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.
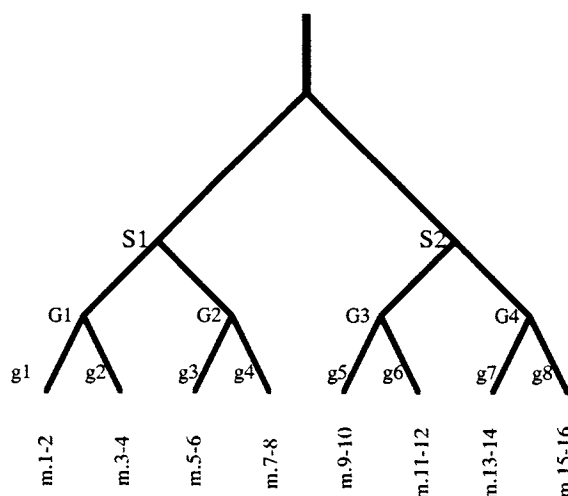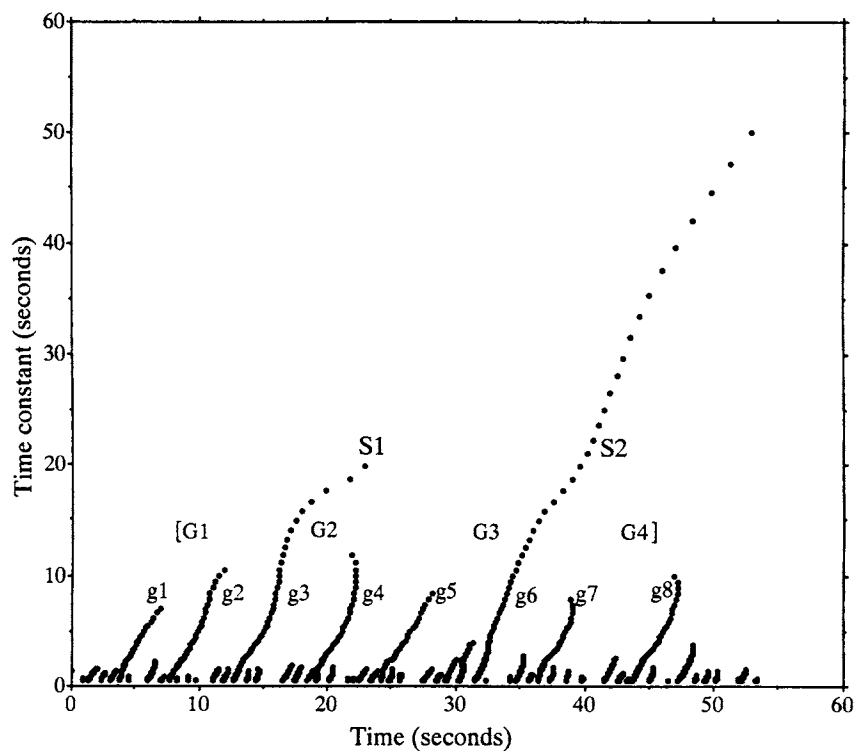
FIGURE 1   Rhythmogram of a professional pianist's performance of the Chopin Prelude Op. 28, No. 7 (top half) together with a Lerdahl and Jackendoff grouping analysis of the music (bottom half) showing the close relationship between the two. Units g1–g8 in both parts of the figure indicate two-measure 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 accen-

tual 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 inter-

vals 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

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 metre. *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.

14

THE PERFORMANCE OF MUSIC

ALF GABRIELSSON

Department of Psychology
Uppsala University
Uppsala, Sweden

I. INTRODUCTION

Music performance is a large subject that can be approached in many different ways. This chapter focuses on empirical research of music performance and related matters. Most of this research is concerned with Western tonal music and mainly art music, an obvious limitation that should be kept in mind. Studies on performance of single notes, intervals, chords and so forth—that is, with no musical context—are usually not discussed, nor are matters related to music teaching; for the latter, see the handbook edited by Colwell (1992). Singing is only cursorily discussed, because it is treated in the chapter by Sundberg (this volume).

The topics addressed follow a kind of chronological order, beginning with the planning of performance, proceeding to various aspects of the performance itself (sight-reading, improvisation, feedback, motor processes, measurements, and models), then to physical, psychological, and social factors that may influence the performance, and finally to performance evaluation.

Some textbooks on music psychology include questions concerning performance. Mursell (1937/1971) compared vocal and instrumental performance and discussed problems in interpretation and technique. C. E. Seashore (1938) summarized extensive measurements of performance made by him and his coworkers, also cited in Lundin (1985). Schoen (1940) described studies of artistic singing. Lundin (1985) discussed learning and remembering music from a behavioristic point of view. Sloboda (1982a, 1985b) discussed performance plans, sight-reading, rehearsal, and expert performance with a cognitive approach. These and other questions on music performance were also elaborated by different authors in the

volumes edited by Sloboda (1988), Jones and Holleran (1992), and Bruhn, Oerter, and Rösing (1993).[1]

II. PERFORMANCE PLANNING

Excellence in music performance involves two major components: (a) a genuine understanding of what the music is about, its structure and meaning, and (b) a complete mastery of the instrumental technique. Gerig (1976, Chap.1) cited a large number of famous pianists and concluded that it is not necessary to argue for the importance of technique, but "the technical objective at the same time becomes *the* means to a far greater end—the projection of a meaningful interpretation" (p. 1). Statements to the same end are found in Bastian (1989, p. 172), LaBerge (1981a, 1981b), Leimer (1932/1972, p. 46), Restle (1981), or Sloboda (1985b, p. 90). In the words of Pablo Casals: "One needs to have a clear idea of the musical thought, and execute it precisely" (cited from Clynes, 1987, p. 203).

In accordance with this, one may distinguish two interrelated steps in the planning of performance:

1. To acquire an adequate mental representation of the piece of music, coupled with a plan for transforming this representation into sound, and
2. To practice the piece to a level that is satisfactory for the purpose at hand.

A. REPRESENTATION AND PERFORMANCE PLAN

The way in which a performer generates a mental representation of the music differs depending on the type of music and the instrument used, further on his or her experience and knowledge, personality, situational demands, and so forth. For a new piece, one may have to start from scratch; in other cases, it may be possible to retrieve or modify a representation that is already stored. The structural characteristics of the piece are crucial, both the overall structure (e.g., hierarchical, chainlike) and the structure of various units. Clarke (1988, pp. 2–5) discussed the differences in structural representations when music is performed from memory, in sight-reading, and in improvisation. In memorized performance the representation is (should be) fairly complete, depending on the length and complexity of the piece, and is just "unpacked" during the course of the performance. However, in sight-reading and in improvisation the representation is by necessity incomplete in various respects.

Palmer and de Sande (1993) attempted to study units in the representation of homophonic and polyphonic music by analyzing errors in piano performance of such pieces. They predicted and found more chord errors in homophonic than in

1. After the completion of this chapter (August 1995), another review "Music Performance" by Caroline Palmer, focusing on interpretation, planning, and movement was published in *Annual Review of Psychology*, 1997, *48*, 115–138. A volume on *The Practice of Performance*, edited by John Rink (Cambridge: Cambridge University Press, 1995), contains both empirical and theoretical papers on music performance.

polyphonic performances and more single-note errors in polyphonic than in ho-
mophonic performances. Furthermore, error intrusions were more often harmoni-
cally congruent with the correct notes in homophonic than in polyphonic perfor-
mances. Errors were fewer in the melodic (most important) voice than in other
voices, especially when it was the highest in pitch level. In other studies, Palmer
(1989a, 1989c, 1992) studied pianists' performance in relation to their intended
phrasing in pieces by Chopin and Brahms. Phrase endings were characterized by a
ritard, indicating that phrases were important constituents in the mental represen-
tation. Analysis of performance errors showed, for instance, that notes deleted in
the performance were placed within phrases rather than at the structurally more
important phrase endings (see further in Section VII,C,9).

Structural components suggested by these studies—melody, voices, chords,
phrases—are, of course, familiar from general music theory. However, Palmer
pointed out that theories that try to model the intuitions of an idealized listener in
terms of what is notated in the score disregard individual interpretive preferences.
The notation is often open to various interpretations regarding the connections and
relative importance of various structural units (cf. also Sloboda, 1994). The cho-
sen interpretation will in its turn influence the (real) listener's structural represen-
tation of the piece.

The emphasis on structural features in the mental representation has been char-
acteristic for cognitive music psychology. Important as they are, however, they do
not exhaust what is contained in most performers' representations (Gabrielsson,
1988, 1994). Every performance involves some kind of intention from the
performer's side concerning what the music should express or convey to the lis-
tener, be it ideas, associations, memories, feelings, body movements, concrete
events, or just patterns of sound. For instance, in a study by Persson, Pratt, and
Robson (1992) concerning the interpretation of an unknown piece of piano music,
some performers represented the piece in terms of images, scenes, things, events,
characters, and moods and used these as directions for how to perform the piece.
They seemed to be more concerned with such types of representation than with
structural matters. Clarke (1993a, 1993b) discussed the possibility of representing
the music by verbal description or in terms of body movements (see Section VI,D).

Shaffer (1992) questioned the idea of identifying musical meaning with musi-
cal structure, because listeners tend to hear, and performers try to convey, moods
and emotions in the music (see also Section VI,D). He suggested that music can
provide an abstract narrative: "we can think of the musical structure as describing
an implicit event and the gestures of musical expression as corresponding to the
emotional gestures of an implicit protagonist who witnesses or participates in the
event. Thus the performer's interpretation can be viewed as helping to define the
character of the protagonist. This in turn determines the patterning of mood over
the event" (p. 265). Shaffer gave the notation of an unknown piano piece by Bee-
thoven to four pianists and asked them to provide their preferred version of it as
well as any other musically valid version. Three of the pianists got the full score,
but one got a score where all expressive markings were deleted. Although the
former pianists showed high consistency in timing and dynamics across their dif-

ferent versions, this was not the case for the latter pianist. This may be taken as evidence of the power of expressive markings to constrain the interpretation, that they "help to crystallize a narrative for the piece" (p. 273). The idea of considering performance both as elucidating structure and inventing a musical character was further elaborated by Shaffer (1995).

Baily (1985, 1990, 1991) pointed out that the representation of music also involves motor processes. Using examples from African and Afghan music, he claimed that music is as much a question of movement as of sound. The spatial properties of an instrument in combination with convenient movement patterns in fingers, hands, and arms may very well be the decisive factors for the shape of a piece of music. In African music, the representation "may be a movement representation rather than an auditory one" (1985, p. 242). He argued that movement properties are important in Western music as well. Composers take the properties of the instruments and the performers' motor skills into account when creating music. This is especially obvious in the case of études. Furthermore, anyone familiar with improvisation knows how much it is determined by movement patterns; see, for example, Sudnow's (1978) book *Ways of the Hand* dealing with jazz piano improvisation. Berliner (cited in Baily, 1991, p. 150) quoted a saxophonist who indicated that "sometimes the ideas come from my mind ... but other times they come from my fingers." Gay (cited in Baily, 1991) pointed out that in much rock music the musical conceptualization is achieved with reference to the spatial layout of notes and the physical structure of the guitar. There are therefore good reasons to "regard auditory and spatio-motor modes of musical cognition as being of potentially equal importance" (Baily, 1991, p. 150).

A related, yet different, point of view was advanced by Truslit (1938). (An English synopsis of this work was provided by Repp, 1993; see also Repp, 1994a.) Truslit argued that the fundamental element of music is motion and therefore the musician should have a proper representation of the motion character of the music to be performed. He distinguished three basic forms of motion in music—"open," "closed," and "winding"—with numerous variants and illustrated them by many examples. Practice should always be governed by considerations of the proper motion character (see also Section VII,B,1). Given the adequate motion representation, many technical problems will solve themselves. The importance of motion characters was also stressed by Becking (1928) and Clynes (1983), who suggested that there is a characteristic "inner pulse" for each composer (see Section VIII,B), and by the current author (Gabrielsson, 1986a, 1988). Repp (1994a) provided a survey of some historical and contemporary approaches to musical motion.

In conclusion, the representation of music may be generated in many different and interrelated ways: in terms of structure, meaning, expression, imagery, moods, spatiomotor patterns, and so on. It is further instructive to consider the historical background to the modern conception of interpretation (Kopiez, 1993).

The generation of a representation and of a performance plan (Mursell, 1937/1971, chapter VII; Sloboda, 1982a) goes hand in hand. The general question is how the structure and the meaning of the piece should be conveyed in a convincing

way. This necessitates considerations about stylistic conventions, performance practice, instrumental characteristics and the like, leading further to specific questions concerning adequate tempo, timbre, dynamics, phrasing, timing, articulation, accents, and the related motor processes. On the whole there is little systematic knowledge about these considerations. Introspections by conductors and performers appear occasionally in books, music magazines, program notes, and in radio or television programs. Many musicians are even resistant to talk about these things contending that the music should talk for itself. A conflict may be felt between the demands of performance tradition and one's own preferred interpretation (Persson et al., 1992).

Although one can in principle distinguish between the representation, the performance plan, and the practice, they usually interact in complex and inter-individually varying ways. Of course, the representation of the piece has (or should have) a profound effect on the way the piece is practiced. On the other hand, practicing the piece may lead to modification of the representation. Ribke (1993) described this process as an interplay between conception and realization, action and reflection, deliberate planning and spontaneous intuition. Wicinski (cited in Miklaszewski, 1989) interviewed 10 eminent Moscow pianists (including Gilels, Neuhaus, and Richter) about their preparation for a performance. Seven of them distinguished three stages. First, acquiring knowledge of the music and developing preliminary ideas about how it should be performed. Second, hard work on technical problems, and third, a fusion of the first two stages with trial rehearsals leading to the final version. However, the three remaining pianists could not identify separate stages but worked in an undifferentiated way from beginning to end. See also the comments by a pianist in Miklaszewski's (1989) own study (see Section II,B,3).

B. PRACTICING

1. Mental Versus Physical Practice

Mental practice of motor skills refers to the covert or imaginary rehearsal of a skill without any muscular movements. Of course, learning to perform music inevitably requires practice on one's instrument. Having acquired a basic technical ability, the question is if mental practice, in one form or another, can be an efficient means in preparing performance of a piece.

The German piano pedagogue Leimer (1932/1972), teacher of the famous pianist Walter Gieseking, is often considered as a spokesman for mental training. Leimer emphasized that the performer must know the score, or appropriate parts of it, by heart before proceeding to practice on the instrument. Effective memorizing could be accomplished by an analysis of the piece in suitable structural units, that is, "chunks" in G. A. Miller's (1956) terminology. When reading a score, one should be able to hear the music with one's "inner ear." The practice on the instrument should be made with intense concentration to every detail in the performance, and therefore the practice sessions should be relatively short. The tempo

should be slow in order to prevent any errors and successively increased until reaching its proper rate.

Leimer frequently referred to Gieseking, who always performed by heart. Gieseking (1963, p. 94) himself said that most of his practice was done away from the piano, either reading the score and memorizing it, or, when rehearsing pieces played earlier, reviewing them in memory and only hinting at finger movements. Obviously this presupposes complete mastery of the required motor skills, and so the training can be mainly mental. The opinions of many other musicians and music pedagogues vary widely as seen in reviews by Rubin-Rabson (1937) and Kopiez (1990a). Some emphasize cognitive training, others exclusively training on the instrument, and still others a combination of both. LaBerge (1981b) suggested that mental rehearsal before playing a section is advantageous for grasping the overall shape of the section; this may be lost if one starts playing at once.

A majority of experiments on mental practice in sports and other motor activities indicate that "mental practice is better than no practice but not as effective as actual practice" (P. Johnson, 1984, p. 237; see also Coffman, 1990; Rosenbaum, 1991, p. 95). With regard to music, only a few studies have been done. Rubin-Rabson (1937) had four groups of piano students practice short compositions from the 17th and 18th centuries. Two groups analyzed the score, either guided by the experimenter or for themselves, for 20 min before practicing on the instrument. The third group practiced directly on the instrument, and the fourth group first listened to a recording of the pieces and then practiced according to one of the aforementioned methods. In a relearning experiment 3 weeks later, the groups that had received analytical training required less time and fewer repetitions to achieve a correct performance by heart. Rubin-Rabson (1941a, 1945) showed that a period of mental rehearsal inserted midway in the practice on the instrument proved to be as efficient as practicing on the instrument up to 100% overlearning. Mental rehearsal added after the performance criterion was reached was not beneficial. However, in a relearning test after 7 months, these differences among the methods were gone. Generally Rubin-Rabson considered mental practice favorable, but it should be applied to units of comfortable length.

S. L. Ross (1985) had trombone students practice an étude, either playing it three times (physical practice), perform it mentally three times trying to "hear" the pitches and to "feel" the movements of the embouchure and the slide (but with no physical movements), or combining physical and mental practice by playing the piece two times with a mental trial in between. The combined practice gave the largest gain scores, in number of correctly performed measures, from pretest to posttest, followed by physical practice. Coffman (1990) investigated the effects of physical, mental, and alternating physical and mental practice on students' learning of a short four-part chordal composition using a synthesizer. Significant differences among the methods appeared only regarding performance time. Physical practice, alone or in alternation with mental practice, proved to be superior to exclusive mental practice with regard to the speed of the performance after rehearsal. Mental practice was better than no practice at all. However, there were no differences among the methods regarding pitch and rhythm errors.

Kopiez (1990a) did two experiments with guitar students who were given a short excerpt from an unknown composition by Krenek. They practiced under one of four conditions: (a) cognitive practice, first listening to a structural analysis of the excerpt, then trying to memorize it in two 5-min sessions; (b) motor practice, practicing in two 5-min sessions with their instrument; (c) cognitive-motor practice, first listening to the structural analysis, then 5 min memorizing without instrument followed by 5 min practicing with the instrument; and (d) motor-cognitive practice, that is, in the reverse order. In the first experiment, there was no significant difference between the conditions regarding correct pitches and rhythm in the performance. In the second experiment, in which a longer excerpt was used, the group given motor practice was the most successful. Kopiez (1990b, 1991) further found that a structural analysis using a score with colors and graphics added worked better than a conventional verbal analysis.

The investigations just reviewed thus gave different results. This is not surprising, because they differ much with respect to music examples, instruments and performers, operational definitions of mental practice, and criterion measures. Coffman (1990) discussed several critical points. It seems that the less advanced the person is on the instrument and the more difficult the music is, the more important is the motor practice. Combination of physical and mental training can be favorable, as seen in some studies above and emphasized in the survey by Freymuth (1993). The distinction between mental and motor practice is in fact not clear-cut. Hale (cited in Freymuth, 1993) found that internal imagery produced activity in the muscles that would be used in the actual movements.

Conductors are by necessity forced to much mental practice. Herbert von Karajan always conducted from memory and used different editions of the score in order to rely solely on his auditory imagination of the music, avoiding visual memories of a particular score (Vaughan, 1986, pp. 210, 244). However, there are other conductors said to rely on photographic memory. Bird and Wilson (1988) studied electroencephalographic (EEG) and electromyographic (EMG) patterns during imagery in novice conductors and their teacher. There were large individual differences, but the teacher and the more skilled novices showed more repeatable EEG patterns than the less skilled students. The teacher displayed EMG patterns during mental rehearsal that resembled those of the actual performance.

2. Memorizing Music

Rubin-Rabson (1939, 1940a, 1940b, 1941b, 1945) studied skilled pianists' memorization of short pieces (eight measures) under different conditions. Some minutes of analytical prestudy were always made before the practicing on the instrument. Criterion measures were trials to reach a flawless performance in the learning session and in relearning 2 weeks later, as well as errors in a transcription of the music. In accordance with results from other areas (P. Johnson, 1984, p. 235), she found that distributed practice (intervals between practice trials were 1 hour or 24 hours) was more efficient than massed practice in relearning. However, this applied only to the less capable performers, who, she assumed, may not grasp the musical structure until a second presentation. There was no difference between

learning the eight-measure periods as wholes or in parts. Overlearning after the performance criterion was reached did not seem worthwhile; rather the extra trials may be saved for later sessions when the performance should be brought to the same level again. Unilateral practice (each hand separately) and coordinated practice (both hands together) both had advantages.

When these results are evaluated, the conditions of the experiments must be remembered. For instance, O'Brien (cited in Lundin, 1985, p. 140) found, not surprisingly, that learning in parts was more efficient for longer pieces. Farnsworth (1969, p. 168) suggested that one should work with as large a portion of the score as constitutes a manageable unit for oneself. Lundin (1985) concluded that which method is best will depend on the size, difficulty, and meaning of the material as well as on the subject's performance level. LaBerge (1981a) suggested that teachers encourage the student to discover larger global relationships in a piece, even if the practice is made in smaller parts.

Some conductors are known to be extremely good memorizers. Toscanini was said to know by heart every note in hundreds of symphonic works, operas, and chamber music (Marek, 1982, p. 414), and almost unbelievable anecdotes are told about his memory feats.

3. Rehearsal Techniques

Reports on observations of the rehearsal process are rare. Gruson (1988) had 40 piano students, representing different grade levels, and three concert pianists practice three unknown pieces, selected to ensure comparable levels of difficulty across grade levels. The practice sessions were recorded on audiotape and analyzed in terms of 20 categories. The most common event was uninterrupted playing, followed by repeating a single note, repeating a measure, and slowing down. The higher the performance level, the more time was devoted to self-guiding speech, playing hands separately, and especially to repeating sections, whereas the repeating of single notes decreased. Thus chunking in larger sections increased with increased skill. Kopiez (1990a, p. 197) pointed to the common experience that transitions between the sections often cause difficulties and therefore must be given extra consideration

Miklaszewski (1989) analyzed video-recorded practice sessions of a pianist preparing performance of a prelude by Debussy. The most frequent activity was playing alternately in fast and slow tempi. Pauses usually appeared after fast playing and were frequently used for considering corrective actions. The piece was practiced in fragments corresponding to structural units, often in another order than in the score. Some fragments were very short because of their complexity, but as the practice progressed, the fragments were spliced into longer sections. The pianist's own comments suggested that he first intended to get a clear idea of the music and of the technical requirements for performing it. He was then interested to see what he was able to perform directly and what parts must be given special practice.

Barry (1992) found that brass and woodwind students in Grades 7–10 who rehearsed a piece with structured and supervised practice performed better on a

posttest than matched students who were allowed free practice. The structured practice involved starting slowly and gradually increasing the tempo, inspecting and "fingering through" the music silently before practicing, marking errors, and practicing such sections slowly. The free-practice students usually did not use any of these strategies. Jones (1990) found that silent fingering produced a positive effect on later piano performance. Rosenthal, Wilson, Evans, and Greenwalt (1988) reported that graduate woodwind and brass students learned to perform a short piece more effectively by either listening to a model performance of the piece or practicing it on the instrument rather than by using silent analysis or practicing by singing the piece.

Which is the most favorable rehearsal technique obviously depends on the size, difficulty, and meaning of the material as well as on the subject's performance level. A common theme is that practice should be organized according to individually suitable structural units that are then expanded to successively larger units. Special attention should be given to transitions between such units as well as to sections where errors crop up. To avoid errors, the tempo should be initially slow enough and then successively increased (however, this may be questioned, see Section VI,A). Some mental practice and silent analysis alternating with the physical practice may be preferable.

Questions concerning practice are also discussed in connection with performers' musical development (see Section X,A).[2]

III. SIGHT-READING

A. GENERAL CHARACTERISTICS

Sight-reading means performing from a score without any preceding practice on the instrument of that score, to perform *a prima vista*. It must apparently involve reading groups (patterns) of notes, because reading one note at a time would not match the required tempo unless the tempo is extremely slow. Sight-reading actually involves a combination of reading and motor behavior, that is, to read note patterns coming up in the score while performing others just read. A good sight-reader is thus a rapid reader, is efficient in transforming the read pattern into appropriate motor acts, and has a good instrumental technique. Sight-reading is more efficient if the music is known, or if it conforms to a certain style that permits anticipation of what is coming next, and further if the music printing is proper with regard to spacing and other aspects.

The above points were confirmed in early studies by Bean (1938), Kwalwasser (cited in Lundin, 1985, p. 281), and Lannert and Ullman (1945). Bean used short tachistoscopic presentations of material to be performed. The professional musicians performed best, managing correct reproduction of about five notes on average. Introspections showed that the good sight-readers identified patterns and

sometimes used guesswork, as also found by C. A. Elliott (1982), MacKnight (1975), and T. Wolf (1976). Lannert and Ullman (1945) tested the ability to read ahead by presenting one measure at a time in a way that forced the pianist to play the measure he had just seen while reading the next measure. They also noted that eye movements from score to keyboard had to be quick in order not to break contact with the score. One of Wolf's pianists remarked that a good tactile feel of the keyboard makes it possible to play without having to look at the keyboard (cf. typing). If performers are not allowed to see the keyboard, they make more errors, especially poor sight-readers (Banton, 1995).

The reading of patterns rather than of individual notes may make one not notice misprints in scores as shown in the "Goldovsky experiment" (T. Wolf, 1976). A poor sight-reader played a wrong, misprinted note (a G instead of G♯ in a C♯ major chord) in a Brahms piece. Goldovsky then asked several pianists to find the misprint. For more than half of them, he had to indicate in which measure or chord the misprint was before it was detected. This is a musical parallel to the "proofreader's error"—expectation overrules perception.

McPherson (1994) found that errors in sight-reading performance by young instrumentalists were in most cases rhythm errors. He suggested that competent sight-readers seek relevant information (key and time signature, phrases, possible obstacles etc.) by scanning the music and mentally rehearsing (e.g., silent singing and fingering) major difficulties before performance. They further maintain high attention during performance to anticipate problems and to observe the musical indications above and below the musical line, and self-monitor the performance in order to correct the performance when errors occur. McPherson (1995) reported a positive correlation between the ability to play by ear and sight-reading and suggested a causal link, meaning that the former would influence the latter. It seems that both abilities presuppose effective chunking and identification of common patterns, and both work better when the style of the music is familiar.

B. EYE MOVEMENTS

Only a few studies on eye movements in music reading have been done, as reviewed by Goolsby (1989, 1994a). Many of the studies were flawed by various technical problems, and the results must be considered with caution. Furthermore, the experimental situation was often awkward for the performers. In the study by Weaver (1943), 15 female pianists performed three short pieces representing harmonic (a hymn), melodic (Bach minuet with two voices), and melody-with-supporting-chords organization. A headrest and a biting board was used to fix the position of the head. Three patterns of eye movements were discerned: (a) vertical, for example, reading each chord in the hymn from treble to bass, (b) horizontal, for example, reading successive melody notes before looking at the bass, and (c) mixtures of these. All three patterns occurred for all pieces and may be related to the structural features of the pieces but in ways that varied both between and within pianists. The amount of reading ahead (the "eye-hand span") was very "elastic" in

size, varying between and within pieces and pianists. The maximum was eight notes ahead.

Goolsby (1994a, 1994b) used a special "eyetracker" that directs a beam of infrared light into the performer's right eye. The reflections are recorded, and the horizontal and vertical components of eye position are sampled every millisecond during performance. Computer reduction programs convert the data into convenient summary files. The equipment was used to study eye movements in two groups of music students (1994a) and in two individual students (1994b)—one skilled and one less-skilled sight-reader—during their vocalization of four melodies differing in notational complexity. The group study showed that skilled sight-readers had more but shorter progressive as well as regressive fixations than less-skilled sight-readers. This suggests that skilled readers direct fixations well ahead of the performance to see what happens later in the melody and then go back to the point of performance. The use of regressive fixations is different in comparison with the skilled reading of text, in which regressive fixations are minimized. Furthermore, the duration of fixations in music reading was 100–200 msec longer in this study than what is typical for corresponding durations in text reading. Goolsby therefore warned against accepting too simple analogies between music reading and text reading. In the individual study, the positions and durations of the successive fixations were given in instructive detail for both subjects. The skilled sight-reader looked ahead, used the time of longer note values to scan about the notation, did not fixate every individual note (e.g., read scalewise notes with a single fixation), and used peripheral vertical vision to perceive dynamic and expression markings. This all indicates an active processing of the music before the performance, which was flawless. Conversely, the less-skilled sight-reader fixated on virtually every note but rarely on expressive markings, did not look very far ahead, and made many errors during the performance.

C. SIGHT-READING AND MEMORIZING

Even excellent musicians may be poor sight-readers (Bean, 1938; Lannert & Ullman, 1945; T. Wolf, 1976), and it is sometimes said that musicians who are good at sight-reading are poor at memorizing, and vice versa. T. Wolf (1976, p. 167) pointed out that sight-reading and memorizing are different processes. The good sight-reader works with rapid and effective chunking using short-term memory. However, in memorizing music one works slowly with awareness and control of each note until the procedures to a large extent become automatic and stored in long-term memory. The goals as well as the means are thus different. One can be good at sight-reading and poor at memorizing, or the converse, but there is no reason to believe that one ability excludes the other. In fact Nuki (1984) found a positive correlation between sight-reading performance and memorization ability. Furthermore, students of composition were better than piano students in memorization. This difference was ascribed to the training in grasping structural features that is involved in studying composition. McPherson (1995) also found a positive

correlation between sight-reading and memorization in clarinet and trumpet students.

D. RELATION TO MUSICAL STRUCTURE

Sloboda (1974, 1976a, 1976b, 1977, 1978a; summaries in Sloboda 1978b, 1982b, and 1985b, chapter 3) conducted the most comprehensive series of experiments on music reading. A common theme was that sight-reading is determined by various structural features in the music. For instance, when the "eye-hand span" (1974, 1977) was studied by requiring musicians to go on playing when the score was suddenly removed, the span seemed elastic, expanding or shrinking to phrase boundaries. The best sight-readers had a span of 6.8 notes. Another example (Sloboda, 1976b) was that deliberately inserted misprints often went undetected by keyboard players, who instead played the musically correct note. The misprints were harder to detect in the middle of phrases than in the beginning or at the end, especially in the upper stave.

When musicians and nonmusicians looked at short (up to 100 msec) exposures of notated patterns, both groups were poor at reproducing the pattern (1976a, 1978a). However, musicians were better at retaining the approximate contour of the pattern. With increasing exposure times up to 2 sec, both groups improved, but nonmusicians peaked at about three notes correctly reproduced, whereas musicians could record six notes. The difference is due to the musicians' coding and storing of the stimulus as a musical pattern, whereas the nonmusicians had to remember it as a purely visual pattern. This interpretation was supported by Halpern and Bower (1982), who found that musicians, but not nonmusicians, achieved better results with "good" melodies than with "bad" or random ones. Asked to divide the notes into groupings, nonmusicians grouped them according to the direction of the stem, upward or downward.

Little research has been done on the effects of musical printing on performance (Ribke, 1993; Sloboda, 1978b). Generally the printing should facilitate the grasping of the musical structure. There are many examples of improper notation—too crowded notes, inconsistent spacing between notes, misplaced notes, and so on. von Karajan noticed during a rehearsal that the orchestra always increased the tempo at the same place and found that the publisher had compressed the notes there in order to get a clean break at the page turn (Vaughan, 1986, p. 209).

Salis (1980) found that good sight-readers were superior in the rapid perception of chords but not in the perception of dot patterns, indicating that the superior pattern recognition of good sight-readers is specific to music notation. W. B. Thompson (1987), working with flutists, gave further support to this idea in finding that sight-reading ability correlated positively with achievement in a music-recall test, but not with results in a letter-recall test. He further confirmed the positive correlation ($r = 0.85$) between sight-reading ability and "eye-performance span." The latter reflects the ability to simultaneously read something and perform something else. The partial correlation between sight-reading ability and eye-performance span, controlling for music-reading ability, was still relatively high ($r =$

0.65). Thus musicians, equivalent in music-reading and in performance skills, may still differ in sight-reading ability if they differ in their ability to perform these tasks simultaneously. A parallel to this occurs in typing (Shaffer 1981, p. 328).

IV. IMPROVISATION

The definition and meaning of improvisation is no simple matter and is discussed in many papers (Andreas, 1993; Clarke, 1992; Nettl, 1974; Pike, 1974; Pressing, 1984, 1988). Even performance of strictly notated music involves a certain degree of improvisation in the individual interpretation. On the other hand, even performers of free jazz cannot avoid using previously stored material. Between these extremes are many intermediate levels.

Pressing (1984, 1988) provided a broad survey of research in different areas pertinent to the study of improvisation, including physiology and neuropsychology, motor control and skilled performance, intuition and creativity, artificial intelligence, oral traditions and folklore, and furthermore references to historical and ethnomusicological surveys, and to teaching texts. His 1988 paper lists about 200 references and should be consulted by anyone interested in research on improvisation.

In this paper, he also presented a formalized, cognitive model of improvisation. Any improvisation is seen as a sequence of nonoverlapping sections, each containing a number of musical events, called event clusters. The improvisation is an ordered union of event clusters. The generation of each cluster is based on previous events, a referent, long-term memory, and current goals. Two methods of continuation are used, associative and interrupt generation. In associative generation, the improviser wants to keep continuity, by means of similarity or contrast, between successive event clusters. In interrupt generation, the improviser breaks off into quite another direction without regard to what has been before. The choice between the two ways is made in relation to a time-dependent tolerance level for repetition. If the present degree of repetition is higher than this level, interrupt generation results; otherwise associative generation goes on.

Pressing (1987) analyzed two short pieces of improvised music with respect to the macrostructure (similar to traditional music analysis) and the microstructure (various aspects of timing and dynamics) and the correlations between those two levels. Such correlations were observed for one of the pieces, and distinct event clusters and classes of event clusters could be identified. Among the most interesting features were examples of categorical production (in analogy with categorical perception) and of three independent underlying temporal mechanisms.

In discussing Sudnow's (1978) description of how he learned to play jazz piano, Clarke (1988) discerned three alternative representations underlying jazz improvisation—hierarchical structure (in traditional jazz improvisation), associative structure (in free jazz), and selection of events contained within the performer's repertory (in bebop). All three principles may operate and interact in any perfor-

mance. Johnson-Laird (1987, 1991) pointed out that musicians have little conscious access to the processes underlying their improvisations and sketched a theory about what the mind has to compute in order to produce an acceptable improvisation. Given that jazz musicians have stored knowledge of principles for rhythmic patterns, melodic contours, and harmonic and metrical constraints, what has to be computed for each note in the improvisation is its onset and offset, its step in the contour and its particular pitch. Preliminary illustrations were given using a computer program that, given a chord sequence as input, generates bass lines or melodies. Although on some occasions there is no choice about which note to play, on other occasions there may be many feasible notes. The program then makes an arbitrary choice, thus leaving room for creativity. Clarke (1992) discussed various problems with this model—its limitation to tonal jazz, its focus on generation of single notes rather than of larger units, and its neglect of the player's and the instrument's physical constraints.

A few empirical studies have been done. Reinholdsson (1987) analyzed timing and dynamics in a drum solo by Roy Haynes and provided a valuable discussion of the difficulties in analyzing performance of non–notation-bound music. Bastien and Hostager (1988) studied how four jazz musicians accomplished a group performance without having played together earlier. The basic idea is that the unavoidable initial uncertainty is reduced by two constraints, common knowledge of structural conventions in jazz and certain social practices. The latter include accepted behavioral norms—a leader decides and communicates each song, the soloist determines the style, each musician gets an opportunity to be the soloist—as well as certain communicative codes, verbal and nonverbal (eye contact, nodding, turning to an individual, hand signals). On the basis of this and a successively expanded knowledge of each other's behaviors during the performance, the group advanced from a somewhat cautious beginning to finally inventing an entirely new song.

Hargreaves, Cork, and Setton (1991) had novice and expert jazz keyboard players improvise to prerecorded "backing tracks" and interviewed them concerning what principles they used. The experts approached each improvisation with an overall plan—for instance, play in the style consistent with the backing track—but were prepared to adopt a new plan depending on what actually happened. They were relaxed about their performances, sometimes fell back on clichés and automatic performance of "subroutines," whereas the conscious control was reserved for the overall planning. In comparison, the novices' plans were nonexistent or very limited. Sági and Vitányi (1988) asked a large sample of persons to sing improvisations to Hungarian poems and to simple harmonic progressions. More than 3000 recordings were analyzed with respect to melodic, harmonic, and rhythmic features. Influences of Hungarian national music, as well as of art music and popular music, were evident and could be related to the different social backgrounds of the participating persons.

The interest in improvisation in music education is increasing, but there is still little research on the topic (Clarke, 1992; Webster, 1992). McPherson (1993) developed a test to measure improvisational ability of high school instrumentalists.

No significant correlation was found between improvisational ability and performance proficiency in the beginning stages of development, but in more advanced stages a significant correlation was found.

Improvisation plays an important role in much music therapy. A music therapist should be able to improvise in order to interact with his client in real time. Bruscia (1987, 1988) provided a broad and systematic account of improvisational models in music therapy.

V. FEEDBACK IN PERFORMANCE

Perceptual feedback in performance may be auditory, visual and proprioceptive, that is, tactile, kinesthetic, and maybe also vestibular (Todd, 1993).

The importance of proprioceptive feedback is obvious, although its precise role is a matter of discussion (see Section VI,C). Hearing-impaired and even deaf musicians (Glennie, 1991) use it successfully in their performance. Vibrotactile feedback from the skin is available to singers and instrumentalists but only below 1000 Hz. Sensitivity in the skin of the hand is best around 250 Hz; it decreases with increasing age (Verrillo, 1992). Most parts of a stringed instrument vibrate during playing and at levels that are clearly detectable (Askenfelt & Jansson, 1992). These vibrations may be used for intonation purposes, especially in ensemble playing at loud dynamics where auditory monitoring is difficult. Kinesthetic feedback from finger forces (e.g., with different types of "touch" in piano playing) also provides essential information.

Kinesthetic feedback, or even motor commands, is perhaps used to monitor performance in advance of auditory feedback, for instance, to feel that a wrong note is played before actually hearing it (Sloboda, 1978b). Heinlein (1930) found that changing proprioceptive feedback and eliminating auditory feedback had adverse effects on pedaling in piano performance (see also in Section VII,B).

Visual feedback can provide information about the instrument and the behaviors of the conductor, the fellow performers, and the audience, including "social feedback." Two or more performers playing together use auditory and visual feedback to be able to coordinate and respond to each other (Shaffer, 1984a, p. 593).

Delayed auditory feedback (DAF) usually has a detrimental effect, in speech as well as in music. Gates, Bradshaw, and Nettleton (1974) used delays from 0.1 to 1.05 sec for keyboard players instructed to play a Bach minuet on an electronic organ as fast and accurately as possible. In comparison with immediate auditory feedback (IAF), most subjects slowed down their performance. This effect was present over the whole range of delays and was most marked at a delay of 0.27 sec. Three strategies could be discerned: to play fast in order to get ahead of the delay, to play slower or pause and wait for the delay to catch up, or to ignore the delay and go on. One player who consistently speeded up performance tried to get the delayed note to coincide with the note he was playing; thus tempo may be decisive for what delay is most detrimental. Gates and Bradshaw (1974) found that performance without any auditory feedback at all was not different from IAF perfor-

mance (see also Banton, 1995). It was hypothesized that auditory imagery can take the place of IAF. Performance with DAF to one or both ears was worse, as well as performance with combined IAF and DAF. DAF disruption may stem from several sources, such as distraction, error repetition, and conflict with expectancies. Varying the speed of the performance may be one way to investigate this further. Today's computerized synthesizers offer convenient possibilities for studying DAF. According to Yates (cited in Gates & Bradshaw, 1974), the DAF effect is notoriously resistant to any reduction with practice. However, this cannot be generally true, because organists can easily learn to perform on pneumatic organs, which involve DAF.

Auditory feedback may refer not only to hearing the sounds from one's own instrument but also the sounds from other performers. Good acoustics in concert halls should include this aspect as well. This was studied by Gade (1986, 1989a, 1989b) with special reference to "support"—the property that makes the musician feel that he can hear himself well and not is forced to play louder than usual—and to "hearing each other" (see also Naylor & Craik, 1988). Increased reverberation and faster tempo delay the perceived attack of a tone due to the smoothing effects of reverberation on envelopes and increased overlapping of successive tones (Naylor, 1992). In order to maintain the tempo and to achieve perceived synchrony, musicians should therefore play a small amount ahead of the beat they hear. With sharp attacks the delay is less, and instruments with sharp attacks may therefore serve as "beat-definers" for the rest of an ensemble.

In studio recordings of popular music as well as in live performance the auditory feedback is usually obtained in headphones. So-called click tracks may be used for (supposedly) better control of timing and "tightness" (Madison, 1991).

Examples of biofeedback to reduce muscle tension in performance are given in Section IX,A.

VI. MOTOR PROCESSES IN PERFORMANCE

A. SOME GENERAL QUESTIONS

Although motor processes are central to music performance, they are still little understood. Sidnell (1981) asked a number of critical questions concerning efficient motor practice, motor memory, the role of proprioception, transferability of motor skills, and application of current motor models. Hedden (1987) reviewed recent reports and found only tentative answers to some of these questions. F. R. Wilson (1992) and F. R. Wilson and Roehmann (1992) pointed out the psychomotor complexity of human music behavior. "From the perspective of the movement scientist, the questions of greatest importance to music educators must be regarded as being entirely out of reach for the foreseeable future" (F. R. Wilson, 1992, p. 93).

The common advice that practicing should start in a slow tempo and then successively be accelerated was questioned by Handel (1986). He referred to experiments in which the perceived grouping of polyrhythmic patterns changes with tempo and suggested that there may be similar motor reorganizations. Practicing at a slow rate may involve another motor pattern than at faster rates and therefore be relatively ineffective. Taubmann (1988) cited evidence from Ortmann's (1929) classic investigations showing that movements in fast piano playing are different from the movements in slow playing. She concluded that the correct shape of the movements in fast playing "must be analyzed and brought into slow playing. In this way, when practicing slowly ... you are playing fast, slowly" (p. 150). Ortmann's photographs and other records are still invaluable assets. He concluded that the movements of two pianists are never exactly alike. Any key on the piano can be reached effectively in a number of ways, and the playing of a key is affected by the playing of preceding and succeeding keys (McArthur, 1989; F. R. Wilson & Roehmann, 1992).

B. MOTOR EXERCISES

There is evidence that note timing changes with different tempi in motor exercises such as performance of scales. C. Wagner (1971) found smallest variability of internote intervals at an intermediate tempo (about 6–9 notes/sec) and increasing variability at slower and faster tempi. Mackenzie and Van Eerd (1990), using a grand piano equipped with infrared key movement detectors (B. L. Wills, MacKenzie, Harrison, Topper, & Walker, 1985), found that variability of internote intervals and velocity of keypress increased with increased tempo in scale playing. The left and right hands differed in keypress velocities, note durations, and overlap between consecutive notes.

Peters (1985) found interference between hands in a rubato-like tapping task, accelerating in one hand while keeping a regular tempo in the other. Performance was better when the acceleration was made by the preferred right hand. Deecke (1995) studied brain potentials in musicians tapping a 2 x 3 polyrhythm (two in one hand, three in the other) and indicated that the supplementary motor area is highly important for controlling bimanual skills like this. Lee (1989) investigated a pianist's performance of left-hand leap patterns. Moore (1992) demonstrated the remarkable speed and precision in skilled pianists' performance of trills (see also F. R. Wilson, 1992). Askenfelt (1986) measured bow motion and bow force in different bowing styles on the violin; Moore (1988) studied bowing techniques and vibrato in cello playing; and Moore, Hary, and Naill (1988) studied trills in cello playing. Davies, Kenny, and Barbenel (1989) found, contrary to common belief, that professional trumpet players do not generally use lower levels of mouthpiece force than less-skilled players. One may be misled by the player's appearance; some professionals "could use massive amounts of force whilst maintaining a madonna-like appearance" (p. 61).

C. THEORIES OF MOTOR SKILL

Four theories of motor skill may be discerned: closed-loop theory, open-loop or motor program theory, schema theory, and the Bernstein approach (P. Johnson, 1984; Kelso, 1982; LaBerge 1981a; Rosenbaum, 1991; Schmidt, 1988; Sheridan, 1984; Wade, 1990). According to the *closed-loop theory*, sensory information produced from the movement is fed back to the central nervous system and compared with an internal referent to check for discrepancies between the intended and the actually produced movement. The system is thus self-regulating by means of proprioceptive feedback. It has been argued that such feedback is too slow to account for rapid movement sequences in music (for example, a trill), but this is a matter of some debate (LaBerge, 1981a; MacKenzie, 1986; Moore, 1992; Schmidt, 1982, 1988). MacKenzie and Van Eerd (1990) suggested that sensory information is used on a note-to-note basis at slow tempi, but it may also be used at faster speeds to modify the performance with regard to larger groups.

Open-loop or *motor program theory* postulates a central or executive control of all movements in a sequence, thus not relying on sensory feedback in the real-time control of movement. All movement parameters are specified in a motor command, and the movement runs to its completion without alteration, an alternative that seems adequate for fast movements. The theory usually assumes a hierarchical structure with higher levels containing abstract representations of movement sequences to which more and more specificity is added at successively lower levels. Sensory feedback may have a role in constructing and maintaining procedures for translating motor commands into muscle actions (Shaffer, 1980) and for modifying the response after the movement has been completed (Sheridan, 1984, p 60); further possibilities were suggested by Schmidt (1988, chapter 8).

Schema theory assumes that there exist abstract representations of classes of motor actions, generalized motor programs, out of which can be generated a wide variety of movements in novel situations. One must not learn and store every single movement, it can be generated from the stored schema (abstracted rules) by applying appropriate parameters in interaction with the demands in a given situation. A recall schema is concerned with the execution of the movement and a recognition schema with the evaluation of the response. Variability of practice and feedback (knowledge of result) strengthen the schema rules.

The *Bernstein approach* is named after the Russian physiologist Nicolai Bernstein. Arguing that the number of combinations of muscle settings for different movements is too large to be managed by a controlling executive ("the degrees-of-freedom problem"), it is emphasized that muscles are not individually controlled but function in muscle linkages or *coordinative structures*. Groups of muscles are constrained to act as functional units, thus reducing the degrees of freedom and thereby also the need for hierarchical organization or higher-level control. Whereas the motor program theories mainly deal with central representations and say little about the construction and work of the effectors, this approach is concerned with the anatomy and function of the muscles and the whole body to see how the

motor actions may arise as necessary consequences of the way the system is designed to function. However, the Bernstein approach is still little discussed in relation to music. Wade (1990) suggested some examples and also discussed the role that cognition may have in relation to this approach.

LaBerge (1981a) proposed a combination of schema theory and the concept of coordinative structures. Voluntary motor conceptualizations are organized as motor schemas, but the motor schemas do not communicate directly to individual muscles but via coordinative structures. The commands sent to the coordinative structures are interpreted as "pieces of advice." The coordinative structures communicate with individual muscles in terms of involuntary reflex-like commands, balancing the pieces of advice with available information about environmental resistances (e.g., the mechanics of the instrument), inertia of the moving limbs, and momentary state of muscle tension. Thus the "top-down" control is replaced by a type of shared control. This framework is used for an interesting hypothetical description of learning musical performance skills in a series of stages encompassing successively larger musical units as the performance of smaller units become more or less automatic (see also LaBerge, 1981b).

D. EMPIRICAL INVESTIGATIONS

Referring to schema theory, Welch (1985a, 1985b) predicted and also found that children with poor pitch singing would perform better in matching a target pitch if they received adequate knowledge of their results (oscilloscope representations of their singing in relation to the target) than children not given knowledge of results. The concept of generalized motor programs (Schmidt, 1982, 1988) was supported by Repp (1994c), who found relational invariance of expressive microstructure across moderately different tempi in performance of a piano piece. However, Desain and Honing (1994) reported evidence against relational invariance in other piano performance (see Section VII,C,3). The increased variability in timing and keypress velocities at faster speeds and the differences between hands found by MacKenzie and Van Eerd (1990) did not support the phonograph record analogy of generalized motor programs (Schmidt, 1988, p. 246). Handel (1986) raised the question of whether motor programs and procedures may be restricted to a relatively narrow range of performance rates and suggested that therefore performance in other tempi would require the construction of new programs.

The most thorough investigations of music performance in relation to motor skill theories have been conducted by Shaffer and his coworkers. The background is a theory of motor programming (Shaffer, 1980, 1981, 1982, 1984b), meaning that a sequence of movements can be coordinated before their execution in a way that ensures fluency, expressiveness, and generative flexibility of the performance. The program is used for realizing a performance plan and in doing this may construct a hierarchy of intermediate representations leading to output via motor commands that provide target specifications for the movement (location in space, force

and manner of movement). The commands are translated into muscle actions by a peripheral computer in the cerebellar-spinal system.

Timing is handled by the motor system and an internal clock. The internal clock acts as a reference and is used for timing of the beat or another appropriate unit. Its rate can be changed to achieve expressive variation in timing (rubato). The motor system itself acts as a timekeeper by translating a given time interval into a (compound) movement trajectory with the corresponding duration. This is used for timing of the subdivisions of the clock interval (beat). The clock is assumed to generate time intervals with lower variance than the motor system can. Time intervals may be generated in concatenation or in hierarchical manner (for example, for the bar, half-bar, quarter-bar). To distinguish between these alternatives one can study variance and covariance properties of the intervals as proposed by Vorberg and Hambuch (1978, 1984). However, both alternatives assume a constant tempo, which may necessitate modifications to fit much music performance. Shaffer also used hierarchic analysis of variance including an unconventional use of F values less than 1.00 to indicate less variation in the timing of a certain unit (for example, bar, half-bar, or beat).

To test the theory, several piano performances were recorded using a Bechstein grand piano equipped with photocells to permit accurate measures of timing and dynamics (Shaffer, 1980, 1981). Thorough analysis of timing in a concert pianist's performance of a Bach fugue indicated that half-beat intervals (quarter notes) were generated by a clock according to a concatenation principle, whereas the timing of short notes within the half-beat interval, assumed to be achieved by learned motor procedures, favored a hierarchic structure. At the same time, however, the pianist varied the tempo, indicating that the clock rate was modulated according to expressive information given in the motor program. Thus the clock provides temporal markers in an "elastic" way serving expressive purposes. A performance of a Chopin étude—playing three in the right hand against four in the left—demonstrated independence of the two hands in dynamics and timing. With regard to timing, considerable rubato was made common to both hands, as well as different between the hands, in an asynchrony letting the melody in the right hand lead or lag in relation to the left-hand accompaniment. Again this indicated a flexible clock for timing of the beats and separate timekeepers for each hand to construct the subdivision of the beat and to allow one hand to temporarily move off the beat. A performance of one of Bartók's Six Dances in Bulgarian Rhythm (with a 3 + 3 + 2 division) also showed similar principles for hand independence in timing. The Chopin étude was performed by the same pianist 1 year later with very similar results (Shaffer, 1984a).

These results were taken to indicate a flexible clock for beat timing and separate timekeepers for each hand responsible for beat subdivisions and off-beat playing. Furthermore, the high degree of reproducibility in repeated performances suggests that the musician's interpretation of the score, in combination with his general knowledge of music theory and musical style, is able to generate appropriate expressive structures in the motor program on different occasions. In a duet

performance of a Beethoven piece (Shaffer, 1984a), the rubato pattern over the piece was similar for the two pianists, as well as in a repeated performance, indicating that it was generated from an agreed interpretation rather than being memorized (one pianist had not played the piece before). Precision was highest at the bar level, and asynchronies between voices and between pianists showed a systematic pattern. It was assumed that differences in note timing between the voices and the pianists, creating a kind of rhythmic interplay, could be made since the bar served as a common meter for both of them.

In a study (Clarke, 1985a; Shaffer, Clarke, & Todd, 1985) of one pianist's three performances of Satie's *Gnossienne No. 5*, that has a regular left hand accompaniment in eighths but a variety of rhythmic subdivisions in the right hand, the analysis indicated beat timing at the eighth note level. The subdivisions within the beat were performed either to fit the beat timing—regarding subdivisions with unequal note values or many equal note values—or temporarily overrode beat timing in order to favor evenness among a small number of equal note values in the subdivision.

Clarke (1985a; 1985b, p. 229) concluded that expressive rhythmic performance from a score involves an underlying structural representation. This consists of rhythmic figures organized around a framework of beats and an expressive system which transforms the structural representation by altering the clock rate, thus affecting the beat rate or tempo, and by modifying the way in which motor procedures subdivide the beat intervals. Of course, the expressive system may operate with other parameters as well, such as dynamics, articulation, intonation, and timbre, depending on the instruments used and the musical context.

Rather than being memorized, the expressive aspects of a performance—that is, the deviations in timing, articulation, dynamics, and so on—are generated from the performer's structural representation of the piece at the time of the performance. This may explain the high reproducibility in repeated performances as well as the ability to change the performance according to an alternative structural representation. Later Clarke modified this view on the basis of experiments (Clarke, 1991; Clarke 1993a, 1993b; Clarke & Baker-Short, 1987) which showed that listeners were able reasonably well to imitate phrases in which structure and expression were put in some kind of conflict ("unnatural" timing). In Clarke (1993a), the timing profile in four short melodies performed by a pianist on a Yamaha MIDI (Musical Instrument Digital Interface) grand piano were transformed, using the POCO environment (Honing, 1992), by inverting the timing profile or by translating it one beat or a measure plus one beat. Ten pianists were asked to listen to these performances and to imitate them as accurately as possible. The imitations were most accurate and stable for the original performance, least accurate and stable for the inverted version, with the two translated versions in between. Moreover, in a separate listening test the original version was judged as the best and the inverted version as worst. The fact that the subjects were able to reproduce some aspects of even the most disrupted versions could be due to pure auditory storage of them in short-term memory (however, this would work only

for very short melodies), or use of a verbal representation (description) of the version to support the auditory image in short-term memory, or some kind of bodily representation of the version. The latter was suggested by the different body movements that could be observed in the imitating subjects' performance of the different versions.

It should be noted that terms such as *expressive deviations* or simply *expression* in the aforementioned studies refer to physical phenomena, that is, deviations in timing, articulation, intonation, and so on in relation to a literal interpretation of the score. This use should be distinguished from a more general meaning of expression in music (Clarke, 1989a; 1991). I prefer to discuss the deviations as (possible) physical *correlates* of expression, since "expression's domain is the mind of the listener" (Kendall & Carterette, 1990, p. 131). Of course, the structure of the piece is in itself a basic correlate of experienced expression.

Shaffer (1989), discussing the possibility of constructing a robot to play a Chopin waltz, concluded that cognitive planning and motor programming may not be enough. Emotional factors must be considered as well. The robot has to be given feelings about himself as performer, about the social context, the musical tradition, and the sensuousness in the physical movements of playing music. Music has inescapable connotations of mood and a player has to be sensitive to these. "The structural interpretation of the piece remains the same but only sometimes the players' own mood allows them to fully catch the mood of the music, and this is felt by both the players and the listeners" (p. 389).

Palmer (1989a, 1989b), in discussing motor programming, asked whether the procedures that govern the translation from intention to performance differ from one performance to another, or if there is a limited set of procedures whose parameters can change. The latter alternative is the more economical, because a larger set of behaviors may be accounted for by a smaller set of procedures. This hypothesis was investigated in several experiments on the use of three timing procedures: chord asynchronies (melody lead or lag), rubato (especially at phrase boundaries), and articulation (legato vs. staccato). The performances were done on an electronic keyboard (1989a) or a computer-monitored Bösenforfer concert grand piano with optical sensors that detect and code movements of each key and foot pedal (1989b). Expert pianists and students were asked to perform pieces by Mozart, Beethoven, Chopin, and Brahms in a musical, a mechanical, and an exaggerated way. The mechanical performances were characterized by less, and the exaggerated performances by more, asynchronization, rubato, and legato performance, in comparison with the musical performance. In successive performances of an unfamiliar work, these timing procedures were usually applied in an increasing way, and the experts used them to a higher degree than the students. When two pianists were asked to perform a piece in two different interpretations, related to the melodic line and to the phrasing, the timing procedures were adapted accordingly. For example, the voice intended as melody tended to precede the other voices, and ritards were made at the phrase endings of the respective interpretation. Listeners, especially pianists, were able to correctly identify the intended

interpretations from these differences in timing procedures, even when intensity variations in the performances were removed by computer editing. All together, the results were considered to support the idea that the same procedures were used with varying parameters in the different situations. This conclusion was supported by Behne and Wetekam (1993), who investigated rhythmically exact versus expressive performance of the theme in Mozart's Piano Sonata in A major (K. 331).

Halsband, Binkofski, and Camp (1994) instructed pianists to perform some pieces according to different rhythmic groupings and recorded their performance on a Yamaha Disklavier. Halsband et al. also recorded the pianists' hand movements by use of light-emitting diodes attached to the wrists and fingers. Both recordings showed that the formation of motor patterns was affected by the prescribed rhythmic grouping and that this process was mainly under left (dominant) hemisphere control.

E. EXPRESSIVE MOVEMENTS

Playing an instrument requires skilled movements. However, performers also move in many other ways not directly related to the generation of sound but rather to the character of the music. Such expressive movements are an important part of the performer-listener communication and were studied in pioneering work by Davidson (1993, 1994, 1995). She used *point-light technique,* that is, illuminated reflective tapes were attached around the performer's head, elbows, wrists, knees, and ankles and on each hip and shoulder. Violinists and pianists then performed music in three different manners: deadpan, projected (as in a public performance), and exaggerated. Video recordings of the performances were presented to observers in three different modes: vision-only (the point-light patterns resulting from the musician's movements), sound-only (the performed music), or both together. Music students were able to distinguish between the three performance manners in the vision-only mode, even better than when listening to the performances or both listening and watching them. Nonmusicians could make reliable discriminations solely in the vision-only mode. Head movements in pianists were essential for distinguishing among the different performance manners. These studies highlight the need for investigating the role that visual information may play in music perception.

VII. MEASUREMENTS OF PERFORMANCE

A. MEASUREMENT PROCEDURES AND DATA ANALYSIS

Measurements of music performance as a rule require advanced technical equipment. One common way is to study the action of the instrument, for instance, to record key depressions and releases on keyboard instruments. In early research, this was done by electromechanical devices or by moving film, nowadays usually by electronic and computer facilities. A related way, common during the early

decades of this century, was to make measurements of "player rolls" generated from performances on special pianos. A different principle is to analyze the sounds that are emitted from the instruments, in direct recordings or stored on phonograph records or tapes. This alternative is applicable to any instrument, including singing, but puts great demands on the analysis equipment, especially in music with many voices overlapping or masking each other (for examples, see Gabrielsson, 1987, or Repp, 1992a).

The variables measured are various aspects of timing, dynamics, and intonation. With regard to timing, measurements may refer to tempo, durations of sound events and groups of sound events, and of nonsound events (rests, other silent intervals), and further the asynchrony between different instruments supposed to perform at the same time. Concerning durations, a distinction must be made between the duration from the beginning of a tone to the beginning of the next tone (internote interval, interonset interval, or d_{ii} = duration in-in) and the duration from the beginning of a tone to its end (d_{io} = duration in-out). The latter measure is useful for describing articulations, such as legato and staccato. In legato, the tone sounds until the beginning of the next tone or almost so ($d_{io} \approx d_{ii}$). In legatissimo, successive tones even overlap, that is, the tone is sustained until some time after the beginning of the next tone ($d_{io} > d_{ii}$), which is common in keyboard performance. In staccato, the tone ends pretty soon ($d_{io} \ll d_{ii}$), and there is a "silence" until the beginning of the next tone (see further in Bengtsson & Gabrielsson, 1983). Most duration measurements refer to the first alternative. When the term duration is used in the following, it will therefore refer to this alternative (d_{ii}). Articulation refers to the other alternative (d_{io}).

A special problem is how to define the beginning of a tone. It can be defined as the point in time when a key is depressed (or at a certain moment of the key depression), or when a change in amplitude or waveform can be seen, or when the amplitude of the tone has reached a certain level. The last alternative is often assumed to reflect better the perceptual onset (Rasch, 1979, 1988; Rose, 1989). Models for predicting perceived onset were discussed in Naylor (1992). Similar problems, although less discussed, pertain to the determination of the end of a tone (Edlund, 1985, p. 92). Differences in measurement techniques and definitions sometimes make it difficult to compare results from different investigations.

A general problem is the wealth of data. There are so many events, and relations between events, to study, that papers on music performance sometimes provide hard reading. C. E. Seashore (1936, p. 30; 1938, p. 240) remarked that description and interpretation of performance data for a single piece would require a volume. (This gives an interesting perspective to the present author's task.) A common way to present data is to use the musical score (if there is one) as a reference and provide values (for instance, durations) for each note, beat, bar, phrase, or other units. The values may be original raw data, as in Figure 5, or be expressed in relation to a norm, for instance, as deviations from a strictly regular, "mechanical" performance as in Figure 4. A discussion about some alternatives was given by Repp (1994b). Data treatment has been much facilitated by comput-

ers, but knowledge about what data are relevant, and what may be discarded, is still far from complete. Nor is there any comprehensive theory to guide us in this search. The statistical treatment therefore often aims at an appropriate reduction of data and may include analysis by means of correlation, covariance, autocorrelation (Desain & de Vos, 1992), analysis of variance, factor analysis, and regression analysis.

A related question concerns how to represent the results of the measurements to give an impression of how the performance sounded. One way is to provide original or modified visual recordings of the sound, which may give a feeling for the variations in pitch, loudness, and timing (Gabrielsson, 1986b, 1987, 1994, 1995; Gabrielsson & Johnson, 1985; C. E. Seashore, 1938). However, they take up much space and require some training to be meaningfully interpreted. Gjerdingen (1988) proposed a representation in which changes in pitch and intensity combine to suggestive motion shapes, which is in good agreement with the view of music as motion (see Section II,A). He further stressed and illustrated the importance of analyzing not only change, but change in the rate of change in pitch and intensity together, which is certainly perceptually relevant.

The following account of performance measurements takes a historical approach. This history has two distinct periods, one beginning in the early music psychology and running up to about 1940, and the other a restart beginning in the 1960s.

B. EARLY INVESTIGATIONS

1. Mostly in Europe

Binet and Courtier (1895) used a small rubber tube below the keys in a grand piano to record the key depressions. When a key was depressed, the tube was compressed generating an air puff that affected a stylus writing on moving paper. They studied the performance of trills, scales, accents, and crescendo-decrescendo and observed clear differences between amateurs and professional pianists. To achieve an accent on a certain tone, a pianist played the preceding tone *detaché*, whereas the tone to be accented was played with more force as well as lengthened and closely tied (legato) to the following tone. Lengthening of accented tones was also observed by Ebhardt (1898), who attached an electromechanical device to a grand piano to record key depressions on a kymograph. Furthermore, when pianists were asked to play a previously performed piece on a dumb piano (without auditory feedback), the tempo became much slower. Ebhardt assumed that this was due to the increased psychic activity required for imagining the piece. Much later Clynes and Walker (1982) reported that seven out of eight musicians performed pieces of music significantly slower when thinking through them than when actually playing them.

Sears (1902) had four organists play five hymns on a small reed organ. The keyboard action was recorded on a kymograph drum via electromechanical principles. There was often overlap between consecutive tones (legatissimo). The durations of tones with the same note value differed within each piece, and the rela-

tions between tones of different note values—such as half note and quarter note—differed from their nominal relations. The duration of measures varied. There was a ritard toward the end of the hymn, accented tones were usually lengthened, triplets were performed with the last tone lengthened, and there was asynchronization between the tones in chords. The inter-individual variation in performance data was considerable. These results recur in many later studies.

It may be a surprise to learn that perhaps most performance studies until today were made in the 1920s and 1930s. Morton (1920) noted difficulties in playing a 3 x 2 polyrhythm and trills. Heinitz (1926) used a stopwatch to analyze seven singers' performance on phonograph records of a *Meistersingerlied* by Wagner. Heinlein (1929, 1930) made a unique study of pedaling in piano performance by analyzing famous pianists' pedal action on a Duo-Art reproducing piano and conducting experiments with four pianists performing Schumann's *Träumerei*. The pedaling differed markedly between the pianists as well as under the different conditions, and no pianist was able to reproduce his pedal action from an earlier performance. The conditions included performing from the score, by heart, pedaling while imagining playing the piece, pedaling while singing the melody, and others. Heinlein concluded that pedaling was dependent on finger pattern, phrasing, speed of rendition, variation in intensity and timbre, extent of tonal anticipation, and type of imagery in recall. It is so highly integrated with the other phases of piano performance that alteration of any single factor may affect the pianist's customary pedaling. Recently Repp (1994c) provided some data on pedaling in *Träumerei*. Taguti, Ohgushi, and Sueoka (1994) found differences in pedaling among pianists and among different expressive intentions. Multidimensional scaling indicated that the deeper the damper pedal was pressed, the more reverberant and warm was the perceived piano sound.

Guttmann (1932) studied conductors' choice of tempo in numerous concerts over a couple of decades. There were obvious variations between conductors (among them R. Strauss, Nikisch, Furtwängler) as well as for the same conductor at different performances of the same work. R. Strauss himself commented on this variability (cited in C. Wagner, 1974, p. 593).

Hartmann (1932) made a long, careful, and instructive analysis of two pianists' performance of the first movement of Beethoven's *Moonlight Sonata*. The measurements were made directly on paper roll recordings for player pianos. Despite the homogeneous rhythmic structure of the movement, the tempo varied considerably. One of the pianists made a ritard at the end, the other did not. The durations of tones with different note values overlapped much; the shortest half note was shorter than the longest quarter note. Both pianists played legatissimo and with asynchrony within chords. One pianist had a fairly consistent asynchronization pattern, playing the bass tone first, then the higher bass tone (the octave above), and then either the melody part or the first tone of the triplet accompaniment.

Truslit (1938; see also Repp, 1993) presented an interesting mixture of intelligent speculation and empirical study. He emphasized the connection between music and motion (cf. Section II,A). The dynamics and agogics of music must be

shaped in accordance with the general laws of motion, which he tried to demonstrate by means of performance measurements. He also claimed that the function of the vestibular organ is the biological basis for experience of musical motion.

2. Research at Iowa University

Around 1930, a large group of researchers headed by C. E. Seashore studied music performance at the University of Iowa. Most of their reports appeared in two volumes edited by C. E. Seashore (1932, 1937), both classics in the music performance literature. Another volume (C. E. Seashore, 1936) was devoted to the vibrato, and examples from many studies appear in his textbook (C. E. Seashore, 1938). However, C. E. Seashore himself rarely appeared among the authors of the research reports. His function seems to have been that of supervisor and coordinator. The studies comprise performance on the piano, the violin, and singing. The data was presented in so-called "performance scores" and because of the amount of data readers are sometimes asked to analyze such scores themselves according to his own interests (Skinner & Seashore, 1937).

a. Piano

Piano performance was investigated by filming the movements of the hammers (Henderson, Tiffin, & Seashore, 1937; C. E. Seashore, 1938, p. 233). Henderson (1937) analyzed two pianists' performance of the chorale section, in $\frac{3}{4}$ time, in Chopin's *Nocturne*, Op. 15, No. 3. In measures containing three quarter notes, the second was relatively shortened. In measures consisting of a half note plus a quarter note, the latter was relatively lengthened. This lengthening was in some cases related to a ritard. In other cases, it was a way of delaying the entrance of the first beat in the following measure, thereby contributing to achieve an accent on this beat. It was further hypothesized that the lengthening was made "in order to achieve more melodic equality between the written long and short notes" (p. 291).

Phrasing was made by temporal as well as dynamic means. Within a phrase, there was first an acceleration followed by a ritard toward the end together with decrescendo, but the pianists did this in varying ways. They also differed considerably with regard to the tempo chosen for the performance. Pedaling was used to facilitate legato playing. All chords were played asynchronously. However, whereas one pianist played the melody before all other notes, the other pianist deliberately played the bass note first, maybe to emphasize its countermelodic character. The spread among the notes in a chord varied from 20 to 200 msec, with relatively more spread on accented chords than on unaccented. Accents were not related to higher intensity, and Henderson discussed several ways of achieving accents, including factors already inherent in the score.

Skinner (cited in C. E. Seashore, 1938, pp. 246–248) found high within-individual consistency in timing in two pianists' repeated performances of the same pieces by Beethoven and Chopin. When one of them was asked to perform the Chopin piece in uniform metronomic time, the variations in durations became much less, but they did not quite disappear.

Vernon (1937) focused on chord asynchronization in performance of works by Beethoven and Chopin by four famous pianists (Bauer, Backhaus, Hofmann, Paderewski), recorded on paper rolls. All used asynchronization, particularly Bauer, who played almost half of all chords with temporal spread; Backhaus did it rarely. The spread was mostly within 30 msec, but Bauer and Paderewski sometimes had spreads up to 200 msec or more. For these pianists, the melody part came after the rest of the chord more frequently than before. Asynchrony was more common in melody chords than in nonmelody chords, indicating that asynchrony was used to mark off melody notes from the accompaniment. It was also used to give emphasis to a certain point in a phrase.

b. Violin

Violin performance was analyzed by means of phonophotograph apparatus (Tiffin, 1932a) using stroboscope technique for recording of frequency (pitch) and a vacuum tube voltmeter connected to an oscillograph for recording the intensity. All was photographed on the same film. Small (1937) analyzed phonograph recordings of performances by Busch, Elman, Kreisler, Menuhin, Seidl, and Szigeti and made direct recordings of other violinists. Pitch vibrato, approximating a sine curve, was present in practically all tones. The typical rate was 6–7 Hz and the range approximately a quarter tone; rate and range were independent of each other. Intensity vibrato was not as frequent and continuous as the frequency vibrato. It had about the same rate as the frequency vibrato, and the phase relations between them varied. There were frequent deviations from the frequency (pitch) indicated in the score. Leading tones were played high, augmented intervals were expanded, diminished intervals contracted. Portamento was often used when going to another pitch to keep the legato of the melody. There were generally considerable deviations from the time values indicated in the score, for measures as well as for individual notes.

c. Singing

The analysis equipment for singing was essentially the same as for violin performance. Schoen (1922) used the Seashore tonoscope to study pitch intonation in phonograph records of five sopranos' performance of the Bach-Gonoud *Ave Maria*, a favorite piece in many Iowa studies. A tone was usually attacked below the intended pitch when it was preceded by a lower tone but on correct pitch when preceded by a higher tone. All singers tended to sing sharp in respect to both pure and equally tempered intonation, and the movement from tone to tone was mostly in the form of glides. R. S. Miller (1937) studied the pitch of the attack in tones that were preceded by a short pause, for artists (such as Alma Gluck and Enrico Caruso) as well as amateurs. A rising glide was the most common type of attack, and the extent of the glide was larger when the preceding tone was of lower pitch.

H. G. Seashore (1937) analyzed nine singers' performance of arias from Handel's *Messiah*, the Bach-Gonoud *Ave Maria*, and other songs, either directly recorded in a studio or taken from phonograph records. This is one of the most

penetrating studies of music performance ever made. The amount of detail is almost overwhelming, and yet much is left to the reader's own analysis. Pitch vibrato was always present. Its rate varied from 5.9 to 6.7 Hz and the extent from 0.44 to 0.61 whole tones, that is, roughly a semitone. Practically no correlation was found between rate and extent. Short tones usually had faster vibrato rates than long tones. Intensity vibrato was present about half the time with about the same rate as pitch vibrato. There were frequent deviations from nominally correct pitch, and the pitch varied in different ways (beside the vibrato) during the tone, often in a rising direction. Such deviations were usually not perceived. The transitions between tones were made with gliding attacks (rising in pitch) and releases, level attacks and releases, and portamento glides, all of them further classified into different types. A rising melodic line was usually accompanied by higher intensity and vice versa. Within phrases, there was mostly a crescendo followed by a decrescendo in a rather symmetric way. In other cases, there was primarily a decrescendo with no or only slight crescendo at the beginning. Ritards were made at the end of each song, and minor ritards occurred before interludes. Pauses between phrases were considerably longer than pauses within phrases. The duration of measures varied much, and there were many subpatterns of temporal deviations within phrases. Short notes tended to be overheld in comparison with long notes. This tendency toward an equalization of tone durations probably reflected a striving for proper legato performance. The performance of each individual singer was described in detail and illustrated both in performance scores and in graphs displaying the results (Figure 1).

d. Vibrato

Vibrato was the phenomenon studied most of all in the Iowa research. Early studies by Schoen (1922) were followed by more detailed analyses by Metfessel (1932) and Tiffin (1932b) of together about 25 professional singers (among them Caruso, Chaliapin, and Gigli) recorded on phonograph records. Their data were similar. Metfessel reported that the pitch vibrato varied in rate from 5.5 to 8.5 Hz with an average of about 7 Hz, and its extent from a tenth to over a whole-tone step, averaging a half step. According to Tiffin, the average rate was 6.5 Hz and the average extent 0.6 of a whole-tone step. There was throughout a certain variation in both rate and extent that was considered to be critical for the artistic quality of the vibrato. In Tiffin's study, the average difference in extent of adjacent vibrato cycles within the same tone was about 0.1 whole-tone step, in rate about 0.45 Hz. Intensity vibrato was present but of less importance. Easley (1932) found that opera singers used broader and faster vibrato when performing opera songs than when performing concert songs.

Using a special apparatus for synthesizing vibrato, Tiffin (1931) found that the perceived pitch of a frequency vibrato was slightly below the pitch of its mean frequency; however, the investigation was limited to only one frequency, 420 Hz. Recently Brown (1991) found that musicians located the pitch of a vibrato higher than nonmusicians, and neither group located the pitch at the mean of the modula-
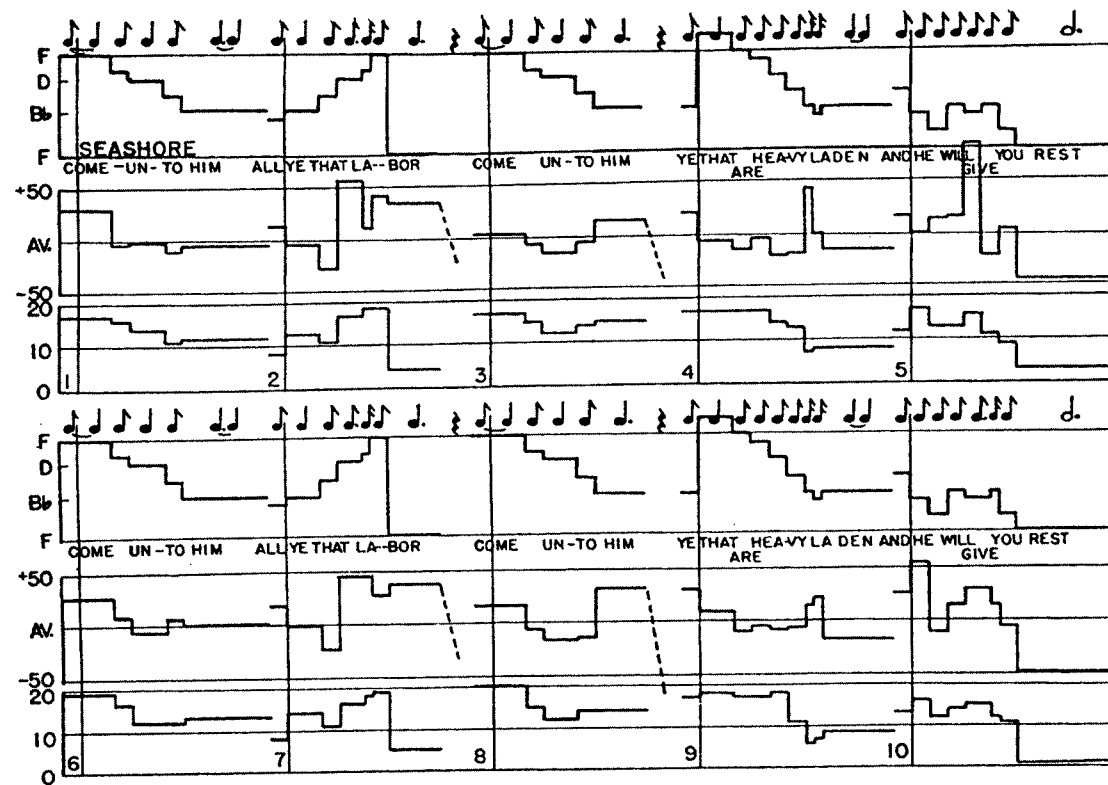
FIGURE 1   Graphical representation of a female singer's performance of the beginning of "Come Unto Him" from *Messiah* by Handel. Top to bottom: pattern of note—length, melody, words, relative duration, tonal power, phrase numbers. (Reproduced from H. G. Seashore, 1937.)