

Chapter 6

Musical Motion in Perception and Performance

Bruno H. Repp

Abstract

A series of experiments has demonstrated that the relative difficulty of detecting a small change in the duration of a single inter-onset interval (IOI) in an otherwise isochronously played musical excerpt is closely related to the relative lengthening of that IOI in a typical expressive performance of the music. These results suggest that musical structures have kinematic implications that not only compel performers to modulate their tempo in certain ways but also induce corresponding perceptual biases in (musically trained) listeners. The perceptual effects may be understood as a form of perceptual-motor interaction. Unlike expressive timing in performance, which is under cognitive control ("interpretation"), the perceptual biases elicited by a piece of music may reflect a precognitive, obligatory response to implied musical motion.

6.1 Introduction

Expressive Timing In Music Performance

The timing of interest in this chapter is not the hierarchically structured regularity of acoustic events that conveys rhythm and meter, that is represented in musical notation, and that has been the primary concern of music theorists and psychologists interested in the temporal structure of music (e.g., Cooper and Meyer, 1960; Povel and Essens, 1985; Jones and Boltz, 1989). Rather, the focus is on systematic *deviations* from this regularity, which are difficult to capture in symbolic notation but are an important aspect of expressiveness in music performance (see Seashore, 1938; Epstein, 1995). Without these deviations, a performance sounds mechanical and artless. At least this is so for music that is intended to be played with expressive timing, the prime example of which is European art music from the nineteenth century.

The variability that constitutes expressive timing is far from random. To be sure, a random component exists, reflecting the limited precision of musicians' motor control as well as, perhaps, uncontrolled variability in their expressive intentions. Typically, however, most of the timing variance among nominally equal inter-onset intervals (IOIs) is due to a musician's stable intentions, as is evident in the high replicability of

timing patterns across repeated performances of the same music (Palmer, 1989; Repp, 1995b).¹ These intentions with regard to timing are rarely explicit and conscious, however; rather, they emerge from a musician's conception of the structure and character of a particular composition (Palmer, 1989; Shaffer, 1995) in combination with acquired (sometimes idiosyncratic) strategies for communicating this conception to listeners.

The precise timing of a performance can be measured—laboriously, and with some margin of error—from the digitized waveform of a sound recording (see, e.g., Gabriellson, 1987; Repp, 1992b), or it can be obtained more quickly and accurately via MIDI (Musical Instrument Digital Interface) if the music was played on a computer-controlled instrument, typically a piano (see Palmer, 1989; Repp, 1995b). One way of representing the timing information is to plot normalized interonset intervals (IOIs) as a function of metrical (score) position.² The resulting graph constitutes the timing pattern or *timing profile* of a performance.

Different artists' performances of the same music have different timing profiles. For the purpose of the present research, it was desirable to obtain a representative or *typical* timing profile (TTP) of the music used in a particular experiment. This requires analysis of a number of performances by different artists. Their average timing profile provides a reasonable estimate of what is typical of these performances, especially if individual differences are not very large, as is often the case with groups of young artists (Repp, 1995b, 1997).

The shape of the TTP for a given piece of music is a complex function of the compositional structure, especially of rhythmic grouping (Lerdahl and Jackendoff, 1983) but also of melodic contour and harmonic progression. These structural factors generate implications—tension and relaxation, continuation and closure—that seem to propel the music forward in time with varying degrees of urgency, and therefore their total effect is often referred to as *musical motion* (see Epstein, 1995; Shove and Repp, 1995). Musicians realize musical motion through physical movement on instruments, and the TTP therefore represents instantiated musical motion—the actual timing of musical events. Several researchers, most notably Todd (1985, 1995), have attempted to devise algorithms that lead from musical structure to typical expressive timing. This will no doubt be an area of intensive investigation in the near future, eventually leading to the successful computer synthesis of expressive music performances.

An example of a TTP is shown in figure 6.1. The musical excerpt is the initial phrase of Chopin's *Etude in E-major*, op. 10, No. 3, terminated with a chord to provide closure. The TTP represents the average IOIs of performances by nine graduate student pianists, each of whom played the excerpt three times on a digital piano (Repp, 1997, in press, a). The initial

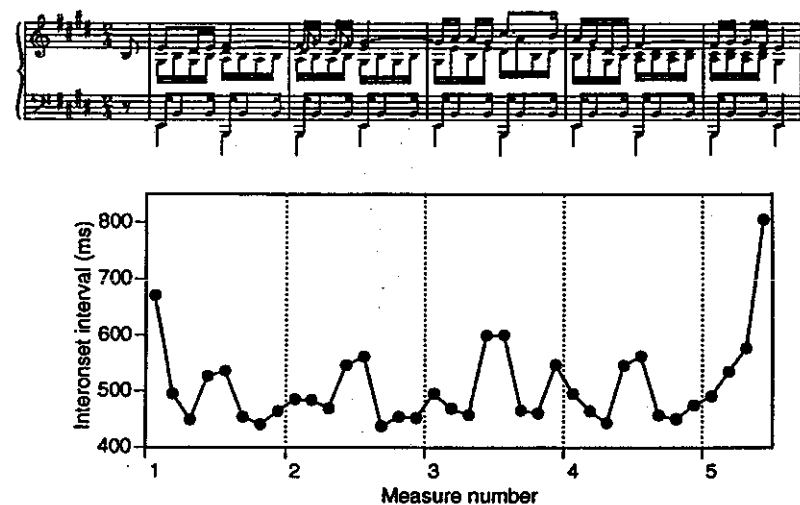


Figure 6.1
A typical expressive timing profile for the initial phrase of Chopin's *Etude in E major*, op. 10, No. 3.

eighth-note upbeat is not included in the graph. All other IOIs represent sixteenth-note intervals, measured (via MIDI) between the onsets of the highest notes in successive positions. There seems to be a "floor" at about 450 ms, corresponding to a basic tempo of about 133 events (or 33 quarter-note beats) per minute, and the longer IOIs constitute expressive lengthenings with respect to this baseline. The initial and final IOIs are lengthened dramatically, marking the boundaries of the phrase. A gradual slowing towards the end of the phrase can be seen in measure 5. Inside the phrase there are four double peaks and one single peak. Each double peak marks the end of a melodic segment: The pianists lengthened the final IOI as well as the IOI between the final long melody note (in the soprano voice) and the following accompanying note (in the alto voice). The single peak at the end of measure 3 emphasizes the anacrusis (upbeat) unique to the fourth melodic segment. This TTP will serve as a reference in the perceptual experiment to be described.

Detecting Deviations from Isochrony

At first blush, the task of detecting small perturbations in the timing of a regular stream of acoustic events seems to have little to do with music perception. It is a psychophysical paradigm for the determination of temporal discrimination thresholds and for the testing of theories about time-keeping mechanisms in the human brain (see, e.g., Schulze, 1978; Hibi,

1983; Hirsh et al., 1990; Monahan and Hirsh, 1990). However, some previous research and informal observations have suggested that the paradigm can be used to reveal more than just a listener's auditory acuity.

Several studies have shown that an increment in one inter-stimulus interval in a regular sequence of tones separated by silent intervals is more difficult to detect at a structural boundary than within a structural unit, both for adults (Fitzgibbons, Pollatsek, and Thomas, 1974) and for infants (Thorpe et al., 1988; Thorpe and Trehub, 1989). In these experiments, the boundary was defined by a change in pitch, so the results are consistent with the phenomenon of perceptual segregation as a function of pitch difference (Noorden, 1975; Bregman, 1990) and with the well-established psychophysical finding that the gap detection threshold increases with the spectral difference between the delimiting sounds (e.g., Perrott and Williams, 1971; Williams and Perrott, 1972; Collyer, 1974; Divenyi and Danner, 1977; Neff, Jesteadt, and Brown, 1982; Formby and Forrest, 1991). In other words, these findings may reflect an elementary principle of auditory organization: Time is more accurately perceived within than between auditory channels or streams.

Similar results have been obtained with more complex materials. Krumhansl and Jusczyk (1990) and Jusczyk and Krumhansl (1993) demonstrated that infants prefer to hear pauses between rather than within phrases of Mozart minuets (as adults undoubtedly would), and analogous findings have been reported for passages of speech containing prosodically marked phrase boundaries (Hirsh-Pasek et al., 1987; Jusczyk et al., 1992). Short pauses are also more difficult to detect at these points (Boomer and Dittman, 1962; Butcher, 1980). A slowing down of the event rate and a lowering of the pitch contour are common correlates of phrase boundaries in both speech and music. These findings, too, could be explained by principles of auditory grouping and segregation.

Not all boundaries are equal in speech and music. Butcher (1980) reported that the relative detectability of boundary pauses in speech was inversely related to the relative depth of the boundary in the hierarchical phrase structure. This observation parallels the finding of Gee and Grosjean (1983) that pause durations in controlled speaking tend to be proportional to boundary depth. In music, too, there is a tendency for larger *ritardandi* in expressive timing to be associated with deeper structural boundaries (Todd, 1985), so one might predict that there will be a corresponding perceptual phenomenon: Hesitations may be more difficult to detect at deep boundaries than at shallow boundaries, and more difficult at shallow boundaries than within melodic units. This prediction was confirmed by Repp (1992c) in the initial experiments of what turned out to be a long series. These experiments went further by probing not just the boundaries but the *entire time course* of a musical structure, to determine whether per-

ception—measured in terms of the relative detectability of hesitations—is modulated as quasi-continuously as is the timing of an expressive performance. This research demonstrated that the *detection accuracy profile* (DAP, percent correct as a function of score position) for IOI increments is inversely related to the TTP. A small increment is the more difficult to detect the longer the IOI tends to be in expressive performance.

This step from a focus on boundaries to the quasi-continuous probing of a whole temporal shape shifted attention from acoustic and auditory determinants of detectability to perception-performance relationships.³ Although lower detectability of hesitations at major boundaries may have more or less obvious psychoacoustic causes, it is less clear whether the acoustic surface structure of music or principles of auditory processing can explain the whole DAP. At the time of writing, the author has completed 19 experiments of similar design that explore this perception-performance parallelism and its possible causes. Here there is space to describe only one of them. The tenth in the series, it occupies a central position and has served as a baseline for all following experiments. (For details omitted here, see Repp, in press, a Exp. 1.)

6.2 Experiment

Method

The music was the Chopin excerpt shown in figure 6.1. Apart from the initial upbeat, which was included but never served as a detection target, the music contains 36 sixteenth-note IOIs. It was synthesized on a Roland RD-250s digital piano under computer control with a fixed baseline IOI duration of 500 ms. All simultaneous tones had synchronous onsets, and successive tones followed each other without any intervening silence, resulting in *legato* articulation throughout. The pedal was not used. The relative intensities (MIDI velocities) of the tones were copied from an expressive performance on the Roland.

The task required the detection of deviations from isochrony. The musical excerpt was presented repeatedly, and each presentation contained between 0 and 4 IOIs (separated by at least four baseline IOIs) whose duration had been changed. Increment and decrement detection were tested in separate parts of the experiment. Each part consisted of 5 familiarization trials and three test blocks of 18 trials each, in the course of which each of the 36 IOIs served as a target once. The amount of duration change decreased across the three blocks, from 42 to 31 to 21 ms. The change was made in the MIDI instructions to the digital piano by lengthening or shortening the note(s) filling the target IOI and delaying or advancing the onsets of all following notes. This amounted to a rhythmic

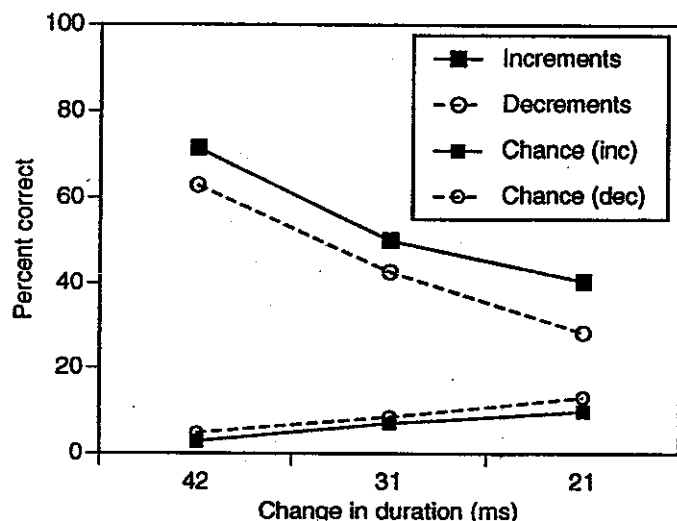


Figure 6.2
Percent correct detection ("hits") as a function of change in IOI duration.

phase shift, perceived as a momentary hesitation or acceleration, with *legato* articulation held constant.

The subjects were 14 musically trained college students. They marked their responses by circling notes in copies of the musical score on prepared answer sheets. Circling of a note adjacent to the correct one was considered a correct response, as such near misses were quite frequent. Other responses were considered false alarms.

Results

Figure 6.2 shows the average percent correct scores ("hits") as a function of the change in IOI duration. Chance level estimates derived from the false alarm rates are shown as well. It can be seen that detection scores declined across blocks but were still better than chance in the most difficult condition. This level of performance was optimal for avoiding floor and ceiling effects in the DAPs. Increment detection was significantly easier than decrement detection [$F(1, 13) = 12.57, p < .004$].

Figure 6.3a shows the DAPs for increment and decrement detection, respectively, averaged across subjects and blocks. It is evident that there was tremendous variation in detection scores as a function of position in the music. The two DAPs are negatively correlated ($r = -0.53, p < .01$). On the whole, decrements were easier to detect where increments were more difficult to detect. As predicted, the DAP for increments is inversely

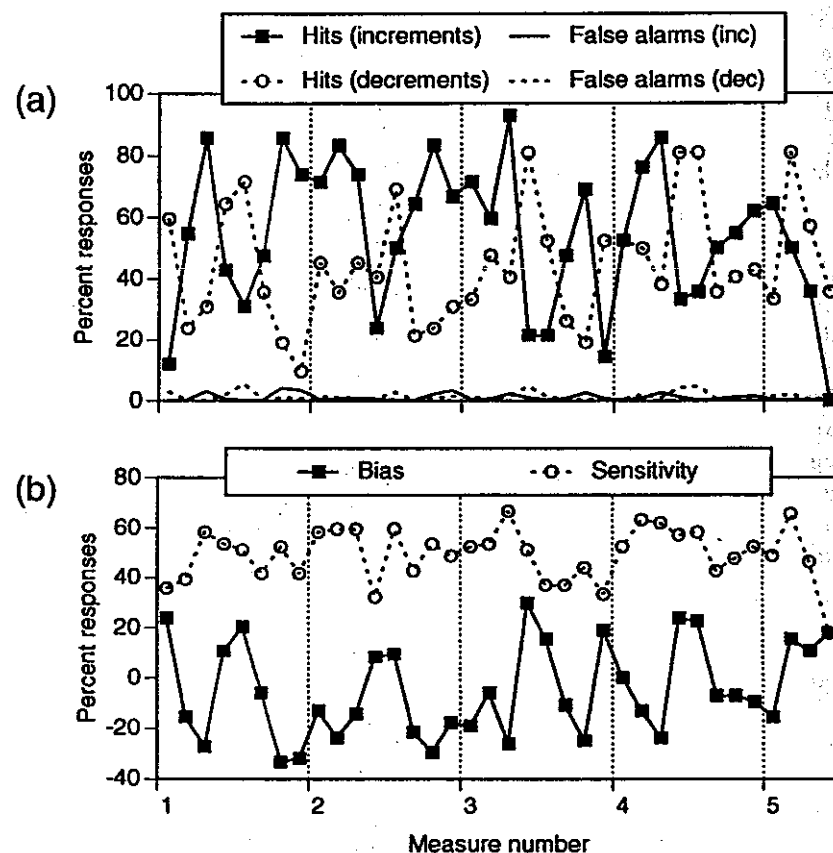


Figure 6.3
(a) The DAPs for increments and decrements; (b) the derived sensitivity and bias profiles.

related to the TTP shown in figure 6.1 ($r = -0.82, p < .001$), whereas the DAP for decrements shows a positive correlation with the TTP ($r = 0.70, p < .001$).⁴ These findings indicate that it is more difficult to detect a duration increment, but easier to detect a duration decrement, in positions that typically exhibit expressive lengthening in performance. Conversely, increments are easier to detect and decrements harder to detect in positions where pianists tend to speed up.

The inclusion of both increment and decrement detection tasks made it possible to distinguish between two hypothetical factors affecting detection accuracy, namely *sensitivity* and *bias*. These concepts are familiar from signal detection theory (see, e.g., Macmillan and Creelman, 1991), but d'

and beta indices could not be computed in the present paradigm because the task did not require a forced choice and because false-alarm responses were too sparsely distributed (see figure 6.3a) to yield reliable probability estimates. The following simple logic seemed applicable: Differences in sensitivity to temporal change across positions in the music should have a similar effect on increment and decrement detection; therefore, the average of the two DAPs yields an estimate of positional variations in sensitivity, a "sensitivity profile." Differences in bias, on the other hand, should affect increment and decrement detection in opposite ways; therefore, the difference between the two DAPs yields an estimate of positional variations in bias, a "bias profile."⁵ Of these two profiles, it is the bias profile that should be correlated with the TTP, whereas the sensitivity profile should exhibit no such relationship. The two profiles are shown in figure 6.3b. The bias profile, computed by subtracting the increment DAP from the decrement DAP and dividing by two, indeed shows a remarkably close match with the TTP in figure 6.1 ($r = 0.85, p < .001$), whereas the sensitivity profile is unrelated to the TTP. These results provide strong evidence for a connection between perception and production of timing in music.

6.3 Discussion

Representativeness of the Data Shown

The overall accuracy results shown in figure 6.2 are typical for this type of task. Somewhat higher accuracy was obtained in unpublished experiments that used monophonic tunes, whereas accuracy was lower with music that contained some IOIs of longer duration (Repp, 1992c: Exp. 2). Elimination of variation in expressive dynamics makes duration differences somewhat easier to detect (Repp, 1992a; in press, a). Naturally, the task becomes much more difficult (but not impossible) when deviations from an expressively timed baseline are to be detected (Clarke, 1989; Repp, in press, b). Subjects without musical training often have difficulty perceiving small timing deviations (Repp, 1995a); therefore, musically trained subjects were used in most experiments. A reliable overall advantage of increment over decrement detection is not always observed, and there are also considerable individual differences in that regard. When the baseline contains variable IOIs, decrement detection is easier overall than increment detection (Repp, in press, b).

The early experiments in this series included only an increment detection task, whereas most later experiments tested both increment and decrement detection. A significant negative correlation between the DAP for increments and some estimate of the TTP has always been obtained:

with different original excerpts from the piano literature (Repp, 1992c), with simplified versions of the Chopin etude excerpt (Repp, in press, a), even with monophonic experimental tunes (Repp, 1992a; 1995a). The monophonic materials yielded DAPs and TTPs with little internal structure, however, so that the perception-performance relationship rested mainly on the final *ritardando* in performance and the corresponding decline in detection scores towards the end of the tune. In order to demonstrate a convincing perception-performance correlation, it is necessary to use music that elicits a finely differentiated timing profile from musicians. This requires a certain degree of rhythmic and harmonic complexity, as well as a tempo, character, and style that encourage expressive tempo modulations.

The inverse relation between the DAPs for increments and decrements and the positive correlation between the DAP for decrements and the TTP are *not* reliable findings. Most subsequent experiments failed to replicate these correlations (Repp, in press, a; in press, b). The bias profile calculated from the two DAPs, however, invariably did correlate with the TTP. There are two ways of interpreting these results. One is that the two hypothetical underlying factors, sensitivity and bias, reinforce each other in the DAP for increments but tend to cancel each other in the DAP for decrements. This is certainly true with respect to the final IOIs, where an increase in bias accompanies a decline in sensitivity (see figure 6.3b). From this perspective, the original DAPs are best ignored and only the derived sensitivity and bias profiles are to be considered. The other possibility is that the detection of decrements is indeed less related to music performance than is the detection of increments. This suggestion has some plausibility in view of the fact that expressive timing is not a symmetric process. Expressive lengthening prolongs tension and conveys emphasis, whereas relative shortening has no particular expressive function and may simply represent a return to some baseline tempo or compensate for lengthening in the vicinity. Listeners' expectations about expressive timing (i.e., their perceptual biases) thus may be formulated in terms of increments rather than decrements. This interpretation has found support in two recent experiments (Repp, in press, b): When the baseline is expressively modulated by the full or attenuated TTP, so that listeners' expectations are satisfied, as it were, the perceptual bias is neutralized in increment but not in decrement detection. Decrements are easy to detect in short IOIs and difficult to detect in long IOIs, even when the changes are proportional to IOI duration.

In theory, it would seem that false-alarm responses should be a direct reflection of perceptual bias, namely the tendency to perceive certain IOIs as longer or shorter than they really are. The frequency of these responses varies dramatically from subject to subject, however, and many subjects

give only very few false-alarm responses. Moreover, as already pointed out, these responses are distributed over many positions, which makes the resulting false-alarm profile (FAP) somewhat haphazard and unreliable. In most experiments the FAP correlated positively with the corresponding DAP, but the correlation did not always reach significance. Similarly, the negative correlation between the FAP for increments and the TTP was significant only in some experiments. The FAP results thus are consistent with the DAP results, but they are far less robust.

Possible Psychoacoustic Factors

Music, even a simple monophonic tune played on the piano, is a rather complex stimulus from the perspective of research in psychoacoustics and auditory scene analysis (Bregman, 1990). It is possible, even likely, that the detectability of changes in IOI duration varies as a function of acoustic surface properties of the music. Even if the shapes of the DAPs (and/or of the sensitivity and bias profiles) could be explained entirely on the basis of auditory processing phenomena, however, the question would remain why there is a relationship between perception and expressive timing in performance. The answer would have to be that performers are subject to the same auditory effects as ordinary listeners, and that expressive timing is a sort of compensatory action to smooth out perceived rhythmic irregularities caused by these effects (Drake, 1993). This seems rather implausible, but let us see first whether the variation in detection scores can be rationalized on a psychoacoustic basis.

Psychoacoustic effects on sensitivity and on bias should be distinguished first. The former affect detectability regardless of the direction of the deviation, whereas the latter are directional; they can also be described as a constant error. Effects on sensitivity are not of particular interest here because they are irrelevant to the perception-performance relationship, which, as we have seen, rests on the variation in perceptual bias. Examples of sensitivity effects are the reduced detection scores in the initial and final positions of a musical excerpt.⁶

The crucial question is whether the positional variations in bias can be accounted for on psychoacoustic grounds. One relevant phenomenon is the "auditory kappa effect" (Shigeno, 1986; Crowder and Neath, 1995). When three successive tones are separated by unequal pitch distances but equal temporal intervals, the temporal interval between the two tones closer in pitch seems to be shorter than that between the two tones farther apart in pitch. In the present experiments, however, there was no evidence for a positive relation between bias and the absolute or directional pitch distance between successive tones. Thus, the auditory kappa effect did not seem to operate, either because of the greater complexity of the stimuli or perhaps because there were no silent intervals between

tones. Instead, significant correlations were found between bias and the absolute pitch height of the highest tone occupying the target IOI (the target tone), and between bias and the directional intensity (MIDI velocity) difference between the two highest tones delimiting the IOI (Repp, in press, a).⁷ In other words, an IOI tended to be heard as relatively *short* when the target tone was high in pitch or when the following tone was more intense than the target tone. These findings do not correspond to known auditory phenomena such as the kappa effect. They could have been due to the musical structure and its implications for performance timing, rather than to some independent auditory interactions between duration, pitch, and intensity.

Of course, it is difficult to separate the acoustic surface structure of music from the underlying structure described in music-theoretic terms, although a manipulation performed on the Chopin excerpt proved enlightening (Repp, in press, a: Exp. 2). In that study, the long melody notes (see figure 6.1) were broken up into repeated sixteenth notes of the same pitch and intensity. This enabled listeners (in theory) to detect deviations from isochrony by tracking the melody tones only, instead of jumping back and forth between melody and accompaniment. In fact, however, the manipulation had little effect on the results, which were quite similar to those shown in figure 6.3. Within the sixteenth notes of the melody, there was no correlation between bias and either pitch height or intensity difference. This finding suggests that the listeners tracked the whole musical structure, not just individual tones in one voice. Admittedly, there may be other incarnations of a bottom-up auditory processing hypothesis, for example one that takes the full spectral content and energy distribution of each chord into account and that models temporal integration (see, e.g., Todd, 1994). For the time being, however, an explanation in terms of structural and musically relevant concepts seems more promising.

Musical Motion and Structure

It seems only parsimonious to couch an explanation of the perceptual bias effects in the same terms as an explanation of expressive timing in music performance. Despite ubiquitous individual differences, expressive timing is governed by certain principles and constraints; and listeners evidently possess implicit knowledge of these principles, which leads them to expect lengthening of IOIs where performers typically linger and shortening of IOIs where performers typically rush. When music is played without any expressive timing, as in the present experiments, IOIs that are expected to be long sound too short (a positive bias, as bias is defined here), whereas IOIs that are expected to be short sound too long (a negative bias). These biases seem to be elicited automatically by the same structural properties that cause performers to modulate their timing.

What are these structural properties? The determinants of expressive timing are not fully understood at present, and a detailed discussion of this complex topic cannot be provided here. In the present Chopin Etude excerpt (figure 6.1), however, the major factor clearly is the segmentation of the melody: Each group of notes is associated with a final *ritardando*, and the final group has extra lengthening. In addition, there is initial lengthening in the first and fourth groups. This is consistent with an archetypal acceleration-deceleration shape within groups, as described in Todd's (1985) model of expressive timing. Todd (1995) has also argued that the precise shape of the timing curve at a coarser grain of analysis is subject to a constraint of linear tempo change, in analogy to forms of physical or biological motion. It would be difficult to apply this idea at the level of detail considered here. The precise timing of small IOIs seems to be modulated by factors in addition to group boundaries, such as melodic contour, harmonic dissonance, and metrical position. Thus, the fact that there is lengthening at the onset of the fourth group (position 3-8) but not at the onsets of the second and third groups, may be due to metrical position (immediately preceding a strong beat) and/or to the greater length of the fourth melodic gesture. The lengthened IOI in the fifth position of each measure may not represent group-final lengthening—after all, it *follows* the final note—but rather may serve to segregate the melody from the accompaniment, which otherwise might be heard as a continuation of the melody with lowered dynamics. At points of melodic re-entry, no lengthening is observed because the melody can hardly be mistaken for an accompaniment. The extent of mid-measure lengthening can be seen to increase from measure 1 to 2 to 3 and to decrease in measure 4; this trend follows the melodic pitch contour and the correlated changes in tension and dynamics.

It is not necessary to have a complete theory of expressive timing to appreciate the parallelism between perception and performance demonstrated in the present research, just as musicians and listeners need not have explicit knowledge of the principles that underly their behavior. The TTP represents the sum and interaction of all structural factors that generate musical motion, as filtered through a typical musician's mind and body. As an empirical measure of musical motion, it is superior to structural descriptions that may include many features that are not communicated (or not communicable) in the typical performer's movements.⁸ It is only a small step from this argument to the claim that the (typical or average) perceptual bias profile, too, is a measure of communicatively relevant musical motion. The high correlation between the TTP and the bias profile suggests that whatever moves the typical musician also moves the typical listener.

Individual Differences

The perceptual bias profile is interesting because it is in some sense an even more direct measure of musical motion than the TTP. It is obtained in a detection task and thus presumably excludes any conscious decision making on the listener's part. In other words, it is an obligatory and automatic response to the music. Musicians, on the other hand, can modify, suppress, and even reverse their kinematic strategies in response to music in order to achieve variety and originality. A listener in the present type of experiment does not seem to have that choice.

In part, musicians' individual differences in expressive timing are not cognitive but organismic in origin; they represent their different body structures, personalities, technical training, and musical experience. The extent and nature of individual differences in perceptual bias profiles are not known. The present experiments never yielded enough responses from single individuals to determine reliable individual profiles that could be compared with each other. Effects of musical experience were examined correlationally in some early experiments (Repp, 1992c), without clear results. Repp (1995a) compared three groups of subjects with different degrees of musical training, but the shapes of their average DAPs for increments did not differ. Thus, there is no evidence yet for significant individual differences in perceptual profiles, but this is an issue that deserves further investigation. It would be particularly interesting to find a relationship between individual differences in perceptual bias and in performance timing among pianist subjects, but such an investigation is difficult, not only because of the large number of sessions required but also because young pianists tend to have rather similar timing profiles (Repp, 1995b, 1997).

If there is indeed less individual variation in the perceptual bias profile than in the timing profile, this has interesting implications for the aesthetics and evaluation of musical performance. The TTP, by virtue of its similarity to the bias profile, may then be understood as a precognitive, pre-interpretative measure of musical motion. This does not mean that a performance whose timing matches this profile is better than others. It may well be perceived as the most naturally timed performance, however, and it may be preferred under certain circumstances. The duality of typicality and individuality with respect to performance aesthetics is discussed in more detail, with preliminary data, in Repp (1997).

Mechanisms of Timing Perception

The detection of deviations from isochrony requires the tracking and prediction of regularly timed events by means of some internal time-keeping mechanism. This could be a mental clock or oscillator, as postulated by many theorists (e.g., Schulze, 1978; Povel and Essens, 1985; Jones and

Boltz, 1989; Desain, 1992; Large and Kolen, 1994), or it could be a memory for interval durations (Keele et al., 1989; Ivry and Hazeltine, 1995). One theoretical question of interest is whether the predictions of this internal time-keeper are directly modulated by the perceptual bias profile or whether the bias has its effect at some later decision stage.

Clearly, a listener could not simply expect events to occur at intervals that match the TTP. This would make it impossible to detect deviations from isochrony with any degree of accuracy. If the internal time-keeper is modulated by expectations, then these modulations must be much smaller than the deviations observed in the TTP. Even so, one would expect accuracy to be lower and false-alarm rates to be higher than they actually were. In fact, many listeners hardly gave any false-alarm responses, and accuracy was not much lower than in a psychophysical discrimination task. This suggests that the internal time-keeper was unmodulated, and that the bias had its effect at a subsequent stage that determined whether or not deviations from isochrony gained access to consciousness.

What kind of timing mechanism do the subjects employ to perform the task? One finding of the present experiments—not discussed so far—may provide a clue. It is the frequent occurrence of “late” responses in the increment detection task, that is, the circling of the note or chord following the correct position in the score. Typically, about one third of all correct responses were late in this sense, whereas “early” responses were far less frequent. Some of the late responses may have been due to subjects’ failure to backtrack in the score; after all, they had to wait for the tone or chord ending an IOI in order to determine whether the IOI was longer or shorter than usual. If that had been the sole cause, however, late responses should have been equally frequent in increment and decrement detection tasks. It turned out, however, that they were much less common in decrement than in increment detection (Repp, in press, a). Thus, it seems that the majority of late responses in the increment detection task had a perceptual cause: The IOI following a lengthened IOI tended to be perceived as lengthened, at least if the actual lengthening was not detected.⁹ Moreover, late responses tended to be relatively more frequent when the following IOI was one in which actual lengthening was easy to detect; thus they were subject to the same perceptual bias that shaped the DAP, which also argues for a perceptual origin of the phenomenon. Late responses in the decrement task not only were much less frequent but also did not show any systematic bias effects (Repp, in press, a).

These results seem more compatible with an oscillatory time-keeping mechanism than with an interval-based memory. An oscillator is perturbed by IOI duration increments or decrements, which amount to phase shifts in the isochronous rhythm. Adjustment to such phase shifts is probably not instantaneous but takes several cycles (Large and Kolen, 1994). When

a target IOI is lengthened, the event that terminates it occurs immediately after a beat (tick, prediction) issued by the oscillator. As a result, the next beat will have to be delayed by lengthening the period of the oscillatory cycle, but unless this adjustment is complete, the next beat will again fall short of the event marking the end of the next IOI, hence the tendency to also perceive it as lengthened. By the same reasoning, one might predict that an IOI following a shortened target IOI should be perceived as shortened, but this did not happen. The reason may lie in the fact that the event marking the end of the target IOI occurs shortly *before* the oscillator issues its beat. This may have the effect of precipitating the beat, thereby leading to an immediate phase shift (assuming that the period of the oscillator remains constant). The difference in the frequency of late responses in increment and decrement detection thus can be explained by the directionality and irreversibility of time.

Unlike an oscillatory process, an interval-based memory need not adjust to a single mismatched IOI. The best strategy would be to hold on to the memory of the baseline interval. A modified target IOI may perturb that memory slightly, but there is no particular reason why the following IOI should be perceived differently (if anything, a contrast effect might be predicted), or why there should be a difference between increment and decrement detection. The present data thus favor an explanation in terms of an oscillatory time-keeping mechanism. It must be added, however, that individual differences in the percentage of late responses were very large. Some subjects gave hardly any late responses, while others gave more than 50%. Thus, individuals may differ in their strategies, as both timing processes seem to be available in principle (Keele et al., 1989).

Whichever mechanism is used in detecting deviations from isochrony, it seems to be subject to the performance-related biases. There is no indication so far that any listener is immune to this bias. It is tempting to speculate that it represents a form of perceptual-motor interaction (Viviani and Stucchi, 1992b).

Viviani and Stucchi (1992a)—and long before them, Derwort (1938)—have shown that a uniform movement of a light point along a path varying in curvature is perceived as nonuniform in velocity (too fast around the narrow curves, too slow along the straight portions), whereas a nonuniform movement varying lawfully with curvature is perceived as constant. Musical motion may be subject to similar laws that can be demonstrated by translating expressive timing into spatial movement (Truslit, 1938; Repp, 1993). The structural properties of music imply nonuniform timing, so that a compensatory effect is observed in perception when the actual timing is uniform. This mechanism may be analogous to that in vision, except that the compensation is not complete, only a tendency. Certainly, no careful listener would judge the TTP of the Chopin

etude (figure 6.1) to be uniform, but a more subtle timing modulation may be perceived in that way (Repp, in press, b). Similarly, there is only weak evidence from false-alarm responses that an isochronous performance sounds nonuniform to listeners. If there is a perceptual-motor interaction here, it is less direct than in vision because it does not actually affect the time-keeping mechanism; the bias must be operating at a subsequent stage in processing.

Toward the Future

Three recent experiments (Repp, in press, c) have provided new information about the nature of the perceptual bias. One experiment showed that repeated exposure to an *atypically* timed performance leaves the bias profile quite unaffected. A second experiment showed that there is no systematic bias when the task is to detect deviations from isochrony in a sequence of clicks while *imagining* the Chopin excerpt and marking the responses in the score. The third experiment did obtain the typical bias profile when the clicks were superimposed on the music, even though the subjects had been instructed to ignore the music and pressed a button rather than marking responses in a musical score. These results show that the bias is contingent on the auditory processing of a musical structure and that it is resistant to modification.

An important question for future research is whether the bias reflects familiarity with classical music and/or expressive performance. To the extent that the perception-performance relationship demonstrated here rests on processes of auditory segmentation and grouping, it may well be independent of specific musical experience (Deliège and El Ahmadi, 1990). This would imply that the TTP, too, is largely the reflection of elementary grouping processes. In other words, there may be a close connection between grouping and timing, as some of the studies cited in the introduction have already suggested. This would be a rather straightforward conclusion, but it is probably not the whole story.

Acknowledgments

This research was supported by NIH grant MH-51230. Thanks are due to Charles Nichols, Linda Popovic, and Lisa Robinson for their assistance and to Steven Boker, Geoffrey Collier, Carol Krumhansl, and David Rosenbaum for helpful comments on the manuscript.

Notes

1. Some of this replicable timing variation may be due to technical difficulties (e.g., extra time required for rapid changes in hand position), although most investigations of expressive timing have focused on slow music in which such technical factors are minimized.

2. Other possibilities are percentage deviations from the average IOI (Gabrielsson, 1987; Palmer, 1989), local tempo estimates (i.e., reciprocal IOIs), or event onset times as a function of real time (see Todd, 1995).
3. A study by Clarke (1989) proved very stimulating. Although his main purpose was to determine the detection threshold for small timing deviations in a few positions of a musical excerpt, selected without particular reference to musical structure, his discussion made extensive reference to expressive performance and inspired the author to undertake his initial experiments (Repp, 1992c).
4. The final IOI was omitted in these correlations because of its extreme length in the TTP and the corresponding floor effect in the DAP for increments.
5. The bias of interest here is a *perceptual bias* (i.e., a constant error), not a response bias. It refers to listeners' tendency to perceive some IOIs a priori as shorter (longer) than others. It is defined as the extent to which an increment is more difficult to detect than a decrement in a given position.
6. The initial IOI is either the first IOI encountered or is shorter than the preceding IOI, so that its deviation from isochrony can be judged only retrospectively, with reference to a fraction of the preceding IOI, or with the aid of memory for a previous presentation of the musical excerpt. Similarly, the final IOI is followed by a longer tone or chord of vague duration (since no event onset follows it) and thus can be judged only relative to preceding IOIs (see also Monahan and Hirsh, 1990). Another factor that may affect sensitivity to changes in IOI duration is the pitch distance between two successive tones, which is known to affect the psychoacoustic gap detection threshold and duration discrimination at auditory group boundaries, as mentioned earlier in this chapter. No evidence for such an effect was obtained (Repp, in press, a), however, probably because of the relatively long duration of the IOIs and because they were filled with decaying sound rather than with silence. One factor that was found to have an effect on sensitivity was the relative intensity of the tones delimiting the IOI. Higher intensity led to higher detection scores (Repp, in press, a).
7. Whenever several tones occurred simultaneously as a chord, it was assumed that the highest tone was perceptually most salient because of its melodic function and higher intensity.
8. It should be kept in mind, however, that timing is only one aspect of expressive performance. Important additional information is conveyed by dynamics and articulation.
9. Circling of two adjacent positions in the music was extremely rare; however, the subjects knew that two modified IOIs could not be adjacent to each other.

References

- Boomer, D. S., & Dittman, A. T. (1962). Hesitation pauses and juncture pauses in speech. *Language and Speech*, 5, 215-220.
- Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.
- Butcher, A. (1980). Pause and syntactic structure. In H. W. Dechert & M. Raupach (eds.), *Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler* (pp. 85-90). The Hague: Mouton.
- Clarke, E. F. (1989). The perception of expressive timing in music. *Psychological Research*, 51, 2-9.
- Collyer, C. E. (1974). The detection of a temporal gap between two disparate stimuli. *Perception & Psychophysics*, 16, 96-100.
- Cooper, G., & Meyer, L. B. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.
- Crowder, R. G., & Neath, I. (1995). The influence of pitch on time perception in short melodies. *Music Perception*, 12, 379-386.

- Delège, I., & El Ahmadi, A. (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.
- Derwort, A. (1938). Untersuchungen über den Zeitablauf figurierter Bewegungen beim Menschen. *Pflügers Archiv für die gesamte Physiologie des Menschen und der Tiere*, 240, 661–675.
- Desain, P. (1992). A (de)composable theory of rhythm perception. *Music Perception*, 9, 439–454.
- Divenyi, P. L., & Danner, W. F. (1977). Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra. *Perception & Psychophysics*, 21, 125–142.
- Drake, C. (1993). Perceptual and performed accents in musical sequences. *Bulletin of the Psychonomic Society*, 31, 107–110.
- Epstein, D. (1995). *Shaping time: Music, the brain, and performance*. New York: Schirmer Books.
- Fitzgibbons, P. J., Pollatsek, A., & Thomas, I. B. (1974). Detection of temporal gaps within and between perceptual tonal groups. *Perception & Psychophysics*, 16, 522–528.
- Formby, C., & Forrest, T. G. (1991). Detection of silent temporal gaps in sinusoidal maskers. *Journal of the Acoustical Society of America*, 89, 830–837.
- Gabrielsson, A. (1987). Once again: The theme from Mozart's Piano Sonata in A major (K.331). In A. Gabrielsson (ed.), *Action and perception in rhythm and music* (pp. 81–103). Stockholm: Royal Swedish Academy of Music Publication No. 55.
- Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411–458.
- Hibi, S. (1983). Rhythm perception in repetitive sound sequence. *Journal of the Acoustical Society of Japan* (E), 4, 83–95.
- Hirsh, I. J., Monahan, C. B., Grant, K. W., & Singh, P. G. (1990). Studies in auditory timing: 1. Simple patterns. *Perception & Psychophysics*, 47, 215–226.
- Hirsh-Pasek, K., Kemler Nelson, D. G., Jusczyk, P. W., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26, 269–286.
- Ivry, R. B., & Hazeltine, R. E. (1995). Perception and production of temporal intervals across a range of durations: Evidence for a common timing mechanism. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 3–18.
- Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96, 459–491.
- Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24, 252–293.
- Jusczyk, P. W., & Krumhansl, C. L. (1993). Pitch and rhythmic patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 627–640.
- Keele, S. W., Nicoletti, R., Ivry, R. L., & Pokorny, R. A. (1989). Mechanisms of perceptual timing: Beat-based or interval-based judgments? *Psychological Research*, 50, 251–256.
- Krumhansl, C. L., & Jusczyk, P. W. (1990). Infants' perception of phrase structure in music. *Psychological Science*, 1, 70–73.
- Large, E. W., & Kolen, J. K. (1994). Resonance and the perception of musical meter. *Connection Science*, 6, 177–208.
- Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- Monahan, C. B., & Hirsh, I. J. (1990). Studies in auditory timing: 2. Rhythm patterns. *Perception & Psychophysics*, 47, 227–242.
- Neff, D. L., Jesteadt, W., & Brown, E. L. (1982). The relation between gap discrimination and auditory stream segregation. *Perception & Psychophysics*, 31, 493–501.
- Noorden, L. P. A. S. van (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology, The Netherlands.
- Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 331–346.
- Perrott, D. R., & Williams, K. N. (1971). Auditory temporal resolution: Gap detection as a function of interpulse frequency disparity. *Psychonomic Science*, 25, 73–74.
- Povel, D. -J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 2, 411–440.
- Repp, B. H. (1992a). Detectability of rhythmic perturbations in musical contexts: Bottom-up versus top-down factors. In C. Auxiette, C. Drake, & C. Gérard (eds.), *Proceedings of the Fourth Rhythm Workshop: Rhythm perception and production* (pp. 111–116). Bourges, France: Imprimerie Municipale.
- Repp, B. H. (1992b). 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. H. (1992c). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition*, 44, 241–281.
- Repp, B. H. (1993). Music as motion: A synopsis of Alexander Truslit's "Gestaltung und Bewegung in der Musik." *Psychology of Music*, 21, 48–72.
- Repp, B. H. (1995a). Detectability of duration and intensity increments in melody tones: A partial connection between music perception and performance. *Perception & Psychophysics*, 57, 1217–1232.
- Repp, B. H. (1995b). Expressive timing in Schumann's "Träumerei": An analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America*, 98, 2413–2427.
- Repp, B. H. (1997). The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception*, 14, 419–444.
- Repp, B. H. (in press, a). Variations on a theme by Chopin: Relations between perception and production of timing in music. *Journal of Experimental Psychology: Human Perception and Performance*.
- Repp, B. H. (in press, b). The detectability of local deviations from a typical expressive timing pattern. *Music Perception*.
- Repp, B. H. (in press, c). Obligatory "expectations" of expressive timing induced by perception of musical structure. *Psychological Research*.
- Schulze, H.-H. (1978). The detectability of local and global displacements in regular rhythmic patterns. *Psychological Research*, 40, 173–181.
- Seashore, C. E. (1938). *Psychology of music*. New York: McGraw-Hill (Reprinted by Dover Publications, 1967).
- Shaffer, L. H. (1995). Musical performance as interpretation. *Psychology as Music*, 23, 17–38.
- Shigeno, S. (1986). The auditory tau and kappa effects for speech and nonspeech stimuli. *Perception & Psychophysics*, 40, 9–19.
- Shove, P., & Repp, B. H. (1995). Musical motion and performance: Theoretical and empirical perspectives. In J. Rink (ed.), *The practice of performance: Studies in musical interpretation* (pp. 55–83). Cambridge, UK: Cambridge University Press.
- Thorpe, L. A., & Trehub, S. E. (1989). Duration illusion and auditory grouping in infancy. *Developmental Psychology*, 25, 122–127.

- Thorpe, L. A., Trehub, S. E., Morrongiello, B. A., & Bull, D. (1988). Perceptual grouping by infants and preschool children. *Developmental Psychology, 24*, 484-491.
- Todd, N. [P. McA.] (1985). A model of expressive timing in tonal music. *Music Perception, 3*, 33-58.
- Todd, N. P. McA. (1994). The auditory "primal sketch": A multiscale model of rhythmic grouping. *Journal of New Music Research, 23*, 25-70.
- Todd, N. P. McA. (1995). The kinematics of musical expression. *Journal of the Acoustical Society of America, 97*, 1940-1949.
- Truslit, A. (1938). *Gestaltung und Bewegung in der Musik*. Berlin-Lichterfelde: Chr. Friedr. Vieweg.
- Viviani, P., & Stucchi, N. (1992a). Biological movements look uniform: Evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 603-623.
- Viviani, P., & Stucchi, N. (1992b). Motor-perceptual interactions. In G. E. Stelmach & J. Requin (eds.), *Tutorials in motor behavior II* (pp. 229-248). Amsterdam: Elsevier.
- Williams, K. N., & Perrott, D. R. (1972). Temporal resolution of tonal pulses. *Journal of the Acoustical Society of America, 51*, 644-647.