ELSEVIER

Cognitive Science 26 (2002) 1–37

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

Perceiving temporal regularity in music

Edward W. Large[a,*], Caroline Palmer[b]

[a]*Florida Atlantic University, Boca Raton, FL 33431-0991, USA*
[b]*The Ohio State University, USA*

Received 15 September 2001; received in revised form 26 September 2001; accepted 26 September 2001

Abstract

We address how listeners perceive temporal regularity in music performances, which are rich in temporal irregularities. A computational model is described in which a small system of internal self-sustained oscillations, operating at different periods with specific phase and period relations, entrains to the rhythms of music performances. Based on temporal expectancies embodied by the oscillations, the model predicts the categorization of temporally changing event intervals into discrete metrical categories, as well as the perceptual salience of deviations from these categories. The model's predictions are tested in two experiments using piano performances of the same music with different phrase structure interpretations (Experiment 1) or different melodic interpretations (Experiment 2). The model successfully tracked temporal regularity amidst the temporal fluctuations found in the performances. The model's sensitivity to performed deviations from its temporal expectations compared favorably with the performers' structural (phrasal and melodic) intentions. Furthermore, the model tracked normal performances (with increased temporal variability) better than performances in which temporal fluctuations associated with individual voices were removed (with decreased variability). The small, systematic temporal irregularities characteristic of human performances (chord asynchronies) improved tracking, but randomly generated temporal irregularities did not. These findings suggest that perception of temporal regularity in complex musical sequences is based on temporal expectancies that adapt in response to temporally fluctuating input. © 2002 Cognitive Science Society, Inc. All rights reserved.

Keywords: Music cognition; Rhythm perception; Dynamical systems; Oscillation

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

PII: S0364-0213(01)00057-X

1. Introduction

The ease with which people perceive and enjoy music provides cognitive science with significant challenges. Among the most important of these is the perception of time and temporal regularity in auditory sequences. Listeners tend to perceive musical sequences as highly regular; people without any musical training snap their fingers or clap their hands to the temporal structure they perceive in music with seemingly little effort. In particular, listeners hear sounded musical events in terms of durational categories corresponding to the eighth-notes, quarter-notes, half-notes, and so forth, of musical notation. This effortless ability to perceive temporal regularity in musical sequences is remarkable because the actual event durations in music performances deviate significantly from the regularity of duration categories (Clarke, 1989; Gabrielsson, 1987; Palmer, 1989; Repp, 1990). In addition, listeners perceive these temporal fluctuations or deviations from duration categories as systematically related to performers' musical intentions (Clarke, 1985; Palmer, 1996a; Sloboda, 1983; Todd, 1985). For example, listeners tend to perceive duration-lengthening near structural boundaries as indicative of phrase endings (while still hearing regularity). Thus, on the one hand, listeners perceive durations categorically in spite of temporal fluctuations, while on the other hand listeners perceive those fluctuations as related to the musical intentions of performers (Sloboda, 1985; Palmer, 1996a). Music performance provides an excellent example of the temporal fluctuations with which listeners must cope in the perception of music and other complex auditory sequences.

The perceptual constancy that listeners experience in the presence of physical change is not unique to music. Listeners recognize speech, for example, amidst tremendous variability across speakers. Early views of speaker normalization treated extralinguistic (nonstructural) variance as noise, to be filtered out in speech recognition. More recently, talker-specific characteristics of speech such as gender, dialect, and speaking rate, are viewed as helpful for the identification of linguistic categories (cf. Nygaard, Sommers, & Pisoni, 1994; Pisoni, 1997). We take a similar view here, that stimulus variability in music performances may help listeners identify rhythmic categories. Patterns of temporal variability in music performance have been shown to be systematic and intentional (Bengtsson & Gabrielsson, 1983; Palmer, 1989), and are likely to be perceptually informative.

We describe an approach to rhythm perception that addresses both the perceptual categorization of continuously changing temporal events and perceptual sensitivity to those temporal fluctuations in music performance. Our approach assumes that people perceive a rhythm—a complex, temporally patterned sequence of durations—in relation to the activity of a small system of internal oscillations that reflects the rhythm's temporal structure. Internal self-sustained oscillations are the perceptual correlates of beats; multiple internal oscillations that operate at different periods (but with specific phase and period relations) correspond to the hierarchical levels of temporal structure perceived in music. The relationship between this system of internal oscillations and the external rhythm of an auditory sequence governs both listeners' categorization of temporal intervals, and their response to temporal fluctuations as deviations from categorical expectations.

This article describes a computational model of the listeners' perceptual response: a dynamical system that tracks temporal structures amidst the expressive variations of music performance, and interprets deviations from its temporal expectations as musically expressive. We test the model in two experiments by examining its response to performances in which the same pianists performed the same piece of music with different interpretations (Palmer, 1996a; Palmer & van de Sande, 1995). We consider two types of expressive timing common to music performance that correlate with performers' musical intentions: lengthening of events that mark phrase structure boundaries, and temporal spread or asynchrony among chord tones (tones that are notated as simultaneous) that mark the melody (primary musical voice). Two aspects of the model of rhythm perception are assessed. First, we evaluate the model's ability to track different temporal periodicities within music performances. This tests its capacity for following temporal regularity in the face of significant temporal fluctuation. Second, we compare the model's ability to detect temporal irregularities against the structural intentions of performers. This gauges its sensitivity to musically expressive temporal gestures that are known to be informative for listeners. Additionally, we observe that some types of small but systematic temporal irregularities (chord asynchronies) can improve tracking in the presence of much larger temporal fluctuations (rubato). Comparisons of the model's beat-tracking of systematic temporal fluctuations and of random fluctuations in simulated performances indicate that performed deviations from precise temporal regularity are not noise; rather, temporal fluctuations are informative for listeners in a variety of ways. In the next section, we review music-theoretic descriptions of temporal structures in music, and in the following section, we describe the temporal fluctuations that occur in music performance.

1.1. Rhythm, metrical structure, and music notation

Generally speaking, *rhythm* is the whole feeling of movement in time, including pulse, phrasing, harmony, and meter (Apel, 1972; Lerdahl & Jackendoff, 1983). More commonly, however, rhythm refers to the temporal patterning of event durations in an auditory sequence. Beats are perceived pulses that mark equally spaced (subjectively isochronous) points in time, either in the form of sounded events or hypothetical (unsounded) time points. Beat perception is established by the presence of musical events; however, once a sense of beat has been established, it may continue in the mind of the listener even if the event train temporarily comes into conflict with the pulse series, or after the event train ceases (Cooper & Meyer, 1960). This point is an important motivator for our theoretical approach; once established, beat perception must be able to continue in the presence of stimulus conflict or in the absence of stimulus input. Music theories describe *metrical structure* as an alternation of strong and weak beats over time. One theory conceptualizes metrical structure as a grid of beats at various time scales (Lerdahl & Jackendoff, 1983), as shown in Fig. 1; these are similar to metrical grids proposed in phonological theories of speech (Liberman & Prince, 1977). According to this notational convention, horizontal rows of dots represent levels of beats, and the relative spacing and alignment among the dots at adjacent levels captures the relationship between the hypothetical periods and phases of the beat levels. Metrical accents are indicated in the grid by the number of coinciding dots. Points at which many beats coincide are called strong beats; points at which few beats coincide are called weak beats. Although these metrical grids are idealized (music performances contain more complex

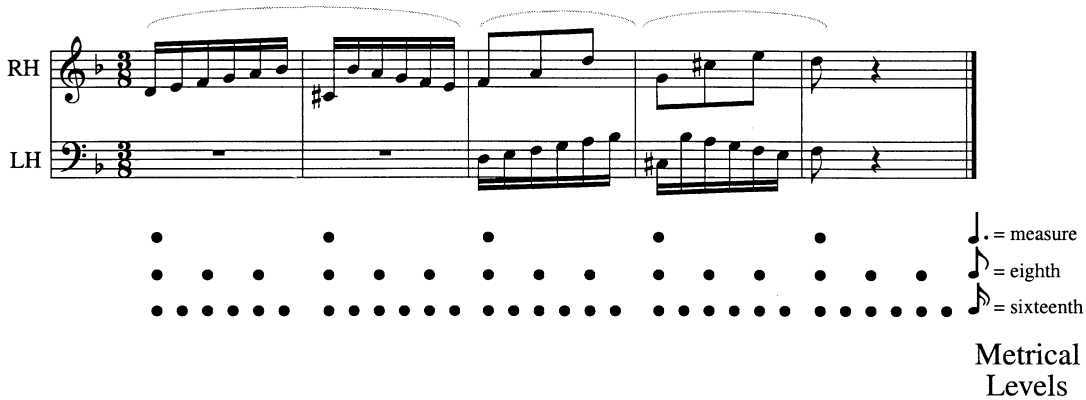

Fig. 1. Opening section from 2-part invention in D-minor, by J.S. Bach. This example shows one of the instructed phrase structures used in Experiment 1 (top); metrical grid notation indicates metrical accent levels (bottom).

period and phase relationships among beat levels than those captured by metrical grids), the music-theoretic invariants reflected in these grids inform our model of the perception of temporal regularity in music.

Western conventions of music notation provide a categorical approximation to the timing of a music performance. Music notation specifies event durations categorically; durations of individual events are notated as integer multiples or subdivisions of the most prominent or salient metrical level. Events are grouped into measures that convey specific temporal patterns of accentuation (i.e. the meter). For example, the musical piece notated in Fig. 1 with a time signature of 3/8 uses an eighth-note as its basic durational element, and the durational equivalent of three eighth-notes defines a metrical unit of one measure, in which the first position in the measure is a strong beat and the others are weaker. Although notated durations refer to event onset-to-offset intervals, listeners tend to perceive musical events in terms of onset-to-onset intervals (or inter-onset intervals, IOIs), due to the increased salience of onsets relative to offsets. Hereafter we refer to musical event durations in terms of IOIs.

In this article we focus on the role of meter in the perception of rhythm. Listeners' perception of duration categories in an auditory sequence is influenced by the underlying meter; the same auditory sequence can be interpreted to have a different rhythmic pattern when presented in different metrical contexts (Clarke, 1987; Palmer & Krumhansl, 1990). To model meter perception, we assume that a small set of internal oscillations operates at periods that are roughly approximate to those of each hierarchical metrical level shown in Fig. 1. When driven by musical rhythms, such oscillations phase-lock to the external musical events. Previous work has shown this framework to provide both flexibility in tracking temporally fluctuating rhythms (Large & Kolen, 1994; Large, 1996) and a concurrent ability to discriminate temporal deviations (Large & Jones, 1999). In the current study, we extend this framework to a more natural and complex case that provides a robust test of the model: multivoiced music performances that contain large temporal fluctuations. Most important, the model proposed here predicts that temporal fluctuations can aid the perception of auditory events, as we show in two experiments. The next section describes what information is available in the temporal fluctuations of music performance.

1.2. Temporal fluctuations in music performance

The complex timing of music performance often reflects a musician's attempt to convey an interpretation of musical structure to listeners. The structural flexibility typical of Western tonal music allows performers to interpret musical pieces in different ways. Performers highlight interpretations of musical structure through the use of expressive variations in frequency, timing, intensity, and timbre (cf. Clarke, 1988; Nakamura, 1987; Palmer, 1997; Repp, 1992; Sloboda, 1983). For example, different performers can interpret the same musical piece with different phrase structures (Palmer, 1989, 1992); each performance reflects slowing down or pausing at events that are intended as phrase endings, similar to phrase-final lengthening in speech. Furthermore, listeners are influenced by these temporal fluctuations; the presence of phrase-final lengthening in different performances of the same music influenced listeners' judgments of phrase structure, indicating that the characteristic temporal fluctuations are information-bearing (Palmer, 1988). Thus, a common view is that temporal fluctuations in music performance serve to express structural relationships such as phrase structure (Clarke, 1982; Gabrielsson, 1974) and these large temporal fluctuations provide a challenging test for the model of beat perception described here.

Temporal fluctuations in music performance may also mark the relative importance of different musical parts or voices. Musical instruments such as the piano provide few timbral cues to differentiate among simultaneously co-occurring voices, and the problem of determining which tones or features belong to the same voice or part over time is difficult; this problem is often referred to as stream segregation (cf. Bregman, 1990). Most of Western tonal music contains multiple voices that co-occur, and performers are usually given some freedom to interpret the relative importance of voices. Performers often provide cues such as temporal or intensity fluctuations that emphasize the melody, or most important part (Randel, 1986). Early recordings of piano performance documented a tendency of pianists to play chordal tones (tones notated as simultaneous) with asynchronies up to 70 ms across chord-tone onsets (Henderson, 1936; Vernon, 1936). Palmer (1996a) compared pianists' notated interpretations of melody (most important voice) with expressive timing patterns of their performances. Events interpreted as melody were louder and preceded other events in chords by 20–50 ms (termed melody leads). Although the relative importance of intensity and temporal cues in melody perception is unknown (see also Repp, 1996), the temporal cues alone subsequently affected listeners' perception of melodic intentions in some performances (Palmer, 1996a). Thus, temporal fluctuations in melody provide a subtle test for the model we describe here.

Which cues in music performances mark metrical structure? Although a variety of cues indicate some relationship with meter, there is no one single cue that marks meter. Melody leads tend to coincide with meter; pianists placed larger asynchronies (melody preceding other note events) on strong metrical beats than on weak beats, in both well-learned and unpracticed performances (Palmer, 1989; 1996a). Performers also mark the meter with variations in event intensity or duration (Shaffer, Clarke & N. Todd, 1985; Sloboda, 1983). Which cues mark meter the most can change with musical context. Drake and Palmer (1993) examined cues for metrical, melodic, and rhythmic grouping structures, in piano performances of simple melodies and complex multivoiced music. Metrical accents and rhythmic groups (groups of short and long durations) were marked by intensity, with strong metrical beats and long notated durations performed louder than other events. However, the performance cues that coincided with important metrical locations changed across different musical contexts. These findings suggest that performance cues alone may not explain listeners' perception of metrical regularity across many contexts. We test a model of listeners' expectancies for metrical regularity that may aid perception of meter in the absence of consistent cues.

1.3. Perceptual cues to musical meter

Which types of stimulus information do listeners use to perceive the temporal regularities of meter? Several studies suggest that listeners are sensitive to multiple temporal periodicities in complex auditory sequences (Jones & Yee, 1997; Palmer & Krumhansl, 1990; Povel, 1981). The statistical regularities of Western tonal music may provide some cues to temporal periodicities. For a given metrical level to be instantiated in a musical sequence, it is necessary that a sufficient number of successive beats be sounded to establish that periodicity. Statistical analyses of musical compositions indicate that composers vary the frequency of events across metrical levels (Palmer & Krumhansl, 1990; Palmer, 1996b), which provides sufficient information to differentiate among meters (Brown, 1992). Although this approach is limited by its reliance on a priori knowledge about the contents of an entire musical sequence, it supports our assumption that musical sequences contain perceptual cues to multiple temporal periodicities, which are perceived simultaneously during rhythm perception.

One problem faced by models of meter perception is the determination of which musical events mark metrical accents. Longuet-Higgins and Lee's (Longuet-Higgins & Lee, 1982) model assumes that events with long durations initiate major metrical units, because they are more salient perceptually than are events with short durations. In their model, longer durations tend to be assigned to higher metrical levels than short durations. Perceptual judgments document that events that are louder or of longer duration than their neighbors are perceived as accented (Woodrow, 1951). Thus, the correct metrical interpretation may be found by weighting each event in a sequence according to perceived cues of accenting. However, duration and intensity cues in both music composition and performance are influenced by many factors in addition to meter, including phrase structure, melodic importance, and articulation (Nakamura, 1987; Palmer, 1988; Sloboda, 1983). Often the acoustic cues to meter are ambiguous, interactive, or simply absent; yet listeners can still determine the meter.

Large (2000a) proposed a model of meter perception in which a musical sequence provides input to a pattern-forming dynamical system. The input was a temporally regular recording of musical pieces (i.e. with objectively isochronous beats; see Snyder & Krumhansl, 2000), preprocessed to recover patterns of onset timing and intensity. Under such rhythmic stimulation, the system begins to produce self-sustained oscillations and temporally structured patterns of oscillations. The resulting patterns dynamically embody the perception of musical beats on several time scales, equivalent to the levels of 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 conflicting information, yet they can also reorganize, given sufficient indication of a new temporal structure. The performance of the model compared favorably with the results of a synchronization study (Snyder & Krumhansl, 2000) that was explicitly designed to test meter induction in music. However, the auditory sequences used from Snyder & Krumhansl (2000) were computer-generated and temporally regular; they contained no temporal fluctuations in the categorical event durations. We describe a model in the next section similar to that of Large (2000a), but applied to more realistic, temporally fluctuating performances.

2. Modeling meter perception

Before we provide the mathematical description of the system, we first provide an intuitive description. The perception of musical beat is modeled as an active, self-sustained oscillation. This self-sustaining feature may be conceived of as a mathematicization of Cooper & Meyer's description of the sense of beat that "once established, (it) 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," (Cooper & Meyer, 1960, p. 3; cf. Large, 2000a). The job of the oscillator is to synchronize with the external rhythmic signal. However, it does not respond to just any onset as a potential beat; it responds only to onsets in the neighborhood of where it expects beats to occur. Thus, it has a region of sensitivity within its temporal cycle whose peak or maximum value corresponds to where the beat is expected. An onset that occurs within the sensitive region, but does not coincide exactly with the peak, causes a readjustment of the oscillator's phase and a smaller adjustment of period. Additionally, the width of the sensitive region is adjustable. Onsets that occur at or very near the peak sensitivity cause the width of the sensitive region to shrink; other onsets within the region but not close to the peak cause the sensitive region to grow. Finally, the coupling of multiple oscillators with different periods gives the system a hierarchical layering associated with musical and linguistic meter.

The current model draws upon earlier work (Large & Kolen, 1994; Large & Jones, 1999) with the important distinction that it combines previous notions of a temporal receptive field (the sensitive region) and an attentional pulse (which determines the perceptual noticeability of temporal fluctuations), using the notion of an expectancy function. The model is a mathematical simplification of Large's (2000a) model, and it addresses beat-tracking in the challenging case of temporally fluctuating music performance. The model is temporally discrete, and captures the behavior of a few oscillators whose periods correspond to the metrical structure of the piece, which is assumed to be known *a priori*. The initial periods of the oscillators, as well as their invariant phase and period coupling relationships, are chosen in advance. Thus, we assume the metrical structure and initial beat period, which are inferred in Large's (2000a) more complete continuous time model. The discrete-time formulation is used here because it offers several advantages compared to its continuous-time cousin; it is economical, and predictions concerning time difference judgements have been fully worked out for this model (Large & Jones, 1999). In this section, we begin by describing the dynamics of a single oscillator, and then describe the coupling of multiple oscillators.

The synchronization of a single oscillator to a periodic driving signal can be described using the well-studied sine circle map (Glass & Mackey, 1988). The sine circle map is a model of a nonlinear oscillation that entrains to a periodic signal, and it uses a discrete-time formalism. A series of relative phase values is produced by the circle map, representing the phases of the oscillator's cycle at which input events occur (in our case, notes). It calculates the relative phase for event $n + 1$, ϕ_{n+1}, in terms of the relative phase of event n, the ratio of the signal's period, q, to the oscillator's period, p, and the coupling of the driven oscillation to the external signal, $-\eta/2\pi \sin 2\pi\phi_n$. The coupling term models synchronization of the oscillator with the signal.

$$\phi_{n+1} = \phi_n + \frac{q}{p} - \frac{\eta}{2\pi} \sin 2\pi\phi_n \qquad (mod_{-0.5,0.5}1) \tag{1}$$

The notation $(mod_{-0.5,0.5}1)$ indicates that phase is taken modulo 1 and normalized to the range $-0.5 < \phi < 0.5$. This means that relative phase is measured as a proportion of the driven oscillator's cycle, where zero corresponds to time of the expected beat, negative values indicate that an event occurred early (before the beat) and positive values indicate that the event occurred late (after the beat).

Two modifications of the sine circle map (Equation 1) allow the model to track the beat in complex rhythms where each event IOI is potentially different and which contain multiple periodicities (Large & Kolen, 1994). First, to handle IOIs of varying sizes, it is necessary to replace the fixed period, q, on the nth cycle, with nth IOI, which is measured by $t_{n+1} - t_n$, where t_n is the onset time of event n. The phase advance, indicated by the clockwise arrow in Panel A of Fig. 2, is the proportion of the oscillator's period corresponding to the nth IOI, that is: $+(t_{n+1}-t_n)/p_n$. Thus, this modification maps the event onset times of the complex rhythmic sequence onto the phase of the internal oscillation.

Second, to account for the model's synchronization with a signal that contains multiple periodicities, we exploit the notion of a *temporal receptive field* (Large & Kolen, 1994), which is the time during which the oscillator can adjust its phase. Events that occur within the temporal receptive field cause a phase adaptation, whereas events that occur outside the temporal receptive field result in little or no phase adaptation. Fig. 2A also illustrates an adjustment to relative phase, $-\eta_\phi X_n F(\phi_n, \kappa_n)$, indicated by the counterclockwise arrow. As described below, the oscillator attempts to synchronize to events that occur near "the beat" (i.e. $\phi = 0$) while ignoring events that occur away from the beat. Together, these modifications yield the following equation, capturing the phase of the internally generated oscillation (the beat) at which each event occurs.

$$\phi_{n+1} = \phi_n + \frac{t_{n+1} - t_n}{p_n} - \eta_\phi X_n F(\phi_n, \kappa_n) \qquad (mod_{-0.5,0.5}1) \tag{2}$$

Here $F(\phi_n, \kappa_n)$ is the coupling function modeling synchronization of the oscillation with a subset of the event onsets in the complex rhythm, η_ϕ is the coupling strength, capturing the overall amount of force that the rhythm exerts on the oscillation, and X_n is the amplitude of

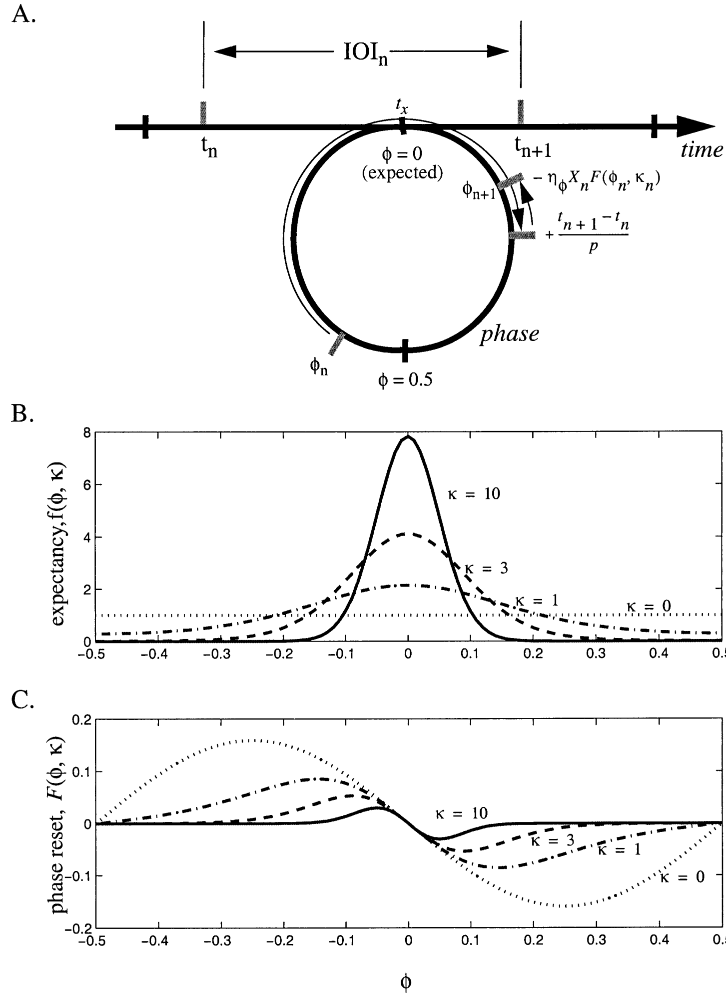

Fig. 2. A) The modified circle map (Equation 2) takes the time of external events (t_n and t_{n+1}) onto the phase of an internal oscillation. The counter-clockwise arrow indicates phase resetting (see text). Effects of kappa (focus parameter) on the expectancy window are shown in Panel B and on phase resetting are shown in Panel C.

the nth onset, capturing the amount of force exerted by each individual event onset. In this paper, X_n is fixed at 1 as a simplifying assumption. κ_n is a focus or concentration parameter that determines the extent of an expectancy function, as shown in Fig. 2B (termed a pulse of attentional energy by Large & Jones, 1999). It models the degree of expectancy for the occurrence of events near $\phi = 0$. High values of κ, shown in Fig. 2B, imply highly focused temporal expectancies, whereas low values of κ, also shown, imply uncertainty as to when events are likely to occur.

Next we define the model's expectancy for when an event will occur, termed the *attentional pulse* by Large and Jones (1999). The attentional pulse is modeled as a periodic probability density function, the von Mises distribution, which is shaped similarly to a

Gaussian distribution but defined on the circle (i.e., phase). Equation 2a defines the pulse, and I_0 is a modified Bessel function of the first kind of order zero that scales the amplitude of the expectancy.

$$f(\phi, \kappa) = \frac{1}{I_0(\kappa)} \exp \kappa \cos 2\pi\phi \tag{2a}$$

Four attentional pulses are shown in Fig. 2B, with different shapes corresponding to different values of κ. Each pulse defines a different temporal *expectancy function*, a region of time during which events are expected to occur, i.e. when expectancy is near maximum. For example, when $\kappa = 10$, expectancy is highly focussed about $\phi = 0$; however, when $\kappa = 0$, expectancy is dispersed throughout the oscillator's cycles. Fig. 2C compares the pulses with their corresponding coupling functions (shown for the same values of κ). The coupling function is the derivative of a unit amplitude-normalized version of the attentional pulse (cf. Large & Kolen, 1994). Thus it shares the same expectancy function with the attentional pulse. The temporal region where events are most highly expected is identical to that over which phase adjustment is most efficient; both are determined by κ. As illustrated by comparison of Figs. 2B & C, when expectancy is near its maximum, phase resetting is efficient; when the expectancy level is near zero, phase adjustment does not occur.

$$F(\phi, \kappa) = \frac{1}{2\pi \exp \kappa} [\exp \kappa \cos 2\pi\phi] \sin 2\pi\phi \tag{2b}$$

The basic idea is that if κ is large—and expectations are highly focussed—the oscillator will synchronize to those events that occur near the expected beat, but other events can move around the circle map without affecting its phase or period. Thus, the temporal receptive field must be wide enough to accommodate temporal variability in the sequence at the corresponding metrical level, while being narrow enough to ignore events that correspond to other metrical levels. Real-time adaptation of κ is incorporated into the model as described in Large & Jones (1999, Appendix 2). The parameter that determines the adaptation rate of focus is η_κ. The basic idea of this procedure is that accurate predictions cause an increase in focus (κ), whereas inaccurate predictions result in decreased focus. Large & Jones (1999) found that κ—as indexed by noticeability of temporal deviations—increased as sequence variability decreased. Attentional focus depends on the variability of the sequence, as predicted by this model.

Phase coupling alone is not sufficient to model phase synchrony in the presence of the complex temporal fluctuations typical of music performance. To maintain synchrony, the period of the oscillation must also adapt in response to changes in sequence rate (cf. Large, 1994; Large & Kolen, 1994; McAuley & Kidd, 1995). The period of event $n + 1$, p_{n+1}, is modeled as

$$p_{n+1} = p_n(1 + \eta_p X_n F(\phi_n, \kappa_n)) \tag{3}$$

in which the coupling function for period is the same as that for phase, but an independent parameter for coupling strength, η_p, is allowed for period adaptation. In all there are three parameters that determine the behavior of each oscillator, phase coupling strength, η_ϕ, period adaptation rate, η_p, and focus adaptation rate, η_κ. These parameter values are chosen to enable stable tracking of rapidly changing stimulus sequences. In general the model tracks well for a relatively wide range of values, where we generally assume that $0 < \eta_\kappa < \eta_p < \eta_\phi \leq 1$.

2.1. Modeling hierarchical metrical structures

Thus far, we have described the model's ability to track individual metrical levels or periodicities. However, musical rhythms typically contain multiple periodicities with simple integer ratio relationships among the phases and periods of the components. To track the metrical structure of musical rhythms, multiple oscillations must track different periodic components, or levels of beats. Furthermore, multiple oscillators must be constrained by their relationships with one another. Specifically, the internal oscillators are coupled to one another so as to preserve certain phase and period relationships that are characteristic of hierarchical metrical structures.

Phase and period coupling behavior is determined by the *relative period* between two metrical levels. Relative period is the number of beats at the lower metrical level that correspond to a single beat period at the higher level. Typical values of relative period in Western tonal music are 2:1 and 3:1 (e.g. Lerdahl & Jackendoff, 1983). Phase and period relationships are maintained by two linear coupling terms, one for phase and another for period. Phase coupling strength is determined by the parameter α_ϕ and period coupling strength by α_p. To simulate uncoupled oscillations, we choose $\alpha_\phi = \alpha_p = 0$; for coupled oscillations $\alpha_\phi = \alpha_p = 1$. When two or more oscillators are coupled in this way, the maximum value of their attentional pulses occur at (very nearly) the same time when they coincide (for further details of internal coupling, see Large & Jones, 1999).

To model expectancy pulses for a multi-leveled metrical structure, we use a mixture of von Mises distributions. This model is general enough to capture any number of metrical levels; in this paper the number is restricted to two. Fig. 3A shows a two-leveled metrical structure modeled as a mixture of two von Mises distributions. The figure illustrates a 3:1 metrical relationship, and the mixture includes one component distribution (shown using dashed lines) for each level of the metrical hierarchy. First, we write the component von Mises distributions using subscripts, as:

$$f_j(\phi) = \frac{1}{I_0(\kappa_j)} \exp \kappa_j \cos 2\pi j \phi \tag{4}$$

and then a mixture of two multimodal von Mises distributions is given by

$$f(\phi, \underline{\kappa}) = \sum_j w_j f_j(\phi) \tag{5}$$

where $\underline{\kappa}$ is the vector of values across j. j is a sequence that gives the period of each oscillator relative to the one below it in the hierarchy. In this paper, $j = \{1, 2\}$ or $j = \{1, 3\}$ (shown in Fig. 3A), indicating binary or ternary ratio relationships between metrical levels typical of Western meters. Thus, each entry in j is the number of beats at the metrical level immediately

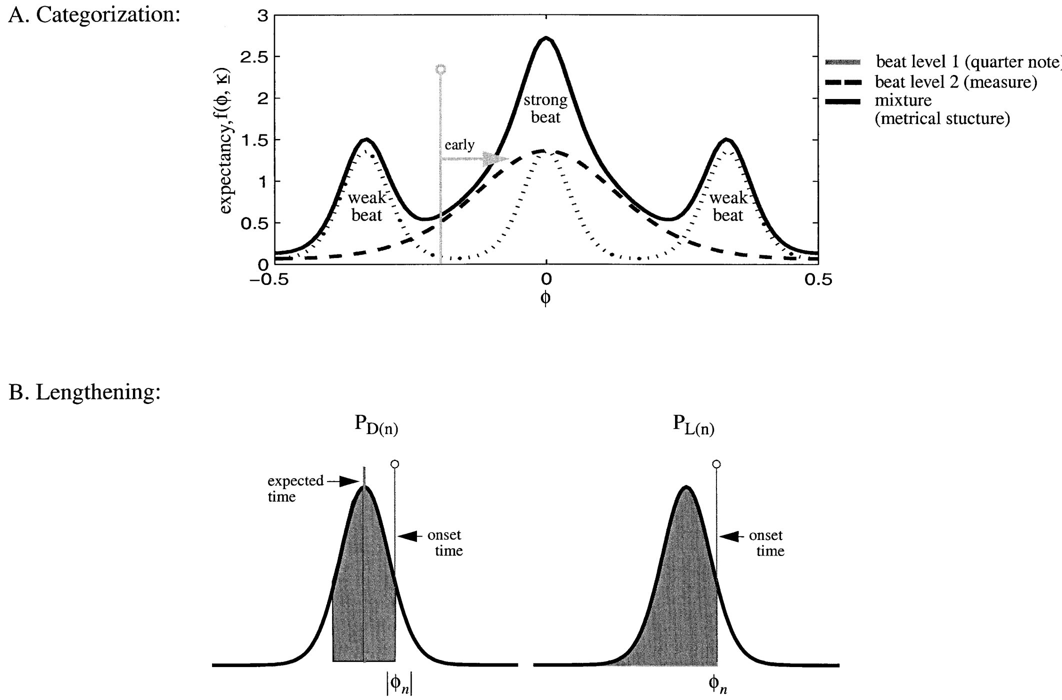

Fig. 3. A) Model expectancies for a ternary meter (3:1 period ratio) based on a mixture of two von Mises distributions (Equation 5, solid line). Component von Mises distributions correspond to a quarter-note beat level (dotted line) and a dotted half-note beat level (dashed line); $\kappa = 1.5$ for each component. B) Shaded area under the curve indicates the probability of perceiving a deviation, P_D, and probability of the event having occurred late in the cycle, P_L, for a single event onset (vertical line).

below, corresponding to a single beat period at the current level. Finally, w_j is the weight associated with each metrical level in j. For all simulations described in this paper, we will consider two-component mixtures with equal weights, $w_1 = w_2 = 0.5$ (the contributions of the two von Mises distributions are equivalent).

2.2. Sensitivity to temporal fluctuations

We model sensitivity to temporal fluctuations in two steps. The first step is the *categorization of each note onset* as marking a particular beat at a particular metrical level; the second step is the *perception of temporal differences* as deviations from the durational categories. Note that we are explicitly hypothesizing the perceptual recovery of duration categories as reflected in the notated score as a prerequisite to the perception of temporal fluctuations. In previous studies of expressive timing in musical sequences (e.g. Clarke, 1985; Palmer, 1996a; Sloboda, 1983; Todd, 1985), it has generally been assumed that durational categories are available to the listener *a priori*. In contrast, we require that our model recover both the duration categories and the expressive timing information.

2.2.1. Categorizing note onsets

As the model tracks events in a musical sequence, it associates each event with either a strong beat (corresponding to a larger metrical periodicity) or a weak beat (corresponding to a smaller metrical periodicity). Additionally, it associates each note onset with a specific pulse at that level. For example, the event shown in Fig. 3A is categorized as a strong beat because the amount of expectancy associated with the oscillator at the measure level (dashed line) is greater than the amount of expectancy associated with the oscillator at the quarter-note level (the dotted line). Multiple onsets associated with the same attentional pulse are heard as a chord. We can make this classification explicit by applying the von Mises model of the attentional pulse. To classify each event onset, we calculate τ_j, the probability that the onset with observed phase ϕ belongs to the j^{th} component of the mixture (i.e. a higher or lower metrical level). This can be calculated as (see also Large & Jones, 1999):

$$\tau_j = \frac{w_j f_j(\phi)}{f(\phi, \underline{\kappa})} \tag{6}$$

This gives the probability that the n^{th} event marks periodicity j, based on the amount of expectancy from oscillator j divided by the total expectancy across oscillators.

2.2.2. Perception of temporal differences

Once an onset has been associated with an attentional pulse, it is possible to explain the perception of temporal fluctuations. Temporal fluctuations are perceived in terms of the difference between an event onset time and the expected time as specified by the peak of an attentional pulse. For example, an event onset may be heard as early, on time, or late, with respect to an individual oscillation (phase, ϕ), and the salience of the deviation depends on the (focus, κ) of the expectancy function. According to our hypothesis, deviations from temporal expectations govern the listener's perception of the performer's musical intentions. In this section, we specify the model's perception of two types of temporal fluctuations: the perception of phrase structure that arises from phrase-final lengthening, and the perception of melody (primary musical part) that arises from the temporal asynchrony of a melody note relative to other notes of a chord.

We first investigate the model's ability to perceive phrase boundaries that are typically marked by large temporal fluctuations, i.e., phrase-final lengthening. We model this as a probability with two components. The first component is the probability that event n will be heard as deviating from its expected time, $P_{D(n)}$; the second component is the probability that event n is heard as occurring late in the cycle, $P_{L(n)}$. Both are shown in Fig. 3B. The product of the two components models the probability $P_{P(n)}$ that an onset n will be perceived as characteristic of phrase-final lengthening, often used by performers to mark phrase boundaries.[1]

$$P_{D(n)} = 2\int_{x=0}^{|\phi_n|} f(x,\ \kappa)\ dx \qquad P_{L(n)} = \int_{x=-0.5}^{\phi_n} f(x,\ \kappa)\ dx \qquad P_{P(n)} = P_{D(n)}P_{L(n)} \tag{7}$$

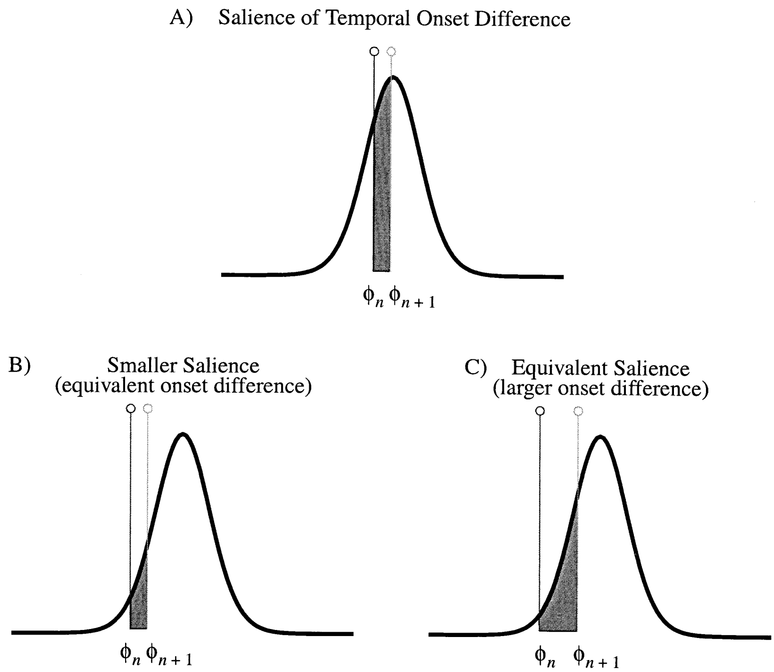

Fig. 4. A) Salience of perceived melody lead, based on modeled probability (shaded area under expectancy curve) of hearing a difference in onset time between two events. B) Smaller salience (less area under curve) results for equivalent onset difference located farther from peak expectancy; C) Equivalent salience (equal area under curve) results for larger onset difference, located farther from peak expectancy.

In other words, the probability that event n will be perceived as marking a phrase boundary, $P_{P(n)}$, has two components: One reflects the salience of a temporal deviation; the other reflects the directionality, or probability that the event is late. We use these probabilities to test the model's ability to perceive phrase-final lengthening in a range of temporally fluctuating performances in Experiment 1.

We next compare the model's ability to simulate the perception of small temporal differences among voice onsets that often coincide with performers' intentions to mark one voice within a chord as melody. We begin with the probability that the first note of a chord is perceived as earlier than the second note of a chord, where a chord is defined as those onsets associated with the same expectancy function. We operationalize this probability as the area under the expectancy curve from the first note to the second note of the chord, as shown in Fig. 4A for two tone onsets at times ϕ_n and ϕ_{n+1},

$$P_{A(n)} = \int_{x=\phi_n}^{\phi_{n+1}} f(x,\ \kappa)\ dx \tag{8}$$

in which onset n is the earliest onset associated with the current expectancy function. The area under the curve, $P_{A(n)}$, represents the salience of the time difference between the first tone onset and the second tone onset. Salience is relative to the expectancy function, because it is the area under the curve. Figs. 4A and 4B depict 2 tones with equivalent amounts of onset difference between them; the chord occurring closest to peak expectancy (4A) is predicted to be more salient. Figs. 4A and 4C depict 2 tones with equivalent salience; the tones occurring farthest from the peak expectancy (4C) require a larger onset difference to be equally salient. We use these probabilities to test the model's ability to perceive the melody in a variety of performances in Experiment 2. Thus, time differences are measured in terms of phase relative to an internal oscillation, and the salience of a time difference depends on amount of expectancy, quantified as a probability: the area under the expectancy function associated with the oscillation.

We examine the model's salience predictions for phrase-final lengthening and melody leads in piano performances in which phrase structure (Palmer & van de Sande, 1995) or melodic structure (Palmer, 1996a) were altered experimentally. Piano performances were collected on a computer-monitored acoustic piano, and the event timing of those temporally fluctuating performances provides a strict test of the model's performance. The model's perception of categorical durations, as well as temporal fluctuations, is systematically tested with performances containing large and small (or no) temporal fluctuations. Experiment 1 describes tests of the model's ability to perceive temporal regularity in performances of the same musical sequence with different phrase structures. Performances of contrapuntal music by J.S. Bach were chosen because they provide a moderate rubato context in which phrasal lengthening is especially salient (i.e., temporally disruptive) (Palmer & van de Sande, 1995). Experiment 2 describes tests of the model based on performances of different melodic structure. Performances of classical music by Beethoven were chosen because they provide a richer rubato context in which large melody leads (temporal asynchronies within chords) are observed (Palmer, 1996a).

3. Experiment 1: horizontal temporal fluctuations (rubato)

The first test of the model concerns the large temporal fluctuations or deviations from a regular beat or pulse in music performance, sometimes called rubato, which are often largest near phrase boundaries. Beat tracking in the presence of rubato provides a challenging test of the model's ability to adapt to a changing tempo. We draw from a study of music performance that examined the effects of phrase structure on temporal fluctuations in piano performances (Palmer & van de Sande, 1995). In this study, performances of polyphonic music by Bach (two- and three-part inventions) which contained multiple voices were collected on a computer-monitored acoustic piano. Pianists performed the same musical pieces in terms of three different phrase structures as marked in different versions of the music notation; in a control condition, there were no marked phrase boundaries. We contrast the model's ability to track in the presence and absence of large temporal fluctuations by comparison among these conditions. The temporal fluctuations in each performance of the different phrase conditions offer a strong test of the beat-tracking model because they contain many large deviations from expected event onsets: events performed two to four times slower than other events (Palmer & van de Sande, 1995). In addition, performances of the same music in which the entrance of one voice was delayed, were found to create larger temporal fluctuations (Palmer & van de Sande, 1995). We include those performances for comparison of the model's ability to track the beat in a variety of temporal fluctuations.

3.1. Methods

3.1.1. Stimuli

Piano performances of 2- and 3-part inventions by Bach, taken from Palmer & van de Sande (1995), provided tests of the model. Opening sections (approximately 3 measures) of two 2-part inventions (D-Major and D-minor) and one 3-part invention (B-flat Major) were used. The three inventions began on the first beat of the measure and contained two voices, composed predominantly of eighth-note and sixteenth-note durations. Each stimulus was presented to pianists with one of 3 different phrase structures marked in notation on each trial. In the fourth phrase condition, no phrase structure was marked on the notation and performers were instructed to apply their own phrase interpretation. Each piece was adapted to include two voice entrances: An additional version of each stimulus was created for each of the 4 phrase conditions, in which the entrance of the second voice occurred one-half measure earlier or later than in the original performance. Thus, there were 8 variants (4 phrase conditions and 2 voice entrances) for each of the three stimuli. The tempi of the 32 performances were moderate to fast; the mean quarter-note IOI was 448 ms (range = 344 ms–692 ms).

An example of one of the musical excerpts and phrasing instructions is shown in Fig. 5. Skilled adult pianists were instructed to practice each stimulus with its phrasing, presented in notation, and then to perform the excerpt from memory (see Palmer & van de Sande, 1995 for further details). The performances chosen for inclusion were based on two criteria: 1) only performances that contained no errors were included; and 2) within that constraint, the three pianists whose performances displayed the most temporal fluctuation and the three whose performances displayed the least were chosen, based on the standard deviations of the sixteenth-note interonset intervals in each performance. This created 144 performances (6 pianists $\times$ 4 phrase conditions $\times$ 2 voice entrances $\times$ 3 excerpts) in all. The amount of temporal fluctuation was computed as the proportion change in each interonset interval relative to the expected IOI, as estimated from the mean sixteenth-note IOI (the smallest notated duration) for each performance. Tempo proportions are shown in Fig. 5 for one of the performances; values greater than 1 indicate a lengthening of an event relative to the global tempo.

3.1.2. Apparatus

The pianists performed the excerpts on a computer-monitored Boesendorfer 290 SE acoustic concert grand piano, and event IOIs (interonset intervals) were collected by computer, with timing resolution of 1.25 ms.

3.1.3. Model simulation

The simulated oscillations tracked the sixteenth-note and eighth-note levels (2 smallest periodicities) of the metrical structure in the music performances. Thus, two oscillations tracked each performance, with a relative period of 2:1, reflecting the duple metrical

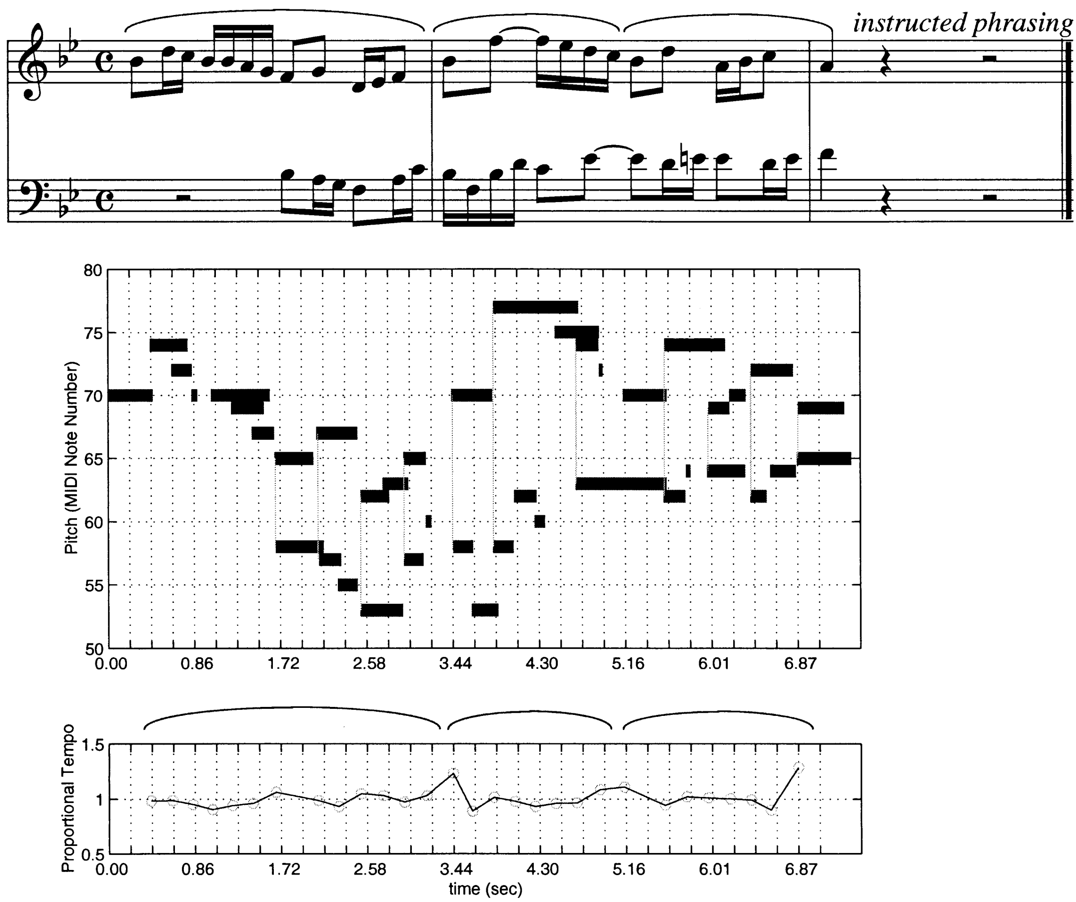

Fig. 5. Sample performance from Experiment 1 of 3-part invention in B-flat Major by J.S. Bach (top) shown with one of the instructed phrase structures, with piano roll notation of event onsets as performed (middle) and calculations of proportional tempo (bottom).

organization of the pieces at this level. Furthermore, the initial period of the sixteenth-note level oscillator was set to match the initial IOI in the performance at the sixteenth-note metrical level; the eighth-note oscillator period was double that of the sixteenth-note oscillator period. The initial phase of each oscillator was set to zero, and an initial value of $\kappa = 3$ was chosen for attentional focus (an intermediate value). Phase coupling strength, η_ϕ, was set to 1.0, period coupling, η_p, was set to 0.4, and the adaptation rate for focus, η_κ was set to 0.2. Simulations of both uncoupled ($\alpha_\phi = \alpha_p = 0$) and coupled ($\alpha_\phi = \alpha_p = 1$) oscillations were run.

Phase, period, and focus adapted as the two oscillations tracked the temporally fluctuating rhythms. The simulation produced a time series of phase, period, and focus values for each oscillator, with each value corresponding to a unique stimulus event. The success of beat-tracking was calculated from the phase time-series: the phase of each stimulus onset relative to the internal oscillations. Stimulus onsets were early $\phi < 0$, on time $\phi = 0$, or late $\phi > 0$, relative to the internal oscillation.

Finally, two measures were calculated for each note onset: *metrical category* and *salience of a temporal difference*. Each onset was categorized as marking either the smaller metrical level (16th-note period) or the larger metrical level (8th-note period), and associated with a particular pulse at that level (see Section II). Salience of the differences from categorical durations were based on the probability that an onset was perceived as a deviation ($P_{D(n)}$) and was perceived as late ($P_{L(n)}$), computed relative to the temporal expectancy function using the von Mises model. The product $P_{D(n)}P_{L(n)}$ gives the probability that the onset marked a phrase boundary, $P_{P(n)}$.

3.2. Results

We report the temporal fluctuations measured in each performance and the model's success in tracking the event onsets within each performance. Both the piano performance timing and the model's tracking performance were analyzed with circular statistics, which are appropriate for signals that contain circular (periodic) components.[2] Relative phase (ϕ) was used to measure both performance timing and the model's tracking performance. Relative phase refers here to the difference between an onset time and an expected time at a particular metrical level, normalized for cycle period (i.e. in angular units). For the performance timing, the normalizing period was the mean beat period at the metrical level of interest,[3] and expected times were computed for each event onset in the performance using the mean beat period. For the oscillators, the relative phase values are produced by Equation 2, so that the normalizing period was the period of the oscillator. Relative phase values ranged from $-.5$ to .5, with negative values indicating that an event occurred earlier than expected, and positive values indicating an event occurred late (zero indicates no difference or perfect synchrony of an oscillator with an event onset). Angular deviation, a measure of variability in relative phase analogous to standard deviation, was used to gauge both performance timing variability and overall oscillator tracking success. Angular deviation values range from 0 to .2241 ($= \sqrt{2}/(2\pi)$), where $0 =$ no variability in relative phase (consistent level of synchrony).[4]

3.2.1. Performances

The angular deviation measures had a mean value across performances of .0830, indicating moderate levels of variability. A repeated-measures analysis of variance (ANOVA) was conducted on the angular deviation measures for each performance by phrase condition (4), metrical level (2), and voice entrance (2), with events as repeated measures. The angular deviation measures were significantly greater at the smaller metrical level than the larger metrical level, ($F(1, 5) = 186.4$, $p < .01$), indicating that pianists used more expressive timing at the sixteenth-note level than the eighth-note level in these excerpts. There were no other significant effects.

The relative phase values for one of the performances are shown in Fig. 6 (top) for the 16th-note level (left) and 8th-note level (right). The points scattered around the circles in Fig.

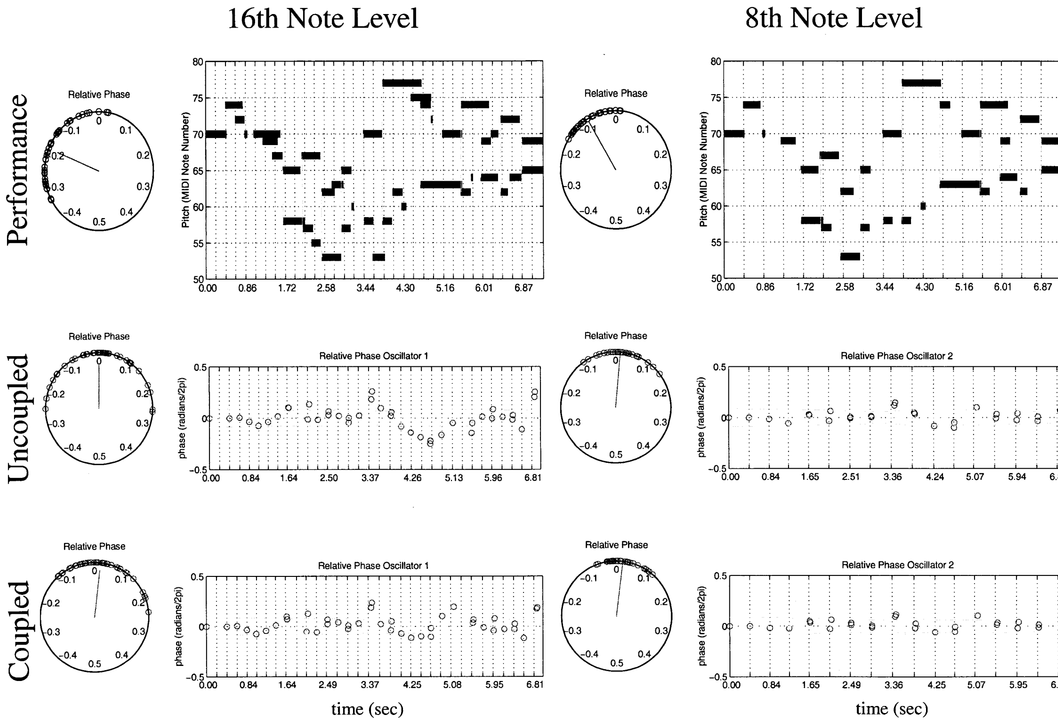

Fig. 6. Relative phase values of performance shown in Fig. 5 and model's relative phase values in Experiment 1. Relative phase values at 16th note level are shown in left column, 8th note level are shown in right column. Circular plots indicate relative values for individual events; spread around the circle indicates angular deviation. First row: relative phase values of performance events (relative to mean period). Middle row: relative phase values of oscillators when uncoupled. Bottom row: relative phase values of oscillators when coupled. For each plot, verticle grid lines indicate the beginning of the cycle relative to which relative phase was calculated. For performance statistics, the first cycle begins at $t = 0$, and average period (inverse of tempo) of each metrical level was used to project cycles forward. For oscillators, the time series of relative phase (ϕ_n) is plotted, and zero phase points were interpolated from the time series.

6 are the relative phase values for each note event; these relative phase plots indicate more angular deviation at the sixteenth-note level than the eighth-note levels. The phrase conditions that contained multiple notated phrases were further analyzed to examine whether the largest timing deviations coincided with notated phrase boundaries. An ANOVA on the relative phase measures for each performance by intended phrase boundary locations (the two locations adjacent to them were also coded as phrase boundaries) and non-boundary locations (remaining events) indicated that events on and around the notated phrase boundaries had larger relative phase values than the remaining events in each phrase ($F(1, 5) = 20.32$, $p < .01$). These nonlinear analyses confirmed Palmer and van de Sande's (1995) linear analyses that showed pianists significantly lengthened events at phrase boundaries relative to other events. Thus, the pianists used larger temporal fluctuations at notated phrase boundaries, also shown in Fig. 5, typical of phrase-final lengthening.

3.2.2. Model

The model's angular deviation measures of relative phase had a mean value across performances of .0801, smaller than the performance variability. This is exactly what one would expect if the oscillators were successfully adapting phase and period to track the ongoing sequences. Thus we conclude that tracking of these performances was good overall. A repeated-measures ANOVA was conducted on the angular deviation measures for each performance by phrase condition (4), metrical level (2), voice entrance (2), and coupling (coupled/uncoupled oscillators) with events as repeated measures. There was a significant effect of metrical level, $F(1, 5) = 61.9$, $p < .01$. Similar to the timing variability differences found in the performances, the smaller metrical level (sixteenth-notes) showed greater variability than the larger metrical level (eighth-notes). The relative phase values for the model are shown in Fig. 6 (bottom 2 rows), at both the 16th-note and 8th-note metrical levels. There was also an effect of phrase condition, $F(3, 15) = 4.1$, $p < .05$; the model tracked the beat better in the natural phrase condition (in which performers were not instructed as to phrase interpretation) than in the experimental phrase conditions. There were no differences in beat-tracking across the voice entrances or interactions; the model tracked the most variable and least variable performances equally well.

In addition, there was a significant effect of coupling, $F(1, 5) = 49.1$, $p < .01$; tracking by the coupled oscillators was better than by the uncoupled oscillators. Fig. 6 shows the oscillators' angular deviation around the relative phase circles, which is smaller in the bottom row (coupled model) than in the middle row (uncoupled model). The coupling advantage was present in the three least variable and the three most variable performances. There was also an interaction of coupling with metrical level, $F(1, 5) = 13.7$, $p < .05$; the coupled oscillator model consistently outperformed the uncoupled model, and more so at the smaller metrical level (the more variable level) than at the larger metrical level. This interaction is also shown in Fig. 6, in the bottom 4 panels. Thus, internal oscillator coupling aided beat-tracking, and more so at metrical levels that contained increased temporal variability. This last effect is what we would expect: Internal coupling propagated the phase adaptations from the higher-level oscillator down to the lower-level oscillator, improving tracking at the lower, more variable, levels.

3.2.3. Comparison of model and performance

We next test the model's ability to detect the large temporal fluctuations seen at phrase boundaries in music performance. The model's ability to detect phrase boundaries was measured as the probability, $P_{P(n)}$ of detecting late events, ranging from zero to one, and is shown for one performance in Fig. 7. A correlation analysis was conducted between the model's probability measures and the performance tempo measures for each event location in all performances except the first and last events; the correlation indicated a modest but significant relationship, $r = .34$, $p < .01$. The same correlation conducted on only the experimental conditions containing multiple phrase boundaries (the most challenging test of the model) indicated similar results, $r = .36$, $p < .01$. Thus, the model tended to detect delays relative to temporal expectation for events at which performers delayed the timing relative to notated categorical durations.

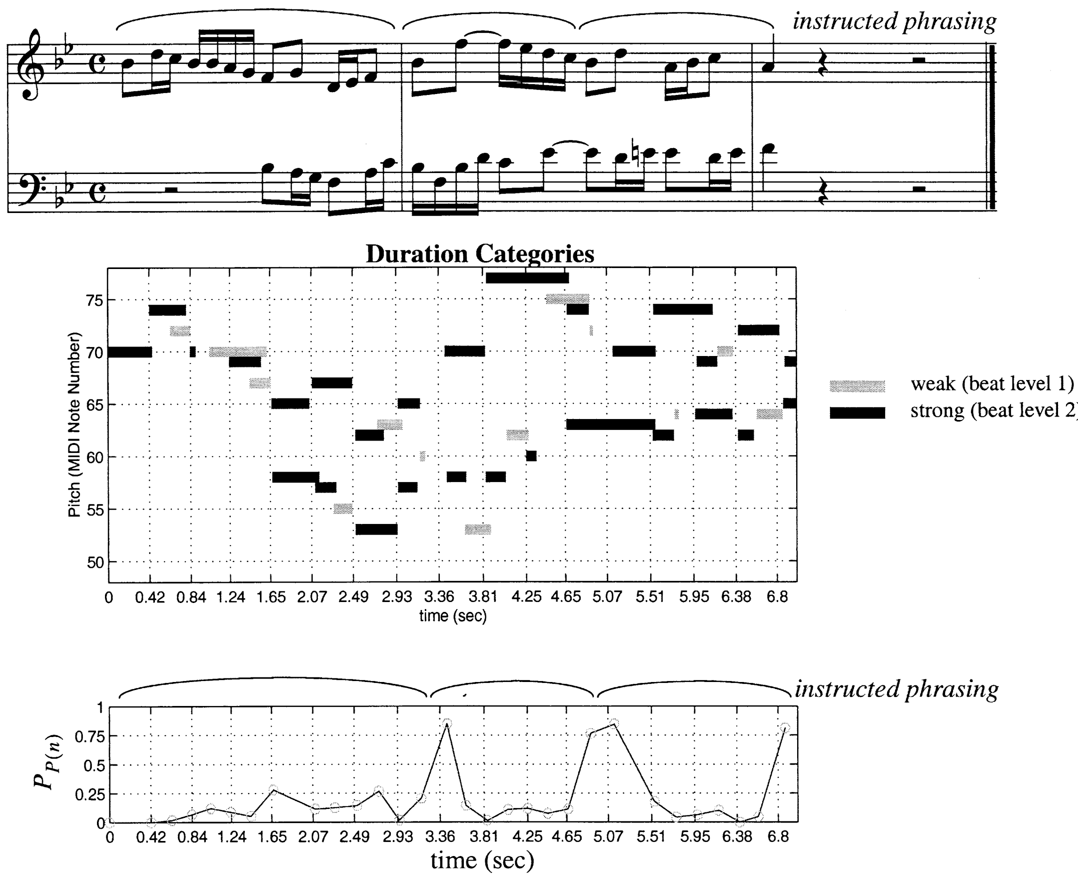

Fig. 7. Model's categorization of event durations and phrasal salience values for the performance shown in Fig. 5, from Experiment 1. Events categorized as 16th-notes (smaller beat level) shown in grey; events categorized as 8th-notes (larger beat level) shown in black (top). Model's probability of detecting phrasal lengthening at each event (bottom).

Next, we evaluated the model's categorical abilities to detect phrase boundaries. A criterion value of the 75th percentile was applied to both the model's probability measures and the performed tempo changes. Thus, events for which model probabilities were greater than .75, and events whose IOIs were greater than the 75th percentile of all performed events, were categorized as locations of lengthening. As before, event locations immediately surrounding notated phrase boundaries were considered part of the phrase boundary. Table 1 shows the number and column percentages of event locations that passed the lengthening criterion for the performances from the experimental phrase conditions (that contained multiple notated phrase boundaries) and the model's salience measures. Both the hit rate (upper left corner) and the correct rejection rate (lower right corner) were higher than expected by chance, as determined by the percentage of total events that were notated as phrase boundaries (binomial test, $p < .01$). A chi-squared test indicated a significant interaction between the model's phrase-detection and the performance lengthening, $\chi^2(1) =$

Table 1
Number of events passing lengthening criterion for performance and model

Model	Performance	
	>75%	<75%
>.75	234 (64%)	120 (12%)
<.75	121 (36%)	935 (88%)

417.3, $p < .01$. Thus, the model was able to detect lengthening more often than chance at locations where performers used lengthening; the fact that the correct rejection rate is greater than the hit rate may reflect the relatively modest amounts of rubato in these performances, typical of performances of Bach's polyphonic music, which drive the model's expectations.

3.3. Conclusions

This experiment provided the first test of a multiple oscillator model tracking temporally fluctuating, multivoiced music performances with high accuracy; the model's beat tracking variability was slightly lower than the amount of stimulus variability. In addition, the model's predictions of phrasal salience increased as performers' use of phrase-final lengthening increased; the correlation between model salience and performance timing indicates that the model's expectations were coordinated enough with the performance to adapt to the temporal fluctuations that marked phrase boundaries. The model's detection of those events likely to be phrase boundaries corresponded overall with those performance locations that contained the most rubato, indicating that the expectancy model can adapt successfully in the face of large temporal fluctuations typical of phrase-final lengthening. These findings demonstrate the plausibility of a perceptual principle—entrained, self-sustained oscillation—for identifying temporal regularity in musical performances, despite large temporal fluctuations. Furthermore, information conveyed by specific types of temporal fluctuations can also be extracted by such a system and used in a meaningful way, rather than simply being treated as noise to be eliminated.

The coupling of oscillators also improved the model's beat-tracking; most important, coupling aided beat-tracking most at metrical levels that contained the most temporal variability. Coupling represents the effect of one metrical level on another, such that oscillators that are tracking successfully can stabilize oscillators that are not tracking successfully. In this way, musical meter can be construed as a framework that generates predictions or expectations about events' relative timing. This perspective concurs with music-theoretic perspectives on Western tonal music that view the relative timing of musical events at least as crucial as the pitch contents of those events (Cooper & Meyer, 1960; Lerdahl & Jackendoff, 1983). A beat-tracking mechanism that relies on internal coupling of different periodicities is also consistent with psychological approaches in which the timing of individual sequence events constrains the timing of surrounding events, due to the hierarchical nature of metrical structure (Large & Jones, 1999; Martin, 1972; Vorberg & Wing, 1996).

4. Experiment 2: vertical temporal fluctuations (melody leads)

In Experiment 2, we address the problem of how listeners track smaller tempo fluctuations (20–50 ms) between individual voices in performance, that often correspond to performers' melodic intentions. We use piano performances of the same music with different melodic intentions from Palmer (1996a) to test further the model's adaptive abilities in the presence of temporal fluctuations. Melody leads provide a robust test of the beat-tracking model for several reasons. First, they provide a pervasive cue as to the interpretive intent of the performer that might enlighten us as to how processes of stream segregation and melody identification occur. Melody leads tend to be larger on metrically strong positions in piano performance (Palmer, 1989, 1996a); these small asynchronies may provide a cue to beat-tracking. Second, the 20–50 ms melody leads in the performances (about 3–6% of the IOI's) provide a more sensitive test of the model's reaction to temporal fluctuations than the larger phrasal lengthening patterns of Experiment 1 (about 200% of the IOI's). The perceptual salience of one voice, relative to other nearby voices as predicted by the model, depends on the amount of expectation that is active during the asynchrony. We compare the model's predicted salience values for each voice with the performed melody leads in performances of the same music with different melodic interpretations. The performances given to the model varied only in temporal cues; other performance cues, such as sustain pedalling (which can influence the perception of event offsets) and intensities, were removed.

Further tests of the model included edited performances in which melody leads were removed, but all other temporal fluctuations were retained. Comparisons of the model's beat-tracking abilities on performances with and without melody leads provide a test of the contribution of the melody leads versus other temporal fluctuations. To ascertain the role of individual voices on beat-tracking, the model was also tested on each voice separately. This comparison of beat-tracking in multivoiced music with the individual (monophonic) parts allows a robust test of whether additional temporal fluctuations added by multiple voices provide performance cues. The model is also presented with performances containing asynchronies created from random temporal fluctuations, to test whether any advantage of chord asynchronies is due to their systematic nature or simply to their temporal variability. Finally, we compare the model's ability to identify the melody with listeners' abilities. Palmer (1996a) reported listeners' ratings of the voice intended as melody for both performances that contained melody leads and for the same performances with melody leads removed. Pianist listeners correctly identified the melody more often when melody leads were present than when they were absent; their ratings provide a test of the model's salience predictions.

4.1. Methods

4.1.1. Stimuli

Performances of the theme (first 8 measures) from a piano Sonata in E-Major, mvmt 3, Opus 109 by Beethoven, were taken from Palmer (1996a). The opening section in 3/4 meter contains 3 voices composed predominantly of quarter-note durations. This excerpt was chosen because two voices could be interpreted as melody: the upper (highest frequency) or

Fig. 8. Opening section of Piano Sonata in E-Major, Opus 109, mvmt 3, by Beethoven, used in Experiment 2. Upper melody interpretation marked 'U'; Lower melody interpretation marked 'L'.

lower (lowest frequency) voice, as shown in Fig. 8. Two performances of each melody interpretation, performed by the same professional pianist, were included in the study (for more details, see Palmer, 1996a). Pedaling and event intensities were removed (the model does not respond to either cue). The tempi of the four performances were similar and slow (mean quarter-note IOI = 1449 ms, range = 1322 ms–1497 ms). In addition to the original performances, synchronous versions of each performance were synthesized by removing all chord asynchronies, setting non-melody chord tone onsets equal to melody tone onsets. Thus, the original (asynchronous) and synchronous versions retained the same tempo pattern of the melody; the synchronous versions had no melody leads, and the asynchronous versions retained the original melody leads.

Finally, four different voicing versions of the original and synchronous performances were created, in which each of the three voices appeared alone (voice 1 (highest-frequency voice), 2, or 3) or all three voices were retained. This allowed us to test effects of individual voices, which retained their original tempo patterns, on the model. The asynchronies associated with the arpeggiated chord in measure 5 (shown in Fig. 8), which necessarily creates a large temporal fluctuation but a fixed melody lead (the highest-frequency voice is performed last) were also removed from the synchronous performances. All voices within the arpeggiated chord in the synchronous performances were preserved in each of the voice conditions, to control for effects of the arpeggiated chord on voice effects. All other timing information (tempo changes, articulations) was constant across synchronous and asynchronous performances. Thus, there were 2 melody interpretations (upper and lower) $\times$ 2 repetitions $\times$ 2 asynchrony versions (original/synchronous) $\times$ 4 voice versions (voice 1, 2, 3, or all voices), yielding a total of 32 performances on which analyses were conducted.

4.1.2. Apparatus and procedure

A professional pianist from the Boston area performed the excerpts on the same computer-monitored piano as in Experiment 1. The pianist was shown the two melodic interpretations, notated U (upper) and L (lower) on the musical score as in Fig. 8, and was asked to perform the excerpt emphasizing the upper or lower voice as melody. In a second performance of each melody interpretation he was asked to perform the excerpt in an exaggerated fashion (to give extra emphasis to the notated melody interpretation). Thus, there were two repetitions of each melody interpretation, yielding four performances (for further details, see Palmer, 1996a).

4.2. Results

The model's beat-tracking performance was compared as before with temporal aspects of the piano performances. Interonset timing measures were computed as in Experiment 1, for each event in each voice. In addition, melody leads were computed (melody onset time minus mean onset of remaining chord tones) for each of the original and synchronous performances that contained all voices. Relative phase and angular deviations for the events in each voice were computed as in Experiment 1 for the quarter-note and dotted half-note metrical levels.

4.2.1. Performances

The mean angular deviation of the relative phase values was .1173, a relatively high deviation value (maximum = .22); thus, the performances were more variable than the Bach performances of Experiment 1, as expected for this musical composition in the Romantic style. This increased variability in relative phase for the Romantic composition is depicted in Fig. 9 (top) for one of the upper melody performances. An ANOVA was conducted on the angular deviation measures with events as repeated measures, and with the four performances treated as a random factor (all performances were performed by the same pianist to control for other stylistic differences). Independent variables included the presence of melody leads or asynchrony (asynchronous (original) or synchronous performances), voices (voice 1, 2, 3, or all voices), and metrical level (quarter-note or dotted half-note). There was a significant effect of asynchrony, $F(1, 48) = 4.4$, $p < .05$; angular deviation measures of timing variability were larger for the performances that retained the asynchronies, as expected. There was also a significant effect of metrical level, $F(1, 48) = 1083$, $p < .01$, with larger deviations at the lowest metrical level (quarter-note). There were no significant interactions of these factors.

4.2.2. Model

Parameters and initial conditions were set as before; relative period was chosen to be 3:1, reflecting the metrical organization of the current piece. The mean angular deviation measure of the model's relative phase values was .0996 across performances, a higher variability measure than was seen for the Bach performances, indicating more difficulty in tracking the Beethoven performances, as expected. An ANOVA on the angular deviation measures by asynchrony (2), voice (4), metrical level (2), and coupling (coupled/uncoupled oscillators)

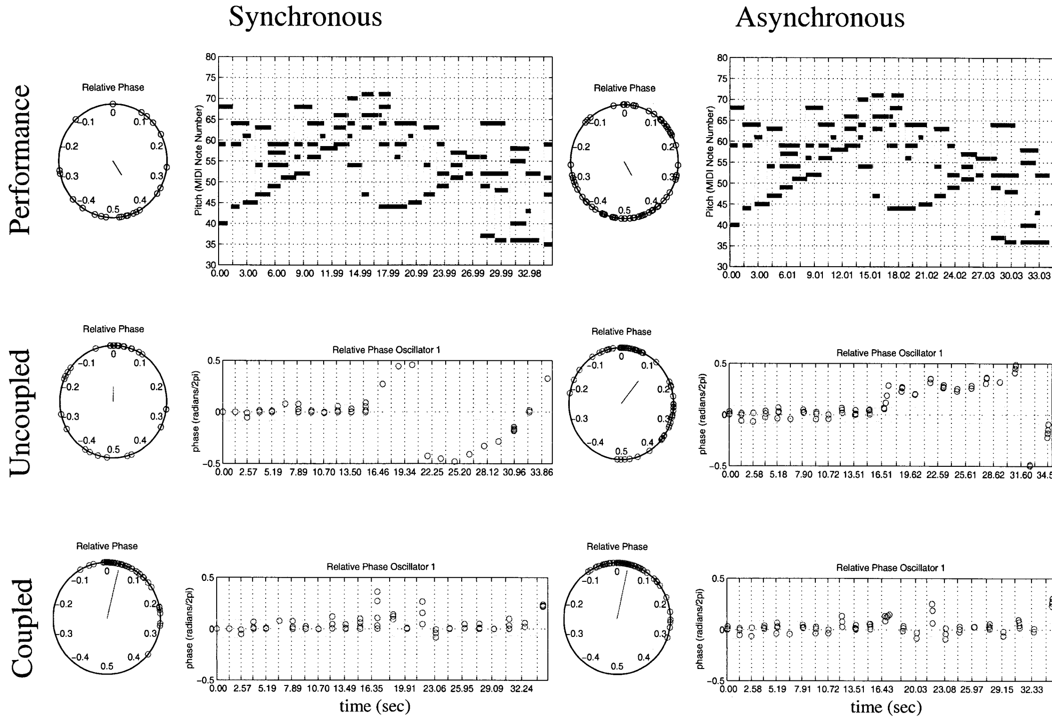

Fig. 9. Relative phase values for one of the performances in Experiment 2 (upper melody interpretation) at quarter-note level (lower beat level). Synchronous performances (with no melody leads) shown in left column, asynchronous performances (with original melody leads) shown in right column. Top row: relative phase values of performance (relative to mean tempo). Middle row: relative phase values of oscillators when uncoupled. Bottom row: relative phase values of oscillators when coupled. For each plot, verticle grid lines indicate the beginning of the cycle relative to which relative phase was calculated. For performance statistics, the first cycle begins at $t = 0$, and average period (inverse of tempo) of each metrical level was used to project cycles forward. For oscillators, the time series of relative phase (ϕ_n) is plotted, and zero phase points were interpolated from the time series.

with events as repeated measures and performances as the random factor indicated significant effects of asynchrony; the model's beat-tracking ability was more precise for performances that contained asynchronies than for synchronous performances, $F(1, 96) = 7.04$, $p < .01$. The presence of chord asynchronies improved the model's beat-tracking in all four of the original performances. Fig. 9 shows an example of the model's beat-tracking in the presence and absence of chord asynchronies at the quarter-note level in one of the upper-melody performance. There was also a significant effect of voices, $F(3, 96) = 5.3$, $p < .01$, with less variability in the presence of all voices and voice 2 alone (inner voice) than for other individual voices. There were no interactions among these factors. Thus, temporal variability associated with melody leads and between-voice differences aided beat-tracking.

There was also a significant effect of coupling, $F(1, 96) = 120$, $p < .01$; the coupled model displayed smaller phase variability than the uncoupled model. There was a significant effect of metrical levels, $F(1, 96) = 227.9$, $p < .01$, with larger angular deviations in phase at the lowest metrical level (quarter-note), which contained more temporal fluctuation in the performances. Finally, there was a significant interaction of coupling with metrical level, $F(1, 96) = 8.7$, $p < .01$; the coupled model outperformed the uncoupled model at both levels but more so at the lower (quarter-note) metrical level. Fig. 9 shows the improvement in the model's beat-tracking with coupling for one of the upper-melody performances; the middle row shows that both oscillators, when uncoupled, are unable to keep track of the beat, indicated by the phase-wrap beginning about halfway through the musical sequence and toward the end of the sequence. The bottom row shows the effects of coupling; both oscillators stay on track for the same sequence. Coupling allowed one oscillator to influence the relative phase values of another oscillator, thus reestablishing coordination after significant temporal perturbations.

The beat-tracking advantage observed for the asynchronous performances may have been due simply to the presence of variability in temporal onsets. To test whether the observed advantage could be a form of resonance, rather than as a result of some systematic relationship between chord asynchronies and rubato, we compared the model's ability to track the synchronous and asynchronous performances with its ability to track performances that contained random perturbations. The onsets within each chord event in the synchronous performances were perturbed with gaussian noise, with mean determined by the original chord onset time, and a standard deviation of either 10, 25, 50, or 75 ms. Thus, mean onset times remained approximately the same across the synchronous and random-noise performances, and the onset times of singleton events (i.e. notes that were not part of a chord) were unchanged. The same comparisons across musical voices, coupled and uncoupled oscillator models, and oscillator levels (periods) were made as before. An ANOVA was first conducted on the model's angular deviation measures across the asynchronous, synchronous, and level 10 random-noise performances. Angular deviation measures were significantly smaller for the asynchronous performances than for either the synchronous or the random-noise performances, $F(2, 144) = 5.1$, $p < .01$. Again, there was an advantage for the coupled model and for the oscillator with the larger period; coupling helped beat-tracking at both levels (F, $1, 144) = 211$, $p < .01$, and more so at the smaller level (that contained more variability), $F(1, 144) = 16.8$, $p < .01$. The same analysis repeated on the 25, 50, and 75 levels of random perturbations indicated the same significant advantage of asynchronous over random-fluctuation performances.

Systematic chord asynchronies may aid beat-tracking because they correlate with other performance features, such as rubato, that cause the oscillators to adjust their expectancies. For example, within a temporally extended chord onset, phase resetting is also extended, holding the phase of the oscillator near zero until the onset is complete. If this happens more often when the tempo is slowing, this should improve the ability to track a large change in tempo. To test this possibility, we correlated the amount of asynchrony measured by chord spread (difference between onset time of last note in chord minus onset time of first note) with the rubato measures for each chord. The correlation was modest but significant, $r = .31$, $p < .01$; the performances contained more temporal spread on chords that deviated more in tempo. Thus, the presence of systematic asynchronies, not simply variability of onset times, provided useful information for beat-tracking.

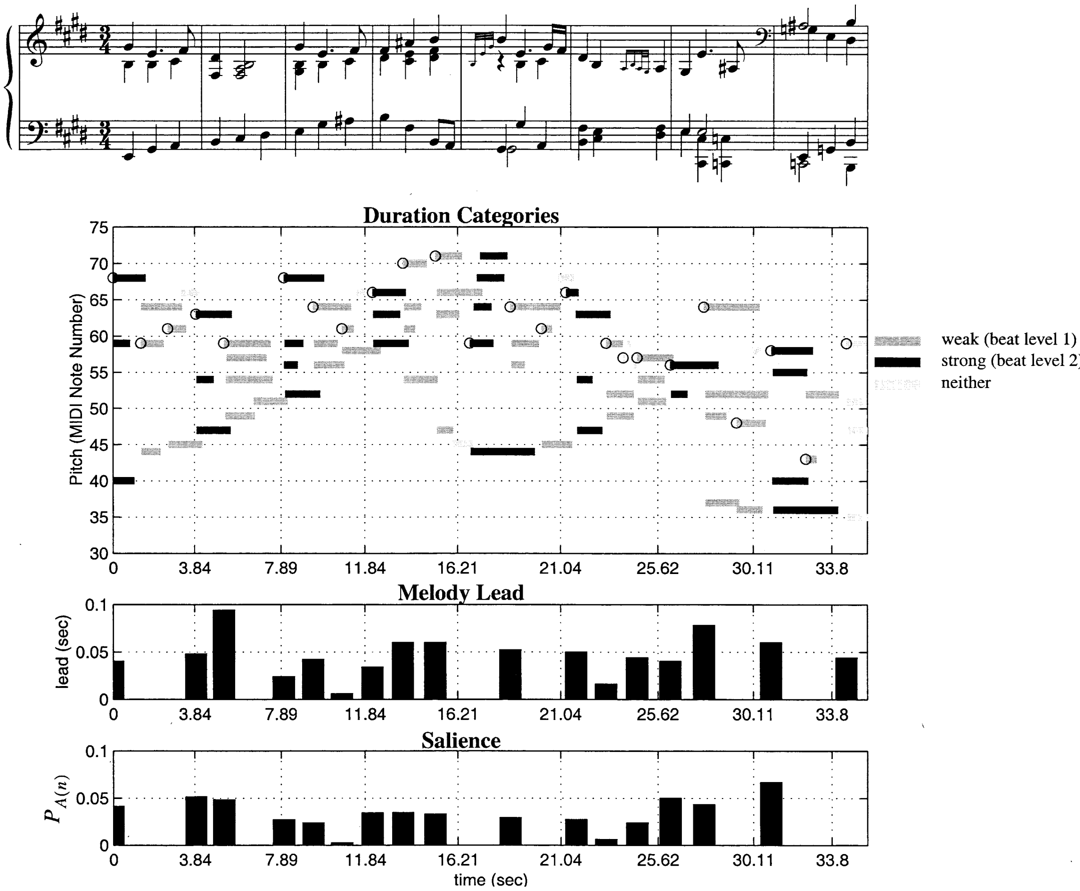

Fig. 10. Model's categorization of event durations and melody salience values for the performance shown in Fig. 9 from Experiment 2. Events categorized as quarter-note level (beat level 1) shown in grey; events categorized as dotted-half note level (beat level 2) shown in black (top). Amount of melody lead in performance (middle) and model's probability of detecting melody lead (bottom), P_A, shown by event.

4.2.3. Comparison of model and performance

We next test the model's ability to detect the melody leads in the piano performances. The perceived difference between melody and non-melody voices within a chord is measured by the area under the curve in the probability density function between two tone onsets associated with a given oscillator, $P_{A(n)}$. This area reflects the probability that the model will recognize the difference between those chord events. Fig. 10 shows the melody salience values for one of the upper melody performances; the circles indicate the first notes of each chord, which the model considers as melody.

Next, we test the prediction that size of the performed melody lead should correlate with the salience measures predicted by the model by correlating the size of melody leads at each event location (Fig. 10B) with the area under the probability density curve defined by those event onsets (Fig. 10C). According to the model's predictions, performers must increase the amount of asynchrony for events that occur farther from the expected onset to make them equally salient. The first event onset was the voice instructed as melody in 81% of all chords in the performances. The model's salience measures and the performance asynchronies were correlated across event locations and performances. The correlation was significant and positive, $r = .88$, $p < .01$, indicating that the model tracked these performances well enough to utilize the small asynchronies that cue melody interpretation. The same correlation was repeated after events for which the melody was not earliest were excluded (19% of all chords); this correlation was also significant, $r = .81$, $p < .01$. Performers tended to display more asynchrony of melody events, the farther away those events were from their expected onset.

4.3. Comparison of model and listener ratings

We next compare the model's melody saliences with listeners' melody ratings. Palmer (1996a) collected 8 musically trained listeners' ratings of which voice was intended as melody (upper or lower melody) for the asynchronous performances and the synchronous performances studied here, as well as additional performances that contained other cues (not examined here). Pedaling and intensity cues were removed or normalized in all performances. Each listener heard each performance twice, and was instructed to indicate which voice (upper or lower melody, notated on music notation) was intended by the performer as melody. We compare the proportion correct responses of pianist listeners (the only listeners whose responses improved significantly in the presence of melody leads in Palmer, 1996a) for the asynchronous performance minus the proportion correct for the synchronous performances. This difference score (from -1 to 1) adjusts for any residual performance cues that may have influenced listeners' responses.

To generate the model's predictions of melody, we calculated the total salience of leads for the voice intended as melody (upper or lower) and the total salience of leads in the non-intended voice (lower or upper). The totals were normalized so that they summed to 1. The normalized value for the intended melody is taken as proportion correct, i.e. the proportion of time the model would correctly choose the intended melody. To produce a difference score for comparison with the listener data, we subtracted from the intended melody value the model's proportion correct score for the synchronous versions, which had no melody leads (which always = .5 in the absence of melody leads). The proportion correct (PC) for listeners and model salience values are shown in Fig. 11. Overall, the model did well when the listeners did well, and the model failed when the listeners failed; however, the model outperformed the listeners somewhat. Although this analysis reflects a small number of performances and further tests are warranted, the similarity between model predictions and listener ratings supports the conclusion that the asynchronies are perceptually useful.

4.4. Conclusions

This experiment confirmed the earlier finding that the oscillator model can track music performances with high accuracy, even in the presence of more extreme temporal fluctuations (rubato) than those of Experiment 1. In addition, the model tracked better in the presence of the chord asynchronies than in their absence, suggesting that even small amounts

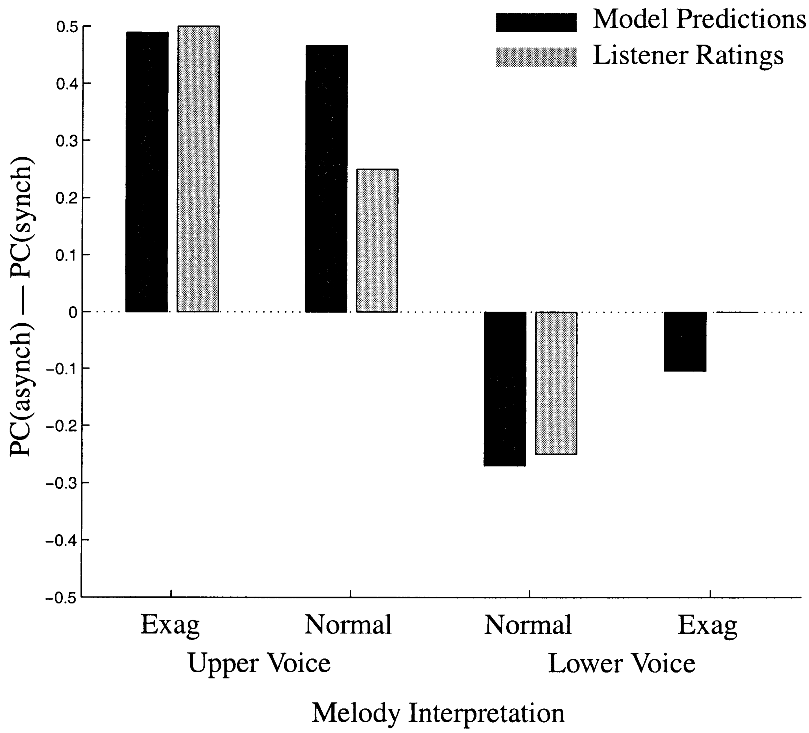

Fig. 11. Model predictions and listener ratings for an experiment (Palmer, 1996a) in which listeners were asked to identify the intended melody for the Beethoven performances.

of temporal variability (30–50 ms) can be perceptually informative. The model's measure of melodic salience increased as performer's use of melody leads increased; the correlation between model and performance timing indicates that the model tracked well enough to pick up the melody leads that can communicate information about musical structure. Furthermore, random temporal fluctuations in tone onsets did not aid the model; only systematic asynchronies enabled the phase-resetting to adapt to tempo change. These results suggest that chord asynchronies can carry information about meter as well as melody, consistent with findings that melody leads are often larger on strong metrical beats (Palmer, 1989, 1996a). Finally, comparison of the model's predictions of melodic salience with listeners' melody identification judgments from Palmer (1996a) suggest also that temporal fluctuations as small as chord asynchronies are informative, and that perception of temporally fluctuating performances is based on expectancies that change in response to the performance.

5. General discussion

We have described a perceptual model that addresses both the categorization of temporally continuous intervals and sensitivity to the temporal fluctuations from those categories found in musical sequences. Our approach assumes that people perceive rhythmic regularity in temporally fluctuating signals, in terms of the activity of a small system of internal oscillations. The psychological persistence of idealized music-theoretic beats is modeled as internal self-sustained oscillations, and the perception of metrical structure is captured by multiple internal oscillations operating at different periods. When driven with a fluctuating musical rhythm, the system of oscillations adapts to the rhythm. It is this adaptive property that accounts for listeners' perception of temporal regularity amidst the temporal fluctuations of performance, a form of perceptual constancy.

These experiments are the first to document the success of a multiple-oscillator model in tracking events in the context of real (human) multi-voiced music performances. A small network of oscillations was able to sustain coordination with temporally fluctuating music performances in Experiment 1, reestablishing coordination after significant temporal perturbations such as those established by phrasal lengthening. The experiments also demonstrated that internal coupling among oscillators improves tracking, particularly at temporally variable metrical levels. One reason for this improvement is that reduced variability at one metrical level allowed oscillations at that level to stabilize the tracking at more variable metrical levels through coupling among oscillators, as was also observed by Large and Jones (1999). We have shown further that even small amounts of temporal fluctuations can improve beat-tracking. In Experiment 2, tracking deteriorated when chord asynchronies were artificially removed from piano performances, indicating that small temporal fluctuations (on the order of 20–50 ms) aid temporal tracking in the face of significant rubato. The tendency for performers to introduce more asynchrony at locations of changing tempo offers a reason for why the model's relative phase measures improved for the systematic chord asynchronies but not for the random asynchronies. Within each chord, many small phase resets can act to hold the phase of the oscillator near zero for the duration of a prolonged (chord) onset. Furthermore, tracking of multiple voices was as good as or better than tracking of individual voices in both asynchronous and synchronous performances, suggesting that temporal fluctuations between voices as well as within voices provide useful perceptual information.

The perceptual salience of performers' phrase structure and melodic intentions was formalized in terms of peak expectancies in an attentional pulse. The notions of a temporal receptive field and attentional pulse are combined in this paper, to predict when listeners are most sensitive to a temporal fluctuation. The model's salience measures correlated strongly with music performance fluctuations. Thus, temporal fluctuations from categorical durations reflect performers' structural intentions and provide meaningful perceptual information; these variations are useful, and not simply noise. Such a view of timing in music performance parallels recent theoretical approaches to speech and music perception which treat stimulus variability not as noise to be normalized, but as information-carrying (Palmer, Jungers, & Jusczyk, 2001; Pisoni, 1997).

Finally, the model categorizes the temporally fluctuating event intervals, associating each note onset with an expectancy pulse at some metrical level. Categorization entailed association of note onsets with strong and weak beats of the metrical structure, as well as a grouping of onsets into perceptual simultaneities (chords). These processes are requisite for listeners to be able to recognize the intentions of performers, who use temporal fluctuations to communicate musical interpretations. As far as we are aware, no other models have been proposed to date that recover duration categories from complex performances.

5.1. Model limitations and future directions

Despite its successes, the model that we have proposed here has some limitations. The most significant is the importance of initial conditions. The initial phases and periods of the oscillators were chosen *a priori*, based on knowledge of the sequence; also, internal coupling parameters were chosen to reflect knowledge of the metrical structure. Thus, the model has the ability to track temporally fluctuating rhythms, but metrical structure and initial beat period cannot be inferred by this model. Another limitation is the fact that this model is driven by discrete-time input such as the MIDI performance recordings used here. This general approach has been criticized for its discrete-time formulation (Scheirer, 1998), that is, as system of discrete-time maps rather than continuous-time differential equations. The criticism is that driving the model with (more realistic) digitally recorded audio signals would require preprocessing of a continuous signal that was sophisticated enough to extract information comparable in quality to MIDI recordings.

Large (2000a) addressed these issues, describing a network of Hopf oscillators for inducing metrical structure, formulated as a system of continuous-time differential equations. The component oscillators of the network have a phase dynamics similar to the current model, but they have also amplitude dynamics. Oscillators compete for activation in the amplitude dimension, and in the end only a few oscillations remain active, embodying the metrical structure of the rhythm. In the current model, a mathematical simplification of Large (2000a), the phase dynamics are a straightforward discretization of the continuous phase dynamics, and the amplitude dynamics are replaced with the assumption that the period of each oscillation can adapt smoothly in response to tempo changes. Thus, the specification of initial phases, periods, and internal coupling is not so much a theoretical problem as it is a limitation imposed by the style of simulation that was chosen here.

A discrete-time formulation was chosen because it offers several advantages compared to its continuous-time cousin. The discrete-time model presented here relied only on onset times, not pitch or amplitude; it did not require all of the information available in acoustic recordings. This limited information is easily recoverable from the types of preprocessed signals used as input to the continuous-time models of Scheirer (1998) or Large (2000a); thus, the choice of continuous or discrete input is not as important for the current study. Second, the ability to work in discrete time with MIDI recordings has the advantage that future modeling of meter perception can make use of the additional continuous information available to the auditory system, without first solving the equally difficult problem of how the auditory system resolves such information. Finally, the continuous-time models of Scheirer (1998) and Large (2000a) have not been applied yet to time discrimination. By contrast, Large & Jones's (1999) discrete approach to time discrimination has been successful in capturing time discrimination behavior and is quite straightforward within a discrete-time framework. The model we described here extends the discrete-time framework to temporal fluctuations that occur naturally in music performance, capturing a level of temporal communication beyond what has been accomplished with other models.

5.2. Temporal categorization and attentional constraints

Some studies suggest that the perception of rhythmic patterns within a metrical context exhibits certain features of categorical perception, including abrupt category boundaries and nonmonotonic discrimination functions (Clarke, 1987). When asked to categorize the ratio of the final two time intervals of a sequence, listeners categorized ambiguous duration ratios (between 1:1 and 2:1) as 2:1 in the context of triple meter, whereas these same ratios were likely to be categorized as 1:1 in the context of duple meter. In a discrimination task, Clarke (1987) discovered nonmonotonic discrimination functions with single peaks at category boundaries, providing evidence for categorical perception. Schulze (1989) also found non-monotonic discrimination functions outside of a metrical context. However, these were not the single-peaked discrimination functions of classic categorical perception; they contained multiple peaks. In addition, Clarke (1987) found excellent within-category discrimination, much stronger than is classically associated with categorical perception (Liberman et al., 1957). These results suggest that some sort of perceptual categorization takes place within a rhythmic context, but the phenomenon is more complex than what has traditionally called "categorical perception."

Clarke (1987) suggested that two processes operate in rhythm perception: one assigns events to duration categories depending on the metrical context, while another interprets deviations from category durations as information-bearing. In order to perceive information in temporal deviations from notated durations, listeners must somehow be able to perceive the durational categories relative to the meter. This interpretation possesses a certain circularity, however. Perceived temporal deviations influence the perception of metrical structure, while metrical structure influences the perception of temporal deviations. How does metrical structure subserve both categorization and discrimination? Which temporal fluctuations force adaptation or structural reinterpretation of those categories?

Our model addresses these questions in a theoretical framework in which meter reflects the operation of a small system of self-sustained oscillations guiding the perception of event durations. This enables categorization of input events through association with specific points in the metrical hierarchy. It is equivalent to categorizing durations, where the categories are provided by the metrical context. Thus, this process provides the information necessary for notating a rhythm according to conventional Western notation. Generalizing the model explored here, Large (2000a) has addressed the issue of how different metrical hierarchies are formed, and evidence from rhythm categorization (Clarke, 1987; Large, 2000b) supports this account.

Two basic properties of this class of dynamical systems support the observed phenomena. First, the self-sustained oscillation exists independently of the signal that originally activated it. Thus, it can function as a generator of temporal expectations that may sometimes contrast with the timing of events in a stimulus pattern. This results in an expectancy violation (Jones, 1976; Meyer, 1956) that may be exploited in musical communication. Second, a self-sustained oscillation entrains to an external signal, so that even as it registers expectancy violations, is also adapts to changes in the exogenous rhythm, modifying future expectations to better conform with what has been experienced in the past.

The mechanism presented here for rhythm perception may not be specific to music.

Similar descriptions of meter have been advanced by linguists and music theorists (e.g. Hayes, 1984; Lerdahl & Jackendoff, 1983; Liberman & Prince, 1977; Selkirk, 1984; Yeston, 1976), in which direct analogies are often made between the rhythmic organization of speech and music. Simple categorical distinctions among timing units in language (e.g. "stress" versus "syllable" timing; Abercrombie, 1967; Pike, 1945) have not received strong empirical support (Hoequist, 1983; Roach, 1982); similarly, timing in music is significantly more complex and flexible than is commonly assumed. It is remarkable that listeners are able to perceive durational categories corresponding to the eighth-notes, quarter-notes, half-notes, and so forth, of musical notation because the actual durations measured in music performance deviate greatly from notated categorical durations (Clarke, 1987; Longuet-Higgins & Lee, 1982). Temporal fluctuations are commonly observed in speech as well, where they are often referred to as time-warping. These temporal perturbations in speech and music can communicate information about various types of structure (Lehiste, 1977; Price et al., 1991; Palmer, 1989; Shaffer, Clarke & N. Todd, 1985). Transient stimulus fluctuations can signal variations on thematic content (given/new distinctions), mark the boundaries of structural units, and communicate affect.

We view the perception of meter as a particular case of the operation of a general attentional mechanism (Large & Jones, 1999). The mechanism as modeled here generates expectancies, is selective for events happening near expected time points (i.e. melody leads), and exhibits in its adaptation a form of attentional capture, adjusting expectations to reflect changes in the stimulus (Large & Jones, 1999). This interpretation leads to additional predictions for auditory attention; for example, events occurring at strongly expected times should have a perceptual advantage, a prediction that has already received empirical support (Palmer & Kelly, 1992; Palmer & Krumhansl, 1990; Jones et al. 1988). Furthermore, many other forms of activity display significant structure in the temporal domain (e.g. Johansson, 1973) that present opportunities for attentional engagement based on temporal synchrony. Thus, musical rhythm may capture the attention of listeners in a compelling way that reflects the nature of attentional processes. The study of rhythm may inform us of one of the most basic acts of human behavior, the act of attending.

Notes

1. This probability is a conjunction of two components, one corresponding to the deviation from expectancy, and the other corresponding to the direction of the deviation. Note that neither integral by itself yields the desired probability.
2. Analyses based on linear statistics yielded the same findings as the circular statistics, and therefore only the circular statistics are reported.
3. The same relative phase value will correspond to a smaller time difference at a higher metrical than at a lower metrical level, because the higher level is normalized to a larger beat period. We report relative phase values at each metrical level to facilitate comparisons across performances and model.
4. Because phase was defined for the performances relative to a hypothetical period equal to the mean IOI for events at a particular metrical level, the mean (and individual)

relative phase values for the piano performances are not directly comparable to those of the model. However, the angular deviations of these values are on the same scale as those of the model and thus are directly comparable.

Acknowledgments

This research was partially supported by NSF Grant SBR-9808446 to the first author and by NIMH Grant R01-45764 to the second author. Thanks to Steven Finney, Armin Fuchs, Mari Riess Jones, Melissa Jungers, Rosalee Meyer, Peter Pfordresher, Betty Tuller, and two anonymous reviewers for comments on an earlier draft and to Zeb Highben for their help in preparing this manuscript.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh Univ. Press.

Apel, W. (1972). *Harvard dictionary of music* (2nd ed.). Cambridge, MA: Belknap Press of Harvard University Press.

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

Bengtsson, I., & Gabrielsson, A. (1983). Analysis and synthesis of musical rhythm. In J. Sundberg, *Studies of music performance* (pp. 27–60). Stockholm: Royal Swedish Academy of Music.

Bregman, A. S. (1990). *Auditory scene analysis: the perceptual organization of sound*. Cambridge, MA: MIT Press.

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

Clarke, E. F. (1982). Timing in the perfomrance of Erik Satie's Vexations. *Acta Psychologia, 30,* 1–19.

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West, *Musical structure and cognition*. London: Academic Press.

Clarke, E. F. (1987). Categorical rhythm perception: an ecological perspective, In A. Gabrielsson, *Action and perception in rhythm and music* (pp. 19–33). The Royal Swedish Academy of Music, 55.

Clarke, E. F. (1988). Generative principles in music performance. In J. A. Sloboda, *Generative processes in music: the psychology of performance, improvisation, and composition* (pp. 1–26). New York: Oxford University Press.

Clarke, E. F. (1989). The perception of expressive timing in music. *Psychological Research, 51,* 2–9.

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

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

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception, 10,* 343–378.

Gabrielsson, A. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology, 15,* 63–72.

Gabrielsson, A. (1987). Once again: the theme from Mozart's piano sonata in A Major (K 331). In A. Gabrielsson, *Action and perception in rhythm and music* (pp. 81–104). Stockholm: Royal Swedish Academy of Music.

Garner, W. R., & Gottwald, R. L. (1968). The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology, 20,* 97–109.

Glass, L., & Mackey, M. C. (1988). *From clocks to chaos: the rhythms of life*. Princeton, NJ: Princeton University Press.

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

Hayes, B. (1984). The phonology of rhythm in English. *Linguitic Inquiry, 15,* 33–74.

Henderson, M. T. (1936). Rhythmic organization in artistic piano performance. In C. E. Seashore, *Objective* *analysis of musical performance*, University of Iowa Studies in the Psychology of Music IV (pp. 281–305). Iowa City: University of Iowa Press.

Hoequist, C. (1983). Syllable duration in stress-, syllable- and mora-timed languages. *Phonetica, 40,* 203–237.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14,* 210–211.

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., & Yee, W. (1997). Sensitivity to time change: The role of context and skill. Manuscript in press, *Journal of Experimental Psychology: Human Perception & Performance, 23,* 693–709.

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

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

Large, E. W. (1996). *Modeling beat perception with a nonlinear oscillator*. In Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society.

Large, E. W. (2000a). On synchronizing movements to music. *Human Movement Science, 19,* 527–566.

Large, E. W. (2000b). Rhythm categorization in context. In *Proceedings of the International Conference on Music Perception and Cognition,* August.

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

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics, 5,* 253–263.

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

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54,* 358–368.

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

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

Martin, J. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review, 79,* 487–509.

McAuley, J. D., & Kidd, G. R. (1995). Temporally directed attending in the discrimination of tempo: further evidence for an entrainment model. *Journal of Acoustical Society of America, 97* (5), 3278.

Meyer, L. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Nakamura, (1987). The communication of dynamics between musicians and listeners through musical performance. *Perception and Psychophysics, 41,* 525–533.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science, 5,* 42–46.

Palmer, C., & van de Sande, C. (1995). Range of planning in music performance. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 947–962.

Palmer, C., & Kelly, M. H. (1992). Linguistic prosody and musical meter in song. *Journal of Memory and Language*.

Palmer, C. (1988). Timing in skilled music performance. Unpublished doctoral dissertation, Cornell University, Ithaca, NY.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception & Performance, 15,* 331–346.

Palmer, C. (1996a). On the assignment of structure in music performance. *Music Perception, 14,* 23–56.

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

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

Palmer, C., Jungers, M. K., & Jusczyk, P. W. (2010). Episodic memory for musical prosody. To appear in *Journal of Memory and Language*.

Pike, K. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix, *Talker variability in speech processing* (pp. 9–32). San Diego: Academic Press.

Povel, D., & Okkerman, H. (1981). Accents in equitone sequences. *Perception & Psychophysics, 7,* 565–572.

Povel, J. D. (1981). Internal representations of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 3–18.

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

Randel, D. M. (1986). *The new Harvard dictionary of music*. Cambridge, MA: Harvard Univ Press.

Repp, B. H. (1990). Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists. *Journal of the Acoustical Society of America, 88,* 622–641.

Repp, B. H. (1992). Probing the cognitive representation of musical time: structural constraints on the perception of timing perturbations. *Cognition, 44,* 241–281.

Repp, B. H. (1996). Patterns of note onset asynchronies in expressive piano performance. *Journal of the Acoustical Society of America, 100,* 3917–3932.

Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal, *Linguistic controversies: essays in linguistic theory and practice in honour of F. R. Palmer*. London: Arnold.

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

Schulze, H. H. (1989). Categorical perception of rhythmic patterns. *Psychological Research, 51,* 10–15.

Selkirk, E. (1980). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Shaffer, L. H., Clarke, E., & Todd, N. P. M. (1985). Metre and rhythm in piano playing. *Cognition, 20,* 61–77.

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

Sloboda, J. A. (1985). *The musical mind*. Oxford: Oxford University Press.

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

Todd, N. P. M. (1985). A model of expressive timing in tonal music. *Music Perception, 3,* 33–59.

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

Vernon, L. N. (1936). Synchronization of chords in artistics piano music. In H. Seashore, *Objective analysis of musical performance* (pp. 306–345). Iowa City: University of Iowa Press.

Vorberg, D., & Wing, A. (1996). Modeling variability and dependence in timing. In H. Heuer & S. W. Keele, *Handbook of Perception and Action, Volume 2: Motor Skills*. (pp. 181–262). London: Academic Press.

Woodrow, H. (1932). The effects of rate of sequences upon the accuracy of synchronization. *Journal of Experimental Psychology, 15,* 357–379.

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