change: The role of context and skill. Journal of Experimental Psychology: Human Perception & Performance, in press.

Kelso, J. A. S., DelColle, J. D., & Schöner, G. (1990). Action perception as a pattern formation process. In *Attention and Performance XIII* (pp. 139–169). Hilldale, NJ: Erlbaum.

Langner, G. (1992). Periodicity coding in the auditory system. *Hearing Research*, *60*, 115–142.

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

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How we track time varying events. *Psychological Review*, *106* (1), 119–159.

Large, E. W., & Palmer, C. (2000). The temporal response to musical performance: Recovering structure spite of and because of temporal deviations. *Cognitive Science* (under review).

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

Longuet-Higgins, H. C., & Lee, C. S. (1982). The perception of musical rhythms. *Proceedings of the Royal Society of London B*, *207*, 187–217.

Mates, J., Radil, T., & Pöppel, E. (1992). Cooperative tapping: Time control under different feedback conditions. *Biological Cybernetics*, *70*, 463–473.

McAuley, J. D. (1995). Perception of time as phase: Toward an adaptive oscillator model of rhythmic pattern processing. Unpublished Ph.D. dissertation. Indiana University.

Murray, J. D. (1989). *Mathematical Biology*. Berlin: Springer.

Palmer, C. (1997). Music performance. *Annual Review of Psychology*, *48*, 115–138.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, *11*, 409–464.

Peper, C. L. E., Beek, P. J., & Van Wieringen, P. C. W. (1995). Multi-frequency coordination of bimanual tapping: Asymmetrical coupling and signs of supercriticality. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1117–1138.

Perko, L. (1991). *Differential equations and dynamical systems*. New York: Springer.

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

Pressing, J., & Jolley-Rogers, G. (1997). Spectral Properties of human cognition and skill. *Biological Cybernetics*, *76* (5), 339–347.

Pressing, J. (1998). Error correction processes in temporal pattern production. *Journal of Mathematical Psychology*, *42*, 63–101.

Pressing, J. (1999). The referential dynamics of cognition and action. *Psychological Review*, *106* (4), 714–747.

Recanzone, G. H., Merzenich, M. M., & Schreiner, C. E. (1992). Changes in the distributed temporal response properties of SI cortical neurons reflect improvements in performance on a temporally based tactile discrimination task. *Journal of Neurophysiology*, *67* (5), 1071–1091.

Repp, B. H. (1999a). Control of expressive and metronomic timing in pianists. *Journal of Motor Behavior*, *31*, 145–164.

Repp, B. H. (1999b). Detecting deviations from metronomic timing in music: Effects of perceptual structure on the mental timekeeper. *Perception and Psychophysics*, *61*, 529–548.

Repp, B. H. (1999c). Relationships between performance timing, perception of timing perturbations, and perceptual-motor synchronization in two Chopin preludes. *Journal of Australian Psychology*, *51*, 188–203.

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, MA: MIT Press.

Scheirer, E. D. (1998). Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, *103*, 588–601.

Semjen, A., Vorberg, D., & Schulze, H. H. (1998). Getting synchronized with the metro-nome: Comparisons between phase and period correction. *Pyschological Research*, *61*, 44–55.

Slaney, M. (1998). *Auditory toolbox: A Matlab toolbox for auditory modeling work, Version 2*. Technical Report #1998-010, Interval Research Corporation.

Snyder, J., & Krumhansl, C. L. (2000). Tapping to ragtime: Cues to pulse-finding. *Music Perception* (under review).

Todd, N. P. M. (1994). The auditory primal sketch: A multi-scale model of rhythmic grouping. *Journal of New Music Research*, *23*, 25–69.

Todd, N. P., Lee, C. S., & O'Boyle, D. J. (1999). A sensory-motor theory of rhythm perception. *Journal of New Music Research*, *28*, 1–24.

Toiviainen, P. (1998). An interactive MIDI accompanist. *Computer Music Journal*, *22*, 63–75.

Van Noorden, L., & Moelants, D. (1999). Resonance in the perception of musical pulse. *Journal of New Music Research*, *28*, 43–66.

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.

von Holst, E. (1937). *The behavioral physiology of animals and man*. London: Methuen & Co Ltd.

Vorberg, D., & Wing, A. (1996). Modeling variability and dependence in timing. In H. Heuer, & S. W. Keele (Eds.), *Handbook of perception and action, Vol. 2: Motor skills* (pp. 181–262). London: Academic Press.

Vos, P. G., Van Dijk, A., & Schomaker, L. (1994). Melodic cues for metre. *Perception*, *23*, 965–976.

Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception and Psychophysics*, *14* (1), 5–12.

Woodrow, H. (1951). Time perception. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology* (pp. 1224–1236). New York: Wiley.

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