musicians need geometrical and mathematical descriptions of subtle differences in timbre?

3. What are the most important pitfalls to be avoided by anyone who wants to conduct controlled and musically valid perceptual studies of timbre?

9

Cognition and Time in Music

Time and Memory

In earlier chapters, there often was some connection to be made between psychophysics and cognition. The connection might have been tenuous, but still we can say that our learned responses to complex relationships of musical pitch and timbre often stem from our immediate and untutored responses to spectral-temporal ingredients and frequency relationships of tones.

The connection does not seem so apparent between the discussions in Chapter 6 and the discussions in this chapter. In Chapter 6, time was always an assigned value applied to some other property, such as the arrival of a sound wave or the comings and goings of spectral ingredients of a tone. These events were usually measured in milliseconds. Yet, we know that we encounter another sort of time as we perceive music. One of the more interesting and mysterious aspects of time in general is that clock time and mental time can be quite independent of one another. Every child (or at least every parent who has traveled with a child) knows that it seems to take much longer to get to a pleasurable event than it does to return home. Dreams also afford a common nonmusical experience of this independence. Most of us have at some time or other had a dream that we are certain took hours to unfold; yet, when we awoke, we found that only a few minutes had "actually" elapsed.[1] Musicians often relate similar experiences as they perform or listen to music (e.g., Kramer, 1987, 1988); some speak of being "transported" during this activity and seem to be alluding to a temporal experience rather than speaking in a romantic, melodramatic vein.

In this chapter, we will review studies of our consciousness of time itself in a musical context and discuss the interplay of time and musical (as opposed to psychophysical) information. This musical information is picked up across expanses of time measured in seconds, minutes, and even longer units, rather than in milliseconds; this means that memory becomes an important component of perception.

Psychologists have written for centuries about the intimacy of the connection between time and memory.[2] Yet, the questions of how our sense of time is altered by what we perceive, how our perceptions are altered by time, and even how we remember are still far from resolved. The discussion in this chapter will proceed from an introduction to general theories of time and of memory to a review of current theories and experimental studies of how perception, time, and music may influence one another.

The Perceptual "Now"

Some acoustical events, such as the droning of an electric fan or the roar of the surf on a beach, last so long and change so little that we seem to make the unconscious decision that they carry no important information, and therefore ignore them. Most often, acoustical events are much briefer, and this is true of most of the events on the perceptual surface of music as well. As we listen to a piece of music, the acoustical pattern that we heard just a few seconds ago is gone forever, and what we hear this instant is probably about to be modified slightly or changed radically in the next fraction of a second.

Suppose that visual information were as discontinuous across time as auditory information is—that it changed as rapidly and as completely. What visual patterns would we be able to recognize? Try to identify the visual pattern in Figure 9–1, in which a narrow "window" passes from left to right in each succeeding frame. The entire figure is shown in Note 3.

If it was difficult to grasp a sensible visual pattern in Figure 9–1, and it probably was, the reader understandably might be impressed that we can make so much sense of our aural experience of music; after all, it is forced to pass through a very narrow temporal window that we call "now." A number of psychologists have made estimates of the size of this perceptual now (commonly called the perceptual present), and these estimates range from a timeless instant up to several seconds in duration.[4] Fraisse (1978, 1982) suggests that the perceptual present for auditory information is flexible, averaging around 2 seconds in duration and seldom extending to more than 5 seconds.

Yet, the analogy of perceiving visual objects through a moving slit is not really satisfactory, for two reasons. First, the only apparent activity within the analogy is the motion of the slit, which serves as a metaphor for the passage of time. The perceiver is assumed to be a passive observer. This does not seem to match our musical listening experience, which is

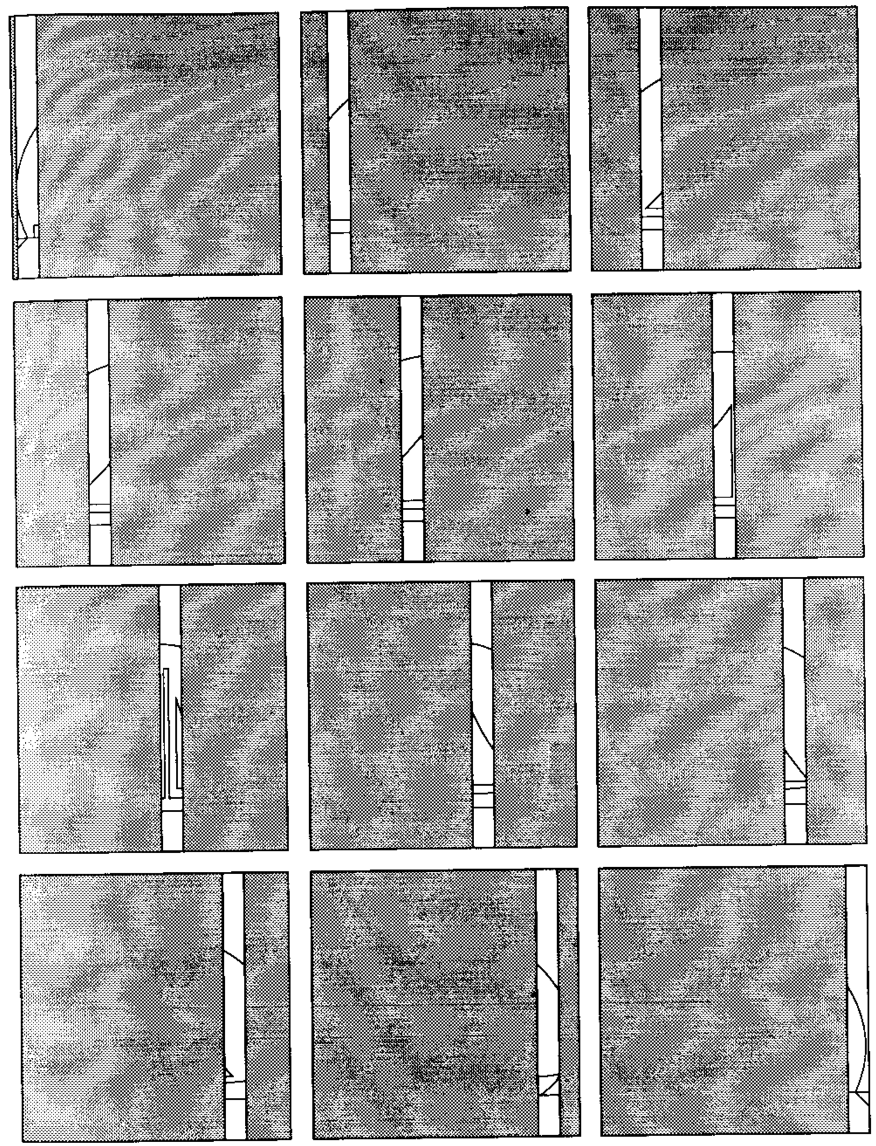

FIGURE 9–1. *Illustration hidden by a grey occluding panel, seen through a slit moving from left to right.*

much more active and complex. At one level, experience tells us that we track the successive events of music as we encounter them while listening to the performance; but at another level, we compare ongoing musical events with our memories of what we have already heard in this piece, making connections with earlier listening experience and making tacit predictions about what we will hear. At yet another level, we draw on our experience with other compositions that share formal and stylistic features with the music in process, and that experience shapes our attention. A second shortcoming of this visual analogy is that the slit is assumed to have a rigid, unchanging width. More likely, the size of the perceptual now depends on the particulars of the test used to measure it—particulars such as the structure of the stimulus material, the context in which the stimulus is presented, and the nature of the response task.

A final comment on the visual analogy of the masking panel with the moving slit: it might be that our musical knowledge guides our attention to the important particulars of what we can hear directly through the time window, so that we are not overwhelmed with minutiae. It might also be that our musical experience allows us figuratively to peek around the corners of the slit or simply to enlarge it, cataloguing what we have heard more effectively and efficiently and steering our anticipations of what may be encountered in the immediate and more distant future within the composition. This loose, metaphorical description of attention and memory will be replaced with a review of more-substantial theories and experimental work in the next section.

Musical Attention and Memory

Since the early 1960s, most of the theoretical and experimental work on perception has to some extent taken the view that attention and memory operate much as computers do: both humans and computers receive, organize, and store information; both recall and restructure that information; and both have limited capacities to do this input, storage, and retrieval work. This has come to be known as the information-processing model of perception and memory.

Information

Before any further discussion of the activity of processing, it would be useful to discuss the concept of information itself. In the late 1940s and early 1950s, mathematicians made rapid advances in the fledgling science of information theory, and George Miller (1956) drew attention to important connections between information theory and the mental operations that psychologists seek to describe. Miller pointed out that when you

want to talk about variance—that is, variations that you detect in the states of things in your environment—you have to express amounts of variance along a specific dimension: differences in warmth might be measured in British thermal units, calories, or degrees on different scales along one dimension, while differences in weight might be measured in ounces, grams, or other units on different scales along a different dimension. Variance, according to Miller, could instead be treated as information not confined to any single dimension, stating that "anything that increases the variance also increases the amount of information" (p. 81).

Information can be measured in **bits,** short for **bi**nary digi**ts**, or yes-no decisions, and these units of measurement are arranged along a neutral scale, independent of any single sensory dimension. To illustrate the concept of the bit, any yes-no choice between two equally likely possibilities amounts to one bit of information. If you were to play a number-guessing game with a person who was allowed to respond only "yes" or "no," or "higher" or "lower," to your question, you would gain one bit of information about that number with each response. For example, suppose that the challenge was to "pick a number [integers only, to speed up the game a bit]; the number is either 1 or 2." You would be only one question away from knowing which number had been picked, and thus would stand to gain one bit of information about the number. Pressing the example further, suppose you were playing this game with someone of diminished capacity who would always choose the same challenge. In this event, there would be a high level of redundancy and a low level of information. If the challenge, on the other hand, was to "pick a number between 1 and 15," the redundancy would be less, and the uncertainty, or **entropy,** would be greater; there would be more information to be transmitted. In this case, you would do best to reduce the number 15 to two equally likely alternatives (for example, above 7 or not above 7) and then split the resulting numeric range into two equally likely alternatives and so on, until you had identified the number.

Miller stated that human observers are limited in the amount of information they can take in without making errors and noted that these limits, or **channel capacities,** were quite similar, regardless of sensory mode. For example, observers begin to make errors in absolute perceptual judgments of more than about eight variations of hue, about five sizes of squares, and about seven locations of vibrations on their chests. Absolute judgments of auditory information become muddled at about the same point: listeners, according to Miller, could reliably discern about five levels of loudness and about six different pitches.[5] Miller distinguished between bits, which defined the channel capacity for absolute judgment, and **chunks,** which defined the channel capacity of immediate memory: a chunk is a mentally organized group of an indefinite number of bits of information. A simple illustration of perceptual chunking (common usage has transformed this noun into a part-time verb) is given in Figure 9–2.

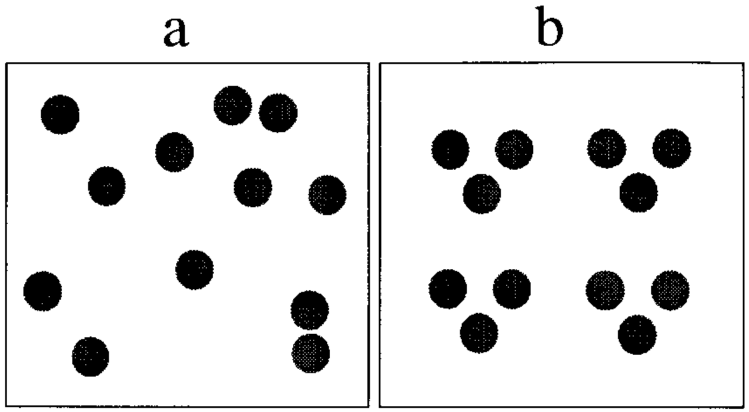

FIGURE 9–2. *Dots arranged (a) in a quasi-random configuration, and (b) in equivalent groupings.*

After covering up Figure 9–2b, ask someone to glance quickly at Figure 9–2a and then look away and try to estimate the number of dots that they saw. Now, do the same with Figure 9–2b. Although both fields contain the same number of dots, the observer most likely had less trouble estimating the number of dots in Figure 9–2b, because they probably grasped immediately that the dots were organized into four groups of three (recall the discussion of Gestalt principles in Chapter 7). By grouping information into chunks, then, the observer is thought to be able to sidestep channel capacity limitations, to perceive and remember much more information than would be possible without grouping. It is commonly assumed that chunking can take place at several levels: several bits of information can be grouped into a perceptual chunk, which can then be grouped with several other chunks to form yet a larger chunk, and so on.

Attention

The theory of chunking begins to explain how we can pick up many small pieces of information that are simultaneously available to one sensory organ and group those pieces of information so that our memory for that sensory information does not become overloaded. The theory does not describe how we select components of the sensory information that are of primary importance, tacitly assigning a role of secondary importance to other portions. For example—do not look back at Figure 9–2 yet—were the groups of three dots in Figure 9–2b arranged such that they formed triangles with the points up, down, to the left, or to the right? The task was to recall numbers of dots, and not to recall their geometrical configurations so clearly that those patterns are easy to recall. How do we make the unconscious decision to focus our attention on some of the information that our sensory organs pass along and to shove other information into the background?

A leading theory of attention within the information-processing line of thought held that perceptual filters admitted some incoming stimulus information and blocked the remainder (e.g., Broadbent, 1958, 1962). The blockage, so the theory went, was caused by a limitation of channel capacity; that is, it was suspected that there was a single channel of access (Welford, 1959, 1960) from the environment to the perceiver's memory and that this channel could transmit only so much information at any given time. The cocktail-party effect (see Chapter 7) seemed to support this single-channel theory, because when competing verbal messages were presented to the two ears and the listener was instructed to track one of those messages, listeners seemed to pick up little of the competing message.[6]

Other studies, however, provided evidence that at least some of the competing sensory input often did get picked up by the perceiver, and this gave rise to parallel-access theories (e.g., Deutsch and Deutsch, 1963), in which two or more stimuli can have parallel access to the perceiver's memory and in which the perceiver selects material for mental processing from information in memory rather than from sensory input.

A number of experiments have demonstrated that competing stimuli can be attended to, but that the competition may slow down the perceiver's reactions to those stimuli. Keele and Neill (1978) give a good summary:

> The early theories of Welford and Broadbent placed selective attention at an early point in processing, between the sensory and memory systems. . . . The fate of irrelevant information, observations of redundancy gain, reaction-time probes inserted during encoding, and the effects of retrieval time on interference all suggested widespread, parallel access to the memory system. In many situations, however, parallel access to memory entails costs. When stimuli are complex and in the same modality, they may mask or merge with one another; or if only one message is relevant and other codes or messages lead to conflict, then it would be to the organism's advantage to filter prior to memory. . . . Thus, rather than viewing attention as always being invested at a particular place in processing, it is better viewed as a control process that can influence the flow of information. When messages interfere with one another, the control process can attenuate one source of information prior to memory, although attenuation is quite likely incomplete. . . . When the messages do not interfere, they can be processed in parallel through at least the memory stage with no cost. This last point is the primary emphasis of this review. Not only may different memory locations

> within a coding system be activated in parallel, but also different codes may develop in parallel. A given word may activate physical, orthographic, phonemic, and semantic codes [p. 41].

Can memory itself serve as the attention-as-control process that influences the flow of information to our minds? Ulric Neisser's view of attention (discussed briefly here and at some length in Chapter 10) holds that the perceiver forms mental representations or **schemata** of objects in the environment and that these schemata are modified in later perceptual encounters with similar objects.[7] Just as important, these schemata guide our later attention by allowing us to anticipate features of the stimulus. This theory holds that memory, attention, and perception are inextricably bound together in an ongoing cycle (illustrated in the next chapter in Figure 10–2) in which schemata guide perceptions, which in turn sample the environment and then shape the schemata that guide later perceptions, and so on.[8]

Jones (e.g., 1976, 1986) argues, however, that none of these theories of attention describes in sufficient detail how attention is organized across time. Jones's rhythmic model of attention departs from the coding/decoding brain-as-computer metaphor, asserting instead that attention is a function of a rhythmical perceiver and of rhythmicities in the environment. This model seems congruent with some recent findings in speech perception research and indeed may describe an aspect of attending that people bring to bear both on speech and on music.[9] Jones and her colleagues (Jones, Boltz, and Kidd, 1982) found that they could systematically direct listeners' attention to extended or broken patterns of pitches in simple melodies by altering the rhythmic patterns in which the pitches were heard. An example of these pitch and rhythmic patterns is given in Figure 9–3.

Pitch strings similar to that shown in Figure 9–3a were played twice in each trial of a listening experiment. The common feature among these pitch strings was that pitches were organized in easily discerned groupings; for example, the sample sequence CDEGABFED in Figure 9–3a is quite clearly broken up into three-note segments that are stepwise internally and bounded from one another by skips. These pitches were laid over rhythmic patterns that were monotonic (Figure 9–3b), that were designed to emphasize the sixth tone in the pitch pattern (Figure 9–3c), that were designed to emphasize the fourth tone in the pitch pattern (Figure 9–3d), or that were irregular (Figure 9–3e).[10] In one-half of the second playings of each of these patterns, one of the tones was changed to another pitch within the C-major set: the change came on either the fourth tone, which marked the beginning of a three-tone pitch group, or the sixth tone, which was within a three-tone pitch group. Listeners were able to identify alterations of tones that began the three-tone groups with significantly higher accuracy when rhythmic context directed listeners' attention to those tones.
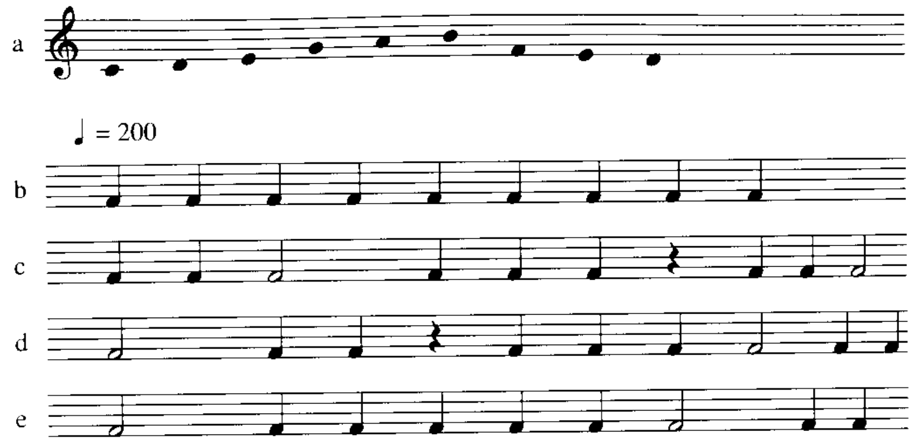
FIGURE 9–3. *(a) A pitch sequence and (b–e) sample rhythmic sequences. After Jones, Boltz, and Kidd (1982). Used with permission.*

Carlsen (1981) found that when listeners are asked to extrapolate melodically from only two tones (representing all ascending and descending intervals of an octave or less), they will tend to agree on how some of these melodic incipits are best extended. He found further that these patterns of concurrence change from one country—and therefore presumably one musical culture—to another. This result seems to support a number of related perceptual theories that address the role that expectations—based on important melodic "prototypes" (e.g., Jones, 1976, 1981, 1982) and "archetypes" (Meyer, 1956, 1967, 1973; Narmour, 1990; Rosner and Meyer, 1982, 1986; cf. the related discussion in Chapter 7)—play in perception. Unyk and Carlsen (1987) extended this work by building melodic-expectancy profiles for individual listeners, based on their sung extrapolations of melodic dyads. For example, listener A might tend to hear an ascending major third and then sing a tone a minor third above that; listener B might tend instead to extend a major third by singing a perfect fourth above the second tone of the dyad. After mixing orders of presentation and repeating this process, Unyk and Carlsen were able to find patterns, of various strengths, of both direction and size of steps or skips that individuals sang in response to the melodic dyad cues. Unyk and Carlsen then composed melodic-dictation problems that either fulfilled or denied the patterns of melodic expectation for these participants and evaluated the accuracy of the participants' dictation of these melodies. They found that listeners did indeed make more errors in dictation when the melodies violated their expectations. This was true even when melodies were quite simple and listeners were fairly advanced in their musical training.

Similarly, Bharucha and Stoekig (1986) have reported that listeners can develop harmonic expectations after having heard only one chord.[11] Dowling, Lung, and Herrbold (1987) asked listeners to track familiar melodies interleaved (see Chapter 7) with competing distractor tones and found that, based on their familiarity with common structural features of melodies within their culture, listeners develop "expectancy windows": that is, pitch-time regions of expectancy within which tones are detected more accurately than are tones outside the regions. These results support Jones's (e.g., 1976, discussed earlier) contention that we not only expect events based on our prior knowledge and immediate perceptual activity but also have some sense of the place(s) and time(s) to expect these events. Schmuckler (1989) has found that both listeners and performers demonstrate fairly consistent melodic and harmonic expectancies: the listeners through ratings of continuations of melodic and harmonic patterns, and the performers through rehearsed, quasi-improvisational continuations of those patterns.

Memory

The processes of mentally storing and retrieving auditory information have received a great amount of discussion in the psychological literature.[12] Various information-processing theories tend to assume that we remember things in two ways, somewhat as computers do. Immediate memory, which is variously called **short-term memory (STM)** or *episodic memory,* is likened to the random-access memory (RAM) or volatile memory in a computer: it tends to be described as very fast-moving and as limited in capacity.[13] **Long-term memory (LTM),** or "semantic" memory, is thought of as more nearly analogous to a computer's magnetic-disk or magnetic-tape mass-storage device. It may take longer to retrieve information from long-term memory, but long-term memory contains much greater storage capacity. Broadbent (1971) wrote of short-term memory as a metaphorical desktop, describing what more than a decade later would become an enormously popular graphic user interface in computers. The processor (the human's mind or the computer's central processing unit) works actively on the job immediately at hand at the desktop level of activated memory (the human's STM or the machine's RAM), using information that can come either from beyond the desk (the human's sensory channels or the machine's input-output devices, such as a keyboard, a "mouse," a joystick) or from "file drawers" inside the desk (the human's LTM or the machine's information-storage devices, such as magnetic tape or hard disks).

What is the dividing line between short-term and long-term memory? Attempts to arrive at a specific time limit have produced inconsistent results, and it seems instead that it would be more useful to think of an operational distinction. Short-term memory would be the "buffer" or

bank that contains all of the information being "processed" at any given time. Long-term memory, then, would be the bank that contains the stored information not being consulted or processed by current mental activity.

There have been numerous objections to the "information processing" theory of memory. One objection often heard some years ago was that computers do their thinking—which includes storing impressions away in memory and retrieving them (or others) again—in a strict serial order: earlier computers (and many in current use) did their work by stepping through programs in strict sequence. Many psychologists argued that humans can, and do, think simultaneously on more than one level—sometimes along parallel or convergent mental paths, at other times along simultaneously diverging lines. The recent development of computers with multiple parallel processors offers a response to these sorts of objections to the strict machine model of mental processing.

Another objection, not so easily answered, is that mass storage units (if they do their job properly) preserve information in an inert form. On the other hand, human memories can decay or, more likely, will become altered (interfered with) by subsequent mental activity. Moreover, a computer's information-storage unit stores away everything fed to it with the same zeal; the disk or tape recording head makes no distinctions about which information is important and which is trivial. Human memories are more sensitive to effects of perceptual salience. For example, beginnings of melodies are usually recalled from long-term memory more easily than middles or endings are. Similarly, it should be easier to recall tonal series with an obvious pattern than without one, as Listening Example 9.1 shows.

Listening Example 9.1: (a) A twelve-member tonal series with an obvious pattern of pitch relations across time, followed by (b) a twelve-member series without an easily discerned pattern. The same pitches, in the same quantities, are present in each example. Which can you recall more completely after a single hearing? The two series are presented twice, and are notated in Note 14, Chapter 9.

The Perception of Time in Music and of Music in Time

Usually, it seems that we are only aware of the passing of time when we have nothing better to do with it: when we are trapped in a doctor's waiting room, have to endure an extended layover in an airport, or must listen to a long piece of music that we do not understand. These are unpleasant situations; fortunately, most of us do not encounter them too often. This raises a question about psychological theories of time perception: Is it worthwhile to study the perception of time in and of itself, or would it be more useful to study the perception of time only as it accompanies the perception of other objects and events?

The historical answer to this question is that experiments on time perception have measured the perception of both filled and unfilled time. These experiments have tested a number of theories of time perception; two types of theories, described categorically as *storage-size* and *attentional capacity* theories, have received most of the experimental attention in recent years. The storage-size theory (cf. Ornstein, 1969) states that perception is linked to memory and that memory-storage needs influence our estimates of perceptual time: a percept containing a large amount of information will require more storage capacity in short-term memory, which then will generate an impression of greater elapsed time. Alternatively, several attentional capacity theories (e.g., Block, 1978; Block and Reed, 1978; Craik and Lockhart, 1972; Hicks, Miller, and Kinsbourne, 1976; Underwood and Swain, 1973) have been proposed. These are based on the beliefs that attention is a limited commodity and that we are always to some extent aware of passing time—that we possess a cognitive "clock" that moves into and out of our focus of attention.[15] The theory, then, is that perceptual activities compete for attentional center stage, and the more we attend to other mental impressions of the stimulus, such as the amount of information or the complexity of information, the less attention we have left to spend on time itself, with a resulting impression of less elapsed time.[16]

Listening Example 9.2: Two excerpts from "Very Small Variations for Sybil on her Third Birthday" (1989) by Thomas Wells; used with permission of the composer.

Listening Example 9.2 demonstrates that listeners may get different impressions of elapsed time, depending on such variables as tempo, rhythmic activity, and the number of formal groups that the listener identifies. Each excerpt in the example is 15 seconds long. Try asking several listeners to estimate the lengths of the two extracts, compare responses, and then read Problem 10 at the end of the chapter.

These theories seem too general to be of much musical use, unless they are re-shaped to address specific attributes of perceived music, where information can exist in varying levels of complexity for a number of different attributes that may be either interrelated or independent of one another. If one 12-second-long segment of music has a simple melody and simple harmonic accompaniment that is arpeggiated in a complex rhythmic pattern, will the information content be equivalent to a 12-second-long segment of music with a disjunct melody and harmonically complex accompaniment that is carried by simple, regular blocked chords? No one knows. Moreover, there is some experimental evidence (Arlin, 1986a, 1986b) that both theories can be supported by the results of a single experiment that involves relatively simple perceptual tasks (discussed in Chapter 10). In summary, time-estimation behavior may be more complex than either type of theory alone can describe.

Temporal Aspects of Musical Perception and Performance

Tempo

Tempo has an obvious relationship to the passage of clock time, and a number of studies have been made of listeners' and performers' abilities to recognize or reproduce given tempi. Clynes and Walker (1982, 1986) have measured the durations of performances of the same piece of music by the same conductor or performer in separate performances weeks, months, and even years apart, and have found that some conductors and performers can be amazingly consistent in the tempi at which they perform the same piece of music on different occasions. For example, Clynes and Walker (1986) compared different recordings of the same compositions, made at public performances by the Sydney (Australia) String Quartet during the years 1975–1979 and found a great deal of stability in the durations of those performances—often a fraction of a single percentage point of difference.

Converging results were obtained by Halpern (1988a, 1988b), who asked one group of test participants to alter the tempi of computer-generated familiar tunes (e.g., "Happy Birthday," "Do Re Mi," "When the Saints Go Marching In") until they were satisfied that the tunes were proceeding at the most appropriate rate. Another group of test participants was required to imagine those tunes—again, at the preferred tempo and to set a metronome to click at the imagined tempo. Halpern found that test participants (selected only on the basis of their familiarity with the well-known tunes used in the study) were quite consistent in their selection of "best" tempi. In a second experiment, she asked participants to imagine each tune at the tempo that seemed most natural to them and then (in one phase of the test) to set the slowest and fastest tempi at which they could imagine the tune and (in another phase) to rank the difficulty of imagining that song at a tempo they progressively dialed faster (to a maximum of 208 beats per minute) or slower (to a minimum of 40 beats per minute). She found in this experiment that the slowest and fastest tempi were directly related to the preferred tempo, suggesting that there are not absolute minima and maxima for musical tempo. Halpern (1988b) concluded:

> These two studies suggest that familiar tunes are represented in memory with a particular tempo, but that the tempo can be mentally transformed up to a certain point. . . . Visual imagery theorists like Kosslyn (1981) propose that the imagery experience is a "surface representation" generated from an underlying "deep structure" in long-term memory. To simplify, the deep structure is an abstract (symbolic) description of the imaged object or scene. In the case of musical imagery, conceiving of the tempo as a symbolic encoding at a deep level has logical appeal. Some form of internal timekeeping mechanism must operate to allow us to judge remembered or ongoing durations or produce events at designated time intervals. Perhaps the deep level of auditory imagery contains a setting for the internal timekeeper attached to a node for a tune title. This setting would be considered a default value that could be changed if necessary [p. 201].

Expressive Timing

A computer can turn out a perfectly precise performance of a musical composition, a flawless performance in which no tones are played in error and in which every tone is heard precisely when the score indicates that it should be. Many people would agree that such a performance ought to be heard only by other computers; it would be roundly criticized by most musicians as lifeless, mechanical, rigid, inhuman. Deviations from a strict, metronomic tempo seem to be expected by the listener as part of the vocabulary of musical expression. Most musicians can probably think of performers or of single performances that span a broad compass of emotionality—from robotic to melodramatic. Although there was some experimental investigation of such timing deviations many years ago (e.g., Henderson, 1936), technological advances have stimulated much more work in recent years.

Three obvious questions we can ask about timing deviations are (a) whether there is enough consistency in these deviations that they can be distinguished from performance errors, (b) whether these timing deviations are best performed under direct conscious control or sound most natural if done unconsciously, and (c) whether expressive timing is related most directly to the individual performer, to the individual passage, to the individual composition, to the entire output of a composer (e.g., a "proper Mozartian" style of timing deviations), or to the entire body of performance literature from a stylistic period (e.g., a "proper Classical" style of timing deviations).

A number of performance studies (e.g., Sloboda, 1983; Clarke, 1985; Shaffer, Clark, and Todd, 1985; Todd, 1985) seem to show that performers will make consistent, nonrandom deviations from metronomic regularity in their performances and that these deviations relate directly to such local-level musical features as metrical stress and phrase endings. Palmer (1989) investigated three types of deviations from metronomic tempo: asynchronous attacks of tones notated as synchronous; rubato; and articulation (specifically, measuring the size of gap in staccato passages and the amount of overlap in legato passages). Palmer found that performances were consistent when judged in relation to any given performer and were systematically related to the respective musical interpretations of different performers. Performers played synchronous groups of notes such as block chords such that melodic tones preceded the others by a subtle (about 20 milliseconds) lead time. Rubato was most evident at phrase boundaries: "Most performers lengthened phrase endings, similar to phrase-final lengthening in speech utterances. . . . Some performers, however, showed the opposite pattern, which indicates that the amount of change (rather than the direction of change) may signal a new musical phrase or idea" (pp. 344–45). Legato overlap was found to be proportional to the durations of preceding and following tones. Palmer suggested that the motor programs for these timing deviations are under only a limited amount of conscious control: "The pianists in this study had indirect control over the timing methods through conscious manipulation of their interpretations or musical intentions, such as choice of melody or choice of phrasing. These conscious intentions were then translated into performance timing through the primarily unconscious application of chord asynchrony, rubato, and overlap methods" (p. 345).[17]

Listening Example 9.3: Comparative renditions of the same performance, using a Bösendorfer computer-controlled piano. (a) Excerpt from Beethoven, Sonata No. 30 in E Major, op. 109, mvt. III; (b) the same performance, but with all notated simultaneities adjusted so that their onsets are actually synchronous; (c) the same performance, but with all key velocities equalized. After Palmer (1988).

Listening Example 9.3 allows the listener to compare the effects of varying key velocity and varying onset times in expressive performance. A pianist, who had taken sixteen years of private instruction and who was currently taking lessons, played the theme of the third movement of the Beethoven Sonata No. 30 in E, op. 109, on a Bösendorfer 290SE computer-actuated grand piano. This performance is presented in Listening Example 9.3a. The computer code produced by the piano showed that the performer played melodic tones on average 54 milliseconds before playing the accompanimental tones notated as simultaneities.[18] Palmer (1988) reported that the performer also tended to play the melody substantially louder than the accompaniment. The piano records key velocities on a scale of 0 to 1,023 units, although typical performances remain within a range of around 400 of these units. The upper-voice (melodic) tones in this particular performance (Listening Example 9.3a) were on average 101 units louder than the accompaniment. Listening Examples 9.3b and 9.3c allow the listener to hear the effect of removing the differences in onset times and in key velocity from the original performance.

Thus far, experimental evidence has tended to indicate that expressive timing has most to do with local features of the music, such as metrical stress and phrase boundaries. Is expressive timing guided as well by more-general musical characteristics such as a composer's style or a broader set of stylistic norms that relate to a historical period? Although our musical intuitions may argue strongly that there must be different styles in expressive performance timings, especially with regard to stylistic periods, there is as yet little solid experimental evidence to bear these intuitions out. There has been some work done on composer-specific patterns of subtle timing deviations, but these do not offer us a clear picture. Clynes (1983, 1987) proposed that there exists a "composer's pulse"—a subtle, cyclical composer-specific pattern of timing and intensity variations that exists in every work by that composer, no matter the differences in tempo, meter, mode, or form. Clynes's primary evidence for the existence of this "composer's pulse" has been a series of computer-controlled renditions of compositions by composers whose musical pulse Clynes seeks to demonstrate. Repp (1990) analyzed oscillographs of nineteen performances of the minuet movement of the Beethoven Sonata op. 31, No. 3, and his oscillograms did not support Clynes's composer's-pulse hypothesis. Rather, Repp found that deviations from metronomic regularity strongly adhered to a common performance standard and could be explained by structural features of the piece at several levels—most noticeably, lengthening at phrase boundaries and at moments of melodic or harmonic tension, and elongation of short tones when they were metrically unstressed and shortening of tones of the same notated duration when they were to receive metrical stress. Repp (1989) and Thompson (1989) both tested the perceptibility of Clynes's pulse hypothesis, and each obtained inconsistent results: although listeners often matched composers with the pulses that Clynes's hypothesis deemed most appropriate (that is, listeners often identified as "Mozartian" musical examples played with the pulse pattern that Clynes asserts is Mozart's), there were some troubling mismatches. Thompson pointed

out that there are some important sources of potential confounding in these studies. For example, Thompson argued that although there might be systematic differences between the rhythmic "microstructure" in Beethoven's and Mozart's compositions, there are also characteristic modes of articulation, rubato, dynamic range, and pedaling (in keyboard compositions) for each of the two composers, and it has not yet been demonstrated that any of these facets of performance is any less important in conveying the composer's individual stylistic "signature" than the microstructural "pulse" is. In sum, it is still an open question whether there are styles of expressive timing specific to individual composers.

Two different discussions in this chapter have come as close to dealing with musical emotion as the scope of this book will permit. The earlier discussion of expectation and the coverage of expressive timing just completed give a hint of a connection between the cognitive and affective domains of musical thought.[19] Although limitations of space prevent an adequate treatment of the topic of emotion in music, the reader is encouraged to consult a number of important source works on this topic. Leonard B. Meyer's work (e.g., 1956, 1967, 1973) is the standard (and wisest) starting point for gaining a theoretical account, and Rosner and Meyer (1982, 1986), Schmuckler (1989), and Sloboda (1989, 1990) have reported relevant experimental work.

Rhythm and Meter

Psychological studies of "rhythm" are often of limited musical use because they measure little more than binary-judgment responses (e.g., same/different, strong/weak) to single-variable (for example, long/short, loud/soft) patterns formed by a few sinusoidal tones.[20] This is not likely to interest musicians who observe much more complex perception and production activity in the aural training classroom and in the performance studio. Nevertheless, a number of studies have produced results that are useful for framing some rough descriptions of rhythmic and metric perception. A study by Povel (1981) demonstrated that given sufficient context, people both with and without extensive formal training in music can accurately tap low-ratio series such as 2:1 (quarter-eighth-eighth) and 3:1 (dotted quarter-eighth-eighth-eighth). Sternberg, Knoll, and Zukofsky (1982) found that their listeners (who were extensively trained and experienced but only three in number) could accurately describe and/or reproduce rhythms that adhered to commonly occurring rhythmic patterns even when beats were divided into fifths, sixths, sevenths, and eighths. On the other hand, the observers had much greater trouble describing and reproducing short-long patterns that occur infrequently in "common-practice era" Western music.

In another study (Handel and Lawson, 1983), observers were asked to tap along with simple polyrhythmic patterns such as two-against-three and three-against-four. Observers tended to tap along with the periodicity points—the points of coincidence—formed by the two rhythms. As the tempo of the pattern was increased, observers would drop the tap rate to the next-slower level so that each tap marked a longer periodicity. When faced with more-complex patterns such as two-against-five-against-seven, observers would either tap along with both of the slower streams, evidently grouping them as a single stream, or they would group the faster two streams, ignore that "busy" group, and tap along with the slowest stream. This result indicates that observers are facile at shifting from one rhythmic level to another, possibly seeking the simplest pattern into which the others will nest. This might be a sensible way of viewing the aural identification of meter, which is much easier to specify on the score than aurally: it may be easy to decide that one hears a duple pattern or that the beat is simple or compound, but unless one is told the unit of beat, the aural identification of meter has to be an educated guess, as is shown in Listening Example 9.4.

Listening Example 9.4: Brief adaptation of Beethoven, Symphony No. 5 in C Minor, mvt. III; the example may be 3/2, 3/4, 3/8, 6/4, 6/8, or 12/8 meter. If you can decide which meter the example is "really" in, what would prevent someone else from arguing just as successfully for another choice? What effect(s) would a change of tempo have on your choice?

Proportionality

Studies of proportionality take a longer-range view of how we might be aware of time passing in music. Some of these studies carry a lingering odor of Pythagorean number worship, in that they tend to assume that certain "well-formed" or "intrinsically interesting" sets of numbers have guided composers as they structure their musical compositions across time. Some of these favored numeric relationships are found in small integers, in the Fibonacci series, and in the Golden Mean (the relationship of three quantities X, Y, and Z in which $X + Y = Z$ and in which $X{:}Y = Y{:}Z$).[21] We can question the perceptual significance of whatever proportional relations may exist in music. For one thing, analytic studies of proportional relations in music tend to take the goal of finding musically significant events at points that coincide with the structure of some numeric series. Quite often, the authors of such studies are willing to overlook the fact that the sought-out significant event occurs not at the temporal point specified by the numeric series, but only in its vicinity. Perhaps more important, the question arises that even if proportions are demonstrably present in the performance score, can they be heard? It would seem doubtful that they can, if one puts any faith in either the storage-size or the attention-capacity theories of time perception described earlier. Regardless of their differences, both theories seem to concur that psychological time and clock time are independent to some degree, because (although the theories may disagree on the manner in which it happens) the amount of information—or the degree to which that information absorbs the attention of the perceiver—can stretch or shrink the perceived time in which the information was conveyed. This should mean that an analysis of proportions based on simple counts of measures or of elapsed time on a stopwatch ignores the changes in density of information that occur within a piece of music—changes that can make a phrase or a section seem to last a shorter or longer time. Proponents of theories of proportionality would counter—as do proponents of levels theories (see below) and pitch-set theories when the argument of perception comes up—that perceptibility is not everything. It is enough, they say, to know that such long-range relationships exist in the music, because these relationships can influence the choices a composer may make, and these relationships may be indirectly important to the way we perceive the music.

Perception of Hierarchical Levels of Musical Structure: Levels Theories

Undoubtedly, the writings of Heinrich Schenker and his followers constitute the best-known body of theories of hierarchical levels of contrapuntal-harmonic relationships in music. A thumbnail sketch of the most fundamental characteristics of the theory is that it takes a "natural law" stance by beginning with the major triad—as revealed in the harmonic series—as its basis.[22] The major triad is the basis of the *Ursatz,* a contrapuntal "background" structure consisting of a scalar upper line (*Urlinie*) and a tonic-dominant bass line (*Grundbrechung*) that spans and lends coherence to a tonal composition. This background level, typically with a descending scale-wise *Urlinie* and a I-V-I or I-III-V-I *Grundbrechung,* exerts a shaping force on a middleground level, which may itself comprise one or more stages. The methodology for deriving separate stages within the middleground is still subject to some debate (Proctor and Riggins, 1986). Similarly, the middleground exerts a strong shaping force on the most melodically and harmonically significant notes in the actual composition (called the "surface" by some writers and the "foreground" by others). This explanation may seem backward in comparison to most descriptions of Schenkerian levels, which tend to begin with the foreground level and systematically reduce it to the background level. The analytic process is described as it is here to underscore the point (cf. Proctor, 1977; Proctor and Riggins, 1986) that the purpose of levels analysis is not to erase notes from scores in the reductionistic pursuit of a simple, overarching control structure; rather, it is to reveal the existence and interrelationships of several levels of tonal process operating simultaneously in musical compositions.[23]

Although Schenker apparently proposed his theory of multiple levels on the basis of logic rather than on the basis of their perceptibility (Schenker, 1935, p. 18), the theory is so influential that it seems that some of its aspects—harmonic elongation and surface-level melodic and contrapuntal trajectories, for example—would have been fully examined in a number of controlled perceptual studies. This does not seem to be the case, however, for experimental studies of the perception of large-scale tonal relationships are quite rare. One, reported by Nicholas Cook (1987), will be discussed later in this chapter. Another study, by Killam and Baczewski (1985), was an ambitious and complicated affair that ended up with complicated results. Killam lured an unsuspecting group of music theorists into a session at a professional conference in 1983 on the pretext of presenting a paper and then passed out blank music paper and requested that they make an on-the-spot transcription of J. S. Bach's Chorale No. 117, "Ich bin's, ich sollte büssen." A tape recording of the chorale was played once, and the 58 participants were instructed to write as much as they could of the soprano and bass lines and any analysis they considered "significant" during the playing of the tape and a silent interval of five minutes afterward. On the average, 35.77 theorists out of 58 correctly transcribed any given note in the soprano line, while only 11.82 correctly transcribed any given bass note. Harmonic analysis was far less accurate, with a mean of 4.32 accurate chord symbols, out of a total of 5.6 per verticality.[24] Killam and Baczewski reported that the participants did work reductively to the extent that they seemed to disregard nonharmonic tones, but not to the extent that they picked up nonadjacent repetitions of the tonic A♭ in the melody. The authors concluded that their results seemed to indicate "some sort of Schenker-Lerdahl reduction" and that these results could not be accounted for either by Bharucha and Krumhansl's (1983) hierarchies of stability or by Deutsch and Feroe's (1981) melodic hierarchy model (cf. Chapter 7).

Recently, a "generative" theory has been advanced by Lerdahl and Jackendoff (1983) as an alternative means of describing long-range time relations of pitch in tonal music. The theory rests primarily on four "components of musical intuition that are hierarchical in nature" (Lerdahl and Jackendoff, 1983, pp.8–9). The first is *grouping structure,* the perceptual grouping of motives, phrases, and sections. *Metrical structure* is the second component, having to do with the perceptual grouping of strong and weak beats on a number of levels. The third component is *time-span reduction,* which involves the perceptual ranking of the significance of each pitch based on its position in the grouping and metrical structure. Fourth and finally, *prolongational reduction* involves the perceptual ranking of each pitch with regard to its expression of melodic and harmonic stasis or momentum. The theory is powered by a number of "well-formedness rules" and "preference rules" that can be applied to these four perceptual components at each hierarchical level. The well-formedness rules describe a stylistically appropriate analytic-perceptual interpretation of a specific component. There may be several well-formed interpretations of a single grouping, and in this event the preference rules, which are derived from Gestalt principles such as proximity and similarity (cf. Chapter 7), may be applied to identify the best choice among these alternatives. Perceptual groupings, according to the theory, arise from perceptual segmentation on the basis of pitch distance (the Gestalt principle of similarity) and temporal distance (the Gestalt principle of proximity), as well as from discontinuities of articulation, dynamics, timbre, and note length. Irregularities of formal structures, such as elided phrases, are described by "transformational rules," which are regarded as peripheral rather than central to the Lerdahl-Jackendoff theory. Figure 9–4 presents a graphic synopsis of the theory, with the musical input at the top and the perceived structure at the bottom.

The report of an early test of these grouping rules has been published by Deliège (1987), who found that listeners (both musicians and nonmusicians) tended to segment the melodic lines of short (three to sixteen notes long) extracts from recorded music performances essentially as the grouping rules predicted. Deliège noted that musicians and nonmusicians did not respond with "radically" different groupings (p. 325), and suggested that the grouping rules be expanded to include segmentation based on changes in harmony and on instrumental and/or sound density (p. 357).

Deliège's study, wisely enough, tested perceptual grouping responses to short extracts and asked listeners to mark a single point of segmentation in each example. The complexity of grouping preference tests would quickly become overwhelming when segmentation judgments were solicited on more than one level or when longer extracts of music were presented. This is probably why there is a virtually nonexistent experimental literature on perception tests that deal with long-range multilevel pitch relationships such as those described in "levels" and "generative" theories of tonal music. Strong doubts have been raised about the value of these music theories as psychological theories. For instance, Rosner and Meyer (1986) argued that analysts such as Schenker and Lerdahl and Jackendoff can graph the hierarchical properties of long stretches of music, but that these are of little use as psychological theories if they have little or no relevance to perception. They point out that

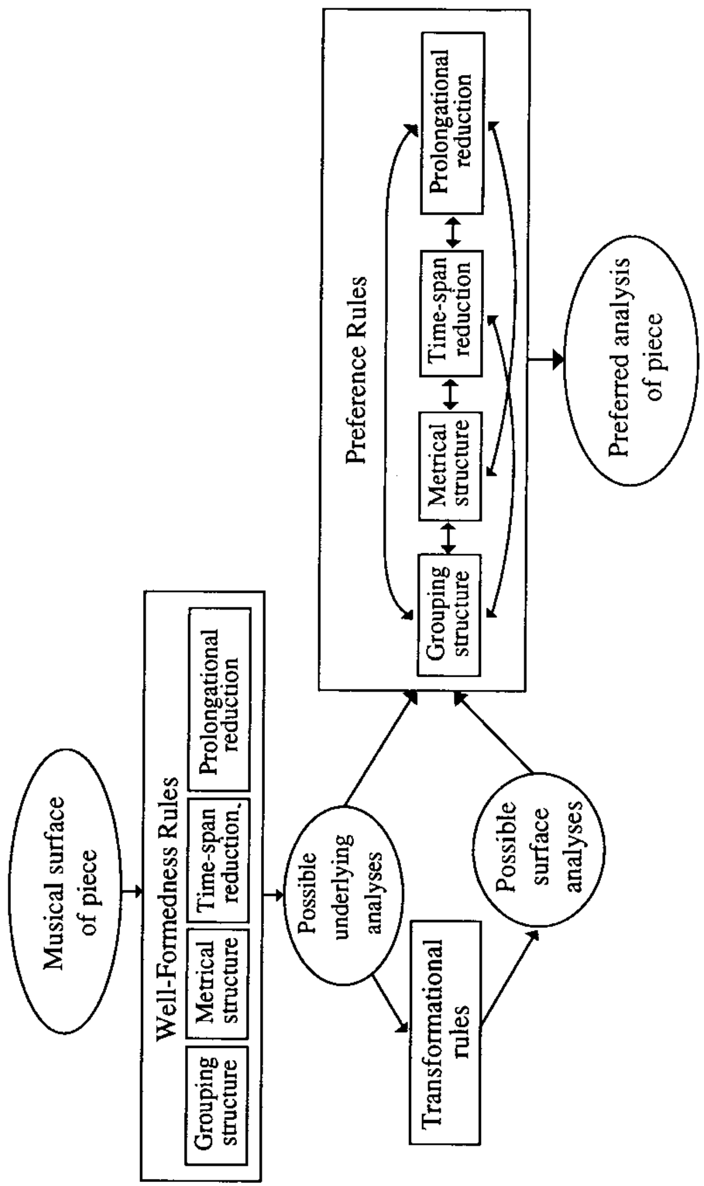

FIGURE 9–4. *Graphic depiction of the generative theory of tonal music proposed by Lerdahl and Jackendoff (1983, p. 10). Used with permission.*

the top node of a grammatical tree is an immediately observed datum: a sentence or an utterance. It represents some incident, occurring over time, which can be entered completely and rapidly into memory. The associated tree decomposes that uppermost node into parts at several lower levels of a strict hierarchy. The lowest nodes in music-theoretic tree structures, however, represent a datum: an actual stretch of music. Quite often, only fragments of it are held faithfully in memory. The lowest nodes in the tree are not decompositions of higher ones. Instead, higher nodes are selections from among lower ones. We therefore cannot believe that the increasingly higher nodes, which represent ever more rarefied selections, form the core of music perception [p. 37].

Regardless of the many problems that accompany controlled tests of the perceptibility of long-range temporal relationships in music, these relationships have been shown to exist in musical scores, and it seems reasonable to suspect that some people might be able to hear them.[25] Clearly, some listeners are much more capable than others of recognizing the harmonic and key levels of temporally remote points in large, multi-movement compositions. Although the entire audience at a performance of, say, the Mahler Ninth Symphony hears the acoustical "surface" of the music at essentially the same time, some listeners will likely be trapped at or near this surface because of their lack of listening expertise; others may mentally dart back and forth from one portion of the surface to another, collecting and synthesizing these impressions and formulating expectations. There is some question, however, as to how much even sophisticated listeners can do to make mental connections among temporally remote points in tonal music: Cook (1987), for example, found that listeners were quite insensitive to whether tonal compositions ultimately returned to the original tonic or whether instead they concluded in a different key.[26]

Cone (1968, especially p. 89) distinguishes "immediate apprehension," the musical "now," from "synoptic comprehension," the connection of the present with the comprehension of past. Cone also discusses the "esthetic surface":

> There is a great deal of music that, while resting comfortably in its extremes and making no demands on a hypothetical environment, nevertheless cannot be structurally comprehended. . . . It is certainly true that what many of us are enjoying (if that is the right word) through long stretches of *Also sprach Zarathustra* or *Siegfried* is surface. Even if one could . . . prove that each Wagnerian act could be analyzed as a structural unity that organically incorporated every smallest detail, that would not prove that each act could be so *heard*. Synoptic comprehension is indeed partly conceptual—

but only partly so. It is still a mode of perception. There is probably a limit in sheer time to what the human ear can take in structurally; when this limit is overstepped, the listener falls back on immediate perception [p. 96].

We have seen earlier that unidimensional perceptual models often do a poor job in describing the complex relationships in music: witness the mel scale (Chapter 3) and the sone scale (Chapter 4). Fraser (1985) argues that the perceptual "now" is hierarchical, consisting of nested social, mental, and biological "nows." It seems inconceivable that our temporal impressions of large expanses of music, or our abilities to remember and anticipate musical relationships across large expanses of time, could be adequately measured across a single dimension or described by a simple cognitive model.

Summary

Our senses give us access to music through a very narrow audible "now." If important acoustical events take place, say, every 10 milliseconds or so, or even if our perceptual sense of present spans as much as 5 seconds, we have direct perceptual access to less than 3 percent of a 3-minute popular song or less than 1 percent of a 10-minute symphonic movement. The rest of the music exists in our memories and in our expectations. A growing number of perceptual theories are suggesting that memory and expectation are closely connected to one another and to the way we perceive: our memory guides our attention, influencing our perceptions; these perceptions reshape our memory, and the cycle continues. For example, listeners apparently build up a mental library of melodic models or archetypes and can, given only the beginning of a melody, form a fair amount of consensus about how the tune should most likely continue. Melodies that fulfill the expectations of the listener are easier to transcribe than are melodies that are no more complex but fail to fulfill the transcriber's expectation. This mutual involvement of memory, perception, and expectation implies that memory is more dynamic, more active, than some "long-term storage" accounts of it suggest.

Our perceptual awareness of time in music deals primarily with short-term relationships, relationships that we can observe directly. For example, it appears that performers and listeners are quite sensitive to tempo. Listeners seem to have a mental library of "good" tempi for familiar tunes. Performers have been observed to play the same composition at virtually the same tempo, in performances that are sometimes years apart. Performers also can be very consistent in the subtle deviations from strict tempo that they add in expressive musical performance. These deviations—at least so far as experiments have tested them—tend to relate to the local musical terrain: lengthening or shortening tones for metrical stress or at phrase endings, and articulating one tone a scant amount early to make it more prominent.

There have been comparatively few experimental studies of long-term relationships either between temporal spans or between pitch levels in music. The few studies that have been carried out have yielded mixed results. It appears that listeners do segment relatively short melodic excerpts into similar smaller groups, and it appears too that "expert" listeners will transcribe at least the soprano line of a fairly complex Baroque-era chorale quite accurately without having to write each note of the melody in strict sequence. But thus far, experimental evidence indicates that listeners without absolute pitch are quite insensitive to pitch relationships that are more remote in time, such as key levels in a sonata or symphony movement.

Problems

1. Miller (1956) stated that variance could be measured, but the measurement had to be made along some dimension: you measure variance of weight in carats, kilograms, or tons, and you measure variance of heat in calories or degrees Celsius or Fahrenheit, but you cannot mix these dimensions. List several acoustical dimensions of sound, along with some of their units of measurement. List several musical dimensions of sound, again along with some of their units of measurement. Are these dimensions on a comparable level of complexity? Which dimensions are immediately perceivable and which are mediated by learning? Is it feasible to discuss bits of information as units of measurement along all of these dimensions?

2. Elaborating on the number-guessing game that was used to illustrate the concept of bits, what would be the fewest questions (barring a lucky guess) with which you could identify a mystery number in the range 1–1,000? Responses to your guesses would be "higher," "lower," or "correct."

3. Cocktail-party effect studies, and in fact most studies of attention that use auditory (rather than visual) patterns, tend to use verbal stimuli such as spoken tracts of prose or nonsense syllables, although a few use tones that are heard simultaneously with flashes of light or with spoken words or numbers. Try to devise a more musical setting in which to demonstrate the cocktail-party effect.

4. What features of the pitch string CDEGABFED (cf. Figure 9–3a) are most important to you as you look for patterns within it? If you are asked to extend the melody by three notes while preserving the

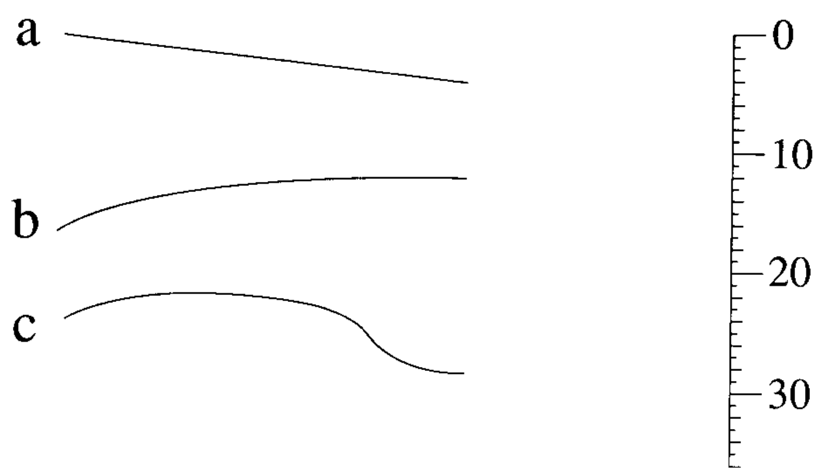

FIGURE 9–5.

pattern within it, what would those notes be? Why? If you were asked to extend the melody by one note, what note would that be? Again, why?

5. To elaborate on the point raised in Problem 4, please examine Figure 9–5. If the straight line a, simple curve b, and complex curve c were extended rightward, they would all arrive at some point on the vertical scale. If you had to extend these lines mentally without using any drawing tools to help you, what would be the numbers of the arrival points on the scale? In which of your mental extrapolations do you have greatest confidence? Is any musical listening activity similar to this visual task? For example, what pitch do you think will be found in the melody in Figure 9–6 in five beats? How much confidence do you have in this extrapolation? Is your guess based on melodic direction? Intervallic pattern and sequence? Rhythmic pattern(s) in the melody? An expectation of what the implied harmony "should" be in five beats? Some combination of these features, or others?

6. As a casual experiment, note how long it takes several people to respond to the question "Are you left-handed?" Then, compare how long it takes those people to respond to the question "How many windows are there on the first floor of the house (apartment, dormitory, etc.) where you now live?" Unless the respondent lives where there are few or no windows, (s)he will probably pause to take a quick mental tour of the place, counting windows. Some memories take more coaxing, depending on mental context. For example, recite the alphabet rapidly, but in reverse order; do not pause until you have reached *a*. Then, rapidly recite the pitch names of the white keys of the piano beginning with G and descending through one octave. Probably, you recited the letters GFEDCBA more quickly as you thought through the scale steps than you did in the less-familiar context of the backward alphabet.
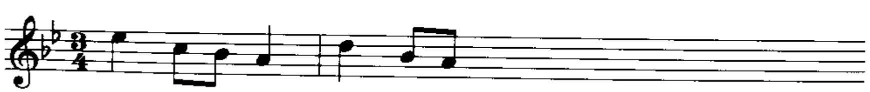
FIGURE 9–6.

7. As we have developed increasingly powerful methods and technologies for storing information over the past several centuries, we have begun to rely less and less on our memories. A few cultures still have oral historians who can perform astounding memory feats as they pass along the stories and wisdom of their people. This was once the norm rather than the exception in all cultures and was once one of the most prized possessions of an individual or culture. One of the best-known methods for developing the memory is now called the method of *loci*.[27] You used this method by making mental attachments between places and images. For example, you might imagine a walk through a garden, a room, or a building. As you walked through, say, a room, you would remember each piece of furniture, each picture and mirror, every memorable feature of the room; you would attach an image of each thing that you wanted to remember to each of these features of the room in the serial order that you walked through it. Then, when you wanted to recall the items, you would take another mental walk through the room and would find each item where you had "left" it earlier.

   Can you think of any type of musical analogue to the method of loci? What about pitch relationships? Do people with absolute pitch unconsciously use something akin to the method of loci?

8. Suppose that you were instructed to design a perceptual measure for tempo. What potentially confounding variables would you have to control?

9. Suppose that you were instructed to design a perceptual measure for meter. First, recall the relationship that may exist between tempo and meter—that a change in tempo may influence the identification of meter. (Consider the similarity between this relationship and the trill-threshold studies that were discussed in Chapter 3.) What other potentially confounding variables would you have to control?

10. Again, each excerpt in Listening Example 9.2 was 15 seconds long. There are several factors, such as tempo, dynamics, rhythmic activ-

ity, and formal structure, that might have contributed to divergence among responses. Can you think of other, less obvious, possibilities? How might you design a test, or a series of tests, that would identify stronger and weaker contributors to listeners' estimates of elapsed time in music? What do you suspect would happen to estimates when listeners were confronted with music with which they were unfamiliar, or that they disliked?

10

The Development of Cognitive Awareness in Music

Milestones

The topic of musical development is probably the most complex to be discussed in this book; at least, it will probably be the chapter that will lead to the most unanswered questions. In fact, we must begin with some very troublesome ones.

First, a few questions about musicality. What is it exactly? Can we define it adequately and then find a reliable objective measure of it? How much of what we call musicality is rooted in perception, and is properly measured by perceptual tests? How much of what we call musicality is appropriately measured instead by performance tests? How much are performance and perception related to one another, so that testing either one separately might be misinforming? Perceptual tests and performance tests can both measure differences in emotional, affective responses as well as cognitive awareness. How are these two aspects of musicality interrelated, and how should our measures of musicality take these aspects into account?

Next, a few questions about development. How much of musicality (or amusicality) is inherited and beyond the influence of music educators? How much of musicality (or amusicality) results from the person's environment and can be altered by formal training? It certainly appears that these are all still wide-open questions. Serafine (1988), for example, says:

> A large part of the musical research on adults and nearly all of that on children was conducted with the use of such tests [as Seashore's