13

Rhythm and Timing in Music

Eric F. Clarke

Music Department
University of Sheffield
Sheffield, United Kingdom

I. INTRODUCTION

The aim of this chapter is to give an overview of research relating to the temporal dimension in music. In its entirety, this constitutes a very large body of work, despite the frequently repeated observation that time in music has received rather less attention than pitch (e.g., Kramer, 1988). This chapter therefore focuses primarily on small- to medium-scale temporal phenomena in music, the domain that would commonly be referred to as rhythm, rather than the larger-scale properties of form. The detailed temporal properties of performed music, often referred to as temporal microstructure, and the relationships between rhythm and movement are also considered.

Although pitch may have had the lion's share of attention as far as both empirical and theoretical work in the psychology of music is concerned, there is nonetheless a considerable amount of material on rhythm in music. Amongst this literature, the work of Paul Fraisse stands out above the research of any other single individual in both its scope and the manner in which it foreshadows the preoccupations of a great deal of the more contemporary work in the area. The particular integrity and character of Fraisse's work makes it difficult to divide into the categories of this chapter, and as a consequence it is presented separately from more recent work. His work has also played an important part in stimulating and informing modern rhythm research, so it is appropriate to begin with it.

II. THE WORK OF PAUL FRAISSE

The special character of Fraisse's work is in part attributable to the fact that it covers both musical issues and the more general field of time perception, and that he comes from a Piagetian tradition in which the relationship between perceptual capacities, sensorimotor organization, and human development is paramount. This results in a rather more holistic view than is found in most current work and incorporates a relationship with biology that is also rare today. Fraisse's work has been widely published (e.g., Fraisse, 1956, 1963, 1978, 1982, 1987), so the present account is confined to a distillation of the principal ideas without presenting the abundant supporting evidence that Fraisse amassed.

Fraisse drew a primary distinction between the perception of time and the estimation of time. The former is confined to temporal phenomena extending to no more than about 5 sec or so, whereas the latter relies primarily on the reconstruction of temporal estimates from information stored in memory. The boundary between these two corresponds to the length of the perceptual present, which he defined as "the temporal extent of stimulations that can be perceived at a given time, without the intervention of rehearsal during or after the stimulation" (Fraisse, 1978, p. 205). Rhythm perception, therefore, is essentially concerned with phenomena that can be apprehended in this immediate fashion and is also closely tied up with motor functioning. In studies of spontaneous tapping, Fraisse observed that by far the most ubiquitous relationship between successive tapped intervals was a ratio of 1:1 (i.e., isochronous or pendular motion). Fraisse regarded this as intimately connected with anatomical and motor properties—most notably the bilateral symmetry of the body, the pendular movements of the limbs in walking and running, and the regular alternation of exhalation and inhalation in breathing. He showed that when subjects are asked to tap rhythmically, they produce a bimodal distribution of ratios between intertap intervals, with peaks around 1:1 and 2:1, whereas when they are asked to tap arhythmically they produce a more continuous distribution of intertap interval ratios, and the frequency with which a ratio appears declines in proportion to its value (i.e., larger ratios are more unlikely). Fraisse describes both arhythmic and rhythmic tapping as a break with the underlying tendency for pendular movement, but whereas there is no structure in the former case, the latter exploits a principle of identity or clear differentiation between time intervals. This principle of equality or differentiation creates two distinct categories of duration, according to Fraisse, which he terms *temps longs* and *temps courts* (long durations and short durations), which are not only quantitatively but also qualitatively different. Short durations do not extend beyond about 400 msec, and 2:1 ratios between successive intervals are found only *between* the two categories.[1] *Temps longs* have the property of true duration according to Fraisse (we are aware, or can become aware, of the passage of time during such an interval), whereas *temps courts* have the character of collection rather than duration: we have no real sense of the passage of time during each event, but are

[1]It would be perfectly possible for a 2:1 ratio to exist within either category: an interval of 75 msec followed by 150 msec would be a 1:2 ratio within *temps courts*, and 500 msec followed by 1000 msec would be the same within *temps longs*. But Fraisse does not find this in his empirical data.

aware of the manner in which numbers of such intervals group together. The distinctions presented here are succinctly expressed by Fraisse in the following passage:

> Rhythmic and arhythmic sequences both consist of a break with the natural tendency to equalize successive intervals of time. Arhythmia is characterized by inequalities between successive durations that decrease in frequency in proportion to their size. Rhythmic structures, on the other hand, consist of the interplay of two types of value of temporal interval, clearly distinct from one another (in a mean ratio of 1:2). Within each type the durations are perceptually equal to one another. The collection of shortest intervals appears, from initial results, to consist of durations less than 400 msec. (Fraisse, 1956, pp. 29–30. Author's translation)

Although 400 msec appears here as the cutoff between the two categories of duration, elsewhere Fraisse cites 600 msec as an important value with analogous properties: it is what Fraisse terms the "indifference interval"—that interval of time for which people's duration estimates are most veridical, showing neither systematic overestimations (as they tend to for shorter durations) nor underestimations (as they tend to for longer durations).[2] Fraisse claims that this has a direct relationship to the duration of the whole perceptual process, corresponding "to the continuation of two percepts with no overlapping and no interval" (Fraisse, 1978, p. 225). The link between "indifference" in perceptual judgment and the threshold between the two categories of duration in rhythmic structures (*temps longs* and *temps courts*), with their respective properties of duration and collection, is made clear here.

To summarize, at the heart of Fraisse's contribution to an understanding of rhythm in music are the following:

1. The perceptual present as the dividing line between the direct *perception* of duration and its *estimation.*
2. The fundamental status of pendular motion and the close association between rhythm and movement.
3. The distinction between rhythmia and arhythmia, based on the distinction between a continuous and a bimodal distribution of duration ratios between successive time intervals.
4. The existence of a categorical distinction between two types of duration (*temps long* and *temps court*) in rhythm, in a mean duration ratio of 2:1, and with the quality of duration and collection, respectively.
5. A threshold between these two categories at a value around 400–600 msec, also associated with "indifference" in perceptual judgments.
6. The operation of two complementary principles (assimilation and distinction) that preserve both the integrity and distinctness of the two categories.

[2]There is considerable variability in the value given by different authors to the indifference interval. This may be due to the different methods that have been used to assess it, or may be because the phenomenon itself is unstable or even artifactual.

III. FORM PERCEPTION

The distinction that Fraisse draws between time perception and time estimation or construction allows a division between rhythm and form to be established. Musical form, understood as the sectional proportions of a work, might conceivably be regarded as part of rhythm in music if one adopts a sufficiently inclusive definition of the term. Indeed, Cooper and Meyer (1960) do precisely that when they present an analysis in which a single set of rhythmic categories is applied to the first movement of Beethoven's Eighth Symphony, ranging from single notes through all levels of the music up to the entire movement. The unbroken continuity of rhythmic notation implies that our response to note-to-note relationships is governed by the same principles and processes as is our response to the relationships between sections of the work, each of which lasts of the order of 5 min. Estimates of the perceptual present, which forms the boundary between direct perception and the memory-dependent processes of construction and estimation, are variable, but a value somewhere around 3–8 seconds is in agreement with a good deal of the available evidence. Crowder (1993), for example, following research by Cowan (1984, 1987), concurs with the proposal that there may be a very short auditory store[3] of around 250 msec, and a longer store, with a period of about 2–10 sec, with the two stores being the behavioral consequence of different perceptual/cognitive processes.

Michon (1978) provides a review of properties of the perceptual present that is useful for a consideration of the relationship between form and rhythm. The primary character of the perceptual present is that the contents of the present are active and directly available, whereas memories must be retrieved—must be transformed from a state of inactive storage to current awareness. This strongly suggests that it is not possible to have any direct apprehension of form, but that a sense of form becomes available only through a retrospective, and in some sense deliberate, act of (re)construction. Further, the extent of the perceptual present is governed by organizational considerations rather than pure duration: although there seems to be an upper limit beyond which the perceptual present cannot be extended whatever the structure of the material concerned, within this upper bound the determination of the contents of the present is primarily a function of perceptual structure, such that the boundary of the perceptual present falls at a natural break in the event structure (see Clarke, 1987a, for further discussion). In fact, despite Michon's protestations that the perceptual present should not be equated with any kind of memory, it looks very much as though the perceptual present should be understood as a temporal view of the contents of working memory

[3]Crowder (1993) argues persuasively for a procedural approach to auditory memory in which "stores" are simply the behavioral consequence, or by-product, of perceptual activity rather than having any anatomical or systematic reality—in the manner of receptacles—themselves. However, as he also points out, there is a large and somewhat contradictory literature on the whole subject of short auditory storage.

(Baddeley, 1986) with all the properties that have been described by research in that area.

Let us turn now from a consideration of the characteristics that divide rhythm from form and examine empirical research that has investigated listeners' sense of musical form. Remarkably few studies have tackled this issue, but Cook (1990) provides an interesting account of some informal tests that he has conducted. Using the first movement of Beethoven's Piano Sonata Op. 49, No. 2, Cook reports that music students:

> frequently predicted that the music would continue for another minute or more when the performance was broken off just before the final two chords. As soon as they heard those chords, of course, they realized that the movement had ended ... [A]s far as these listeners were concerned, the conclusion was not implied by anything that had come before - the recapitulation, for instance, or the coda. Furthermore ... a majority of the listeners failed to observe the repetition of the exposition, or else believed the repeat to be a modified one. (Cook 1990, pp. 44–45)

A similar study using the first movement of Webern's Symphony Op. 21 showed a similar lack of awareness of the most basic formal features of the music—in this case the literal repeat of the exposition of the movement. By contrast, Deliège and Ahmadi (1990) demonstrated that listeners were quite successful in picking up the formal articulation of music and that musicians and nonmusicians differ very little in their capacity to apprehend the basic formal divisions of even quite challenging music.

Clarke and Krumhansl (1990) rather more directly investigated listeners' sense of the medium- to large-scale temporal structure of two contrasting pieces of piano music: Stockhausen's Klavierstück IX and Mozart's Fantasie in C minor K. 475, each of which has a total duration of about 10 min. The study required a perceptual segmentation of each piece and demonstrated a very high level of agreement between the highly trained musicians who performed this part of the study (many of whom were professional composers or performers) in the location, strength, and structural characteristics of the segment boundaries. For the remainder of the study, the subjects were music students who, in separate experiments, listened to each piece twice and then made judgments about the duration, structural characteristics, and original location in the piece of a number of 30-sec extracts taken from the music. The results showed that listeners had a surprisingly good sense of where an extract came from in the overall scheme of the music, although they tended to judge extracts from both the beginning and the end of the piece as being relatively later in the music than their true position, as compared with extracts taken from the middle. It is as though the music appeared to move quickly towards the middle, then to become rather static, and finally to move quickly again near the end. Interestingly, this pattern was the same for both pieces despite their dramatically different stylistic characteristics. Deliège (1993), in a related study with music by Boulez, found a similar pattern of location judgments for musician listeners, and a rather flatter (and more veridical) profile for nonmusicians. It remains to be seen whether this pattern of systematic departure from veridicality should be attributed to some rather general processing consideration or whether it is attributable to properties of the music: it is not implausible that the structure of a great deal of music might reflect a very general scheme in which material is introduced over approximately the first third of the work, developed over the middle third, and then driven toward closure over the final third. Although there are good reasons to be cautious, the general literature on time perception (e.g., Michon, 1985) suggests that in the context of organized and goal-directed stimulus materials, time passes more quickly (i.e., durations are underestimated), which would make sense of the pattern found in the Mozart, Stockhausen, and Boulez.

IV. RHYTHM PERCEPTION

One of the problems that has hampered work in rhythm perception is that until comparatively recently there was no systematic and generally agreed definition of rhythm itself. Important though Cooper and Meyer's book on rhythm was in stimulating interest in rhythmic structure in music (Cooper & Meyer, 1960), they adopted a very broad approach to rhythm that did little to focus the concept. Quite apart from its importance in other respects, a significant contribution of Lerdahl and Jackendoff's *A Generative Theory of Tonal Music* (Lerdahl & Jackendoff, 1983) was its clarification of the elements of rhythmic structure in music, in particular the distinction between grouping and meter. They pointed out that rhythm in the tonal/metric music of the Western tradition consists of two independent elements: grouping—which is the manner in which music is segmented at a whole variety of levels, from groups of a few notes up to the large-scale form of the work—and meter—which is the regular alternation of strong and weak elements in the music. Two important points were made in this definition: first, although the two elements are theoretically independent of one another, the most stable arrangement involves a congruence between them such that strong points in the meter coincide with group boundaries. Second, the two domains deal respectively with time spans (grouping) and time points (meter): grouping structure is concerned with phenomena that extend over specified durations, whereas meter is concerned with theoretically durationless moments in time. Todd (1994a, 1994b) has pointed out that the two perspectives are directly analogous to the adoption of frequency-domain and time-domain approaches to pitch, with frequency corresponding to meter and wavelength corresponding to grouping. This also strikingly reveals the complementarity between grouping and meter.

The remainder of this section will therefore be divided between research that has focused on grouping or segmental structure in music and research on meter. Few studies have investigated the relationship between grouping and meter, despite Lerdahl and Jackendoff's insistence that it is in the interactions between the two that the power and interest of rhythmic structures lies (Lerdahl & Jackendoff, 1983, pp. 25 ff.).

A. GROUPING

As part of their theory of tonal music, Lerdahl and Jackendoff (1983) proposed a set of principles to account for segmental structure in music. Although there are some precedents for this (e.g., Nattiez, 1975; Ruwet, 1972; Tenney & Polansky, 1980), Lerdahl and Jackendoff's account is by far the most systematic and as a consequence has been the object of empirical investigation (Deliège, 1987). Lerdahl and Jackendoff propose that grouping is essentially a hierarchical property of music, and in their Grouping Well-Formedness Rules, they outline the formal conditions for hierarchical structure. Coupled with these, the Grouping Preference Rules (GPRs) describe the conditions that determine which of the very large number of possible hierarchical segmentations of any passage of music are actually likely to be perceived by listeners. The preference rules do not rigidly determine the segmentation of any particular passage, but specify the various forces acting in any musical context, which may reinforce one another or compete, resulting in different segmentations for different listeners. The GPRs themselves consist of three components: formalized Gestalt principles (principles of proximity in time, or change in pitch, duration, loudness, or articulation); more abstract formal concerns (principles of symmetry and the equivalence of variants of the same segment or passage); and principles relating to pitch stability.

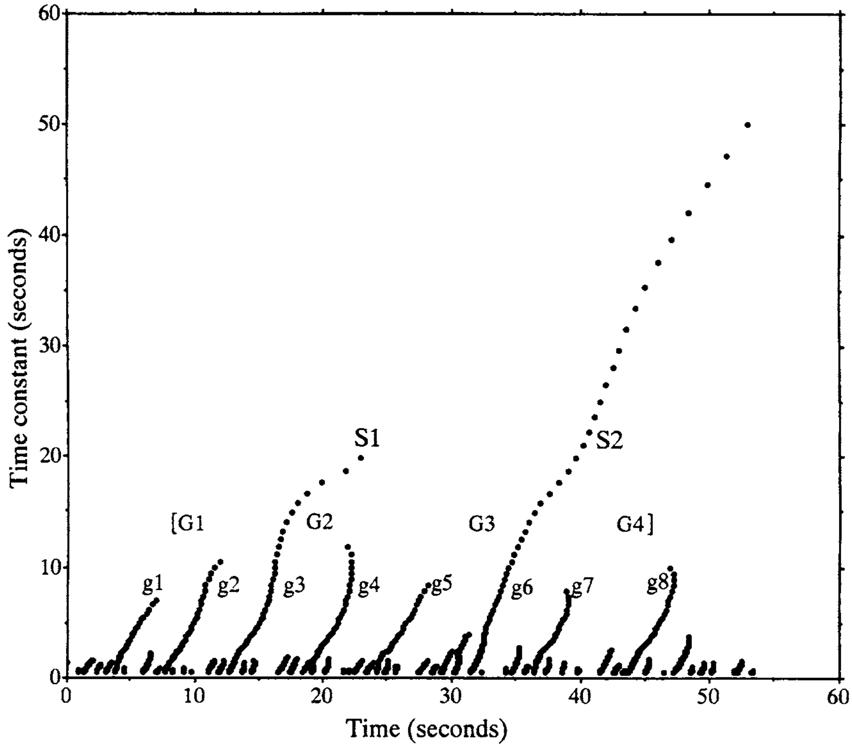
Lerdahl and Jackendoff offer no empirical evidence for the operation of these rules, relying on their own musical intuitions to guide them. However Deliège (1987) has investigated their empirical validity in the context of both highly reduced experimental materials and extracts of real music from Bach to Stravinsky. Her investigation demonstrated the validity of the predictions made by the GPRs, and provided some evidence for the relative strength of the different rules by setting them in conflict with one another. As Deliège herself observes, a great deal more work would need to be done to establish anything approaching a definitive rank ordering of the rules. Equally, she is quick to point out that a rule with a low ranking is not a poor rule: there may simply be intrinsic differences between rules (possibly relating to processing distinctions, such as between primary event structure and more cognitive organizing processes; cf. McAdams, 1993) that put them into different bands within the ranking, but leave unaffected the importance of each rule in the particular circumstances to which it applies.

A decade before the appearance of Lerdahl and Jackendoff's theory, Garner (1974) had investigated structure and segmentation properties in temporal patterns as part of his wider study of the processing of information and structure in spatial and temporal materials. His research, as well as that of Handel (1989), although confined to materials that are a considerable distance from real music, represents an important link between the line of research that has developed into auditory scene analysis (see Bregman, 1990) and the more cognitive work on musical rhythm that is discussed here.

Todd (1994a) has developed a model of rhythmic grouping that converges toward solutions that are often very similar to those offered by Lerdahl and Jackendoff, but is based on rather more explicit perceptual processes and has close parallels to documented properties of the auditory system. The central principle of Todd's approach is the idea that the functioning of the auditory system can be seen as the operation of a number of energy-integrating low-pass filters with differing time constants. At the lowest level, individual events (which are of course always spread out in time) are detected by virtue of filters with relatively small time constants, integrating acoustical energy over durations of the order of milliseconds or a few tens of milliseconds. At a somewhat higher level, small groups are detected as relatively discrete packets of integrated energy over periods of around a second. Larger, and hierarchically superordinate, groups are detected by virtue of integrators using exactly the same process, but with correspondingly longer time constants. Peaks in the output of these low-pass filters can be identified by looking for zero crossings in the second derivative of the filter output, and if these peaks are plotted across all the filters in a multiscale assembly, a representation of rhythmic events at a number of levels, and the grouping relationships between events, is obtained. Todd terms the resulting diagram, which is very similar to a more conventionally derived tree diagram, a *rhythmogram* and has shown that rhythmograms of live performances of music bear a striking resemblance to tree diagrams that depict grouping analyses (such as those developed by Lerdahl and Jackendoff). An attractive feature of Todd's model is that, because it is based on energy integration, it is sensitive to any changes in the acoustical signal that have consequences for the integrated energy level. This includes note duration, pitch, intensity, and even timbre and vibrato,[4] so that the written value of any note (rhythmic value, pitch, notated dynamic) and any expressive treatment that it receives in performance (rubato, vibrato, timbre, local intensity) all contribute in an undifferentiated manner to the integrated energy level that is output from the filter. The virtue of this is that it avoids having to distinguish between "score-based" properties of a musical event, and "expressive" properties in considering rhythm perception—a distinction that is anyway meaningless for all those musical cultures that do not use notational systems (which is, of course, the majority of world music).

As an illustration of Todd's model, Figure 1 shows a rhythmogram for a performance of the Chopin Prelude Op. 28, No. 7, together with a conventional grouping structure analysis of the music, based on Lerdahl and Jackendoff's (1983) theory. The two analyses, although arrived at in fundamentally different ways, show a remarkable level of agreement: the eight most prominent "branches" on the rhythmogram (g1–g8) correspond to the eight lowest level branches on the group-

[4] A note with a sharp timbre, which therefore contains high levels of upper partials, will have a greater level of integrated energy than the same pitch with a duller timbre. Similarly the frequency modulation that vibrato introduces will increase the integrated energy level of a note above that of an otherwise identical but nonvibrato note.



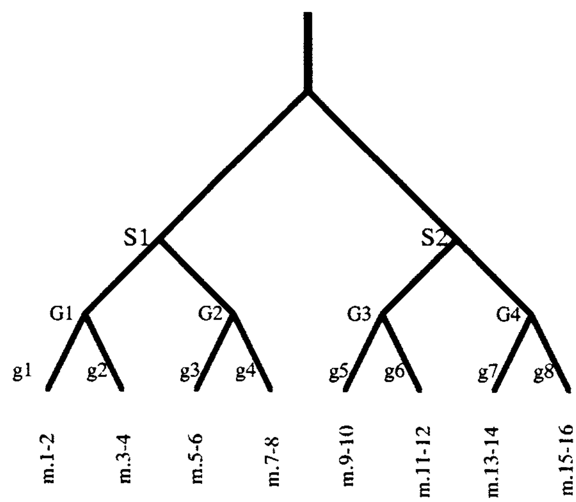


FIGURE 1 Rhythmogram of a professional pianist's performance of the Chopin Prelude Op. 28, No. 7 (top half) together with a Lerdahl and Jackendoff grouping analysis of the music (bottom half) showing the close relationship between the two. Units g1–g8 in both parts of the figure indicate two-measure groups; G1– G4 represent four-measure units (which are not well shown in the rhythmogram); and S1 and S2 represent the two eight-measure sections that constitute the whole prelude.

ing analysis, and those eight certainly group together into two large-scale sections (S1 and S2) of four each, with some evidence that they even pair up at an intermediate level (G1–G4). Thus an analysis that has been arrived at by relatively abstract music theoretic criteria applied to the written score is closely mirrored in an analysis that uses low-pass filtering applied to a standard audio recording of a professional performer. There are undoubtedly properties to which Todd's model is not sensitive—most obviously melodic and harmonic structure. However, to the extent that performers convey these properties by expressive means, even these will exert their influence on the rhythmogram, albeit "by the back door" rather than directly. What is interesting and provocative about the model, however, is the amount of grouping and sectional structure it can recover from real performances despite its "knowledge-free" approach. It suggests powerfully that rather more structural information is available within the acoustical signal itself than has hitherto been recognized and that processes in the more peripheral parts of the auditory system may be more important for rhythm perception than was at one time believed.

B. METER

The area of rhythm research that has attracted the most consideration since about 1980 is meter—undoubtedly a reflection of the dominant influence of metrical structure in the music of the Western tradition and in the overwhelming majority of popular music. This research has taken two forms: computational models of one sort or another and empirical investigations. Just as Lerdahl and Jackendoff performed an important service in first clarifying the relationship between grouping and meter and then specifying some of the conditions that govern the formation of musical groups, so they have also provided some valuable principles relating to the perception of meter.[5] First, they distinguish between three kinds of accent: phenomenal accents (points of local intensification caused by physical properties of the stimulus such as changes in intensity, simultaneous note density, register, timbre, or duration); structural accents (points of arrival or departure in the music that are the consequence of structural properties such as tonality—the cadence being the most obvious example of such an event); and metrical accents (defined as time points in music that are perceived as accented by virtue of their position within a metrical scheme). In general terms, perceiving meter is characterized by Lerdahl and Jackendoff as a process of detecting and filtering phenomenal and structural accents so as to discover underlying periodicities. These constitute the rates of repetition (cf. Yeston, 1976) that define the meter and confer metrical status on regularly recurring phenomenal (and structural) accents. Although this describes the outline of a process whereby physical attributes of the stimulus may result in the sense of meter, Lerdahl and Jackendoff make no attempt

[5]Some authors have preferred to talk of meter induction rather than meter perception in order to emphasize that meter is a construct that has no reality in the stimulus itself (they regard it as an abstraction from stimulus properties).

to specify in any detail all the factors that may contribute to the process or the relative weight of the various contributions made by the elements that they do discuss.

Of the empirical investigations into meter, the vast majority have been concerned with the influence of durational factors on meter. An important precursor here is the work of Fraisse (see Section II), whose identification of the importance of small integer ratios between durations and the role of assimilation and distinction in rhythm perception have left their mark. Equally, Deutsch (1986) showed in a duration comparison task that systematic distortions of short-term memory for duration could be attributed to the interfering effect of integer, or near integer, relations between a series of interpolated events and the durations to be compared. Povel (1981), starting from a position based on Fraisse's work, developed a model that demonstrates important elements of many of the more recent accounts of meter perception. His experiments showed that Fraisse's principle of assimilation and distinction around small integer ratios was generally verified with a battery of systematically structured short rhythmic sequences, but that in certain contexts an integer ratio could nonetheless prove difficult for a subject to perceive and reproduce.[6] The condition that created this difficulty was when a sequence could not be parsed into a simple repeating structure in which individual durations were organized into higher level units in multiples of two or three (but not an alternation of these). From this evidence, Povel formulated a "beat-based" model that proposed that the perception of rhythmic sequences depends on two steps: first, the segmentation of the sequence into parts of equal length (beats), based on the detection of regularly occurring accents; second, the identification of individual events as specific subdivisions of these beats into a small number (usually only two or three) of equal parts, or parts relating to one another in a ratio of approximately 1:2. The model allows a number of hierarchically nested levels (although not to the degree of depth that is suggested by Martin, 1972, for example) and specifies that the first level of subdivision below the level of the beat must consist only of divisions into two *or* three, and not a mixture or alternation. This constraint is motivated by Povel's empirical finding that subjects find it difficult to reproduce rhythms that involve alternations or mixtures of triplets and duplets at the same hierarchic level.

The model is developed further and adopts a more explicitly metrical character in Povel and Essens (1985), the essential idea of which is that a rhythmic sequence induces an internal clock with a period that captures the primary metrical level of the sequence. This is, in other words, a model of meter induction, the main driving force in the model being the identification of points of accentuation and the extraction of an underlying regularity from these. It is important to note that the sequences that Povel and Essens use in their empirical work consist of tone bursts with identical amplitude and duration but separated from one another by variable interonset intervals (all of which are whole number multiples of a 200-msec time base). All accentuation that their model handles arises out of three aspects of the

[6] The experimental task was to synchronize with a presented sequence and then to continue it for 17 repetitions after the stimulus ceased.

way in which events are grouped together[7]: (a) isolated events acquire an accent, (b) the first and last events of a run of three or more equally spaced events acquire accents, and (c) the second of a pair of events acquires an accent. The basis for these assumptions of perceptual accentuation are the results of Garner (1974), who found that the first and last events in a run acquired a perceptual salience, and the findings of Povel and Okkerman (1981), who demonstrated that the second of a pair of identical tone bursts occurring in reasonably close proximity is judged louder than the first. The model developed by Povel and Essens processes any rhythmic sequence *as a whole*, establishing the position of all accents as defined by the three principles just outlined, searching for a pattern of periodicities in the accents and weighing the evidence and counterevidence (accents that support or conflict with any particular periodic interpretation) for any candidates for the meter of the sequence. The predictions of the model correspond well to empirical evidence, but there must be doubts about the realism of a model that requires the whole sequence to be input before it starts to make any metrical interpretation: listeners often start to make a metrical interpretation within a few events of a sequence starting.

A model that is also intended to establish the meter of a pitchless sequence of equal-intensity events, but which makes rather more realistic processing assumptions, is described by Longuet-Higgins and Lee (1982). The fundamental principle in this model is that after two onsets (O_1 and O_2) have been detected, a third onset (O_3) is predicted to occur at the same time interval after the second event as the second is after the first (a principle of isochronous continuation). Confirmation of this prediction (by the arrival of an event at or near to O_3) causes the system to jump up a level in the emerging metrical hierarchy and to make a new prediction that an event will arrive at a time interval beyond O_3 that is equal to the interval between O_1 and O_3—that is, a principle of isochronous continuation at a periodicity that has twice the duration of the first level. The process continues to construct a binary metrical hierarchy in this way as long as events continue to confirm predictions. Longuet-Higgins and Lee recognize that the sense of meter does not extend up to indefinitely large time spans, and put what they recognize is an arbitrary stop on this process once the time between events exceeds about 5 sec (cf. the duration of the perceptual present). The simple principle outlined here turns out to be remarkably effective at parsing metrical structures, but it is also obvious that something more is needed if it is to make sense of nonbinary sequences (confined here to the single alternative of ternary sequences—a reasonable simplification for the vast majority of Western music) and sequences that do not begin on the strong beat of a meter. In broad terms, a single simple principle is used to handle both of these requirements—namely, the "accent-attracting" character of long notes.[8] Longuet-Higgins and Lee's model causes the strong beat of the meter (particularly

[7]This illustrates the close connection between grouping and meter, despite their theoretical independence, that was noted earlier.

[8]An alternative way to express this is to recognize that duration is one of the determinants of phenomenal accent (see Lerdahl & Jackendoff, 1983).

in early stages of any sequence) to be shifted, under carefully specified conditions, to any note that is significantly longer than its surrounding neighbors. This will either result in a ternary level in the meter or cause one or more short notes at the start of a sequence to function as upbeats to the first main beat (at the first longer note) of the meter. One implication of this assumption about the function of long notes is that in the absence of pitch, dynamic, or timbral information, a sequence of isochronous events will be heard as starting on a strong beat and organized in a purely binary meter—a consequence supported both by intuition and empirical evidence. Indeed Vos (1978) found a strong bias toward binary metrical interpretations (including sequences that were in fact in a ternary meter) even when the task presented listeners with commercially recorded extracts of music by Bach in which harmonic, melodic, and dynamic information were all available.

Closely related models of meter perception are presented by Johnson-Laird (1991) and Lee (1991). The former adopts the view of meter as a generative grammar for rhythms proposed earlier by Longuet Higgins (1979) and explores various issues relating to families of metrically related rhythms, syncopation, and phrase structure in music. Lee provides a very full exploration of the nature of a number of competing models of meter perception, including that of Lerdahl and Jackendoff (1983), two different models by Longuet-Higgins and Lee, and the Povel and Essens model (1985). Having thoroughly examined the strengths and weaknesses of all these models, Lee conducted a number of experiments designed to test critical differences predicted by the various models. The results (which failed to confirm any single existing model) are used to create a modification of the earlier work with Longuet-Higgins, which handles metrical counterevidence better than before. There are four differences between this and earlier models: (a) it has a variable responsiveness to metrical counterevidence resulting in a certain amount of flexibility in its treatment of duration sequences; (b) it is rather more conservative about the position of the downbeat, placing it on the first event unless there is powerful counterevidence (a direct reflection of Lee's empirical findings); (c) the model is capable of metrical subdivision (moving to lower levels of the metrical hierarchy than the starting level); and (d) the model takes account of the effects of tempo. The heart of Lee's approach is summarized as follows:

> a) Every metre is associated with a (possibly culture-specific) pattern of 'strong' and 'weak' beats (henceforth termed the 'canonical accent-pattern' of the metre).
> b) The metrical grouping chosen by a listener at each successive level of a metrical hierarchy will preferably be one whose canonical accent-pattern is consistent with the natural accent-pattern of the portion of sequence under consideration: that is, the events occurring on the strong beats, t1 and t3. will preferably be no 'weaker' (perceptually less salient) than the one occurring on the intervening weak beat t2, which in turn will preferably be no weaker than any event occurring between t1 and t3.
> c) (Major) syncopations and weak long notes contradict the canonical accent-pattern and are hence avoided. (Lee, 1991, p. 121)

In a paper that provides a wide-ranging survey of a great deal of the previous work on rhythm perception, Parncutt (1994) proposes a theory of meter perception based on the salience of different possible pulse trains, and the perceived accentual strength of individual pulses, in music. The model, which as with the others discussed thus far is primarily concerned only with the relative onset times and durations of events (and not with their pitch, loudness, timbre, harmonic function, etc.), can be summarized as consisting of the following series of processes: (a) individual event durations are converted into phenomenal accents; (b) an absolute tempo factor (which recognizes the particular salience of periodic phenomena with a period of around 700 msec) and pattern-matching process select the most salient pulse trains (i.e., sequences of isochronous phenomenal accents)—the single most salient level of pulsation being identified as the tactus (the level at which a listener is most likely to tap his/her foot); (c) the three or four most salient and mutually consonant pulse trains are superimposed to create a metrical hierarchy for the sequence, with an associated overall salience. There is considerably more to the model, which has 'extensions' that handle issues such as expressive timing deviations, categorical rhythm perception, and the possible contribution of dynamic and other kinds of accent in real music, but its great strengths are its systematic simplicity (it is largely based on known processes of one kind or another, such as loudness adaptation or the identification of isochronous sequences), and the fact that it is virtually entirely spelled out as a quantitative model, thus permitting systematic and rigorous testing. Parncutt's own empirical data provide the first step in this direction.

If Parncutt's model is based on principles that are very close to classical psychophysics, Desain (1992) provides an equally intriguing approach that owes its origins to connectionism. The model emerges as an extension of Desain and Honing's (1989) connectionist approach to what they describe as the "quantization problem"—the extraction of discrete durational values in reasonable relationships with one another (essentially equivalent to standard Western rhythmic notation) from a string of continuously variable durations (equivalent to the raw data of human performance). Desain and Honing's connectionist quantizer uses the fundamental principle of the stability of small integer ratios, but does so in a simple network that considers not only the ratios between the individual intervals of the rhythmic sequence but also the compound intervals formed by summating two or more basic durations. Pairs of intervals (whether basic intervals, compound intervals, or a mixture of the two) are then 'steered' toward integer ratios through a number of iterating adjustments of the original noninteger values. The model is realized as a process such that the interonset intervals of a performed sequence are passed to the network one by one, are processed within a window of a certain size, and then shifted out of the network as quantized durations. Desain's (1992) model develops from the quantizer by considering how any new time interval is handled by the quantizer relative to the immediately preceding context. By holding the context fixed, which would normally mutually adjust with the new interval, ('clamping' it, to use Desain's term), the predictability of the next event onset can be assessed by considering how much adjustment or steerage is applied to the time interval formed between the last event of the clamped context and the new event. If the new event arrives at a time entirely consistent with the pattern of onset intervals formed by the context, then no steerage will be applied and the event can be considered to have occurred at an entirely predictable moment. If the interaction between the clamped context and the new event generates steerage toward a later moment in time, this demonstrates that the new event has occurred earlier than the context would predict. The converse is true for an event whose interaction with the context generates steerage toward an earlier point in time. Desain equates steerage with the inverse of expectancy: an event to which a positive steer is applied has occurred earlier than expected; an event with a negative steer is later than expected; and an event that generates no steer has occurred exactly at the point that the prior context would lead one to expect.

By plotting steerage/expectation curves for the positions of large numbers of hypothetical event positions following different prior rhythmic sequences, Desain demonstrates that rhythmic sequences that imply different metrical organizations generate distinct curves. Furthermore, these curves clearly show hierarchical metrical properties (mirrored in empirical results obtained by Palmer and Krumhansl, 1990), even though there are no explicitly metrical principles built into the model itself. Relying solely on the principle of the stability of small integer ratios between adjacent intervals (simple or compound), it nonetheless gives rise to archetypally metrical behavior as a result. Meter is, in other words, an emergent property of the model rather than a deliberate feature of its design—a characteristic that seems consistent both with meter's fundamental importance as a perceptual framework for music and listeners' abilities to acquire a sensitivity to meter at a very early age and without explicit formal instruction (Hargreaves, 1986; Moog, 1976).

A similarly subsymbolic model for meter perception is presented by Large and Kolen (1994), although it is based on rather different principles. The starting point for this approach is resonance theory, in which the behavior of oscillatory units that continuously adjust their phase and period to the rhythmic characteristics of a stimulus sequence is used to model the human response to rhythm. The fundamental idea is that neural units with differing natural resonances, and 'tuned' to be more or less sensitive/restrictive in terms of what they will adjust to, adapt to the periodicities of external stimulus events. These units exist at a number of levels/timescales, with their hierarchical relationships reflecting the hierarchical nature of meter itself. The authors show that a system using six oscillators covering a resonance range from 600 msec to 2560 msec will successfully track the meter of a fairly complex piece of real musical performance. In particular, oscillators at around 900 msec and 1800 msec (corresponding to the two primary levels of a binary metrical hierarchy) locked on and tracked tempo changes successfully, while the other oscillators (appropriately) did not lock on because they did not correspond to any significant level of periodic activity. Large and Kolen point out that their model (in common with all the models discussed in this section) pays no attention to phenomenal accents, although they briefly suggest that this might be incorporated in a simple manner by allowing accented events in the stimulus sequence to cause greater adjustments of the phase and period of relevant oscillators.

Once again using a largely knowledge-free and strongly biologically motivated approach, Todd (1994a, 1994b; Todd & Brown, 1996) has proposed a model for meter perception arising out of the filter-based model for grouping described earlier. In essence, Todd proposes that a frequency-domain multiscale filter system exists in parallel with the time-domain multiscale filtering that is responsible for detecting the grouping properties of rhythmic structures. Whereas the time-domain filters are low-pass, and provide information about the onset, stress, and grouping of events, the frequency-domain filters are band pass and provide information about tempo and meter. The output of the band-pass filters is simply a set of periodicities (a spectrum, but with the very low frequencies—typically below 10 Hz—that identify this as a rhythmic, rather than a pitch, phenomenon), which in itself does not specify a meter, even if it strongly constrains the set of possible meters. Todd proposes that a culture-specific top-down process then interprets the pattern of periodicities as a particular meter by recognizing one of a limited number of specified metrical patterns in the spectrum.

Finally, meter is not just an auditory and cerebral phenomenon: it also has an important motor component, and both Parncutt and Todd are concerned to build this into their models. This motor component is most evident in the way in which people tap their feet or dance to metrical music, the level in the metrical hierarchy at which they synchronize their foot tapping being commonly called the 'tactus' (e.g., Lerdahl & Jackendoff, 1983, p. 21). Parncutt (1987, 1994) takes account of this in his calculation of pulse-train salience by weighting each pulse train with a bell-shaped function centered on 600 msec, which has been cited as the modal spontaneous tapping period (see Fraisse, 1982; cf. also Fraisse's "indifference interval" discussed earlier), and also observes the similarity between the principle of this approach and that adopted by Terhardt, Stoll, and Seewann (1982) in calculating pitch salience in complex signals. In a closely related manner, Todd's model regards the tactus as the result of combining the output of the band-pass filter bank with a sensory-motor filter that has fixed and relatively narrow tuning characteristics and represents the intrinsic pendular dynamics of the human body (Todd & Lee, 1994). The output of the band-pass filter bank is fed through this sensory-motor filter, with the magnitudes of the individual filter outputs being modified by this secondary filter. The strongest output from this combined system will be that band-pass filter whose frequency lies closest to the center frequency of the sensory-motor filter[9]—a kind of second-level tuning system. The center frequency for the sensory-motor filter is none other than 600 msec again. Todd actually proposes a second, parallel, sensory-motor filter with a center frequency of 5 sec that also acts on the output of the band-pass filter bank, and which reflects the rate at which listeners and performers sway their bodies in listening, playing, and dancing.

[9]It is possible that a very strong output from a somewhat more distant band-pass filter would "win" over a closer but weaker output, but the spacing of the band-pass outputs actually makes this unlikely. However the basic principle is that the "winner" is calculated as the product of magnitude and proximity.

Before moving on from this section on rhythm perception, it is important to note that in addition to the distinction between grouping and meter, rhythm has been observed by a number of authors to have another dual aspect—temporal and accentual. Cooper and Meyer (1960), for example, deal with rhythm primarily in terms of accent structures, although they incorporate temporal factors into both their definition of accent and grouping of accents. By contrast, the vast majority of empirical work on rhythm has focused on temporal matters, almost to the exclusion of accent. A smaller literature does exist, however, that has considered the nature and variety of different kinds of accent and the ways in which accentual and temporal factors mutually influence one another (e.g. Dawe, Platt, & Racine, 1993; Drake & Palmer, 1993; Povel & Okkerman, 1981; Thomassen, 1982; Windsor, 1993; Woodrow, 1909). Useful though this work has been, there has been little attempt so far to integrate what is known about the perceptual functioning of anything other than temporal accents with existing or new models of meter perception. This is unfortunate both because purely temporal models of meter perception (such as those reviewed in this chapter) are unrealistic in the demands that they make by deriving meter from temporal information alone and because such models tend to project a static and one-dimensional view of meter, rather than the more dynamic and fluid reality that is the consequence of the interplay of different sources of perceptual information (temporal and accentual).

V. TIMING IN MUSIC

If there is one principle on which virtually all rhythm research agrees, it is that small integer ratios of duration are easier to process than more complex ratios. And yet this principle is at first sight deeply paradoxical when the temporal properties of real musical performances are examined, for almost nowhere are small integer ratios of duration found—even in performed sequences that are manifestly and demonstrably simple for listeners to interpret. The answer to this apparent paradox is that a distinction must be made between the structural properties of rhythm (which are indeed based on the principle of small integer ratios) and their so-called expressive properties—continuously variable temporal transformations of the underlying rhythmic structure (Clarke, 1985). These temporal transformations, referred to by some authors (e.g., Clynes, 1987, 1983; Repp, 1992a) as expressive microstructure, are what the term "timing" identifies, and there has been considerable attention paid to the nature and origins of these timing properties in performed music, as well as a rather smaller literature on their perceptual consequences for listeners.

A. PERCEPTION OF TIMING

The distinction between rhythm and expressive timing is only psychologically plausible if some mechanism can be identified that is able to separate the one from the other. That mechanism is categorical perception, and both Clarke (1987b) and Schulze (1989) have demonstrated its existence empirically. In general terms, the idea is that listeners assign the continuously variable durations of expressive performance to a relatively small number of rhythmic categories. The pattern of these categories constitutes the rhythmic structure of the sequence, and the departure of each duration in the original performance from its appropriate categorical target value is understood as expressive timing. Clarke (1987b) and Parncutt (1994) have suggested that categorical perception may be confined to a single categorical distinction—between a 1:1 ratio and a 2:1 ratio, or even more generally to the distinction between even (1:1) and uneven (*N*:1) divisions of a time span. Even such a simple categorical system can be quite powerful, however, particularly when the interaction between categorical distinctions and the prevailing metrical context is considered. Two durations in a relationship of inequality in a duple meter are likely to be interpreted as a 3:1 (or possibly 7:1) ratio, whereas the same inequality in a triple meter will be interpreted as 2:1 or 5:1. The interdependence between type of division (even/uneven) and meter (triple/duple) brings with it an expressive component, because the same objective pair of durations interpreted as 2:1 in one meter and 3:1 in another must imply different expressive departures from the target (canonical) values. For example, the sequence 600 msec–400 msec–1000 msec might be interpreted in a duple meter as a 1:1:2 pattern with an expressive lengthening of the first event. In a triple meter, the same sequence might be interpreted as a 2:1:3 pattern with an expressive lengthening (slowing) over the second and third events.[10] Thus the raw durational information specifies perceptual information in three interdependent dimensions: meter (duple/triple), division (even/uneven), and expression.

Desain and Honing (1989) have presented a connectionist model for the recovery of underlying rhythmic structure from continuously variable performed durations. The model takes small integer ratios as its target categories and steers the individual durations (and compounds formed from pairs of adjacent durations) toward their nearest target integers. The smaller the integer, the stronger is the steering, so that 1:1 and 2:1 exert the most powerful influence on the behavior of the system. The model is successful in correctly interpreting even quite difficult sequences (e.g., correctly interpreting the same duration in two different sequential positions as two different rhythmic values), and the behavior of the system appears to mimic the limited amount of empirical data on categorical perception quite closely.

If categorical perception is one way to account for the distinction between rhythmic structure and expressive timing, it remains to demonstrate the abilities of

[10]The overwhelming majority of Western music is based on the distinction between duple and triple meters, or compounds of duple and triple. This in no way precludes other metrical possibilities, which would exert their influence on the relationship between rhythmic structure and expressive timing in a suitably altered manner. For example the 600–400–1000 sequence in a quintuple meter would be interpreted as an inexpressive (i.e., metronomic) 3:2:5 pattern.

listeners to detect expressive timing in rhythmic sequences. Kendall and Carterette (1990) showed that listeners were successful in picking up the expressive intentions (neutral, normal, exaggerated) of performers conveyed by means of timing (and dynamics), on a variety of instruments, and that there was no difference between musicians and nonmusicians in their ability to make this perceptual judgment. Sloboda (1983) showed that listeners (musicians) who were asked to distinguish between two metrical variants of a tune were able to do so with a degree of success that directly reflected the clarity of metrically related expression (timing, dynamics, and articulation) in the original performances. In his study, the performance data of more expert performers showed a more clearly distinguished pattern of expressive features for the two metrical variants of the melody, and the subsequent listeners were more successful in distinguishing between the two metrical variants in their performances than in the performances of less expert performers. The magnitude of the expressive changes recorded in these data varies considerably, but in a study that focused on timing changes alone, Clarke (1989) showed that listeners are sensitive to changes in duration of as little as 20 msec in simple isochronous sequences, and those listeners were still able to detect the lengthening of a single duration by as little as 50 msec in a melody that had base durations of 350 msec and had continuously modulated expressive changes in tempo.

Listeners' sensitivities to expressive timing are variable, however, and Repp (1992b) has shown that there are strong structural constraints on listeners' perceptions of expressive timing. In his study, various degrees of lengthening, ranging from approximately 23 msec to 45 msec (corresponding to between 7% and 13% of a 341-msec eighth note) were applied to each of the 47 possible eighth-note durations in a metronomic rendering of the first part of a minuet by Beethoven, and listeners were then asked to indicate where they heard a lengthening. The overall percentage of correct identifications was well above chance even for the smallest amount of lengthening, bearing out the evidence that listeners have a high level of sensitivity to timing perturbations, but more significant than this was the pattern of changes in percentage correct responses across the extract. There was remarkable variation in this profile, ranging from 0% to 90% correct, with the peaks and troughs mirroring the phrase structure of the melody, and having a strongly negative correlation with the timing profile of a separately analyzed expert performance of the music. At phrase boundaries, where the expert performer showed a tendency to lengthen the duration of eighth notes, listeners' abilities to detect a timing perturbation in the expressionless performance declined markedly and did so in proportion to the structural importance of the boundary. By contrast, in the middle of phrases, listeners' detection scores reached their peak values. This is strong evidence to show that listeners' unconscious parsing of the musical structure, and the expectations that follow concerning the likely expressive treatment of the music by a performer, have a striking effect on listeners' abilities to detect timing changes in the music.

B. EXPRESSIVE TIMING IN PERFORMANCE

In recent years, a considerable body of research has built up aimed at specifying the principles that govern expressive performance in music (e.g., Clarke, 1988; Gabrielsson, 1988; Palmer, 1989; Repp, 1992a; Todd, 1989, 1985; Shaffer, 1981; Sundberg, 1988; see also Chapter 14, this volume). This research has used a mixture of empirical measurement and computer simulation to explore the ways in which human performers expressively transform various musical dimensions (e.g., tempo, loudness, intonation) in performance, with a particular focus on timing, and has identified a number of recurring characteristics that indicate a particular model of the origin and control of expression. The critical evidence is that expressive timing can be extremely stable over repeated performances that may sometimes span a number of years (Clynes & Walker, 1982), is found even in sight-reading performances (Shaffer, 1981), and can be changed by a performer at a moment's notice (Clarke, 1985). Taken together, these observations mean that expressive timing cannot possibly be understood as a learned pattern that is applied to a piece each time it is played, but must be generated from the performer's understanding of the musical structure.[11] Any other model imposes memory demands on a performer that are completely implausible psychologically and is unable to account for the mixture of stability and flexibility that has already been mentioned. The stability of performances over time is thus understood as the stability of a performer's representation of the musical structure; the existence of expression in sight-reading performances is the consequence of a performer's emerging representation of the music as he or she reads and organizes it; and changes in expression, either spontaneous or the result of instruction, are a consequence of the fact that musical structures can be interpreted in a variety of different ways.

In principle, every aspect of musical structure contributes to the specification of an expressive profile for a piece, but a number of authors have shown that phrase structure is particularly salient. Todd (1985) details a model that takes the hierarchical grouping structure of the music as its input and gives a pattern of rubato as its output on the basis of an extremely simple rule (Todd, 1985, 1989). The resulting timing profiles compare well with the profiles of real performances by professional players, as Todd's own data and subsequent data collected by Repp (1992a) have shown. A number of other studies have also shown rule-like correspondences between various aspects of musical structure and expression (e.g., Clarke, 1988; Shaffer & Todd, 1987; Sloboda, 1983; Sundberg, 1988). In a study of 28 performances of a short piano piece by Schumann, taken from commercial recordings by many of this century's greatest pianists, Repp (1992a) showed that there was a remarkable degree of commonality underlying the expressive timing profiles of the performances, despite the idiosyncratic nature of some of the performers. He

[11]Note that this does not mean that expression cannot be learned, only that such learning takes place on a foundation of musical structure as perceived, or conceived, by the performer.

also showed convincingly that at more surface levels of the timing profiles, increasing diversity between the performers is found, thus confirming the intuition that performers agree substantially about the deeper structural levels of a piece of music, and impose the stamp of their own individuality by manipulating the finer details of structure and its expressive implementation.

One way to investigate the generative model of performance expression is to use an imitation (reproduction) task: the idea is that pianists hear performances of short melodies and then have to imitate, as precisely as possible, all the nuances of timing by playing the melody back on a piano. Some of the "performances" that they hear are real performances by a pianist; others are versions that have been transformed in various ways so as to disrupt the relationship between structure and expressive timing. One such transformation inverts the expressive timing profile of the melody, so that passages in which the player originally accelerated now decelerate by the same proportion, and vice versa. Another transformation translates the pattern of timing along the melody by some specified amount, so that (for instance) the expressive timing associated with the first note (or more strictly the first interonset interval) is transferred onto the fourth note, the second onto the fifth, the third onto the sixth, and so on. The purpose of these different transformations is to introduce differing degrees of disruption into the relationship between structure and expressive timing, so as to assess how this disruption influences the pianists' abilities to imitate what they hear. A strictly generative theory would predict that only when there is a reasonable relationship between structure and expression can a pianist successfully imitate a performance, because only then can the performer grasp a structure/expression pairing sufficiently to regenerate (imitate) it.

The results of experiments (Clarke, 1993; Clarke & Baker-Short, 1987) largely bear out this prediction, pianists being more inaccurate and inconsistent when they try to imitate an expressive timing profile that does not maintain a conventional relationship with the musical structure, the degree of inaccuracy and inconsistency being directly related to how disrupted or abnormal the relationship is. Similarly, a separate group of musicians who were asked simply to listen to the same sequences and judge their quality as performances gave ratings that followed the same pattern: real performances were rated best, and the more disrupted was the relationship between structure and expression, the lower were the listeners' ratings.

Although these findings lend strong support to a generative model of expression, there is also evidence from the imitation attempts that performers are at least partially successful in their attempts to imitate the disrupted versions. There are various possible explanations of this: pianists may create some kind of direct auditory "image" of the performance, which they then try to match as closely as possible when they make their imitation attempt; or they may try to remember some kind of verbal blueprint for the peculiar performance (e.g., "speed up toward the end of the first phrase, slow down during the middle of the second phrase, and then rush the end"); or they may encode the performance in terms of some type of body image—a kind of mental choreography that will recreate the performance out of an image of the movements (real or imagined) that might produce, or express the shape of, such a performance. There is informal evidence that a movement component of this sort is involved: some of the pianists participating in the study were observed to move in quite striking ways when attempting to imitate the more disrupted versions, using movements that seemed to express and capture the awkward timing characteristics of the music. Whatever the strategy used, it is clear that a model that portrays expressive timing simply as the outcome of a set of expressive rules is unrealistically abstract and cerebral, and that the reality is far more practical and corporeal. The body is not just a source of sensory input and a mechanism for effecting output: it is far more intimately bound up with our whole response to music—perceptual and motor.

VI. RHYTHM, TIMING, AND MOVEMENT

As far back as the ancient Greeks, writers have remarked on the close relationship between music and human movement (see Barker, 1989). In a study using factor analysis and multidimensional scaling methods, Gabrielsson (1973a; 1973b) showed that listeners' responses to rhythms (either descriptive adjective ratings, or similarity ratings between pairs of rhythms) showed the presence of a perceptual dimension that Gabrielsson termed "movement character." Although a structural dimension and an emotional dimension also played an important part in listeners' judgments, for some listeners the movement character was so strong as to constitute the primary dimension used in making similarity judgments. These listeners described how in trying to assess the similarity of pairs of rhythms, they tried "to 'find out the similarity by movements of the body' and/or 'by imagining which movements I could do to the music'" (Gabrielsson, 1973a, p. 173).

More recently, the work of a German author, Alexander Truslit, has been brought to light (see Repp, 1993). Truslit carried out experimental research in the 1930s showing that different kinds of movement instruction (or movement image) given to performers resulted in performances with measurably different timing properties, from which he went on to claim that a very small number of general movement types, perhaps no more than three, underpinned the kinematic basis of performance. Independently of this, other researchers have shown that the pattern of timing that performers spontaneously use follows the temporal curve of objects moving in a gravitational field (Feldman, Epstein, & Richards, 1992; Sundberg & Verillo, 1980; Todd, 1992), suggesting that performances that sound natural do so because they mimic the behavior of physical objects moving in the real world. Todd (1992) has shown that a model of performance timing and dynamics based on the speed and force of movement of objects moving under the influence of gravity can give a good account of the expression found in spontaneous musical performances. In a similar way, Repp (1992c) compared a family of parabolic timing curves (which produce a close approximation to the linear tempo function used by Todd) with other timing functions in a study that tested listeners' preferences for different versions of a melodic gesture from a Schumann piano piece. He found that listeners preferred the parabolic curves over the other timing functions, and among the parabolic curves themselves, preferred parabolas that maintained a symmetrical shape (rather than being skewed right or left).

The relationship between rhythm and movement can be conceptually separated into rhythm and timing seen as the consequence of movement, and rhythm and timing seen as the source of, or motivation for, movement. The first of these two relationships is primarily an issue of motor control: timing information can either be seen as the input to a motor system, which then produces some kind of temporally structured behavior, or timing can be seen as the consequence of the intrinsic characteristics of the motor system and the body itself. This is part of the wider question of whether temporal control in behavior should be seen as regulated by an internal clock of some kind (see, e.g., Luce, 1972), or as the temporal expression of the intrinsic dynamics of a system (e.g., Kugler, Kelso, & Turvey, 1980). Shaffer (1982, 1984) has proposed that these options are not mutually exclusive and that the two principles operate at different levels. In music, the primary level of timing, and the level at which some kind of internal clock exerts its influence, is the tactus (that level of the metrical structure at which a listener might tap his/her foot, or a conductor beat time). Subdivisions of the beat (i.e., individual notes) are not directly timed, but are produced by overlearned motor procedures that specify movement patterns that have as their consequence a definite timing profile. The timing properties are thus the consequence of movement rather than a control parameter in their own right: "Note timing is, in effect, embodied in the movement trajectories that produce them" (Shaffer, 1984, p. 580). Equally, time periods greater than that of the primary timing level are produced by concatenations of beat periods rather than by means of some higher level clock. A hierarchy of clocks is therefore not involved, despite the multileveled nature of the timing profiles that are characteristic of expert expressive performance.

Turning now to rhythm as the source of, or motivation for, movement, the most striking and obvious evidence for this close relationship comes from the ancient association between music and dance—an association that in contemporary popular culture remains as vital as at any time. Equally, there is a very long history of work songs, and here it is clear that one of the primary functions of rhythmic singing is to coordinate the activities of people working together (Blacking, 1987, chapter 3) and through this rhythmic action to optimize the efficiency and economy with which energy is expended (Bernstein, 1967). With this deep-seated association in the history of music and human action, it is hardly surprising that music with a periodic rhythmic structure tends to elicit accompanying movements, whether these are explicitly dance movements or less formalized responses and whether the music is intended to be dance music or not. Todd (1993, 1995) has proposed that the auditory system interacts directly with two subsystems of the motor system, one responsible for relatively rapid periodic phenomena (foot tapping) and the other associated with slower and less strictly periodic movements (body sway and whole body movement generally). These interactions are responsible for our strong motor response to music, and the two subsystems embody the distinction between genres of dance in which the dancer's center of moment (Cutting, Proffitt, & Kozlowski, 1978) remains essentially fixed while the limbs move periodically and relatively rapidly around it (disco dancing), and those in which the center of moment itself moves, with more limited movement of the limbs around that center (ballroom dancing).[12]

VII. SUMMARY

The temporal dimension of music offers an enormous diversity of issues to the psychology of music. Having distinguished between rhythm and form on the basis of the length of the perceptual present, and the attendant distinction between perception and construction, this chapter has adopted a view of rhythm that sees it as the interaction between meter and grouping. Although there are theoretical accounts of the conditions for the formation of grouping structures, comparatively little work has explored this empirically. There is a greater diversity of models to account for meter perception, ranging from rule-based systems to connectionist and auditory models, although once again empirical investigations have been rather more limited in both quantity and scope.

The term *timing* is used in this chapter to refer to the temporal microstructure that is characteristic of performances of music and is widely regarded as the generative consequence of a performer's conception of musical structure. The relationship between this continuously variable component and the discrete categories of rhythmic structure is discussed both conceptually and perceptually, with categorical perception playing a crucial role. Finally, the significance of the relationship between rhythm, timing, and movement is considered in both perception and performance. In this, as elsewhere in research in rhythm and timing, there has been a shift from a rather abstract symbolic approach to perception and production toward an outlook that takes more account of properties of the auditory and motor systems, and of the body in general, or makes use of subsymbolic principles, which require fewer explicit rules to be built into the models.

Inevitably in a field of this size, many issues have not been tackled: no consideration has been given here to the aesthetic consequences of different kinds of temporal organization, or to the perceptual and performance implications of historical changes in the temporal organization of music, or to rhythm and timing in ensemble coordination, or to the relationship between rhythmic organization and

[12]Somewhat controversially, Todd (1993) has claimed a still closer link between rhythm and movement by suggesting that the vestibular apparatus may be directly stimulated by sound and that the corresponding sense of motion, or "strong compulsion to dance or move" (Todd, 1993, p. 381) is the direct result. Some evidence from animal studies shows that the necessary anatomical links may exist, but a behavioral expression of such a proposed vestibular-motor link has yet to be demonstrated in humans.

trance or other altered states of mind, or to the development of children's and adults' perceptual and performance skills in the domain of rhythm and timing. But with the exception of the last item in this list, these are all areas that have received very little consideration in the psychological literature and thus offer the enticing prospect of still greater diversity in future rhythm research.

ACKNOWLEDGMENTS

I am grateful to Richard Parncutt, Neil Todd, and the Editor for their helpful comments on an earlier draft of this chapter. I thank Neil Todd for providing the data necessary to produce Figure 1.

REFERENCES

Baddeley, A. (1986). *Working memory*. Oxford: Clarendon Press.

Barker, A. (1989). *Greek musical writings: Vol. 51. Harmonic and acoustic theory*. Cambridge: Cambridge University Press.

Bernstein, N. (1967). *The coordination and regulation of movements*. Oxford: Pergamon.

Blacking, J. (1987). *'A commonsense view of all music': Reflections on Percy Grainger's contribution to ethnomusicology and music education*. Cambridge: Cambridge University Press.

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

Clarke, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 209–236). London: Academic Press.

Clarke, E. F. (1987a). Levels of structure in the organisation of musical time. *Contemporary Music Review, 2* (1), 211–238.

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

Clarke, E. F. (1988). Generative principles in music performance. In J. Sloboda (Ed.), *Generative processes in music* (pp. 1–26). Oxford: The Clarendon Press.

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

Clarke, E. F. (1993). Imitating and evaluating real and transformed musical performances. *Music Perception, 10,* 317–343.

Clarke, E. F., & Baker-Short, C. (1987). The imitation of perceived rubato: A preliminary study. *Psychology of Music, 15,* 58–75.

Clarke, E. F., & Krumhansl, C. (1990). Perceiving musical time. *Music Perception, 7,* 213–252.

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg (Ed.), *Studies of music performance* (pp. 76–181). Stockholm: Royal Swedish Academy of Music.

Clynes, M. (1987). What can a musician learn about music from newly discovered microstructure principles (PM and PAS)? In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 201–233). Stockholm: Royal Swedish Academy of Music.

Clynes, M., & Walker, J. (1982). Neurobiologic functions of rhythm, time and pulse in music. In M. Clynes (Ed.), *Music, mind and brain: The neuropsychology of music* (pp. 171–216). New York: Plenum.

Cook, N. (1990). *Music: Imagination and culture*. Oxford: Oxford University Press.

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

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96,* 341–370.

Cowan, N. (1987). Auditory memory: Procedures to examine two phases. In W. A. Yost & C. S. Watson (Eds.), *Auditory processing of complex sounds* (pp. 289–298). Hillsdale, NJ: Erlbaum.
Crowder, R. (1993). Auditory memory. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 113–145). Oxford: The Clarendon Press.
Cutting, J. E., Proffitt, D. R., & Kozlowski, L. T. (1978). A biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 357–372.
Dawe, L. A., Platt, J. R., & Racine, J. R. (1993). Harmonic accents in inference of metrical structure and perception of rhythm patterns. *Perception & Psychophysics, 54,* 794–807.
Deliège, I. (1987). Grouping conditions in listening to music: An approach to Lerdahl & Jackendoff's grouping preference rules. *Music Perception, 4,* 325–360.
Deliège, I. (1993). Mechanisms of cue extraction in memory for musical time. *Contemporary Music Review, 9,* 191–205.
Deliège, I., & Ahmadi, A. E. (1990). Mechanisms of cue extraction in musical groupings: A study of perception on *Sequenza VI* for viola solo by Luciano Berio. *Psychology of Music, 18,* 18–44.
Desain, P. (1992). A (de)composable theory of rhythm perception. *Music Perception, 9,* 439–454.
Desain, P., & Honing, H. (1989). The quantization of musical time: A connectionist approach. *Computer Music Journal, 13* (3), 56–66.
Deutsch, D. (1986). Recognition of durations embedded in temporal patterns. *Perception & Psychophysics, 39,* 179–186.
Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception, 10,* 343–378.
Feldman, J., Epstein, D., & Richards, W. (1992). Force dynamics of tempo change in music. *Music Perception. 10,* 185–204.
Fraisse, P. (1956). *Les structures rythmiques.* Louvain, Paris: Publications Universitaires de Louvain.
Fraisse, P. (1963). *The psychology of time.* New York: Harper & Row.
Fraisse, P. (1978). Time and rhythm perception. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception: Vol. 8* (pp. 203–254). New York: Academic Press.
Fraisse, P. (1982). Rhythm and tempo. In D. Deutsch (Ed.), *The psychology of music* (pp. 149–180). New York: Academic Press.
Fraisse, P. (1987). A historical approach to rhythm as perception. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 7–18). Stockholm: Royal Swedish Academy of Music.
Gabrielsson, A. (1973a). Similarity ratings and dimension analyses of auditory rhythm patterns. II. *Scandinavian Journal of Psychology, 14,* 161–176.
Gabrielsson, A. (1973b). Adjective ratings and dimension analyses of auditory rhythm patterns. *Scandinavian Journal of Psychology, 14,* 244–260.
Gabrielsson, A. (1988). Timing in music performance and its relation to music experience. In J. Sloboda (Ed.), *Generative processes in music* (pp. 27–51). Oxford: The Clarendon Press.
Garner, W. R. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.
Handel, S. (1989). *Listening: An introduction to the perception of auditory events.* Cambridge, MA: MIT Press.
Hargreaves, D. J. (1986). *The developmental psychology of music.* Cambridge: Cambridge University Press.
Johnson-Laird, P. N. (1991). Rhythm and metre: a theory at the computational level. *Psychomusicology, 10,* 88–106.
Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8,* 129–164.
Kramer. J. (1988). *The time of music.* New York: Schirmer Books.
Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behaviour.* New York: North Holland.
Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science, 6,* 177–208.

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

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

Longuet-Higgins, H. C. (1979). The perception of music. *Proceedings of the Royal Society of London. B 205,* 307–322.

Longuet-Higgins, H. C. and Lee, C. S. (1982). The perception of musical rhythm. *Perception*, 11, 115-128.

Luce, G. G. (1972). *Body time*. London: Temple Smith.

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

McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146–198). Oxford: The Clarendon Press.

Michon, J. A. (1978). The making of the present: a tutorial review. In J. Requin (Ed.), *Attention and performance VII* (pp. 89–111). Hillsdale, NJ: Erlbaum.

Michon, J. A. (1985). The compleat time experiencer. In J. A. Michon & J. L. Jackson (Eds.), *Time, Mind and Behaviour* (pp. 20–52). Berlin: Springer.

Moog, H. (1976). *The musical experience of the pre-school child* (C. Clarke, Trans.). London: Schott.

Nattiez, J.-J. (1975). *Fondements d'une sémiologie de la musique*. Paris: Union Générale d'Editions.

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

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 3–18.

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

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

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

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

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

Repp, B. (1992a). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America, 92,* 2546–2568.

Repp, B. (1992b). Probing the cognitive representation of musical time: structural constraints on the perception of timing perturbations. *Cognition, 44,* 241–28 1.

Repp, B. H. (1992c). A constraint on the expressive timing of a melodic gesture: Evidence from performance and aesthetic judgment. *Music Perception, 10,* 221–243.

Repp, B. H. (1993). Music as motion: a synopsis of Alexander Truslit's (1938) *Gestaltung und Bewegung in der Musik. Psychology of Music, 21,* 48–73.

Ruwet, N. (1972). *Langage, musique, poésie*. Paris: Seuil.

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

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

Shaffer, L. H. (1982). Rhythm and timing in skill. *Psychological Review, 89,* 109–123

Shaffer, L. H. (1984). Timing in solo and duet piano performances. *Quarterly Journal of Experimental Psychology, 36A,* 577–595.

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

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

Sundberg, J. (1988). Computer synthesis of music performance. In J. Sloboda (Ed.), *Generative processes in music* (pp. 52–69). Oxford: The Clarendon Press.

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

Tenney, J., & Polansky, L. (1980). Temporal Gestalt perception in music. *Journal of Music Theory, 24,* 205–241.

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

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

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

Todd, N. P. (1989). A computational model of rubato. *Contemporary Music Review, 3,* 69–89.

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

Todd, N. P. (1993). Vestibular feedback in musical performance: Response to *Somatosensory Feedback in Musical Performance* (Ed. Sundberg and Verillo). *Music Perception, 10,* 379–382.

Todd, N. P. (1994a). The auditory "primal sketch": a multiscale model of rhythmic grouping. *Journal of New Music Research, 23,* 25–70.

Todd, N. P. (1994b). Metre, grouping and the uncertainty principle: A unified theory of rhythm perception. *Proceedings of the Third International Conference for Music Perception and Cognition* (pp. 395–396). Liège, Belgium: ICMPC.

Todd, N. P. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America, 97,* 1940–1949.

Todd, N. P., & Brown, G. (1996). Visualization of rhythm, time and mètre. *Artificial Intelligence Review, 10,* 253–273.

Todd, N. P., & Lee, C. S. 1994. An auditory-motor model of beat induction. In *Proceedings of the International Computer Music Conference,* (pp. 88–89). Aarhus, Denmark: International Computer Music Association.

Vos, P. G. (1978). *Identification of metre in music* (Internal report 78 ON 06). Nijmegen, The Netherlands: University of Nijmegen.

Windsor, W. L. (1993). Dynamic accents and the categorical perception of metre. *Psychology of Music, 21,* 127–140.

Woodrow, H. (1909). A quantitative study of rhythm. *Archives of Psychology, 14,* 1–66.

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