Music Perception
Summer 1994, Vol. 11, No. 4, 409–464

A Perceptual Model of Pulse Salience and Metrical Accen in Musical Rhythms

RICHARD PARNCUTT
McGill University

In Experiment 1, six cyclically repeating interonset interval patterns (1, 2:1, 2:1:1, 3:2:1, 3:1:2, and 2:1:1:2) were each presented at six different note rates (very slow to very fast). Each trial began at a random point in the rhythmic cycle. Listeners were asked to tap along with the underlying beat or pulse. The number of times a given pulse (period, phase) was selected was taken as a measure of its perceptual salience. Responses gravitated toward a moderate pulse period of about 700 ms. At faster tempi, taps coincided more often with events followed by longer interonset intervals. In Experiment 2, listeners heard the same set of rhythmic patterns, plus a single sound in a different timbre, and were asked whether the extra sound fell on or off the beat. The position of the downbeat was found to be quite ambiguous.

A quantitative model was developed from the following assumptions. The phenomenal accent of an event depends on the interonset interval that follows it, saturating for interonset intervals greater than about 1 s. The salience of a pulse sensation depends on the number of events matching a hypothetical isochronous template, and on the period of the template—pulse sensations are most salient in the vicinity of roughly 100 events per minute (moderate tempo). The metrical accent of an event depends on the saliences of pulse sensations including that event.

Calculated pulse saliences and metrical accents according to the model agree well with experimental results ($r > 0.85$). The model may be extended to cover perceived meter, perceptible subdivisions of a beat, categorical perception, expressive timing, temporal precision and discrimination, and primacy/recency effects. The sensation of pulse may be the essential factor distinguishing musical rhythm from nonrhythm.

Introduction

Imagine that you are walking down a quiet city alley and enter a jazz club. The door opens and suddenly you hear the music. In just a second or two you have a strong impression of the "feel" of the music—in particular, the way it "swings"—its "beat" or "pulse." If instead you had been confronted with some Ghanaian percussion music, or a disco number, or Ravel's *Bolero,* or a movement of a Bach Motet driving inexorably forward in 3/4 time toward a glorious climax, you would have received an entirely different qualitative impression of the music's beat. But the experience would have been similar in two respects: the strength of the impression of beat (swing, pulse), and the remarkably short time required to perceive it.

Yeston (1976) referred to an isochronous sequence of similar-sounding events as a *rhythmic stratum,* or *level of motion* (also called *time level* by Jones & Boltz, 1989, *level of pulsation* by Palmer & Krumhansl, 1990, and *rhythmic level* by Rosenthal, 1992). In Yeston's theory, a rhythmic level may be defined by note onsets, accents of all kinds (dynamic, melodic, harmonic, timbral, textural, phrasal), and repetition of a temporal pattern. Integral to the theory is the idea that a single rhythmic sequence can evoke several rhythmic levels at the same time. A listener may focus attention on any one of these, or switch attention from one to another at will (Jones, Boltz, & Kidd, 1982; Jones, Kidd, & Wetzel, 1981).

Yeston divided the various rhythmic levels evoked by a rhythmic sequence into one, fastest level or *pulse level,* and slower (higher) *interpretative levels.* Interpretative levels whose periodicity is slower than that of the barlines are referred to as *hypermeters* by Rothstein (1989); straddling the conventional borderline between rhythm and form, they play an important role in most tonal-rhythmic music. In the present study, the term *pulse sensation* is used to describe *all* rhythmic levels spontaneously evoked in the mind of the listener. The term pulse sensation is intended as a blanket term for "beat," "swing," "rhythmic level," and so on.

The widespread use of *rubato* (especially in Western classical music) indicates that a musical beat need not be exactly isochronous, or equally spaced in time, to be strongly felt (Shaffer, 1981). In general, however, stricter timing tends to induce a stronger feeling of pulse, and the strength of the beat evoked by a sequence of sounds tends to decrease as the sequence deviates from isochrony (cf. Ehrlich, 1958). Thus, African percussion music (as described by Agawu, 1987; Jones, 1959; Koetting, 1986; Rowlands, 1991) generally evokes a stronger feeling of pulse than does European classical music.

Rhythmically organized auditory patterns are easier to encode, recall, and reproduce than similarly organized visual patterns (Garner & Gottwald, 1968; Glenberg, Mann, Alman, Forman, & Procise, 1989; Handel & Buffardi, 1968). A feasible explanation is that presentation times are encoded more accurately in the auditory case (Glenberg & Swanson, 1986). This may in turn be due to the confinement of the sensation of beat or pulse to the auditory modality (as demonstrated, e.g., by Grant &

Dedicated to Prof.Dr.-Ing. E. Terhardt on the occasion of his 60th birthday.

Requests for reprints may be sent to Richard Parncutt, Faculty of Music, McGill University, 555 Sherbrooke West, Montreal, Quebec, Canada H3A 1E3.

LeCroy, 1986). Perhaps auditory rhythms can be perceived by encodi their constituent events relative to an underlying musical beat or pulse, b visual "rhythms" cannot. This view would be consistent with observatio of Glenberg and Jona (1991) and Schab and Crowder (1989) that t advantage of auditory over visual rhythms diminishes for tasks involvi relatively long durations, because the perceptual salience of pulse sens tions diminishes as their period increases from about 600 to 2000 (Fraisse, 1982).

Schulze (1978) investigated the sensitivity of listeners to various kin of timing deviation in isochronous sequences. His results suggested th the listeners synchronized an *internal time-keeper* with the sequence, d tecting irregularities by monitoring discrepancies between expected an actual event times. Wing and Kristofferson (1973a, 1973b) applied th concept of a timekeeper to a model of periodic motor timing. Shaff (1981) and Shaffer, Clarke, and Todd (1985) analyzed the timing of perfo mances of piano music and concluded that timing was constrained by central timekeeper capable of adjusting its rate in response to expressiv features of the music (rubato). Povel (1984) hypothesized that listene match an isochronic *temporal grid* (called a *template* in the present mode to a rhythmic sequence, choosing the grid in such a way as to direct match a maximal number of tones and to account for the other tones wit the greatest ease (principle of economy). Povel and Essens (1985) an Drake and Gérard (1989) assumed that listeners generate an *internal cloc* or *time base* while listening to a rhythmic pattern.

The terms internal clock and pulse sensation refer to the same phenom non, but differ in emphasis. "Internal clock" alludes to an underlyi neurophysiological mechanism; "pulse sensation," to the experience of th listener. Pulse sensations may be experienced during rhythmic *perceptio* (listening), rhythmic *action* (performance), or both. In this sense, the ide of pulse sensation is consistent with an ecological approach to perceptio (Gibson, 1979), in which perception and action are regarded as inextrica bly linked and sensations are regarded as byproducts of the interactio between an organism and its environment.

SERIAL VERSUS PERIODIC GROUPING

Sound events in speech may be structured either *serially* (concatenated in associative chains), *hierarchically,* or both (Martin, 1972). In music this distinction may be described in terms of *figural coding* or *grouping* on the one hand, and *meter* on the other (Bamberger, 1978, 1982; Cooper & Meyer, 1960; Deutsch, 1982; Jones, 1976; Lerdahl & Jackendoff, 1983; Rosenthal, 1989; Upitis, 1987; Vuori, 1991). A problem with these terms is that meter itself may be regarded as a kind of grouping: It groups events

into equivalence classes, such as all *n*th beats of a bar, by analogy to pitch classes or chroma (Benjamin, 1984). Jones (1987b) avoided the semantic ambiguity of the term grouping by distinguishing between *horizontal* (serial) and *vertical* (metrical) components of rhythmic production. In the present study, the terms *serial* and *periodic* temporal grouping are used, and meter is regarded as a form of periodic grouping.

The relative importance of periodic and serial grouping depends on the listener. For infants, serial grouping is generally more important than periodic grouping (Bamberger, 1980; Drake, Dowling, & Palmer, 1991), whereas the rhythmic perception and performance of adult musicians reflects an organization in which periodic grouping predominates (Smith, 1983). But the effect of age and training on perceptual grouping does not influence all aspects of rhythmic activity. Povel (1981), for example, failed to find significant differences between the results of musicians and nonmusicians in a rhythmic imitation task.

Serial grouping depends primarily on the serial proximity in time, pitch, and timbre of temporally adjacent events (Bregman & Pinker, 1978; Fraisse, 1956; Martin, 1972; Miller & Heise, 1950; van Noorden, 1975). Gaps between serial groups tend to be the largest gaps available in a given sequence (Garner, 1974; Handel, 1974; Vos, 1977). According to Lerdahl and Jackendoff (1983), serial grouping in music includes "motives, themes, phrases, periods, theme-groups, sections, and the piece itself" (p.12); and "only contiguous sequences can constitute a group" (p. 355).

Periodic grouping usually depends on the relative timing and perceptual properties of *nonadjacent* events (Martin, 1972). Periodic grouping may be divided into two stages, here called *pulse sensation* and *perceived meter.* The second stage involves the simultaneous perception of different pulses (or *multiple temporal periodicities:* Palmer & Krumhansl, 1990). The result is a regular alternation of strong and weak beats, corresponding to the generally accepted definition of meter (e.g., Lerdahl & Jackendoff, 1983, p. 12). For example, if the tempi of two pulses stand in the ratio 1:2, then a binary pattern of strong and weak beats results.

Cooper and Meyer (1960) claimed that serial groups arrange themselves around accented events, implying that serial grouping depends on periodic grouping. Similarly, Benjamin (1984) suggested that both accents and serial grouping influence metrical organization, implying that serial grouping influences periodic grouping. Lerdahl and Jackendoff (1983) advanced the simpler (and hence more testable) hypothesis that serial and periodic grouping are independent and must therefore be kept separate: "groups do not receive metrical accent, and beats do not possess any inherent grouping" (p. 26). Similarly, Povel (1984) distinguished between the organization of elements (serial grouping) and the organization of intervals (periodic grouping), remarking that "though presumably occurring in parallel, they may to a greater or lesser extent be incompatible" (p. 332).

The assumption of independence of serial and periodic grouping is fundamental to the present study. Serial and periodic grouping are assumed to depend in separate and distinct ways on the timing (interonset intervals [IOIs]) and other perceptual properties—pitch, loudness, timbre, and articulation—of sound events. Accordingly, the model presented below emulates periodic grouping and metrical organization in simple rhythms without taking serial grouping into account.

The remainder of this paper may be divided into two parts. The first part describes an experimental investigation of periodic grouping in a specific set of rhythms. Experiment 1 concerns the salience of pulse sensations; Experiment 2, the strength of metrical accents. The experiments provide data to enable the development and testing of a quantitative model of pulse salience and metrical accent, which will be described in the second part.

Experiment 1: Pulse Salience

Different rhythmic levels or pulse sensations are rarely heard to be equal in significance (Krebs, 1987). Normally, one level is heard as primary and acts as a frame of reference for the (conscious) perception of other levels. Jones and Boltz (1989) called this the *referent time level.* According to Lerdahl and Jackendoff (1983, p. 21), "The listener tends to focus primarily on one (or two) intermediate level(s) in which beats pass by at a moderate rate . . . Adapting the Renaissance term, we call such a level the *tactus.*" The present experiment investigated the tactus of some simple rhythms.

The perceptual (and hence musical) significance, strength, or prominence of a pulse sensation may be called its *salience.* The tactus is then the pulse sensation with the highest salience. Musical experience suggests that the salience of pulse sensations depends on both tempo and rhythmic pattern. For example, some rhythmic patterns "swing" more than others; and pieces may lose their swing, or simply feel wrong, if played too slow or too fast. Variations in tempo can also affect subjective accentuation. Michon (1974) found that relative temporal stresses (durational accents) in a performance of the *Vexations* by Erik Satie varied as a function of performance tempo, implying a dependency of subjective temporal structure on tempo. Clarke (1982) related these variations to variations in grouping structure, observing that the music tended to be segmented into fewer serial groups at faster tempi.

The present experiment investigated effects of rhythmic pattern and tempo on periodic grouping by measuring the salience of the various pulse sensations evoked by a range of rhythms. Listeners tapped along with each rhythm at equally spaced intervals, in a similar way that a jazz musician or listener taps along to a complex jazz rhythm. Pulse salience was then estimated by counting the relative number of times each pulse-train response was selected.

Related experiments were performed by Handel and Lawson (1983) and Povel and Essens (1985, Experiment 3). Handel and Lawson investigated the perception of polyrhythms consisting of two concurrent pulse trains, distinguishing the two stimuli by playing them at different pitches. Povel and Essens measured the apparent simplicity of concurrently presented rhythms and pulse trains, again at different pitches. The present study differs in that the rhythm presented in each trial consisted entirely of identical sounds. The only other sound heard during the experimental trials was that of the computer's space key being struck or depressed. This sound is unlikely to have influenced the results, because by the time it was heard, the listener had already decided on a pulse response, and was in the process of tapping it out. The model presented later in this study is similarly restricted to rhythms composed of physically identical sound events.

METHOD

Listeners

Twenty-two students (from the University of Stockholm) and researchers (from the Department of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm) took part. Their musical experience (defined as the number of years regularly practicing and/or performing a musical instrument, including voice) covered a wide range: minimum 0 years, maximum 30 years, mean 12 years.

Apparatus

Stimuli were drum strokes on various percussion instruments, digitally recorded on a commercially available drum machine (Roland TR-505 Rhythm Composer). Timing of rhythms was controlled via Musical Instrument Digital Interface (MIDI) by a Le_Lisp program on a Macintosh II computer. Rhythms were amplified and reproduced over a small loudspeaker in a sound-isolated room.

Stimuli

As shown in Figure 1, six different rhythmic patterns were crossed with six different tempi, making 36 trials in all. The first four of the rhythmic patterns (*pulse, waltz, march, swing*) were metrically relatively simple and unambiguous. The last two (*skip, cross*) were more ambiguous: They may be perceived in either 3/2 or 6/4 meter (that is, as either 3 groups of 2 beats, or 2 groups of 3) depending on tempo (cf. Handel & Oshinsky, 1981). One of the aims of the present experiment was to investigate this ambiguity.

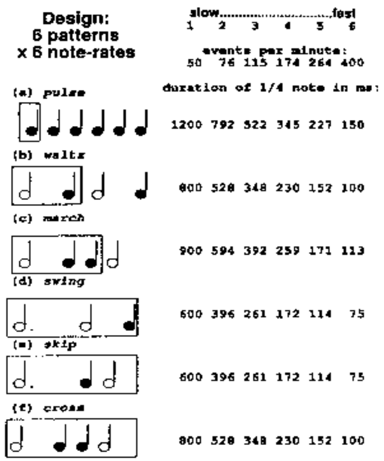

Fig. 1. Reference table for stimuli used in the experiments. Each pattern is given a name for ease of identification. The boxes define rhythmic cycles or measures. The name *cross* refers to a cross-rhythm of 1/2-notes and 3/4-notes.

Tempi were specified by the number of *notes* or events per unit time, as this allowed the full range of musical tempi to be covered more satisfactorily for each pattern than specifying the number of *beats* per unit time. The experimenter adjusted the note rates to cover as wide a range as possible within the musical and physical constraints of the experiment, by subjective evaluation of each of the six patterns when played at different rates. The note rates that were adopted were logarithmically equally spaced and ranged from 50 to 400 notes per minute, or 0.83 to 6.7 notes per second. The corresponding IOIs of the nominal 1/4-note beats in each pattern are shown in Figure 1.

Early events in a rhythmic sequence are more likely to be interpreted as downbeats than are later events (cf. Longuet-Higgins & Lee, 1982). In the present experiment, a starting point effect was undesirable, as the aim was to investigate the perceptual properties of cyclically repeating rhythms as a function only of rhythmic pattern and note rate. The starting-point effect may be softened by starting a sequence at a subliminal intensity and gradually increasing the level (Vos, 1973). But this does not eliminate the effect, as it is still generally possible to define the first audible event. Alternatively, the sequence may be started at a very fast tempo and then made to decelerate toward the intended tempo (Royer & Garner, 1970). This was not a practical alternative in the present study, where tempo was itself an important independent variable. Here, the starting-point effect was avoided simply by starting rhythms at random temporal points in the cycle. For example, pattern (b) *waltz* was presented in either of two forms: 1 2 1 2 1 2 . . . . or 2 1 2 1 2 1 . . . . (where 1 = 1/4-note, 2 = 1/2-note in Figure 1). The probability of a pattern starting on a certain note was made proportional to the interonset interval that would otherwise have preceded it. So *waltz* was twice as likely to start with a 1/4-note than with a 1/2-note.

The attention of the listener was maintained by randomly selecting the timbre (instrument) to be used in each trial from sounds available on the drum machine (labeled low conga, high conga, timbale, low cowbell, high cowbell, hand clap, ride cymbal, bass drum, snare drum, low tom, mid tom, high tom, rim shot, closed high-hat, open high-hat). Timbre remained constant throughout each trial. The loudness of the various instruments had been equalized by ear before the experiment.

Procedure

In each trial, listeners were asked to press or tap the space bar of a computer keyboard in time with the underlying beat of the rhythm. They were permitted to tap in any way they wished; most used the index finger, middle finger, or both together, of their dominant hand. A tap was deemed to have coincided with a particular 1/4-note (as notated in Figure 1) if it fell in a temporal category of width one 1/4-note, centered on the onset of the note. In other words, temporal category boundaries for taps were set midway between 1/4-note onsets. This procedure is consistent with the categorical perception of rhythmic patterns as described by Clarke (1987) and Schulze (1989).

In each trial, tap times were recorded until such time as four consecutive tap times were equally spaced, indicating a consistent pulse response. At this point, the stimulus stopped, and the response was recorded. No feedback was given. Trials were presented in a different random order for each listener. The experiment was preceded by a short practice session.

Raw data were the *periods* (time between taps) and *phases* (relative to the nominal start of the rhythm) of the tapped pulses, in 1/4-note units. The terms period and phase correspond to the terms *unit* and *location* of Povel and Essens (1985). Other possible terms for period are *grid interval* (Povel, 1984), *subdivision* (Povel & Essens, 1985), *metrical unit* (Essens & Povel, 1985), *beat interval* (Summers, Hawkins, & Mayers, 1986), and *cardinality* (Krebs, 1987).

RESULTS AND DISCUSSION

Figure 2 shows the number of times each pulse response was selected (plain text). The results are compared with predictions according to the model (italics). The comparison between the results and predictions will be discussed later.

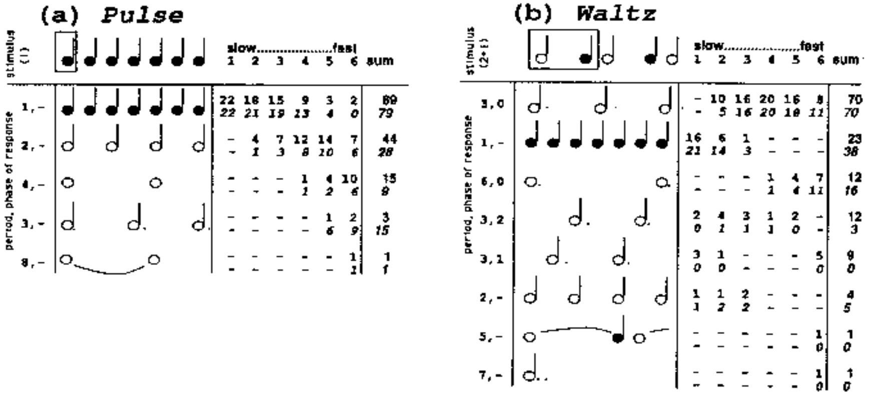

Fig. 2. Results of Experiment 1 (pulse salience) for all 161 selected pulses. In the table on the right side of each panel, plain text denotes numbers of times each pulse was selected; italics denote corresponding calculated values according to Equation 7 with free parameter values $\tau = 550$ ms, $i = 1.6$, $\mu = 760$ ms, $\sigma = 0.23$, and $j = 2.0$. Each column corresponds to a particular combination of rhythmic pattern and note rate, or to one of the 36 trials in the experiment.

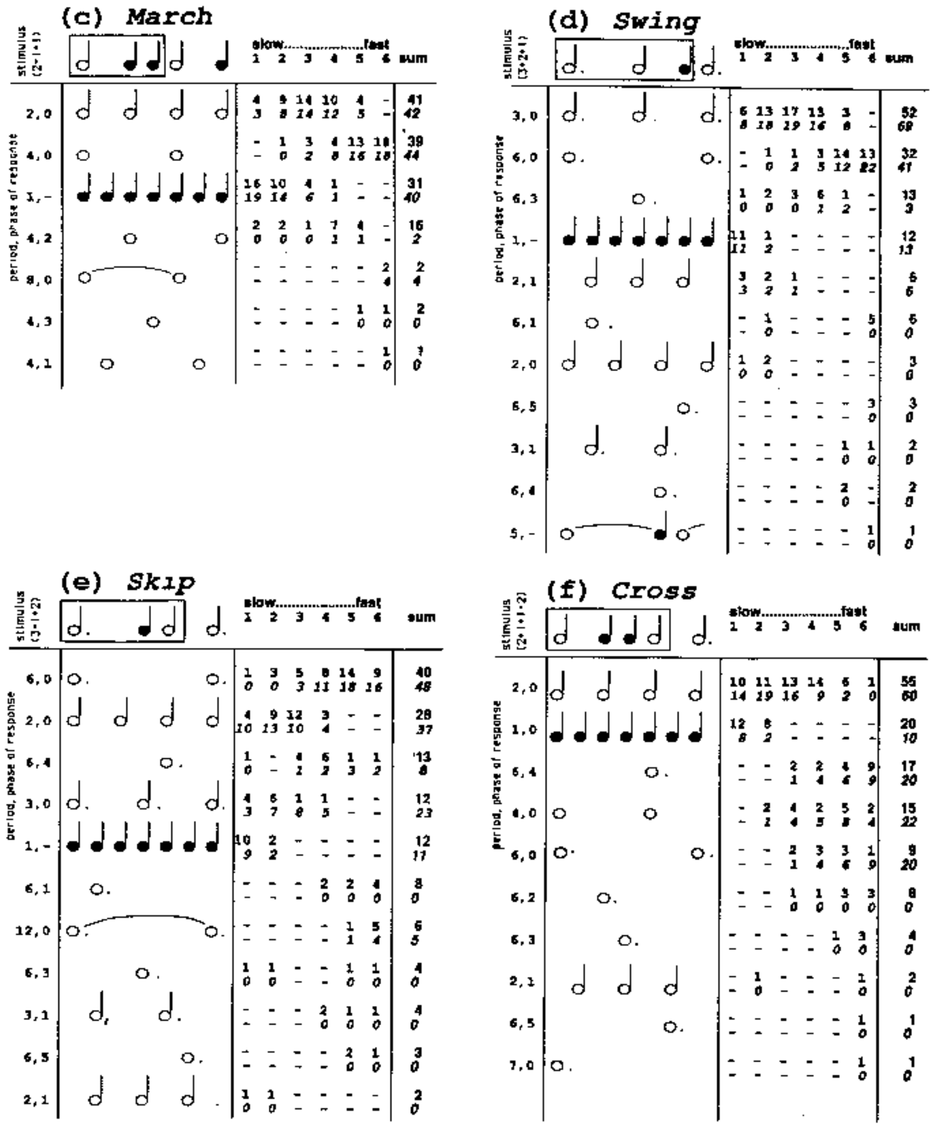

Fig. 2. (Continued) Calculations for each trial are normalized (multiplied by a constant) so as to add up to 22, the number of listeners. The numbers 1–6 (top right) are rate numbers (see Figure 1). The periods and phases shown on the far left are expressed in nominal 1/4-note beats. Phases are shown only if necessary for the unambiguous specification of pulse responses; otherwise, a dash (-) is marked.

Each response was specified by four parameters: the rhythmic pattern and note rate of the stimulus, and the period and phase of the response. Each of the 36 trials of the experiment allowed for a number of different possible pulse responses (periods and phases). In all, 161 different pulse responses were collected. The number of different pulse responses as a function of pattern was 17, 24, 23, 29, 37, and 31; as a function of rate, 22, 28, 23, 24, 29, and 35. These summary data tentatively suggest that (1) the main pulse sensation or tactus evoked by each combination of rhythmic pattern and note rate was quite ambiguous and (2) the ambiguity of the tactus increased as the rhythmic patterns became more complex. In addition, (3) ambiguity showed a small but inconsistent tendency to increase as note rate increased.

Figure 2a shows results for the six different presentation rates of pattern (a) *pulse.* At the slowest rate (no. 1), all 22 listeners tapped in time with every sound event (that is, with period = one 1/4-note beat). At the second rate, 18 listeners tapped in time with every sound event (period = 1 beat) and 4 with every second event (period = 2 beats). The higher the tempo of the isochronous sequence, the more pulse sensations it evoked—that is, the higher was its rhythmic ambiguity. This is consistent with the observation (e.g., Handel & Lawson, 1983) that when a rhythm is played more slowly, responses will tend to be faster relative to the rhythm, and vice-versa. In other words, pulse responses tend to gravitate toward a moderate tempo.

The perceptual grouping of isochronous sound events at moderate to fast tempi is called *subjective rhythmization.* In general, groups of four occur more often than groups of three (see references under Enhancement of pulse salience below). Figure 2a suggests that this effect is quite general and independent of tempo.

In the *waltz* rhythm (Figure 2b) at a very slow tempo (rate 1), 16 of the 22 listeners tapped with a period of one 1/4-note. Of the remaining six listeners, three tapped on the "implied" or "missing" events between the notes (period 3, phase 1), two tapped in time with the 1/4 notes (period 3, phase 2), and one tapped a cross-rhythm (period 2). No one tapped in time with the 1/2-note onsets—the music-theoretical downbeat (period 3, phase 0). At faster rates, the proportion of listeners tapping the music-theoretical downbeat increased until, at the highest rates, some tapped with every other downbeat (period 6, phase 0). Responses with periods 5 and 7 were clearly errors; in the interests of objectivity, however, they were not removed from the data.

Vos, Collard, and Leeuwenberg (1981) played recorded performances of Bach Preludes and asked listeners to tap at equal time intervals as they listened. The location of the downbeat was found to be quite ambiguous:

Some listeners tapped in synchrony with the notated barlines, while others tapped at the same rate but out of phase with the barlines. A similar effect was found in the present results for the *march* rhythm (Figure 2c), for responses with period 4 at slow to moderate note rates. For rates 1–4, results were low for both phase 0 and phase 2. For rates 5 and 6, however, phase 0 (the music-theoretical downbeat) predominated over phase 2. A similar effect of tempo was observed in the *swing* (Figure 2d) and *skip* (Figure 2e) rhythms at period 6, phases 0 and 3 (or 4). In general, the results indicate that long events are more important than short events in fast rhythms, but not necessarily in slow ones.

In the last three rhythmic patterns, a tendency was observed for some listeners to "drag," or tap a little later than expected (see results with period 6, phase 1). This observation is consistent with the idea of a motor delay between intended and actual responses, as incorporated into the theory of self-paced periodic tapping of Wing (1973) (described by Vorberg & Hambuch, 1978). In other tapping experiments, however, listeners have been found to *anticipate* events (Fraisse, 1980, 1987). In the present experiment, it appears that some listeners simply had insufficient time to catch up with fast rhythms before their response was collected and the stimulus ceased.

Figure 3 shows the distribution of the periods of all 22 × 36 = 792 pulse trains selected in Experiment 1. The histogram was determined by calculating logarithms to the base 10 of the periods in milliseconds and then allocating them to categories of width 0.1. The distribution has a mean of −0.149 (corresponding to a period of 710 ms) and a standard deviation of 0.224 (corresponding to the range 420–1190 ms). The bell-shaped curve corresponds approximately to the *existence region of pulse sensation* (formulated in Eq. 6 of the model presented below). The low results at −0.05 on the horizontal axis (for logarithms between −0.1 and 0.0, or periods between 790 and 1000 ms) appears to be an artifact of the experimental procedure, due to the relatively low number of possible responses in this range (remembering that the same set of note rates was used for all rhythmic patterns).

Experiment 2: Metrical Accent

Sequences of identical events spaced at different IOIs may produce *subjective accents*. Subjective accents have been investigated for isolated pairs of tones (Buytendijk & Meesters, 1942) and for sequences of tones containing two alternating time intervals (Povel & Okkerman, 1981). The aim of the present experiment was to investigate subjective accentuation in

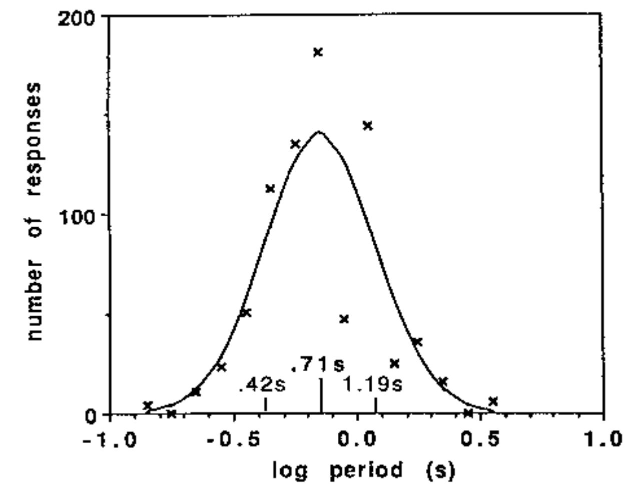

Fig. 3. Further results of Experiment 1 (pulse salience). Crosses: distribution of the periods of all 22 × 36 = 792 pulse responses. Vertical axis: number of pulse responses falling in categories of width 0.1 (in $\log_{10}$ of pulse period in seconds). Curve: The corresponding normal distribution.

simple rhythmic patterns. The same set of six patterns was examined as in Experiment 1.

In music theory, different beats in a measure are assigned different accent strengths (see, e.g., Lerdahl & Jackendoff, 1983). In a 4/4 measure, for example, the first 1/4-note beat is the strongest, the third beat is the second strongest, and the second and fourth beats are the weakest. In a 6/4 measure, the second-strongest beat is the fourth of the six 1/4-note beats (phase = 3); but in a 3/2 measure, the second-strongest beat is the fifth 1/4-note beat.

Palmer and Krumhansl (1990) measured the relative salience of the different beats in standard musical meters by the following method. They first presented a low-pitched pulse train with a period in the range 1.7–4.8 s and asked their listeners to imagine each pulse as the first of two, three, four, or six isochronous beats. After four pulses, a higher-pitched probe tone was sounded between the pulses, corresponding to one of the imagined beats. The listeners then rated how well the probe tone fit the imagined metrical context. The results were consistent with music-theoretical notions of metrical accent and confirmed that their listeners had quite detailed and stable knowledge of typical Western meters.

The present experiment differed from that of Palmer and Krumhansl in that it aimed to measure perceptual properties of *real-time* (rather than imagined) temporal sequences. Listeners in the present experiment were not asked to imagine the subdivision of the bar; instead, the subdivision was suggested directly by the temporal sequence they heard.

METHOD

Listeners

Twenty-one people took part, all of whom had participated in the previous experiment.

Stimuli

The same six rhythmic patterns were presented as in Experiment 1, but only one note rate was used: 150 notes per minute. As before, each trial began at a random point in the cycle. After 8–10 notes of the rhythm, a single target note was played in a different timbre (bass drum). The target coincided with one of the beats in the cycle, but not necessarily with one of the notes. The rhythm then continued without interruption until the listener responded.

The experiment tested each individual 1/4-note beat in the cycle of each of the six rhythmic patterns in Figure 1. The cycle lengths of the six patterns were 1, 3, 4, 6, 6, and 6 beats, so the total number of trials was 26. Instrumental timbres were selected randomly from the same set as before, with the exception of the bass drum, which was reserved for the target beat.

Procedure

Listeners were asked to indicate whether the target was on or off the beat, using a four-point rating scale. Labels for the scale were provided for all listeners in both musical and nonmusical language. The musical labels were: 0, very syncopated; 1, quite syncopated; 2, on a weak beat; 3, on a strong beat. The nonmusical labels were: 0, off the beat (sure); 1, off the beat (not sure); 2, on the beat (not sure); 3, on the beat (sure). Listeners were free to choose whichever labels they preferred. The experiment began with a short practice session.

RESULTS AND DISCUSSION

In subjective rhythmization, an isochronous sequence evokes a pulse sensation whose period is longer than that of the stimulus. It follows that some events in isochronous sequences may sound off the beat. This expectation was borne out in the single result for pattern (a) *pulse*, shown in Figure 4a, which was significantly lower than the maximum of the response scale.

The results for the other rhythmic patterns (Figures 4b to 4f) agreed with music-theoretical expectations, with some minor but interesting exceptions. Results for pattern (c) *march* peaked at the third 1/4-note beat instead of the first, contradicting the idea that events followed by longer IOIs have stronger accents. A similar effect had been suggested in Experiment 1 (see Figure 2c): At rate 4 (174 events per minute, which is close to the rate of 150 events per minute used in Experiment 2) and period 4 (corresponding to the cycle length), four listeners tapped on the first beat (phase 0), whereas seven tapped on the third (phase 2). It appears in this case that the IOI *preceding* the event had a slightly greater effect on its subjective accent than did the IOI *following* the event. According to Povel and Essens (1985, p. 415), the initial and final events in a cluster of three events have subjective accent; here, the accent on the initial event of the group was stronger than the accent on the final event.

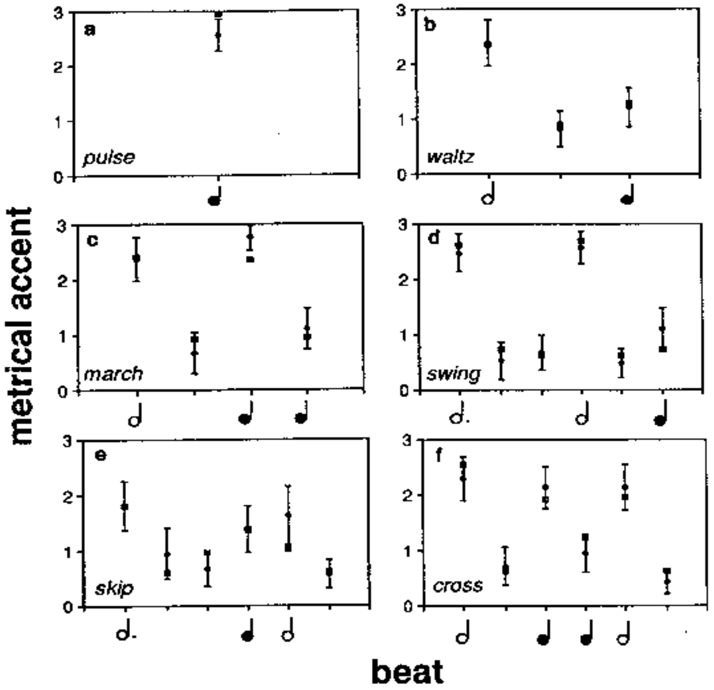

Fig. 4. Results of Experiment 2 (metrical accent) for rhythmic patterns (a) *pulse*, (b) *waltz*, (c) *march*, (d) *swing*, (e) *skip*, and (f) *cross*. The horizontal axis of each panel is marked with the note values of each pattern as shown in Figure 1. Points: Mean responses (21 data each). Bars: 95% confidence intervals of means. Squares: Calculations according to Equation 8, with free parameters τ = 180 ms, i = 1.7, μ = 660 ms, σ = 0.14, and j = 1.9. The 26 calculations have been linearly adjusted to have the same mean and standard deviation as the 26 mean responses.

Results for *swing* (Figure 4d) at the first and fourth 1/4-note beats were not significantly different from each other, that is, the downbeat was about equally likely to fall on either beat. This is consistent with the results of Experiment 1 for *swing* at rate 4, period 6, phases 0 and 3 (Figure 2d). The downbeat of *skip* was similarly ambiguous, falling on either the first or the fifth beat (cf. Figure 2e, rate 4, period 6, phases 0 and 4); and the downbeat of *cross* was perceived with about equal likelihood on beats 1, 3, or 5 (cf. Figure 2f, rate 4, period 6, phases 0, 2, 4).

These results of Experiment 2 only apply for a note rate of 150 notes/min. The results of Experiment 1 suggest that differences between event saliences would have been greater, that is, the meter of the rhythms would have been less ambiguous, at faster tempi.

Model

The need for a quantitative, computational model of the perception of meter was recognized by Simon (1968). Since then, models of meter have been developed by Longuet-Higgins and Steedman (1971), Longuet-Higgins (1976), Longuet-Higgins and Lee (1982, 1984), Povel and Essens (1985), Lee (1991), and Rosenthal (1992). The models of Longuet-Higgins and Lee simulated real-time listening by beginning at the start of a rhythm and testing hypothetical metrical interpretations at each new event. By contrast, the present model, like the models of Povel and Essens and of Rosenthal, looks at a whole rhythmic sequence at once, testing all possible metrical interpretations (clocks, rhythmic levels) for goodness of fit. Tentative extensions of the present model to real-time processing are proposed below under the heading Primacy, recency, and the psychological present.

A central assumption of the present model is the inherent ambiguity of the underlying pulse (tactus) and meter of a rhythm. The model does not output a single solution, but instead considers many possible pulse and meter sensations, estimating the relative importance or salience of each. In this respect it is similar to the models of Povel and Essens (1985), Lee (1991), and Rosenthal (1992). Metric ambiguity is particularly important in the case of hypermeter—meter that is perceived but not notated and whose period exceeds that of a notated measure (Rothstein, 1989).

The present model is based on an earlier version (Parncutt, 1985), which in turn was based on an exposition of the theory of pulse sensation (Parncutt, 1987). The original model had been tested by comparing its predictions with music-theoretical analyses and intuitions. Modifications to the present model were inspired by a combination of experimental results and theoretical considerations. Appropriate mathematical forms were chosen according to the subjective requirements of theoretical clarity and parsimony and the objective requirements of computational simplicity and efficiency. Changes to the model were adopted only if they enhanced correlation coefficients between predictions and experimental data. The same procedure also enabled values of free parameters to be estimated.

The model is primarily intended to account for the results of the two experiments just described and to link them together, predicting metrical accent from pulse saliences. The model also suggests explanations for various further rhythmic phenomena. The perception of meter may be assumed to involve the simultaneous perception of concurrent pulses, and the salience of a perceived meter may be modeled by adding calculated pulse saliences. Other phenomena that may be clarified by the present approach include limitations on the number of allowable subdivisions of a beat, the categorical perception of time in rhythms, the perceptibility of small temporal deviations in isochronous sequences, expressive timing in musical performance, and the temporal development of pulse salience and meter. Finally, the theory suggests a concise definition of musical rhythm.

A schematic representation of the model is shown in Figure 5. The input to the model is a cyclically repeating pattern of IOIs in milliseconds. The sound events making up the rhythm are assumed to be physically identical. The model begins by assigning to each event in a rhythmic cycle a *phenomenal accent,* calculated according to a *saturation function* relating phenomenal accent to IOI (see Eq. 3, Figure 6 below). Next, the perception of pulse is simulated by a pattern-matching routine (see Eq. 4), and the tempo dependence of pulse salience is accounted for in terms of an *existence region* of pulse sensation (see Eq. 6). This leads to a prediction of pulse salience (see Eq. 7), defined as a measure of the probability of occurrence of a given *tactus* or underlying beat. Calculated pulse saliences are compared with the results of Experiment 1. The results of Experiment 2 are predicted by estimating the *metrical accent* at each (actual or implied) rhythmic event, by summing the calculated saliences of all pulse sensations converging on that temporal category (see Eq. 8). Further applications of the model are discussed at the end of the paper.

INPUT

ASSUMPTION: *The relationship between physical and perceived time is essential linear.*

On the basis of a thorough literature survey, Allan (1979) found that perceived time is in general almost proportional to physical time, suggesting that no distinction need be drawn between physical and perceived time in a model of rhythm perception. Time is measured in the present model either absolutely (in milliseconds) or relative to a basic time unit b in milliseconds. The values for b used in the above experiments were the 1/4-

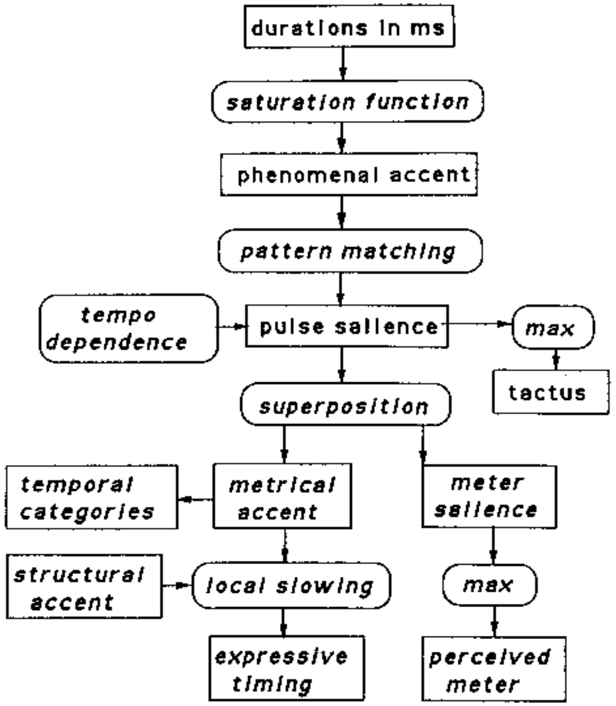

Fig. 5. Schematic representation of the model.

note beats given in Figure 1. Absolute time in milliseconds is denoted t; relative time in basic time units, T.

The model is presently restricted to cyclically repeating rhythms. The length of a cycle is denoted either c (in ms) or C (in basic time units). For example, pattern *pulse* in Figure 1 has C = 1, *waltz* has C = 3, *march* has C = 4, and the other patterns all have C = 6; c additionally depends on the number of events per minute in a given trial. The cycle is assigned an arbitrary starting point where $t = T = 0$. In Figure 1, this point corresponds to the first note in each pattern.

The basic time unit is assumed to be perceived as a *temporal category*. One rhythmic cycle contains C such categories. In the trivial case of an isochronous sequence, all temporal categories are represented by sound events. In nonisochronous rhythms, some categories are "empty." The model begins by assigning to each "full" time category an IOI indicating the number of categories (basic time units) between that category and the next "full" category. Intervening "empty" categories are assigned IOI = 0.

For example, the *waltz* rhythm of Figure 1 had IOI (0) = 2, IOI (1) = 0, and IOI (2) = 1. The IOIs add up to the cycle length:

$$\sum_{T=0,\,C-1} \mathrm{IOI}\ (T) = C \qquad (1)$$

PHENOMENAL ACCENT

ASSUMPTION: *Phenomenal accents may be induced by changes in IOI, loudness, timbre, and pitch and by combinations thereof.*

According to Lerdahl and Jackendoff (1983), rhythmic accents are of three kinds: phenomenal, structural, and metrical. Phenomenal accents may be evoked by changes in IOI (Povel & Essens, 1985; Povel & Okkerman, 1981), loudness (dynamic accent), timbre (e.g., Wessel, 1979), melodic contour (Thomassen, 1982; Woodrow, 1911), and implied harmony (Smith & Cuddy, 1989). Jones (1987a) referred to the superposition of different patterns of accentuation as *joint accent structure* and to the combination of different kinds of accent at a given temporal position as *accent coupling*. In general, the phenomenal accent A_o at temporal category T may include contributions from durational accent A_d, dynamic (loudness) accent A_l, timbral accent A_t, and pitch accent A_p (where A_p has melodic and harmonic components), as well as possible interactions between these parameters:

$$A_o(T) = A_d(T) + A_l(T) + A_t(T) + A_p(T) + \text{interactions} \qquad (2)$$

Of the various contributions to phenomenal accent, durational accent appears to have the greatest impact on metrical organization (cf. Palmer, 1989). Several pieces of evidence support this claim. An increase of 2 dB in the level of a tone in an isochronous sequence is sufficient to produce a dynamic accent (Thomassen, 1982), but an increase of about 4 dB is needed to produce a dynamic accent strong enough to balance an durational accent (Povel & Okkerman, 1981). Young children can reproduce temporal but not dynamic structures of musical rhythms (Gérard & Drake, 1990), suggesting that the temporal structure of simple rhythms is more perceptually salient than their dynamic structure. Finally, the variability of IOIs in tapped patterns is considerably less than the variability of intensities (Brown, 1911)—suggesting that phenomenal accent is more sensitive to changes in IOI than to changes in loudness.

Phenomenal accent may also be influenced by *articulation*. The articulation of an event may be defined as the ratio of the onset-to-offset interval (OOI) to the IOI after the event: articulation = OOI/IOI. According to this definition, articulation approaches a value of one in legato and about one-half or less in staccato. Musical experience suggests that articulatory accents produced by playing particular notes legato rather than staccato (e.g., in the articulation of two-note phrases) are generally weaker than accents due to variations in IOI (cf. Sloboda, 1983, and studies cited below under expressive timing). In other words, the auditory system tends to respond to short-duration events in rhythmic sequences as if they lasted until the onset of succeeding events.

In the present version of the model, phenomenal accent depends only on IOI. In this regard, the present model is similar to other published models of rhythm and meter (Longuet-Higgins & Lee, 1982; Povel & Essens, 1985; Rosenthal, 1992). The stimuli in the above experiments were also limited to variations in IOI; loudness, timbre, pitch, and physical event duration were held constant in each trial.

DURATIONAL ACCENT, ECHOIC STORAGE, AND MINIMUM DISCRIMINABLE IOI

ASSUMPTION: *The subjective accentuation of an event increases with the IOI that follows it.*

In general, the greater the IOI after an event, the greater is (1) its phenomenal accent or perceptual salience; (2) its contribution to the formation of pulse sensations; and (3) its probability of being perceived as a metrical accent or downbeat (cf. Longuet-Higgins & Lee, 1982; Rosenthal, 1992; Thomassen, 1982; Vos, 1977).

Massaro (1970) demonstrated that the perceptual salience of a tone can be reduced by presenting a second tone within the time span of echoic storage. He concluded that the second tone interrupted the processing of the first. This suggests that events followed by longer IOIs are more salient than events followed by shorter IOIs. Povel and Okkerman (1981) found this idea to be consistent with their findings on subjective accentuation.

ASSUMPTION: *Durational accent increases with IOI for small values of IOI and saturates as IOI approaches and exceeds the duration of the echoic store (auditory sensory memory).*

In Experiment 1, it was observed that the downbeat of a cyclic pattern (the event initiating the longest IOI) became less ambiguous as the rhythm was played faster: At faster tempi, phenomenal accent was usually greater for events preceding longer IOIs, whereas at the slower tempi, no consistent relationship was evident between phenomenal accent and IOI. Massaro's concept is consistent with these results, given that the IOIs in Experiment 1 were predominately greater than the duration of the echoic store at very slow tempi and less at very fast tempi.

Echoic storage has been referred to by many different names, among them echoic memory (Neisser, 1967), precategorical acoustic storage (Crowder & Morton, 1969), memory for nonattended auditory material (Glucksberg & Cowen, 1970), preperceptual auditory store (Massaro, 1970), brief auditory storage (Treisman & Rostron, 1972), nonverbal memory trace (Deutsch, 1975), and primary or immediate memory (Michon, 1975). For the present purposes, these diverse terms are assumed to refer to essentially the same kind of "memory." Note that echoic storage is not really memory at all in the conventional sense of the word—rather, it may be regarded as a lingering of sensory experience in the absence of conscious cognitive processing. Michon (1975) described it as a precognitive buffer store for incoming information.

Most estimates of the duration of the echoic store vary between 0.5 and 2.0 s (Crowder, 1969, 1970; Guttman & Julesz, 1963; Kubovy & Howard, 1976; Rostron, 1974; Treisman, 1964; Treisman & Howarth, 1959; Treisman & Rostron, 1972). Other studies have arrived at both lower estimates of the duration of echoic storage, in the vicinity of 250 ms (Massaro, 1970, 1974; Plomp, 1964), and higher estimates such as 3 s (Fraisse, 1982) and even 5 s (Glucksberg & Cowen, 1970).

The variability in published estimates of the duration of echoic storage may be due to differences in experimental method. The duration of memory storage is somewhat flexible, depending on the kind of sound being stored, and the context in which it is heard (Michon, 1978). Another source of uncertainty is the arbitrariness of the point in time at which memory is deemed to be exhausted. Suppose, for the purpose of argument, that the salience of sensory material in echoic memory decays exponentially with a half-life of 1 s (cf. Kubovy & Howard, 1976). The perceptual salience of an event in the sensory store would then fall to 1/2 its original value after a period of 1 s, 1/4 after 2 s, 1/8 after 3 s, and 1/32 after 5 s. Which of these times represents *the* duration of echoic memory is a matter of definition.

An alternative explanation for the relationship between accent and IOI involves familiarity with the auditory environment. Consider an environment in which sound events (produced, for example, by collisions between objects) are distributed at random time intervals of the order of 1 s. The succession of sound events perceived by a human observer in such an environment will be nonrandom, for the following reasons. First, relatively loud sounds will tend to mask relatively quiet sounds that are either simultaneous or close to them in time. Sounds that are completely masked may be regarded as perceptually irrelevant. So perceived time intervals in the vicinity of louder sounds will tend to be longer than perceived time intervals in the vicinity of quieter sounds. Second, loud sounds often begin with a fast, percussive onset, followed by a sustain/decay segment during which other, quieter sounds may be masked. The sustain segment is typically prolonged by the acoustic resonance of the sound source and by reflection of the radiated sound from objects and surfaces in the vicinity. So perceived time intervals *following* loud sounds will tend to be longer than perceived time intervals *preceding* loud sounds. Moreover, forward masking is stronger than backward (Moore & Glasberg, 1983; Zwicker & Fastl, 1972), further enhancing the positive correlation between the IOI following an event and the loudness of the event.

The saturation of durational accent for IOIs exceeding roughly 1 s may be related to the typical or average duration of perceived IOIs that follow loud sounds in the everyday auditory environment, including the sounds of speech and music. Given that loud sounds tend to mask quiet sounds, perceived IOIs may be influenced by physical parameters such as the decay times of typical physical resonators and typical time intervals between direct and reflected sound. The sum of these two contributions would typically lie in the range of, say, 0.1–2.0 s, in both musical and nonmusical auditory settings.

Yet another possible explanation for the saturation of durational accent with increasing IOI is suggested by data on the perceived loudness of single tones of various durations. Extrapolating from experimental results of Plomp and Bouman (1959), Roederer (1979, pp. 88–89) pointed out that the loudness of relatively short tones (lasting less than about half a second) depends on their onset-to-offset duration, increasing steadily as duration increases. The loudness of such tones may be understood to be determined by the total physical power they deliver to the ear. For durations exceeding about 1 s, loudness approaches a constant value that depends on intensity (physical power per unit time) and is independent of duration. The loudness of a single tone thus varies with its duration in much the same way as the durational accent of a short rhythmic event varies with the IOI following that event. Moreover, the loudness of higher-frequency tones saturates more quickly than the loudness of lower-frequency tones, suggesting that the saturation time for durational accent may depend similarly on the pitch of the sound event that initiates the IOI.

ASSUMPTION: *The durational accent of small IOIs depends on an effective IOI that is roughly 50–100 ms shorter than physical IOI.*

In their system of rules for music performance, Friberg, Frydén, Bodin, and Sundberg (1991) lengthened extremely short notes. The extra time was borrowed from neighboring (longer) notes, hence the rule name *social duration*. In addition, a minimum duration of 50 ms was specified.

The procedures adopted by Friberg et al. may be interpreted as compensating for the *minimum discriminable IOI*. Consider a typical isochronous sequence. As the period of the sequence approaches 50–100 ms (10–20 Hz), the events of the sequence eventually cease to be discriminable. At the threshold of discriminability, the IOI perceived to follow each event will effectively be zero. If the speed of the sequence is then gradually reduced so that individual events again become audible, the durational accent of each event will increase monotonically as the difference between the actual IOI (or period) and the minimum discriminable IOI increases. In other words, the relationship between durational accent and IOI will not pass through the origin.

The minimum discriminable IOI appears to be closely related to the *perceptual onset delay* of a sound. Morton, Marcus, and Frankish (1976) defined the *P-center* of a speech vowel as the perceptual moment of occurrence of a word or phoneme. The P-center generally lags behind the physical onset of a sound by a time interval that may be related either to acoustic factors, such as physical duration and amplitude envelope (Howell, 1988; Marcus, 1981), or to articulatory timing (Fowler, 1979). In the case of musical tones, perceptual onset delay has been related to the shape of the rise portion of the tone. Schütte (1978) proposed a temporal integration model, in which perceptual onset occurs when the area under the subjective amplitude envelope passes a specific fraction of its maximum value. Vos and Rasch (1981) suggested that perceptual onset of a tone occurs when the amplitude envelope passes a threshold situated 6–15 dB below maximum amplitude, for maximum amplitudes in the range of 20–70 dB (respectively) above masked threshold.

For rhythmic sequences made up of physically identical sound events (such as the stimuli in the above experiments), the perceptual onset of each event is presumably delayed by a constant time interval. In such sequences, perceived IOIs probably do not depend on perceptual onset delays, but correspond exactly to physical IOIs. In very fast rhythms, however, the minimum discriminable IOI may be influenced (or even determined) by the perceptual onset delay of each event. This hypothesis is feasible given that (1) perceptual onset delays and the minimum discriminable IOI typically have about the same order of magnitude (some 50–100 ms), and (2) both phenomena may be explained by a temporal integration process. In real musical rhythms, of course, sound events are not physically identical, and perceptual onset delays vary from one event to the next. Such variations are not accounted for in the present version of the model.

In the present model, the above three assumptions are accounted for as follows:

$$A_d(T) = [1 - \exp\{-ioi(T)/\tau\}]^{i}, \qquad (3)$$

where $A_d(T)$ denotes the durational accent of the event occurring at time T, exp is the natural exponential function, *ioi* (T) is the IOI following the event expressed in milliseconds, τ is the saturation duration (assumed proportional to the duration of the echoic store), and i is the so-called *accent index*, accounting for the minimum discriminable IOI. The variables τ and i are the first two free parameters in the model.

Note that Equation 3 could be expressed slightly differently by substituting 2^x for exp (or e^x). If i were equal to 1, then τ would correspond to the half-life of the echoic store. The natural exponential is preferred here, as it corresponds to standard mathematical usage.

Equation 3 is graphed in Figure 6 using parameter values found to produce a good fit between calculations according to the model and experimental results: $\tau = 500$ ms and $i = 2$ (see Table 1 below). The graph suggests that durational accent saturates at an IOI of about 1 s, in agreement with most experimental findings on the duration of the echoic store. The further one moves to the left on the graph (i.e., the faster the tempo of a rhythm), the greater the difference in phenomenal accent between long and short IOIs (i.e., the greater the difference between the salience of corresponding pulse sensations), in agreement with the results of Experiment 1.

The concave-upward shape of the left portion of the curve in Figure 6 accounts for the effect of minimum discriminable IOI. The effect could have been accounted for more directly in Equation 3 by setting $i = 1$ and replacing $ioi\ (T)$ by the expression max $\{ioi(T) - ioi_0, 0\}$, where ioi_0 is the minimum discriminable IOI in milliseconds. This modification would have the advantage that τ does not interact significantly with ioi_0 (whereas τ interacts with i in Eq. 3 as formulated above). However, the modification would necessitate an extra routine to deal with events separated by an interval of less than ioi_0. If two such physical events combined to form a single perceptual event, when would its perceptual onset occur? In the absence of experimental data on this issue, the parameter ioi_0 was excluded from Equation 3.

An alternative possible explanation for the shape of the curve in Figure 6 involves metrical ambiguity. Increasing the value of the parameter i has

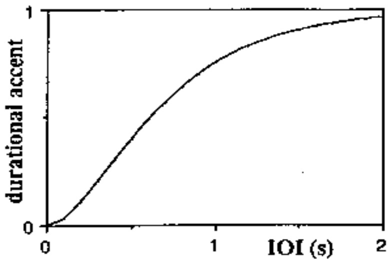

Fig. 6. Dependence of phenomenal accent on interonset interval (IOI) according to Equation 3 with $\tau = 500$ ms and $i = 2$.

the effect of sharpening the perceptual difference between long and short events, thereby sharpening the difference between the salience of different pulse responses and making metrical interpretations less ambiguous. According to this explanation, i should increase with musical experience.

Does the IOI *preceding* an event influence its durational accent? Povel and Essens (1985) proposed that an event in a nonisochronous sequence of physically identical sounds is heard as accented, or perceptually marked, if it satisfies one of the following three criteria: (1) it is relatively isolated, (2) it is the second of a cluster of two events, or (3) it is first or last in a cluster of three events. They then assigned subjective accents of two units to events satisfying the above criteria, and one unit to others. Consistent with Equation 3, point (2) implies that an event is accented if the IOI that follows it is greater than the IOI that precedes it. Point (3) additionally implies that, under certain conditions, the IOI preceding an event may enhance its accent, which is inconsistent with Equation 3. Point (3) was confirmed by Drake, Dowling, and Palmer (1991), who found that the first and last events in a rhythmic group were accented in performances of simple melodies by children and adults. Equation 3 is nevertheless consistent with Lee (1991), who compared various models of rhythm perception and found that models in which accent was primarily determined by the IOI following an event performed considerably better than others.

While formulating the model, attempts were made to incorporate the effect of the IOI preceding an event on the event's durational accent, but these amendments failed to produce any overall improvement in the correlation between calculations and results of the experiments. This result may simply have been due to the particular set of rhythmic stimuli used in the experiments. The effect of the previous IOI may be incorporated into a future version of the model by testing a wider range of rhythmic patterns, such as those tested by Drake (1993), Povel and Essens (1985), or Summers, Hawkins, and Mayers (1986).

According to Povel and Okkerman (1981), events preceding longer IOIs are more accented only for IOIs greater than about 250 ms whose relative difference is greater than 12–15%. They found that if two successive events are followed by small IOIs of almost the same magnitude, then the sound followed by the *shorter* IOIs is perceived as accented. This effect was not incorporated into the present model, for two reasons. First, this kind of accent is quite weak (Povel, 1984). Second, IOIs in musical rhythms that differ by less than 10–15% are generally perceived as categorically identical (Clarke, 1987; Fraisse, 1956). The present model is limited to sequences in which IOI differences are perceived as categorically different.

PULSE PERCEPTION AS PATTERN RECOGNITION

ASSUMPTION: *The perception of pulse involves the spontaneous recognition of sequences that are perceived (categorically) to be isochronous.*

A pattern-recognition approach to pulse perception was described by Parncutt (1987) and Rosenthal (1992). In everyday visual scenes, a familiar pattern may be recognized even if parts of the pattern are missing, for example if a close object obscures part of a distant object. In the case of pulse perception, recognition is possible even if elements of the pulse are missing—for example, if there is a rest at the start of a measure of music, or if part of a pulse train performed by one instrument is masked by another instrument. Musical pulse sensations are generally illusory in that they do not necessarily indicate the presence of single periodic sound sources. In this regard, a pulse sensation is analogous to a musical chord that is perceived to blend or fuse into a single sound, even though it was created by three or more different sources.

The perception of pulse trains in musical rhythms may be conceptualized as a process by which sound onsets are matched against the elements of an *isochronous template.* A given template may be fully specified by two pieces of information: its period *P* and phase *Q*. In the present model, both *P* and *Q* are measured in basic time units *b*.

Phase is defined as the temporal position of an event in a pulse relative to some arbitrary reference time, such as the nominal start of a given rhythmic cycle. At the reference point, phase *Q* is set equal to 0 (instead of 1, for the first beat of the bar), as this simplifies the mathematics. The *waltz* rhythm of Figure 1, for example, evoked pulse sensations of period 3, both "on the beat" (beat = 1, phase = 0) and "off the beat" (beat = 3, phase = 2).

DEFINITION: *The salience of a pulse sensation is a measure of the probability that a listener will tap out that pulse when asked to tap the tactus (main or underlying beat) of a rhythmic sequence.*

This operational definition of pulse salience corresponds to the procedure of Experiment 1. The definition is reminiscent of the psychoacoustic definition of the pitch of a complex sound as the frequency of a pure tone judged to have the same pitch (American Standards Association, 1960). It is also reminiscent of the definition of pitch salience as the probability that a listener will select a pure tone of the corresponding frequency in a pitch-matching experiment (cf. Parncutt, 1989; Terhardt, Stoll, & Seewann, 1982).

The definition of pulse salience may be made more precise by requiring that listeners tap with their dominant hand and that they use the index

finger, middle finger, or both together. This additional specification is not crucial, as rhythmic imitation performance is largely independent of motor constraints such as specific combinations of hands and fingers (Keele, Pokony, Corcos, & Ivry, 1985; Michon, 1967; Mishima, 1965; Summers, Bell, & Burns, 1989).

ASSUMPTION: *Each pair of events in a rhythmic sequence initially contributes to the salience of a single pulse sensation, in proportion to the product of their phenomenal accents; and contributions to the salience of a pulse sensation from different pairs of events add linearly.*

This principle, proposed by Parncutt (1987), produces sensible results when incorporated into the present model. The word "initially" alludes to the assumption (below) that pulse sensations can enhance the salience of other, consonant pulse sensations.

According to the above assumption, as few as two events can produce a sensation of pulse. This conflicts with Krebs' claim (1987, footnote 16) that at least three regularly recurring events are needed to evoke a level of motion. Krebs pointed out that two equal IOIs are needed to imply isochrony, and that three events are needed to define two IOIs. According to this line of reasoning, a sequence of alternating 1/2 and 1/4 notes (the *waltz* sequence in Figure 2b) would never imply a pulse with a period of 1/4 note. Given that many listeners in Experiment 1 tapped at 1/4 note intervals in response to the *waltz* sequence, Krebs' hypothesis may be rejected in favor of the principle quoted above.

It is possible that the 1/4 note responses to the *waltz* rhythm in Experiment 1 were artifacts of the experimental method (Drake, personal communication). Listeners may have experienced no pulse sensation at all in response to *waltz* and chosen the 1/4-note pulse simply because it was the easiest to tap out. However, all the stimuli in that experiment were played at tempi found in real music, suggesting that all evoked at least one pulse sensation; and listeners were free to tap out any pulse sensation they wished. In any case, if the above definition of pulse salience is valid, then pulse salience is *defined* by the results of Experiment 1.

The above assumptions are incorporated into the following formulation of the goodness of fit, or *pulse-match salience* S'_m, of a pulse of period *P* and phase *Q*:

$$S'_m = \sum_{n=0,N-1} A_o \left(\mathrm{mod}_C \{nP + Q\}\right) A_o \left(\mathrm{mod}_C \{[n+1]P + Q\}\right), \qquad (4)$$

where $N = \mathrm{lcm}\,\{P, C\}$.

Here, Σ denotes summation, A_o is phenomenal accent (Eq. 2), mod_C means "modulo base C" (e.g., mod_3, in the case of *waltz* in Figure 1), lcm stands for lowest common multiple, and *N* is the minimum number of summation terms needed to estimate the salience of a pulse of period P evoked by a rhythm of cycle length C.

Povel and Essens (1985) estimated the salience of a pulse sensation (which they called the strength of an induced clock) by a method similar to the above pattern-matching procedure. A difference between the two formulations is that theirs was based on the amount of *counterevidence* a clock meets in an actual sequence—the number of clock ticks falling on unaccented or missing events (called *ghost events* by Rosenthal, 1992). Later in the same paper, they suggested that clocks are actually induced by positive evidence, that is, by accented events, as in the present model. Rosenthal (1992) determined pulse sensations (rhythmic levels) by a method similar to the present one, but did not explicitly evaluate a salience value for each pulse candidate.

It may be possible to improve Equation 4 in a future version of the model by reference to data on the *discrimination threshold for isochrony*—the sensitivity of listeners to small deviations from temporal regularity—on the assumption that this threshold is proportional to the salience of the corresponding pulse sensation (see the section below entitled Temporal precision and discrimination). Schulze (1989) investigated the discrimination threshold for isochrony as a function of the number of isochronous events in a sequence, for pulse periods in the range 50–300 ms. Sequences deviated from physical isochrony only at the last event. The threshold decreased as the number of events increased, approaching a minimum value of roughly 10 ms for sequences comprising more than about five beats (cf. 20 ms in musical contexts: Clark, 1989). Drake and Botte (in press) presented sequences in which the only deviation from isochrony was a slight change of tempo near the position of a missing event at the middle of the sequence. Their results imply that the relative just noticeable difference (JND) approaches an asymptotic minimum of 2–3% for quasi-isochronous sequences comprising four to eight sound events (referred to in the paper as two to four intervals). The relative JND was roughly constant over a wide range of periods (200–1000 ms) and sank as low as 1.5% for sequences of 8–12 events (four to six intervals) in the range 300–800 ms.

The next stage of the model involves the *normalization* of pulse-match salience so that it equals one in the simple case where the stimulus is an isochronous sequence and the pulse sensation corresponds exactly to the stimulus. In the experiments described above, this occurred when pattern (a) *pulse* was presented at a slow tempo, and listeners tapped at the same rate as the stimulus (i.e., without grouping). The purpose of the normalization is to remove any tempo dependence from calculated pulse saliences. This makes it easier to introduce explicit tempo dependence (Eq. 6). The normalization is effected as follows:

$$S_{\mathrm{m}} = \frac{S'_{\mathrm{m}}}{N\,(1 - \exp\{-p/\tau\})^{2i}}, \tag{5}$$

where N is defined as in Equation 4, and $p = Pb$ is the pulse period in milliseconds. (The expression in parentheses is derived from Eq. 3.)

MODERATE TEMPO AND THE EXISTENCE REGION OF PULSE SENSATION

The most salient pulse sensation (tactus) evoked by a rhythm depends on its tempo (Handel, 1986; Handel & Oshinsky, 1981). The present model formulates this dependency quantitatively by means of a function called the existence region of pulse sensation.

DEFINITION: *The existence region of pulse sensation is a range of periods within which isochronous sequences are perceived to be musically rhythmic (or to imply movement).*

ASSUMPTION: *The median, or (logarithmic) center, of the existence region of pulse sensation corresponds to a moderate musical tempo. In general, the closer the tempo of a pulse to moderate tempo, the greater the salience of the corresponding pulse sensation.*

Jones and Boltz (1989) referred to a *dominance region* that "reflects natural attending limits that are independent of ratio-complexity constraints . . . Specifically, the dominance region involves a set of related time levels around the referent beat period . . . Dominant region limits are relative not absolute time limits in that they are expressed in terms of hierarchical levels" (p. 473). The present concept of existence region of pulse sensation is similar to the concept of dominance region of Jones and Boltz, with the exception that the existence region is assumed to depend directly on *absolute* period.

The existence region of pulse sensation has an interesting analog in the area of pitch perception. Plomp (1967) and Ritsma (1967) found that the pitch of complex tones in speech and music is usually determined by harmonics other than the fundamental. Their results implied that the degree to which a given harmonic influences the overall pitch of a complex tone depends on both its absolute frequency and its harmonic number. Terhardt et al. (1982) treated effects of absolute frequency and effects of harmonic number as independent, associating effects of harmonic number with a pattern-matching (*subharmonic coincidence*) routine, and effects of absolute frequency with a bell-shaped *spectral frequency weight* function. The existence region of pulse sensation is similar to the spectral frequency weight function in that it is assumed to depend only on absolute period.

The shape and extent of the existence region of pulse sensation region may be determined in three different ways.

1. Spontaneous, personal, motor, or mental tempo. Subjects tap out isochronous sequences (Braun, 1927; Fraisse, 1963; Fraisse, Pichot, & Clairouin, 1949; Guttmann, 1931; Harrell, 1937; Harrison, 1941; Meumann, 1894; Miles, 1937; Miyake, 1902; Patterson, 1916; Reymert, 1923; Rimoldi, 1951; Seashore, 1899; Squire, 1901; Stern, 1900; Wu, 1935). Ideally, no prior suggestion is made about tempo; alternatively, tempo may be specified as moderate.
2. Preferred tempo. Pulse trains (e.g., metronome ticks) are played to listeners, who then indicate which rates they prefer or find most natural (Frischeisen-Köhler, 1933; Wallin, 1901, 1911, 1912).
3. Tactus. Listeners tap out the tactus (main or underlying beat) of a musical excerpt or rhythm, as in Experiment 1 (see also references on subjective rhythmization under Enhancement of pulse salience).

In each case, responses are limited to much the same range of periods. Perusal of the experimental literature permits the following generalizations (cf. Fraisse, 1982).

1. The most salient pulse sensations have a *moderate tempo* of about 100 isochronous events per minute, or a period of about 600 ms.
2. The region of greatest pulse salience (dominance region) lies between about 2/3 and 3/2 times moderate tempo: 67–150 events per minute, or 400–900 ms.
3. Pulse sensations cease to exist beyond about 1/3 or three times moderate tempo: 33–300 events per minute, or 200–1800 ms (cf. Bolton, 1894; Fraisse, 1956, 1982; Handel & Lawson, 1983; MacDougall, 1903; Michon, 1978).

Regarding the lower limit on the tempo of a pulse sensation, Lerdahl and Jackendoff (1981) pointed out that pulse sensations whose periods exceed two written measures are generally imperceptible. Such levels may be regarded as belonging to the domain of musical form rather than rhythm. Regarding the upper limit, the literature suggests that isochronous sequences faster than about 300 events per minute may sound rhythmic, but the pulse sensations produced by such sequences probably have periods corresponding only to multiples of the period of the original signal (subjective rhythmization) rather than the period itself.

The dependence of pulse salience on tempo is accounted for in the model by a *pulse-period salience* function S_p, a normal or Gaussian (bell-shaped) function of the logarithm of the pulse period p similar to that displayed in Figure 3 above:

$$S_p = \exp\{-\tfrac{1}{2}[\tfrac{1}{\sigma}\log_{10}(\tfrac{p}{\mu})]^2\}. \qquad (6)$$

Here, μ denotes *moderate pulse period* and is typically about 600 ms; and σ is the standard deviation of the logarithm of pulse period, a measure of the width of the existence region, and typically has a value of about 0.2. The variables μ and σ are the third and fourth free parameters of the model. The function may be regarded as a kind of band-pass filter, admitting only pulse sensations that lie within a given range of periods.

The bell-shaped function defined in Equation 6 is quite broad, with a relatively poorly defined peak. The model is thus relatively unaffected by small variations in μ and σ. Most published experimental results on spontaneous motor tempo, the tempo of pulses perceived in response to rhythms, and preferred tempo are consistent with this function.

Fraisse (1956, 1982) asked people to tap at equal time intervals and to group the taps into threes or fours. He reported that intertap periods were considerably shorter than the usual 600 ms, averaging 420 ms (for groups of three) and 370 ms (for fours). Similar results had been obtained in earlier studies (MacDougall, 1903; Miles, 1937; Miyake, 1902; Wallin, 1901). These findings may be explained with reference to the existence region of pulse sensation. When subjects are asked to group taps into threes or fours, they presumably choose tempi that will maximize the *aggregate* salience (rhythmicity?) of all pulse sensations evoked by the sequence. The most salient pulses in Fraisse's experiment presumably corresponded to the tapping period and the group length. Due to the concave-downward (bell) shape of the existence region (Eq. 6), the sum of the saliences of these two pulses is maximal when the two periods lie on either side of peak of the curve—that is, when one pulse has a period of considerably less than 600 ms, the other considerably more than 600 ms.

The existence region of pulse sensation is assumed here to be a continuous function of pulse period. A recent study of synchronization (Collyer, Broadbent, & Church, 1992) suggests that this assumption may not be entirely valid. In that study, listeners tapped in time with an isochronous sequence and continued tapping after the stimulus was turned off. Systematic deviations from the original tapping periods during the continuation phase suggested the existence of temporal categories in motor control. This effect seems unlikely to affect musical applications of the present model.

ASSUMPTION: *The above two contributions to pulse salience—goodness of fit between the rhythm and the corresponding isochronous template, and the tempo of the pulse—are essentially independent of one another.*

The overall salience S of a pulse sensation is estimated in the model by multiplying its pulse-match salience S_m by its pulse-period salience S_p:

$$S = [S_m S_p]^j. \tag{7}$$

The symbol j is the so-called the *pulse-salience index,* the model's fifth and final free parameter. Like i in Equation 3, j generally exceeds one. It exaggerates differences between pulse saliences, thereby reducing the ambiguity of the tactus of a rhythm, and may therefore be expected to increase with listeners' confidence about the location of the underlying beat of a rhythmic sequence.

The salience of pulse sensations may also be affected by motivic repetition (Lee, 1991; Lerdahl & Jackendoff, 1983; Longuet-Higgins & Lee, 1982, 1984; Steedman, 1977). The repetition of temporal patterns may be accounted for systematically in a model of rhythm perception by the mathematical technique of autocorrelation (Brown, 1993). However, no effect of motivic repetition was evident in the present analysis of the results of Experiment 1: The salience of pulse sensations whose period corresponded to the cycle length did not consistently exceed predictions (see Figure 2). Perhaps the effect of motivic repetition only affects pulse perception in more interesting (noncyclic) rhythmic patterns or in melodic motives that are defined by both rhythm and contour.

METRICAL ACCENT

According to Lerdahl and Jackendoff (1983, p. 17), "Phenomenal accent functions as a perceptual input to metrical accent." The present model clarifies this relationship by breaking it up into two stages. First, phenomenal accents (here restricted to durational accents) determine the salience of pulse sensations. In the second stage, pulse saliences determine the strength of metrical accents (Parncutt, 1987).

Longuet-Higgins and Lee (1984) have noted that "The weight of a given note or rest is the level of the highest metrical unit that it initiates" (p. 430). According to Rosenthal (1992), the salience of an event is "an appropriate tally of properties of the note that make it more likely to be considered a downbeat" (p. 68). In the present model, metrical accent is estimated according to the following principle.

ASSUMPTION: *The metrical accent of an event or temporal category (including a missing event such as a rest at the start of a bar) may be estimated by combining the saliences of all pulse sensations including that event or temporal category.*

Specifically, metrical accent A_m at time T is estimated by linear addition of the saliences of all the pulse sensations converging on time T:

$$A_m(T) = \sum_{P,Q} S_p(P, Q) \tag{8}$$

where $\mathrm{mod}_C\{Q + nP\} = T$ for some integer n.

In practical applications of Equation 8, computation time may be saved by limiting pulse periods to a maximum of about 2000 ms. Note also that results as a function of phase Q repeat cyclically with a period of $\mathrm{hcf}\{P, C\}$, where hcf stands for highest common factor. Computation time may be saved by restricting calculations to phases Q in the range 0 to $\mathrm{hcf}\{P, C\} - 1$.

The described concept of pulse sensation includes *hypermeters,* that is, pulse sensations whose periods exceed that of the notated meter (Rothstein, 1989). So the above formulation of metrical accent also includes Rothstein's concept of *hyperaccent.* In the present approach, hypermeters must be fast enough to contribute directly to the music's rhythmic feel (that is, more than 30–40 beats per minute) if they are to be regarded as pulse sensations. Slower hypermeters may subdivide the music into formal chunks that happen to be of approximately equal length, but do not imply rhythmic movement.

MODELING EXPERIMENTAL RESULTS

The above model was tested by comparing calculations with the results of the experiments reported above. Values for the free parameters of the model were estimated by an automatic procedure implemented in the language Le_Lisp. The procedure adjusted parameter values by small steps until the Pearson correlation coefficient r between measurements and calculations was a maximum. The results of this procedure are shown in Table 1. These results were independent of initial parameter estimates for a range of feasible sets of initial values.

In Experiment 1 (pulse salience), measurements were compared with calculations according to Equation 7. Computation time was reduced by considering only those pulses (periods, phases) that had been selected in the experiment (161 data points). Calculated values were first normalized so that, for each trial of the experiment, their sum (like the sum of the experimental data) was 22, the number of listeners in the experiment. Then all five free parameters (τ, i, μ, σ, j) were gradually and independently varied until the correlation between calculations and experimental data was optimal ($r = .88$). This value is high in view of the relatively large number of degrees of freedom in the comparison: 161 (data points) − 36 (normalization) − 5 (free parameters) − 2 (correlation) = 118. The calculated pulse saliences from which this value was determined appear in italics in Figure 2.

TABLE 1

Optimal Values of the Free Parameters

Parameter	τ	i	μ	σ	j	r
1. Pulse salience						
a. Individual pulses	550	1.6	760	0.23	2.0	0.88
b. Distribution of periods	—	—	710	0.22	—	—
2. Accent						
a. Phenomenal	200	2.0	—	—	—	0.87
b. Metrical	180	1.7	660	0.14	1.9	0.95
3. "Typical"	500	2	700	0.2	2	—

NOTE.—Definition of parameters: τ is the saturation duration, or the duration of the echoic store, in milliseconds (Eq. 3); i is the accent index (Eq. 3); μ is the moderate pulse period in milliseconds (Eq. 6); σ is the width of the existence region of pulse sensation, expressed as a standard deviation of log periods (Eq. 6); j is the pulse index (Eq. 7); and r denotes the coefficient of correlation obtained in each case between calculations and experimental results, by using the listed parameter values.

The optimal values of the free parameters for Experiment 1 (Table 1, line 1a) make sense in terms of their interpretations in the model. In particular, the value of the optimal pulse period found by this procedure (μ = 760 ms), and the relatively large value for the standard deviation of its logarithm (σ = 0.23), are consistent with most literature on the center and width of the existence region of pulse sensation, and the saturation duration τ (550 ms) is consistent with most literature on the duration of the echoic store.

The 26 results of Experiment 2 (metrical accent) were compared directly with calculated metrical accents according to Equation 8. The results are shown in line 2b of the table. Consistent with the lower number of degrees of freedom in this comparison [26 (data points) − 5 (free parameters) − 2 (correlation) = 19], the correlation coefficient was higher than before (r = 0.95). The optimal values of the parameters i, μ, and j were about the same as for Experiment 1. The saturation duration τ and existence-region width σ were considerably smaller than in Experiment 2, presumably because of the restricted range of tempi in that experiment—all rhythms were played at the same rate of 150 notes per minute. The IOIs in Experiment 1 more thoroughly covered the durational accent saturation curve (Eq. 3) and the pulse sensation existence region (Eq. 6); so the values of τ and σ estimated from the results of Experiment 1 are more likely to be musically representative or typical.

The results of Experiment 2 were also compared directly with durational accents according to Equation 3, in order to obtain additional estimates of the first two free parameters τ and i (line 2a in Table 1). The resultant correlation coefficient was considerably lower for phenomenal than for metrical accent (.87 versus .95), suggesting that the measured

event saliences were indeed metrical rather than phenomenal in nature—in accordance with the instructions given to listeners in the experiment, which made direct reference to downbeats and syncopation.

The "typical" values of the free parameters in the bottom line of Table 1 are intended for music-theoretical applications of the model.

Extensions and Applications

The model as set out and tested above is sufficiently coherent, parsimonious, and accurate for a range of music-theoretical applications. However, it could be changed in many ways to satisfy future requirements or as new theoretical possibilities emerge. The model may also be extended to account for a range of rhythmic phenomena not dealt with in the described experiments. Some possible extensions are set out below.

RHYTHMIC CONSONANCE AND PERCEIVED METER

Yeston (1976) described two strata (pulse sensations) as *rhythmically dissonant* if their note values were not whole multiples or factors of each other. Krebs (1987) referred to such a relationship as a *type A dissonance,* and went on to define *type B dissonances* as nonaligned (or displaced) combinations of pulse trains whose periods were whole multiples or factors of each other. These definitions were based on Hlawicka's (1958) distinction between *Verwechslungen* and *Verschiebungen,* respectively. Krebs cited several examples of type A and B dissonances from Western art music.

Common type A dissonances are *polyrhythms* such as 2 against 3, 3 against 4, and so on. Studies on the perception of polyrhythms (Beauvillain & Fraisse, 1984; Handel & Oshinsky, 1981) have shown that one of the two dissonant pulses can act as a perceptual frame of reference for the other, but that this happens only at relatively fast tempi—specifically, for reference pulse periods smaller than about 500 ms. This finding may be explained in terms of the existence region of pulse sensation. Consider, for example, a 2 × 3 cross rhythm of 1/4 and 1/6 notes. The superposition of these two pulses may produce an additional pulse sensation of 1/12 notes, here called an *interaction pulse.* For one of the two main pulse trains to be perceived as a frame of reference, it should evoke a considerably stronger or more salient pulse sensation than the interaction pulse. Because of the bell shape of the existence region of pulse sensation, this is possible only if the period of the reference pulse is smaller than the moderate pulse period (600–700 ms), so that both reference and interaction pulses lie on the same (left) side of the peak. If this is the case, the interaction pulse will be much weaker than the reference pulse.

ASSUMPTION: *A prerequisite for the perception of meter is the concurrent perception of consonant pulses.*

According to Yeston (1976), meter may be regarded as an outgrowth of the interaction between two or more consonant pulse sensations (or rhythmic strata). For example, a 3/4 meter may be specified by a pulse of 1/4 notes (the beats) and a pulse of 3/4 notes (the measures or downbeats), where every third 1/4-note coincides with a 3/4-note. This combination satisfies Lerdahl and Jackendoff's (1983) metrical well-formedness rule no. 2 (p. 69): "Every beat at a given level must also be a beat at all smaller levels."

A consonant set of pulses may be conceived as a hierarchical structure. Lerdahl and Jackendoff rreferred to specific elements of hierarchical structures, including consonant rhythmic levels, as *subordinate* or *superordinate* to each other. Rosenthal (1992) used the equivalent terms *parents* and *children* (for pulses two or three times slower or faster than a given pulse), and *ancestors* and *descendants* (for more distant consonant pulses). For example, the 1/4-note pulse in a 3/4 meter is a *child* of the 3/4-note pulse; an 1/8-note pulse would be a grandchild. Following Rosenthal's terminology, a perceived meter may be regarded as a *family* of consonant pulse sensations.

HYPOTHESIS: *The salience of a perceived meter, or the probability of a given metrical interpretation, is proportional to the sum (or some other aggregate) of the salience of the pulse sensations that make it up. The most likely meter to be perceived is the one with the highest predicted salience.*

This aspect of the model has not yet been systematically tested. An example of the procedure using a simple summation is shown in Parncutt (1987); see also Rosenthal (1992).

ENHANCEMENT OF PULSE SALIENCE

HYPOTHESIS: *The salience of a pulse sensation may be enhanced by the presence of a parent or child pulse sensation, that is, a consonant pulse sensation two or three times slower or faster. In other words, there is mutual salience enhancement among consonant pulse sensations.*

This hypothesis is capable of explaining the following two rhythmic phenomena. First, in *subjective rhythmization,* subjective grouping by fours is more common than grouping by threes. Subjective rhythmization, or the perceptual grouping of isochronous sound events at moderate to fast tempi, has a long history in experimental psychology (Bolton, 1894; Duke, Geringer, & Madsen, 1991; Fraisse, 1956; MacDougall, 1903; Mach, 1886; Meumann, 1894; Miner, 1903; Miyake, 1902; Temperley, 1963; Vos, 1973; Woodrow, 1951). Subjective rhythmization appears to involve both serial and periodic grouping, where the two types of grouping are in

phase—the start of each serial group corresponds to a periodic group (downbeat). In the case of grouping by fours, subjective accents occur twice per group, on the first and third elements. This implies that subjective grouping by fours involves a family of three consonant pulse sensations, of periods 1, 2, and 4 beats—as in 4/4 meter—whereas subjective grouping by three involves only two consonant pulse sensations, of periods 1 and 3 beats—as in 3/4 meter. The enhancement of pulse salience due to the presence of other, consonant pulse sensations would therefore be expected to be greater in the case of subjective grouping by fours, making grouping by fours more likely than grouping by threes (as in Experiment 1).

The predominance of binary/quaternary over ternary grouping is explicit in the rhythmic theory of Lerdahl and Jackendoff (1983) and evident in experimental data obtained from both adults and children by Drake (1993). In the present model, the predominance of binary groupings is implicit: It is not stated directly, but arises from a combination of two related assumptions. The first assumption is that pulse sensations are confined to the existence region of pulse sensation. The number of consonant pulse sensations that may exist simultaneously within this region is greatest if the relationship between the pulses is binary (e.g., periods 200, 400, 800, 1600 ms). The second assumption is that consonant pulse sensations enhance one another.

An alternative explanation for the predominance of binary or quaternary groupings over ternary groupings is cultural conditioning. Binary groupings (2/4, 4/4 meter) are more common in our musical culture, and hence more familiar, than ternary groupings (Apel, 1970). The above account sheds some light on the origin of this difference.

The second phenomenon that may be explained by pulse salience enhancement is the following. A sequence of alternating IOIs in the ratio 2:1 appears to evoke a pulse sensation of period 1, but alternating IOIs in the ratio 3:1 (e.g., a dotted rhythm) do not. In other words, an isochronous sequence may be perceptually subdivided by a single additional event into three parts (2:1), but not four or more parts (3:1 or more). This claim is supported by two pieces of evidence, stemming from research on the categorical perception of rhythmic patterns and performance timing.

Regarding categorical perception, Fraisse (1956) found that listeners generally distinguished only two time categories in musical rhythms—long and short, in the perceived ratio 2:1—and categorized IOIs as the same (i.e., both long, or both short) if the ratio of their IOIs was less than about 1.55:1. He also found that sequences of 2:1-related IOIs were judged more rhythmic than other sequences. Such sequences are easier to encode and to reproduce than other sequences (Drake & Gérard, 1989; Essens & Povel, 1985; Povel, 1981; Summers et al., 1989). Summers et al. (1986) found that musicians could reproduce repetitive patterns of two IOIs specified only by their ratios, but that 2:1 was produced more accurately than 3:1, 3:1 than 3:2 and 4:3, and 4:1 than 3:2.

Regarding performance timing, the performance rules of Sundberg (1988) imply that alternating IOIs (notated) in the ratio 3:1 sound more musical or natural if the contrast between the two IOIs is heightened, or made to exceed 3:1; whereas alternating IOIs in the ratio 2:1 sound more musical or natural if the contrast between the two IOIs is softened, or made smaller than 2:1. Similar conclusions may be drawn from earlier measurements of performance timing (e.g., Gabrielsson, Bengtsson, & Gabrielsson, 1983; Seashore, 1938).

The difference between 2:1 and 3:1 ratios in alternating IOI sequences may be accounted for as follows.

HYPOTHESIS: *The salience of a pulse sensation in which no more than two successive events match real events is negligible unless the pulse sensation is enhanced by a salient parent or child pulse sensation.*

According to this hypothesis, alternating IOIs in the ratio 2:1 imply a pulse of period 1, supported by the parent pulse of period 3; but alternating IOIs in the ratio 3:1 do not imply a pulse of period 1 (as the present model would predict), as there is a gap of two generations between period 1 and period 4. For example, a dotted rhythm of alternating 3/8-notes and 1/8-notes does not, according to this hypothesis, imply a pulse of 1/8-notes, unless one or more rhythmically consonant 1/4-notes are present in the immediate context (bridging the gap between the two outer pulse "generations"). In the absence of 1/4-notes, the fastest pulse sensation in this case is probably the 1/2-note pulse.

The model used to simulate the results of the above experiments ignored this effect and may thus have predicted some pulse sensations that are actually imperceptible. However, the salience of such pulse sensations according to Equation 4 is generally quite low, due to the relatively low phenomenal accent assigned to shorter IOIs by Equation 3.

The model of Povel (1981) allowed only two different subdivisions of a pulse period: a number of equal intervals, or two intervals that relate as 2:1. Subdivisions in the ratio 3:1 were disallowed, unless equal (1:1) subdivisions were also prsent (e.g., 2:1:1, or 3:1:2:2). These restrictions have essentially the same effect as the above hypothesis. The above hypothesis has the advantage that it relates the effect of allowable beat subdivisions to the more general effect of pulse salience enhancement.

CATEGORICAL PERCEPTION

Rhythm perception is categorical (Sloboda, 1985): Sounds heard in a metrical context are perceived to correspond to specific (notable) beats (Povel, 1981). As discussed earlier, the distinction between even and uneven subdivisions of a timespan is also perceived categorically (Clarke, 1987; Schulze, 1989).

The assignment of rhythmic events to temporal categories is a complex process. For example, Clarke (1987) investigated the position of the category boundary between alternative interpretations of rhythmic figures (such as dotted, ternary, and binary subdivisions of a beat) and found that the boundary tended to shift so as to favor perceptual judgments that conform to the prevailing metrical context. Desain and Honing (1989) modeled the categorization of musical time by a connectionist system.

The rhythms tested in the present study comprised IOIs in metronomically exact, small-integer ratios, and the model presented above was tested only against the results of these experiments. Given the categorical nature of rhythm perception, the model may be assumed also to apply to freer or more musical renditions of simple rhythms, provided that temporal deviations remain within category boundaries. It is nevertheless possible that small deviations from mechanical timing may affect pulse salience. Effects of this kind may be accounted for explicitly in an event-based model such as Desain (1992). The present, pulse-based model does not account for such effects.

In the above formulation of the model, it was assumed that every event in a musical rhythm corresponded to a temporal category. Essentially the same assumption was made by Lerdahl and Jackendoff (1983): "Every attack point must be associated with a beat at the smallest level of metrical structure." (Metrical Well-Formedness Rule no. 1, p. 69). However, two phenomena—the tendency for listeners to perceive all IOI ratios as either 1:1 or 2:1 (Fraisse, 1956), and idiosyncratic differences between the timing of 2:1 and 3:1 ratios in musical performance (Sundberg, 1988)—suggest that Lerdahl and Jackendoff's assumption may be invalid. Given that an alternating sequence of IOIs in the ratio 2:1 can evoke a pulse of period 1 unit but a sequence in the ratio 3:1 cannot (unless elements of IOI = 2 units are also present), and that the event of unit IOI has metrical accent in the 2:1 case but not in the 3:1 case, it is reasonable to hypothesize that the shorter event corresponds to a temporal category in the 2:1 case but not in the 3:1 case.

HYPOTHESIS: *There is a one-to-one correspondence between metrical accents and temporal categories.*

Consider two sequences of alternating IOIs, one in the nominal ratio 3:1 (notated as a dotted rhythm) and the other in the nominal ratio 7:1 (double-dotted rhythm). The previous hypothesis on pulse salience enhancement implies that, for both of these sequences, the shorter IOI has no metrical accent. The present hypothesis further suggests that, in both cases, the shorter IOI is not perceived as a temporal category. In other words, there is no categorical perceptual difference between cyclically repeating dotted and double-dotted rhythms. Double-dotted rhythms are of course distinguishable from dotted rhythms (they sound "sharper"), but the difference is perceived on a continuous scale.

EXPRESSIVE TIMING

Clarke (1987) distinguished between meter and beat subdivision, which are perceived and performed categorically, and *expressive transformations* (temporal deviations from mechanical timing, or metronomic note values), which are continuous. Expressive transformations depend in a complex fashion on the nominal pitches and time values of a musical score. The quantization of event salience in the present model may serve as a basis for a new approach to the simulation of expressive transformations.

HYPOTHESIS: *Expressive timing may be modeled by a gradual slowing of local tempo in the vicinity of metrical and structural accents. The amount of slowing in the vicinity of an accent increases as the strength of the accent increases.*

Musical experience suggests the following relationship between exact musical timing and event salience: In musical performance, important or remarkable events (e.g., events preceding relatively long IOIs, on rhythmically strong beats, at the starts of phrases, at harmonic dissonances, or at phrase or structural boundaries) may be slightly *delayed* relative to metronomic tempo by *micropauses* (cf. Bengtsson, 1987; Bengtsson & Gabrielsson, 1983; Gabrielsson, 1974; Palmer, 1989; Repp, 1990; Seashore, 1938; Shaffer & Todd, 1987; Sundberg, Askenfelt, & Frydén, 1983; Thompson, Sundberg, Friberg, & Frydén, 1989). Usually, the more important the phrase or structural boundary, the longer the micropause. In other words, the duration of the pause depends on the perceptual salience of the event starting the next phrase. This may be estimated from the number of hierarchical levels whose boundaries coincide at that point (Todd, 1985).

Research on deviations from metronomic tempo has also shown that relatively important events in a musical passage tend to be *lengthened,* and played *more loudly* or *more legato* (Clarke, 1988; Shaffer, 1981; Sloboda, 1983; Stetson, 1905; Todd, 1985; Vos, 1977). In musicology, lengthening a note beyond its nominal IOI (i.e., beyond its notated value) is said to give the note *agogic accent* (Riemann, 1884). Agogic accent should not be confused with durational accent (as defined in Eq. 3), which is assumed to depend only on physical IOI, regardless of nominal IOI (i.e., regardless of musical notation and context).

Important events are thus both delayed and lengthened in musical performance. This is equivalent to saying that *local tempo* (otherwise known as instantaneous, or micro, or linear tempo) slows in the vicinity of important events.

A continuous function of local tempo against time may be formulated mathematically by analogy to the velocity and momentum of moving objects (Kronman & Sundberg, 1987; Sundberg & Verrillo, 1980; Todd, 1992). Such a function may possibly be used as a heuristic for producing musically acceptable performances. Of course, local tempo depends on the events of the rhythm and cannot be perceived independently of them (Desain & Honing, 1991; Gibson, 1975).

In musical performance, variations in local tempo may be either perceptible or imperceptible. The delay of important events (i.e., events bearing strong metrical or structural accents) is a case in point. The existence of imperceptible delays was demonstrated by Drake, Botte, and Gérard (1989), who presented an isochronous sequence followed by a longer event and found that considerable delays in the onset of the longer event (up to 5% of the IOIs of the preceding events) could not be detected. Another example of imperceptible variations is the softening of durational contrasts in performances of alternating IOIs in a nominal 2:1 ratio. In general, performance timing deviates systematically from metronomic timing, even when performers are asked to play mechanically or without expression (Clarke, 1985; Seashore, 1938)—although "mechanical" performances are in some cases preferred to performances by live musicians (Wapnick & Rosenquist, 1991).

When tempo variations are perceptible, they are called *rubato.* Rubato generally involves a gradual or continuous variation in tempo (Robert, 1981). It can function both as a vehicle of musically expression and as a clarifier of musical structure (Clarke, 1982, 1985; Gabrielsson et al., 1983; Palmer, 1989; Shaffer, 1981; Sloboda, 1983).

A slowing of local tempo in the vicinity of accents could explain why 2:1 ratios tend to be softened in music performance whereas 3:1 ratios (dotted rhythms) are sharpened (as noted above). In the 2:1 case, both events have metrical accent, the greater accent falling on the onset of the longer IOI. A slowing of local tempo in the vicinity of the greater accent would cause the IOI ratio to be reduced (Parncutt, 1987). The ratio may be formulated $(2+\delta):(1+\epsilon)$, where both δ and ϵ are small positive real numbers. When $\epsilon = \delta$, the ratio is less than two. Musical experience suggests that the magnitude of the delay of important events is usually greater than the magnitude of their lengthening, so that $\epsilon > \delta$, causing even more softening of the 2:1 ratio.

In the 3:1 case, the event initiating the shorter IOI does not belong to a pulse sensation and so has no metrical accent (provided no other IOI ratios such as 3:1:2:2 occur in the same context). The temporal position of the unaccented event is determined only by temporal proximity to its accented neighbor. Musical experience suggests that the perceived relationship between the two events may be strengthened simply by reducing this distance. An interesting analogy in the area of musical intonation is the sharpening of the leading tone in performance (Nickerson, 1949; Ward, 1970). Apparently, sharpening the leading tone strengthens its melodic or voice-leading relationship with the tonic scale degree, thereby strengthening the overall (melodic and harmonic) relationship between tonic and dominant harmonies (Parncutt, 1989). Similarly, sharpening the 3:1 IOI ratio may yield a "stronger" rhythmic feel. This effect may underlie the baroque performance convention whereby rhythms notated as dotted (nominally 3:1) are often performed as if they were double-dotted (nominally 7:1).

Expressive timing in music performance depends on internal representations of structure (Clarke, 1985, 1988, 1989; Palmer, 1989; Povel, 1984; Todd, 1985, 1992). However, analytical representations of rhythmic structure have so far failed to produce consistently good artificial performances of a reasonable range of pieces. Rosenthal (1992) has suggested that a successful algorithmic model of pulse sensation and perceived meter may aid in the creation of more successful structural representations of rhythm, and hence improve models of timing. For example, algorithmic formulations of phenomenal and metrical accent may be used to model the temporary slowing of local tempo in the vicinity of perceptually important events, as described above. If this approach is successful, the formulations could then be incorporated into a set of rules for the automatic, yet musically satisfying, performance of rhythms, along the lines of the performances rules described by Sundberg (1988).

TEMPORAL PRECISION AND DISCRIMINATION

HYPOTHESIS: *The precision with which an event can be timed, and the precision with which a deviation in the timing of an event can be detected, depend on the metrical accent of that event.*

Two pieces of evidence support this hypothesis. First, experimental data on performance timing suggest that the timing of metrically stronger beats is more precise (i.e., less variable) than the timing of metrically weaker beats. For example, the data of Shaffer (1981) and Shaffer et al. (1985) imply that beats belonging to the underlying tactus are timed more accurately than other beats. The second piece of evidence concerns the existence region of pulse sensation. This region corresponds quite well with the region of greatest sensitivity in synchronization and durational discrimination tasks. Fraisse (1967) presented listeners with two interpolated

pulse trains, each of which was isochronous, and asked them to adjust the time interval between the two sequences to produce isochrony (at double the tempo of each separate sequence). In another experiment, listeners were asked whether the sequence was synchronous or not. Performance was optimal for periods in the vicinity of 500–600 ms (typical range: 200–900 ms), with discrimination thresholds in the vicinity of 4–5% (typical range: 2–8%). Similar experiments on duration discrimination and isochrony were performed by Eisler (1975) and Michon (1964). Fraisse (1982) concluded that synchronization is possible only within the range 200–1800 ms, corresponding to the boundaries of the existence region of pulse sensation.

PRIMACY, RECENCY, AND THE PSYCHOLOGICAL PRESENT

Real musical rhythms are not usually cyclic, but change as they unfold in time. Consequently, the relative saliences of the various pulse sensations evoked by a musical rhythm also vary with time. The model described above is limited to cyclically repeating rhythms and so does not account for such variations. The model may be extended to cover ordinary, noncyclic rhythms by accounting for *primacy* and *recency* effects, according to which the events near the start of a rhythm and near the observation time have a greater influence on the salience of pulse sensations (and hence on perceived meter) than intervening events. Both primacy and recency effects are assumed here to involve the duration of the psychological present.

ASSUMPTION: *The perception of pulse is confined to a limited time span known as the psychological present.*

Lerdahl and Jackendoff (1983, p. 21) pointed out that "Metrical structure is a relatively local phenomenon." The period in time over which metrical structure exists is called the *psychological present* (Fraisse, 1963), otherwise known as the *span of consciousness* (Wundt, 1874), or *specious present* (James, 1950). The psychological (or perceptual) present may be regarded as a continuous time interval comprising all real-time percepts and sensations simultaneously available to attention, perception, and cognitive processing. It enables a temporal sequence to be processed as a unit. It may be regarded as a nonverbal short-term memory possessing a high degree of structure and organization (Deutsch, 1975). Its duration may be measured in various ways, for example by testing the ability of listeners to reproduce sound sequences of various durations. A typical estimate of the duration of the psychological present is 4 s, with experimental results ranging from about 2–8 s (Fraisse, 1963; Michon, 1978). Fricke (1989) proposed a period of 7–15 s. It should be stressed that "because the present is so highly adaptive, no fixed parameter values can be expected to describe it adequately" (Michon, 1978, p. 89). The duration of the psychological present also depends on the time intervals between successive percepts—the faster the presentation rate, the shorter the memory span. The maximum number of successive sounds that can be perceived as a unit has been estimated at 25 (Dietze, 1885; Fraisse, 1982).

In the perception of language, the psychological present enables the different parts of a sentence or clause to be related to each other and hence understood (Fraisse, 1982; Wallin, 1901). In music, the psychological present enables events in a phrase to be related to each other. It appears that both serial and periodic grouping are confined to the psychological present.

ASSUMPTION: *Recent events in a rhythm have more influence on the salience of pulse sensations than earlier events.*

The effect of the psychological present on the temporal development of pulse salience may be modeled simply by giving more weight to more recent events, a procedure that is consistent with the enhanced recall of recent auditory events in rhythmic sequences (as found, for example, by Glenberg & Swanson, 1986). In a first approximation, all phenomenal accents may simply be multiplied by a factor of $\exp\{-t_1/\psi_1\}$, where t_1 is the time elapsed between a given event and the observation time, and ψ_1 is the effective duration of the psychological present in the case of the recency effect.

ASSUMPTION: *Events near the start of a sequence have a greater influence on the salience of pulse sensations than later events.*

The perceived meter of a sequence is generally established quite early. According to Lee (1991), "listeners initially assume that the downbeat falls on the first note and the 'second beat' on the second note, and are reluctant to revise their assumptions" (p. 98). Longuet-Higgins and Lee (1982) noted that early events "allow the establishment of metrical hypotheses which can then be challenged but not necessarily overthrown by later evidence" (p. 118) (see also Longuet-Higgins & Steedman, 1971; Steedman, 1977). In other words, pulse sensations and perceived meter are established quite quickly and are subsequently quite resilient in the face of potentially disruptive syncopations. Only when the degree or strength of syncopation passes a certain threshold will the perceived meter change.

The duration of the primacy effect would appear to be of the same order of magnitude as the duration of the psychological present. In a first approximation, the effect may be modeled by multiplying phenomenal accents by an expression of the form $1 + k \cdot \exp\{-t_0/\psi_0\}$, where k is a measure of the extent of the primacy effect, t_0 is the time elapsed between the start of the sequence and the event, and ψ_0 is a measure of the duration of the psychological present in the case of the primacy effect. Note that ψ_1 and ψ_0 are not necessarily equal. They will each need to be determined experimentally and can be expected to vary as a function of the perceived rate of the music.

A DEFINITION OF MUSICAL RHYTHM

The word rhythm may be used to describe any form of temporal periodicity observable in the physical universe, from molecular vibrations (period: about 10^{-13} s) right up to the "rhythm" of expansions and contractions of the universe between hypothetical big bangs (period: about 10^{18} s). By comparison to this inconceivably broad spectrum, the periodicities of musical rhythms are confined to a very narrow range—roughly 200–1800 ms.

Musical rhythm has been defined in various ways. Plato defined rhythm as order in movement—a definition adopted by Fraisse (1982). Aristoxenes defined rhythm as organization of time spans (Eggebrecht, 1967). Martin (1972) defined rhythm as "temporal patterning" (p. 487). Scholes (1970) gave a more general definition of rhythm: "everything pertaining to what may be called the time side of music." The Riemann Musiklexikon (Eggebrecht, 1967) was more specific, referring to integrated temporal principles of organization and performance in dance, music, and poetry and emphasizing the dichotomy between the foreground sequence of rhythmic elements (serial grouping) and the constant underlying movement (periodic grouping). According to Shaffer et al. (1985, p. 63), "Musical rhythm arises as the interaction of several factors, the time values of notes and rests forming the melodies, the patterning of metrical accents, the hierarchy of melodic groupings and the harmonic movement within these groupings."

A problem with these definitions is that they do not clearly distinguish musical rhythm from nonrhythm in specific cases. Most definitions refer to temporal organization, but temporal organization—at least in the form of serial grouping—is common to both speech and music.

Martin (1972), Jones (1976), and Desain (1992) stressed the importance of *expectancies* in rhythm perception; however, expectancies also occur in nonrhythmic sequences. For example, a particular word may be expected at a particular temporal position in a sentence. This is not a rhythmic effect in the musical sense, even though the expected temporal position may be specified quite precisely. In this light, it is inappropriate to define musical rhythm in terms of expectancy.

Musical rhythms often include IOIs that approximate simple ratios of small integers. However Povel and Essens (1985) showed that sequences composed entirely from IOIs in simple integer ratios (e.g. 1:2 or 1:3) do not necessarily evoke pulse sensations (internal clocks) and so are not necessarily rhythmic in the musical sense. In any case, a definition of rhythm based on approximations to simple ratios of small integers would be plagued by the arbitrariness of the concepts of approximation and simplicity.

Periodic perceptual grouping appears to represent the most appropriate basis for a definition of musical rhythm. Periodic grouping is typical in music, but relatively unusual in speech—with the possible exception of mechanical renditions of poetry. Periodic grouping, or pulse sensation, is thus capable of distinguishing musical rhythm from the rhythm of regular speech. Consider the following concise definition.

DEFINITION: *A musical rhythm is an acoustic sequence evoking a sensation of pulse.*

This definition is reminiscent of the generally accepted definition of a tone as a sound that evokes a sensation of pitch (American Standards Association, 1960). Both definitions are clear and simple, and they correspond closely to the meanings of the terms tone and rhythm in everyday language.

A sensation of pulse may be regarded as a trivial form of expectancy. Once a pulse sensation is established, events are "expected" at equal time intervals. A definition of rhythm based on pulse sensation thus automatically incorporates the notion of expectancy.

The existence of pulse sensations in musical rhythms may be demonstrated by analyzing the timing of rhythmic production or musical performance. Inspired by the work of Lashley (1951), Martin (1972) defined rhythm as "relative timing between adjacent and non-adjacent elements in a behavior sequence" (abstract), implying that the temporal position of each sound event in a rhythmic sequence may be determined by *all* other events in the psychological present—not just immediately preceding and following events. In an analysis of performances of Algerian and Caribbean music, Kluge (1990) found that adjacent IOIs within a phrase were negatively correlated—if one was lengthened, its neighbor was shortened, and vice-versa—however, this was not the case *between* phrases. These results are consistent with the existence of pulse sensations within limited time spans. The negative correlation between timing deviations of adjacent events suggests that a strong pulse feel was maintained by monitoring the timing of nonadjacent events.

HYPOTHESIS: *The greater the (aggregate) salience of the pulse sensation(s) evoked by a sequence, the greater the sequence's overall rhythmicity.*

Povel (1984, p. 317) considered that "the most outstanding characteristic of a rhythm is what may be called its rhythmical value [i.e., its rhythmicity], which tentatively can be described as the amount of tension that accompanies the perception of a rhythm." Experimental results of Povel (1981), Essens and Povel (1985), Summers, Hawkins, and Mayers (1986), and Drake and Gérard (1989) further suggest that the greater the rhythmicity, the easier it is to understand (perceive, process, encode) and subsequently to reproduce a temporal sequence. In short, rhythmic sequences (i.e., sequences implying a metrical framework) are easier to reproduce than nonrhythmic sequences.

Conclusions

The results of this experimental part of the study may be summarized as follows:

1. The underlying beat (tactus) of a cyclically repeating sequence of identical sounds was generally quite *ambiguous*. Each given rhythm evoked a number of different pulse sensations. The position of the downbeat in each cycle was similarly ambiguous.
2. The relative salience of the different pulse sensations evoked by a rhythm depended to a considerable extent on both rhythmic pattern (the sequence of IOI ratios) and note rate (the number of notes per unit time).
3. Rhythmic ambiguity tended to increase with increasing rhythmic complexity.
4. The likelihood of an event belonging to the tactus (or acting as a downbeat) increased with the IOI following the event, implying that events followed by longer IOIs were perceived as more accented. However, this effect was only consistently observed at faster tempi, that is, among relatively short IOIs. At slower tempi, the reverse effect was sometimes observed.
5. The tempi of most pulse responses lay in the vicinity of *moderate tempo* (approx. 600–700 ms). The tactus tended to gravitate toward this tempo regardless of the actual note rate of the rhythm.
6. Subjective grouping in isochronous sequences occurred more commonly by fours than by threes over a wide range of musical tempi.

The model developed and tested in this paper was based on the following assumptions:

1. Durational accent increases with the IOI that follows an event, but saturates as that IOI approaches about 1s, corresponding to the duration of the echoic store.
2. Durational accent approaches zero as IOI approaches the minimum discriminable IOI (typically, 50–100 ms).
3. The perception of pulse in musical rhythms may be regarded as a form of pattern recognition.
4. The salience of a pulse sensation increases with the phenomenal accent and number of events matching that pulse.
5. The salience of pulse sensations is greatest in the vicinity of 100 events per minute (600 ms) and falls to zero for tempi outside the approximate range 33–300 events per minute (200–1800 ms).
6. The effect of tempo on pulse salience is largely independent of the effect of matching events.
7. The metrical accent of a rhythmic event depends on the saliences of the pulse sensations including that event.

A literature survey suggested the following additional hypotheses:

1. The salience of a perceived meter may be estimated by summing the saliences of the consonant pulse sensations of which it is composed.
2. The salience of a pulse sensation may be enhanced by other, consonant pulse sensations. Enhancement of pulse salience may explain certain perceptual limitations on beat subdivision.
3. Temporal categories in rhythm perception are determined by metrical accents.
4. Temporal precision and discrimination are enhanced by metrical accentuation.
5. Expressive timing may be modeled by a slowing of local tempo near metrical and structural accents. The degree of slowing depends on the strength of the accents.
6. Variations in pulse and meter salience as a function of time in real, noncyclic rhythms may be modeled by assuming that the perception of pulse is confined to the psychological present (recency effect) and by assigning extra weight to events near the start of the sequence (primacy effect).
7. A musical rhythm may be defined as a temporal sequence evoking a sensation of pulse.

The present model is nonhierarchical, assuming the concurrent and largely independent existence of pulse sensations of different periods and

saliences. According to Monahan and Hirsch (1990), hierarchical representations of meter are unnecessary for explanations of temporal discrimination of events in simple rhythms. More generally, Swain (1986) has pointed out that any hierarchical theory of music perception is subject to known limitations in the amount of information that humans are able to accept and process. The success of the present model in accounting for the results of the above experiments suggests that hierarchical representations of meter may also be unnecessary for an explanation of pulse and accent strength, at least in relatively simple rhythms.[1]

References

Agawu, V. K. The rhythmic structure of West African music. *Journal of Musicology,* 1987, *5,* 400–418.

Allan, L. G. The perception of time. *Perception & Psychophysics,* 1979, *26,* 340–354.

American Standards Association. *USA standard acoustical terminology.* New York: American Standards Association, 1960.

Apel, W. *Harvard dictionary of music,* 2nd ed. Cambridge, MA: Harvard University Press, 1970.

Bamberger, J. Intuitive and formal musical knowing: Parables of cognitive dissonance. In S. S. Madeja (Ed.), *The arts, cognition and basic skills.* New Brunswick, NJ: Transactions, 1978.

Bamberger, J. Cognitive structuring in the apprehension and description of simple rhythms. *Archives of Psychology,* 1980, *48,* 177–199.

Bamberger, J. Revisiting children's drawings of simple rhythms: A function for reflection-in-action. In S. Strauss (Ed.), *U-shaped behavioral growth.* New York: Academic Press, 1982.

Beauvillain, C., & Fraisse, P. On the temporal control of polyrhythmic performance. *Music Perception,* 1984, *1,* 485–499.

Bengtsson, I. Notation, motion and perception: Some aspects of musical rhythm. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music.* Stockholm: Royal Swedish Academy of Music, 1987, pp. 69–80.

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

Benjamin, W. A theory of musical meter. *Music Perception,* 1984 *1,* 355–413.

Bolton, T. L. Rhythm. *American Journal of Psychology,* 1894, *6,* 145–238.

Braun, F. Untersuchungen über das persönliche Tempo [Investigations into personal tempo]. *Archiv für die gesamte Psychologie,* 1927, *60,* 317–360.

Bregman, A. S., & Pinker, S. Auditory streaming and the building of timbre. *Canadian Journal of Psychology,* 1978, *32,* 19–31.

---

[1] I am grateful to Johan Sundberg, Erik Jansson, Anders Askenfelt, Sten Ternström, and Gunilla Carlsson for providing a friendly working environment, discussions, help, advice, and computer facilities. Alf Gabrielsson advised on the design and running of the experiments. For valuable comments on earlier drafts of the manuscript, I thank Jamshed Bharucha, Chris Childs, Annabel Cohen, Carolyn Drake, Bradley Frankland, Gunter Kreutz, Justin London, David Rosenthal, and an anonymous reviewer. The research was supported by fellowships from the Swedish Institute and from the Natural Sciences and Engineering Research Council of Canada.

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

Brown, W. Temporal and accentual rhythm. *Psychological Review,* 1911, *18,* 336–346.

Buytendijk, F. J. J., & Meesters, A. Duration and course of the auditory sensation. *Commentationes Pontificia Academia Scientiarum,* 1942, *6,* 557–576.

Clarke, E. F. Timing in the performance of Erik Satie's 'Vexations'. *Acta Psychologica,* 1982, *50,* 1–19.

Clarke, E. F. Some aspects of rhythm and expression in performances of Erik Satie's "Gnossienne No. 5." *Music Perception,* 1985, *2,* 299–328.

Clarke, E. F. Categorical rhythm perception: An ecological perspective. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music.* Stockholm: Royal Swedish Academy of Music, 1987, pp. 19–34.

Clarke, E. F. Generative principles in music performance. In J. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation and composition.* Oxford: Science Publications, 1988.

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

Collyer, C. E., Broadbent, H. A., & Church, R. M. Categorical time production: Evidence for discrete timing in motor control. *Perception and Psychophysics,* 1992, *51,* 134–144.

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

Crowder, R. G. Improved recall for digits with delayed recall cues. *Journal of Experimental Psychology,* 1969, *82,* 258–262.

Crowder, R. G. The role of one's own voice in immediate memory. *Cognitive Psychology,* 1990, *1,* 157–178.

Crowder, R. G., & Morton, J. Precategorical acoustic storage (P.A.S.). *Perception & Psychophysics,* 1969, *5,* 365–373.

Desain, P. A (de)composable theory of rhythm perception. *Music Perception,* 1992, *9,* 439–454.

Desain, P., & Honing, H. Quantization of musical time: A connectionist approach. *Computer Music Journal,* 1992, *13*(3), 56–66.

Desain, P., & Honing, H. Tempo curves considered harmful. *Proceedings of the 1991 International Computer Music Conference.* San Francisco: Computer Music Association, 1991, pp. 143–149.

Deutsch, D. The organization of short-term memory for a single acoustic attribute. In D. Deutsch & J. A. Deutsch (Eds.), *Short-term memory.* New York: Academic Press, 1975, pp. 108–151.

Deutsch, D. Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music.* New York: Academic Press, 1982.

Dietze, G. Untersuchungen über den Umfang des Bewußteins bei regelmäßig aufeinanderfolgenden Schalleindrücken [Investigations into the span of consciousness for regular successions of sound impressions]. *Philosophische Studien,* 1885, *2,* 362–393.

Drake, C. Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics* 1993, *53,* 25–33.

Drake, C., & Botte, M-C. Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics* (in press).

Drake, C., Botte, M-C., & Gérard, C. *A perceptual distortion in simple musical rhythms.* Paper presented at the International Conference on Psychophysics, Cassis, France, 1989.

Drake, C., Dowling, J., & Palmer, C. Accent structures in the reproduction of simple tunes by children and adult pianists. *Music Perception,* 1991, *8*(3), 315–334.

Drake, C., & Gérard, C. A psychological pulse train: How young children use this cognitive framework to structure simple rhythms. *Psychological Research,* 1989, *51,* 16–22.

Duke, R. A., Geringer, J. M., & Madsen, C. K. Performance of perceived beat in relation to age and music training. *Journal of Research in Music Education,* 1991, *39*(1), 35–45.

Eggebrecht, H. H. *Riemann Musiklexikon.* Mainz: Schott, 1967.

Ehrlich, S. Le mécanisme de la synchronisation sensori-motorice. *Année Psychologique,* 1958, *58,* 7–23.

Eisler, H. Subjective duration and psychophysics. *Psychological Review,* 1975, *82,* 429–450.

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

Fowler, C. A. "Perceptual centers" in speech production and perception. *Perception & Psychophysics,* 1979, *25,* 375–388.

Fraisse, P. *Les structures rythmiques.* Louvain: Publications Universitaires de Louvain, 1956.

Fraisse, P. *The psychology of time.* New York: Harper & Row, 1963.

Fraisse, P. Le seuil différentiel de durée dans une suite régulière d'intervalles [The differential duration threshold in a regular series of intervals]. *Année Psychologique,* 1967, *1,* 43–49.

Fraisse, P. Les synchronisations sensori-motorices aux rythmes. In J. Requin (Ed.), *Anticipation et comportement.* Paris: Editions CNRS, 1980.

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

Fraisse, P. A historical approach to rhythm as perception. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music.* Stockholm: Royal Swedish Academy of Music, 1987, pp. 7–18.

Fraisse, P., Pichot, P., & Clairouin, G. Les aptitudes rythmiques: Etude comparée des oligophrènes et des enfants normaux. *Journal de Psychologie Normale et Pathologique,* 1949, *42,* 309–330.

Friberg, A., Frydén, L., Bodin, L., & Sundberg, J. Performance rules for computer-controlled contemporary keyboard music. *Computer Music Journal,* 1991, *15*(2), 49–55.

Fricke, J. P. Merkmale und Bedingungen des Sprachlichen in der Musik [Characteristics and conditions of linguistic aspects of music]. In J. P. Fricke (Ed.), *Die Sprache der Musik.* Regensburg: Bosse, 1989, pp. 171–188.

Frischeisen-Köhler, I. Feststellung des weder langsamen noch schnellen (mittelmäßigen) Tempos [Determination of moderate (neither slow nor fast) tempo]. *Psychologische Forschung,* 1933, *18,* 291–298.

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

Gabrielsson, A., Bengtsson, I., & Gabrielsson, B. Performance of musical rhythm in 3/4 and 6/8 meter. *Scandinavian Journal of Psychology,* *24,* 193–213.

Garner, W. R. *The processing of information and structure.* New York: Wiley, 1974.

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

Gérard, C., & Drake, C. The inability of young children to reproduce intensity differences in musical rhythms. *Perception & Psychophysics,* 1990, *48,* 91–101.

Gibson, J. J. Events are perceivable, but time is not. In J. T. Fraser & N. Lawrence (Eds.), *The study of time,* Vol. 2. Berlin: Springer-Verlag, 1975.

Gibson, J. J. *The ecological approach to visual perception.* Boston: Houghton Mifflin, 1979.

Glenberg, A. M., & Jona, M. Temporal coding in rhythm tasks revealed by modality effects. *Memory & Cognition,* 1991, *19,* 514–522.

Glenberg, A. M., & Swanson, N. C. A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition,* 1986, *12,* 3–15.

Glenberg, A. M., Mann, S., Alman, L., Forman, T., & Procise, S. Modality effects in the coding of rhythms. *Memory & Cognition,* 1989, *17,* 373–383.

Glucksberg, S., & Cowen, G. N., Jr. Memory for non-attended auditory material. *Cognitive Psychology,* 1970, *1,* 149–156.

Grant, R. E., & LeCroy, S. Effects of sensory mode input on the performance of rhythmic perception tasks by mentally retarded subjects. *Journal of Music Therapy*, 1986, **23**, 2–9.

Guttman, N., & Julesz, B. Lower limits of auditory periodicity analysis. *Journal of the Acoustical Society of America*, 1963, **35**, 610.

Guttmann, A. Das Tempo und seine Variationsbreite [Tempo and its range of variation]. *Zeitschrift für angewandte Psychologie*, 1931, **40**, 65.

Handel, S. Perceiving melodic and rhythmic auditory patterns. *Journal of Experimental Psychology*, 1974, **103**, 922–933.

Handel, S. Tempo in rhythm: Comments on Sidnell. *Psychomusicology*, 1986, **6**, 19–23.

Handel, S., & Buffardi, L. Pattern perception: Integrating information presented in two modalities. *Science*, 1968, **162**, 1026–1028.

Handel, S., & Lawson, G. R. The contextual nature of rhythmic interpretation. *Perception & Psychophysics*, 1983, **34**, 103–120.

Handel, S., & Oshinsky, J. S. The meter of syncopated auditory polyrhythms. *Perception & Psychophysics*, 1981, **30**, 1–9.

Harrell, T. W. Factors influencing preference and memory for auditory rhythms. *Journal of General Psychology*, 1937, **16**, 427–469.

Harrison, R. Personal tempo and the interrelationship of voluntary and maximal rates of movement. *Journal of General Psychology*, 1941, **24**, 343–379.

Hlawicka, K. Die rhythmische Verwechslung [Rhythmic ambiguity]. *Musikforschung*, 1958, **11**, 33–49.

Howell, P. Prediction of P-center location from the distribution of energy in the amplitude envelope: I. *Perception & Psychophysics*, 1988, **43**, 90–93.

James, W. *Principles of psychology*. New York: Dover, 1950.

Jones, A. M. *Studies in African music*. London: Oxford University Press, 1959.

Jones, M. R. Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, 1976, **83**, 323–355.

Jones, M. R. Dynamic pattern structure in music: Recent theory and research. *Perception & Psychophysics*, 1987a, **41**, 621–634.

Jones, M. R. Perspectives on musical time. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music*. Stockholm: Royal Swedish Academy of Music, 1987b, pp. 153–176.

Jones, M. R., & Boltz, M. Dynamic attending and responses to time. *Psychological Review*, 1989, **96**, 459–491.

Jones, M. R., Boltz, M., & Kidd, G. Controlled attending as a function of melodic and temporal context. *Perception & Psychophysics*, 1982, **32**, 211–218.

Jones, M. R., Kidd, G., & Wetzel, R. Evidence for rhythmic attention. *Journal of Experimental Psychology: Human Perception & Performance*, 1981, **7**, 1059–1073.

Keele, S. W., Pokony, R. A., Corcos, D. M., & Ivry, R. Do perception and motor production share common timing mechanisms: A correlational analysis. *Acta Psychologica*, 1985, **60**, 173–191.

Kluge, R. Zeitliche Bezugssysteme als Basis rhythmischer Stile [Temporal reference frames as a basis for rhythmic styles]. In H. Braun (Ed.), *Probleme der Volksmusikforschung*, 1990, pp. 47–60.

Koetting, J. What do we know about African rhythm? *Ethnomusicology*, 1986, **30**, 6–31.

Krebs, H. Some extensions of the concepts of metrical consonance and dissonance. *Journal of Music Theory*, 1987, **31**, 99–120.

Kronman, U., & Sundberg, J. Is the musical retard an allusion to physical motion? In A. Gabrielsson (Ed.), *Action and perception in rhythm and music*. Stockholm: Royal Swedish Academy of Music, 1987, pp. 57–68.

Kubovy, M., & Howard, F. P. Persistence of pitch-segregating echoic memory. *Journal of Experimental Psychology: Human Perception & Performance*, 1976, **2**, 531–537.

Lashley, K. S. The problem of serial order in behavior. In L. A. Jeffries (Ed.), *Cerebral mechanisms in behavior*. New York: Wiley, 1951.

Lee, C. S. The perception of metrical structure: Experimental evidence and a model. In P. Howell, R. West, & I. Cross (Eds.), *Representing musical structure*. London: Academic Press, 1991, pp. 59–127.

Lerdahl, F., & Jackendoff, R. On the theory of grouping and meter. *Musical Quarterly*, 1981, **25**, 45–90.

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

Longuet-Higgins, H. C. The perception of melodies. *Nature (London)*, 1976, **263**, 646–653.

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

Longuet-Higgins, H. C., & Lee, C. S. The rhythmic interpretation of monophonic music. *Music Perception*, 1984, **1**, 424–441.

Longuet-Higgins, H. C., & Steedman, M. J. On interpreting Bach. In B. Meltzer & D. Michie (Eds.), *Machine intelligence*, Vol. 6. Edinburgh: Edinburgh University Press, 1971.

MacDougall, R. The structure of simple rhythm forms. *Psychological Review*, Monograph Supplements, 1903, **4**, 309–416.

Mach, E. *Die Analyse der Empfindungen* [Analysis of sensations], 1st ed. Jena: Fischer, 1886. (2nd ed., 1911)

Marcus, S. M. Acoustic determinants of perceptual (P-center) location. *Perception & Psychophysics*, 1981, **30**, 247–256.

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

Massaro, D. W. Preperceptual auditory images. *Journal of Experimental Psychology*, 1970, **85**, 411–417.

Massaro, D. W. Perceptual units in speech recognition. *Journal of Experimental Psychology*, 1974, **102**, 199–208.

Meumann, E. Untersuchungen zur Psychologie und Aesthetik des Rhythmus [Investigations into the psychology and aesthetics of rhythm]. *Philosophische Studien*, 1894, **10**, 249–322, 393–430.

Michon, J. A. Studies on subjective duration. 1. Differential sensitivity in the perception of repeated temporal intervals. *Acta Psychologica*, 1964, **22**, 441–450.

Michon, J. A. *Timing in temporal tracking*. Soesterberg, NL: Institute for Perception TNO, 1967.

Michon, J. A. Programs and "programs" for sequential patterns in motor behaviour. *Brain Research*, 1974, **71**, 413–424.

Michon, J. A. Time experience and memory processes. In J. T. Fraser & N. Lawrence (Eds.), *The study of time*, Vol. 2. Berlin: Springer-Verlag, 1975.

Michon, J. A. The making of the present: A tutorial review. In J. Requin (Ed.), *Attention and performance*, Vol. 7. New York: Academic Press, 1978.

Miles, D. W. Preferred rates in rhythmic response. *Journal of General Psychology*, 1937, **16**, 427–469.

Miller, G. A., & Heise, G. A. The trill threshold. *Journal of the Acoustical Society of America*, 1950, **22**, 637–638.

Miner, J. B. Motor, visual and applied rhythms. *Psychological Review*, Monograph Supplements, 1903, **5**, 1–106.

Mishima, J. *Introduction to the morphology of human behavior: The experimental study of mental tempo*. Tokyo: Tokyo Publishing, 1965.

Miyake, I. Researches on rhythmic action. *Studies from the Yale Psychological Laboratory*, 1902, **10**, 1–48.

Monahan, C. B., & Hirsch, I. J. Studies in auditory timing: 2. Rhythm patterns. *Perception & Psychophysics*, 1990, **47**, 227–242.

Moore, B. C. J., & Glasberg, B. R. Growth of forward masking for sinusoidal and noise maskers as a function of signal delay: Implications for suppression in noise. *Journal of the Acoustical Society of America*, 1983, **73**, 1249–1259.

Morton, J., Marcus, S., & Frankish, C. Perceptual centers (P-centers). *Psychological Review*, 1976, ***83***, 405–408.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Nickerson, J. F. Intonation of solo and ensemble performance of the same melody. *Journal of the Acoustical Society of America*, 1949, ***21***, 593–595.

Noorden, L. van. *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Institute for Perception Research, Eindhoven, The Netherlands, 1975.

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

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

Parncutt, R. *Algorithm for event, pulse and metre salience in musical rhythms*. Poster presented at the International Conference on Event Perception and Action, Uppsala, Sweden, 1985.

Parncutt, R. The perception of pulse in musical rhythm. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music*. Stockholm: Royal Swedish Academy of Music, 1987, pp. 127–138.

Parncutt, R. *Harmony: A psychoacoustical approach*. Berlin: Springer-Verlag, 1989.

Patterson, W. M. *The rhythm of prose*. New York, 1916.

Plomp, R. Rate of decay of auditory sensation. *Journal of the Acoustical Society of America*, 1964, ***36***, 277–282.

Plomp, R. Pitch of complex tones. *Journal of the Acoustical Society of America*, 1967, ***41***, 1526–1533.

Plomp, R., & Bouman, M. A. Relation between hearing threshold and duration for tone pulses. *Journal of the Acoustical Society of America*, 1959, ***31***, 749–758.

Povel, D. J. Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception & Performance*, 1981, ***7***, 3–18.

Povel, D. J. A theoretical framework for rhythm perception. *Psychological Research*, 1984, ***45***, 315–337.

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

Povel, D. J., & Okkerman, H. Accents in equitone sequences. *Perception & Psychophysics*, 1981, ***30***, 565–572.

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

Reymert, M. L. The personal equation in motor capacities. *Scandinavian Science Review*, 1923, ***2***, 177–222.

Riemann, H. *Musikalische Dynamik und Agogik*. Hamburg, St. Petersburg, 1884.

Rimoldi, H. J. A. Personal tempo. *Journal of Abnormal and Social Psychology*, 1951, ***46***, 283–303.

Ritsma, R. J. Frequencies dominant in the perception of the pitch of complex tones. *Journal of the Acoustical Society of America*, 1967, ***42***, 191–198.

Robert, W. Chopin's *tempo rubato* in theory and practice. *Piano Quarterly*, 1981, ***113***, 42–44.

Roederer, J. G. *Introduction to the physics and psychophysics of music*, 2nd ed., 2nd reprint, corrected. New York: Springer-Verlag, 1979.

Rosenthal, D. A model of the process of listening to simple rhythms. *Music Perception*, 1989, ***6***, 315–328.

Rosenthal, D. Emulation of human rhythm perception. *Computer Music Journal*, 1992, ***16***(1), 64–76.

Rostron, A. B. Brief auditory storage: Some further observations. *Acta Psychologica*, 1974, ***38***, 471–482.

Rothstein, W. *Phrase rhythm in tonal music*. New York: Schirmer, 1989.

Rowlands, L. *Rhythmic relationships in Ghanaian music and dance*. Unpublished M. Litt thesis, University of New England, Armidale NSW, Australia, 1991.

Royer, F. L., & Garner, W. R. Perceptual organization of nine-element auditory temporal patterns. *Perception & Psychophysics*, 1970, ***7***, 115–120.

Schab, F. R., & Crowder, R. G. Accuracy of temporal coding: Auditory-visual comparison. *Memory & Cognition*, 1989, ***17***, 384–397.

Scholes, P. A. (Ed.). *Oxford companion to music*. London: Oxford University Press, 1970.

Schulze, H. The detectability of local and global displacements in regular rhythmic patterns. *Psychological Research*, 1978, ***40***, 173–181.

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

Schütte, H. Ein Funktionsschema für die Wahrnehmung eines gleichmäßigen Rhythmus in Schallimpulsfolgen [A model for the perception of isochrony in sequences of sound pulses]. *Biological Cybernetics*, 1978, ***29***, 49–55.

Seashore, C. E. Motor ability, reaction-time, rhythm, and time-sense. *University of Iowa Studies in Psychology*, 1899, ***2***, 64–84.

Seashore, C. E. *Psychology of music*. New York: McGraw-Hill, 1938.

Shaffer, L. H. Performances of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology*, 1981, ***13***, 326–376.

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

Shaffer, L. H., & Todd, N. P. The interpretative component in musical performance. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music*. Stockholm: Royal Swedish Academy of Music, 1987, pp. 139–152.

Simon, H. A. Perception du pattern musical par AUDITEUR. *Sciences de l'Art*, 1968, ***5***(2), 28–34.

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

Sloboda, J. A. *The musical mind: The cognitive psychology of music*. Oxford: Clarendon, 1985.

Smith, J. Reproduction and representation of musical rhythms: The effects of musical skill. In D. R. Rogers & J. A. Sloboda (Eds.), *Acquisition of symbolic skill*. New York: Plenum, 1983.

Smith, K. C., & Cuddy, L. L. Effects of metric and harmonic rhythm on the detection of pitch alternations in melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 1989, ***15***, 457–471.

Squire, C. R. A genetic study of rhythm. *American Journal of Psychology*, 1901, ***12***, 492–589.

Steedman, M. J. The perception of musical rhythm and metre. *Perception*, 1977, ***6***, 555–569.

Stern, W. Das psychische Tempo (Psychological tempo). In *Über die Psychologie der individuellen Differenzen*. Leipzig: Barth, 1900.

Stetson, R. H. A motor theory of rhythm and discrete succession. *Psychological Review*, 1905, ***12***, 250–270, 292–350.

Summers, J. J., Bell, R., & Burns, B. D. Perceptual and motor factors in the imitation of simple temporal patterns. *Psychological Research*, 1989, ***51***, 23–27.

Summers, J. J., Hawkins, S. R., & Mayers, H. Imitation and production of interval ratios. *Perception & Psychophysics*, 1986, ***39***, 437–444.

Sundberg, J. Computer synthesis of music performance. In J. A. Sloboda (Ed.), *Generative processes in music*. Oxford: Clarendon, 1988.

Sundberg, J., Askenfelt, A., & Frydén, L. Musical performance: A synthesis-by-rule approach. *Computer Music Journal*, 1983, ***7***, 37–43.

Sundberg, J., & Verrillo, V. On the anatomy of the retard: A study of timing in music. *Journal of the Acoustical Society of America*, 1980, ***68***, 772–779.

Swain, D. The need for limits in hierarchical theories of music. *Music Perception*, 1986, *4*, 121–148.

Temperley, N. M. Personal tempo and subjective accentuation. *Journal of General Psychology*, 1963, *68*, 267–287.

Terhardt, E., Stoll, G., & Seewann, M. Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America*, 1982, *71*, 679–688.

Thomassen, M. T. Melodic accent: experiments and a tentative model. *Journal of the Acoustical Society of America*, 1982, *71*, 1596–1605.

Thompson, W. F., Sundberg, J., Friberg, A., & Frydén, L. The use of rules for expression in the performance of melodies. *Psychology of Music*, 1989, *17*, 63–82.

Todd, N. P. McA. A model of expressive timing in tonal music. *Music Perception*, 1985, *3*, 33–58.

Todd, N. P. McA. The dynamics of dynamics: A model of musical expression. *Journal of the Acoustical Society of America*, 1992, *91*, 3540–3550.

Treisman, A. M. Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning & Verbal Behavior*, 1964, *3*, 449–459.

Treisman, A. M., & Howarth, C. I. Changes in threshold level produced by a signal preceding or following the threshold stimulus. *Quarterly Journal of Experimental Psychology*, 1959, *11*, 129–142.

Treisman, A. M., & Rostron, A. B. Brief auditory storage: A modification of Sperling's paradigm applied to audition. *Acta Psychologica*, 1972, *36*, 161–170.

Upitis, R. Children's understanding of rhythm: The relationship between development and music training. *Psychomusicology*, 1987, *7*, 41–60.

Vorberg, D., & Hambuch, R. On the temporal control of rhythmic performance. In J. Requin (Ed.), *Attention and performance*, Vol. 7. New York: Academic Press, 1978.

Vos, J., & Rasch, R. The perceptual onset of musical tones. *Perception & Psychophysics*, 1981, *29*, 323–335.

Vos, P. G. *Pattern perception in metrical tone sequences*. Unpublished thesis, University of Nijmegen, 1973.

Vos, P. G. Temporal duration factors in the perception of auditory rhythmic patterns. *Scientific Aesthetics*, 1977, *1*, 183–199.

Vos, P. G., Collard, R. F., & Leeuwenberg, E. L. What melody tells about meter in music. *Zeitschrift für Psychologie*, 1981, *189*, 25–33.

Vuori, M. Figural and temporal structures in children's sight-reading performance. *Canadian Journal of Research in Music Education*, 1991, *33*, 201–206.

Wallin, J. E. W. Researches on the rhythm of speech. *Studies from the Yale Psychological Laboratory*, 1901, *9*, 1–142.

Wallin, J. E. W. Experimental studies of rhythm and time. *Psychological Review*, 1911, *18*, 100–131, 202–222; 1912, *19*, 271–298.

Wapnick, J., & Rosenquist, M.-J. Preferences of undergraduate music majors for sequenced versus performed piano music. *Journal of Research in Music Education*, 1991, *39*(2), 152–160.

Ward, W. D. Musical perception. In J. V. Tobias (Ed.), *Foundations of modern auditory theory*. New York: Academic Press, 1970.

Wessel, D. L. Timbre space as a musical control structure. *Computer Music Journal*, 1979, *3*(2), 45–52.

Wing, A. M. *The timing of interresponse intervals*. Hamilton, Ontario: Department of Psychology, McMaster University, 1973, Technical Report No. 56.

Wing, A. M., & Kristofferson, A. B. The timing of interresponse intervals. *Perception & Psychophysics*, 1973a, *13*, 455–460.

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

Woodrow, H. The role of pitch in rhythm. *Psychological Review*, 1911, *18*, 54–71.

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

Wu, C. F. Personal tempo and speed in some rate tests. *Psychological Abstracts*, 1935, *9*, 1709.

Wundt, W. *Grundzüge der physiologischen Psychologie* [Foundations of physiological psychology]. Leipzig: Engelmann, 1874.

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

Zwicker, E., & Fastl, H. Zur Abhängigkeit der Nachverdeckung von der Störimpulsdauer [On the dependence of forward masking on masker duration]. *Acustica*, 1972, *26*, 78–82.