Cognition, 44 (1992) 241–281

Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations*

Bruno H. Repp
Haskins Laboratories, 270 Crown Street, New Haven, CT 06511–6695, USA

Received June 29, 1991, final revision accepted February 6, 1992

Abstract

Repp, B.H., 1992. Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. Cognition, 44: 241–281.

To determine whether structural factors interact with the perception of musical time, musically literate listeners were presented repeatedly with eight-bar musical excerpts, realized with physically regular timing on an electronic piano. On each trial, one or two randomly chosen time intervals were lengthened by a small amount, and the listeners had to detect these "hesitations" and mark their positions in the score. The resulting detection accuracy profile *across all positions in each musical excerpt showed pronounced dips in places where lengthening would typically occur in an expressive (temporally modulated) performance. False alarm percentages indicated that certain tones seemed longer* a priori, *and these were among the ones whose actual lengthening was easiest to detect. The detection accuracy and false alarm profiles were significantly correlated with each other and with the temporal microstructure of expert performances, as measured from sound recordings by famous artists. Thus the detection task apparently tapped into listeners' musical thought and revealed their expectations about the temporal microstructure of music performance. These expectations, like the timing patterns of actual performances,*

Correspondence to: Bruno H. Repp, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511–6695, USA.

*This research was supported by NIH grant RR-05596 to Haskins Laboratories. Some of the results were presented at the 121st Meeting of the Acoustical Society of America in Baltimore, MD, and at the "Resonant Intervals" Interdisciplinary Music Conference in Calgary, Alberta, Canada, both in May 1991. For helpful comments on an earlier version of the manuscript, I am grateful to Carol Fowler, Carol Krumhansl, Mari Riess Jones, and two anonymous reviewers. Additional insights were provided by Patrick Shove in many discussions.

derive from the cognitive representation of musical structure, as cued by a variety of systemic factors (grouping, meter, harmonic progression) and their acoustic correlates. No simple psycho-acoustic explanation of the detection accuracy profiles was evident. The results suggest that the perception of musical time is not veridical but "warped" by the structural representation. This warping may provide a natural basis for performance evaluation: expected timing patterns sound more or less regular, unexpected ones irregular. Parallels to language performance and perception are noted.

Introduction

In every society, people who listen to music have certain tacit expectations about how it ought to be performed. Professional musicians on the whole aim to satisfy these expectations in their performances, while also exploring imaginative deviations from the norm. This applies especially to the tonal art music of Western culture, with which this article is concerned. The primary cause of the mutual attunement of performer and listener presumably lies in the musical structure apprehended by both. It is the musical structure that calls for certain basic properties of a performance, without which it would be considered crude and deficient. Stylistic variation and individual preferences appear as quantitative modulations of these basic properties.

The present research is restricted to the temporal aspect of music performance, which is arguably the most important of the many physical dimensions along which performances vary. It is common knowledge that musical notes are not meant to be executed with the exact relative durations notated by the composer; rather, performers are expected to vary the intervals between tone onsets according to the expressive requirements of the musical structure. Music that is executed with mechanical precision generally sounds dull and lifeless, and this is particularly true of the highly individual and expressive music written during the nineteenth century. Objective measurements of the "timing microstructure" of expert performances, usually by pianists, have amply documented that deviations from exact timing are ubiquitous and often quite large (e.g., Clarke, 1985a; Gabrielsson, 1987; Hartmann, 1932; Henderson, 1936; Palmer, 1989; Povel, 1977; Repp, 1990b; Shaffer, 1981).[1]

These deviations are by no means random or unintended. Individual performers can replicate their own timing microstructure for a given piece of music with high precision (see, for example, Henderson, 1936; Repp, 1990b). Only a small

[1] I am not concerned here with the variation in tone duration as such (i.e., in onset–offset intervals), which signal differences in articulation (i.e., *legato* vs. *staccato*), an important dimension in its own right. "Timing microstructure" here refers exclusively to variation in tone onset–onset intervals, regardless of whether these intervals are filled or partially empty.

proportion of the variance is not under the artist's control. There are also significant commonalities among the performance timing patterns of different artists playing the same music, despite considerable individual differences (Repp, 1990b, 1992). Certainly, musical interpretation is far from arbitrary. To a large extent, artists follow certain implicit rules in translating musical structure into timing variations. Musical listeners, in turn, presumably expect to hear variations that follow those rules, within broad limits. If these expectations could be measured, they might provide a window onto the listener's cognitive representation of musical structure, just as the actual performance microstructure informs us about the artist's structural conception (cf. Palmer, 1989; Todd, 1985).

The objective study of the principles that underlie systematic timing variations in serious music performance has barely begun. Some facts are already well established, however. One is that most timing deviations are *lengthenings* rather than shortenings relative to some hypothetical underlying regular beat, and lengthenings also are larger in size than shortenings.[2] The functions of lengthening are manifold (Clarke, 1985b), but perhaps the most important function is the demarcation of structural boundaries. Lengthening commonly occurs in performances not only at the ends of major sections, where composers may have prescribed a *ritardando* in the score, but also at the ends of subsections and individual phrases. The amount of lengthening tends to be proportional to the structural significance of the boundary; this regularity has been captured in Todd's (1985) formal model of timing at the phrase level. That model generates a timing microstructure for bar-size units by additively combining prototypical timing patterns nested within the levels of a hierarchical phrase structure; this leads to greater lengthening where boundaries at several levels coincide. The resulting timing pattern essentially reflects the *grouping structure* of the music (cf. Lerdahl & Jackendoff, 1983) and, with adjustment of some free parameters, can approximate actual performance timing profiles quite well (Todd, 1985).

Another important function of lengthening is to give emphasis to tones that coincide with moments of harmonic tension or that receive metric accent. Thus, to some extent lengthening is also determined by the *harmonic and metric structure* of a piece. A lengthened tone also delays the onset of a following tone, which may then be perceived as relatively more accented, especially on instruments such as the piano, where the intensity of tones decays over time. While this last function may be specific to keyboard instruments, the phenomena of (phrase-) final lengthening and emphatic lengthening are also well documented in speech (e.g., Carlson, Friberg, Frydén, Granström, & Sundberg, 1989; Lindblom, 1978) and appear to be very general prosodic devices. There is also a parallel with

[2]This assertion is based on the observation that the frequency distribution of the actual durations of nominally equal tone inter-onset intervals in performed music is strongly skewed towards long values (Repp, 1992). Research remains to be done to show what point in that distribution corresponds to the perceived underlying beat, but that point is likely to be near the peak of the distribution.

breathing and pausing patterns in speech production, which are likewise governed by prosodic structure (Grosjean & Collins, 1979; Grosjean, Grosjean, & Lane, 1979).

The focus of the present study, however, is not on the performer but on the listener. Moreover, it is not on the more obvious perceptual function of timing microstructure, which is to convey or reinforce various structural aspects of the music (see, for example, Palmer, 1989; Sloboda, 1985). Rather, the aim of this research was to demonstrate that, just as lengthening is implemented more or less consistently by professional performers (as part of the "art of phrasing"), so listeners with musical inclinations expect it to occur in certain places. There may be individual differences among listeners, just as there are among performers, with regard to the preferred extent of the expected lengthenings, but there may well be substantial agreement as to their location (unless the music is structurally ambiguous). If musical listeners did not have expectations about the microstructure of music performance, their ability to express preferences for certain performances over others could not be explained.[3] What is not known, however, is whether these expectations interact with ongoing music perception. That is, do listeners perceive the timing microstructure of music veridically and then compare it against some internal standard (either remembered literally from previous exposures to performances of the music or generated on-line according to some implicit cognitive rules), or do the expectations elicited by the structure of ongoing music interact with and "warp" listeners' perception of musical time?

To address this question, a simple technique was devised to probe listeners' temporal expectations. Several researchers concerned with music performance synthesis have observed informally that musically appropriate timing variations sound regular (as long as they are relatively small), whereas musically inappropriate variations are perceived as distortions (e.g., Clynes, 1983, p. 135; Sundberg, 1988, p. 62). Accordingly, the present experiments used a task which required listeners to detect a temporal perturbation (a small lengthening) in an otherwise isochronous performance. The hypothesis was that *listeners would find it more difficult to detect lengthening in places where they expect it to occur*, particularly at the ends of structural units, in strong metric positions, and at points of harmonic tension. The null hypothesis was that detectability of lengthening would not vary systematically across the musical excerpt – or if it did, that the variation could be explained by the influence of primitive psycho-acoustic factors. (Such possible factors will be considered in the General discussion.) If the null hypothesis were rejected and the pattern of detection performance reflected structural properties of the music, this would confirm that time perception in music is contingent on the perception of musical events. It would further reveal the level of detail in

[3]Although performances usually differ along many dimensions, Repp (1990a) has provided some evidence that listeners can express consistent preferences among performances distinguished by timing microstructure alone.

listeners' expectations of timing microstructure, and thereby the listeners' cognitive grasp of the musical structure.

Two similar experiments were conducted, each using different musical materials. The materials were chosen from the piano literature, because it was considered important to present listeners with music sufficiently complex and engaging to guarantee its processing as a meaningful structure rather than as a mere sequence of tones, even after many repetitions. The standard psychophysical procedure of extensive training and prolonged testing of individual subjects was deliberately avoided to retain a semblance of natural music listening. The group of listeners was treated as a single super-subject, representative of the average (musical) listener. Their detection performance, expressed in terms of an *accuracy profile* across each musical excerpt, was compared to the timing microstructure profiles of real performances by great artists. It was hypothesized that a significant correlation would exist between the two profiles, with lengthening in performance corresponding to dips in the accuracy profile, if listeners' expectations derive from a cognitive representation of musical structure similar to that in an artist's mind.

EXPERIMENT 1

Musical material

The musical excerpt represented the first eight bars of the third movement of Beethoven's Piano Sonata No. 18 in E-flat major, Op. 31, No. 3. Its choice was influenced by the fact that it had been the subject of a detailed performance analysis (Repp, 1990b). The score is reproduced in Figure 1.

The piece is a minuet in 3/4 meter, and the tempo indication (omitted in Figure 1) is *allegretto e grazioso* (i.e., moderately fast and graceful). As can be seen, the predominant note value is the eighth-note, which served as the temporal unit for the purpose of this experiment. Bars and eighth-notes are numbered above the score in Figure 1. (Disregard the small two-digit numbers for the time being.) Altogether, there are 47 eighth-note intervals in the excerpt (2 in bar 0, 6 in each of bars 1–7, and 3 in bar 8). Their onsets are always marked by at least one note, except for interval 6 in bar 0, which contains only a sixteenth-note.

Horizontally, the music can be divided into three strands: a melody in the upper voice, which exhibits a complex rhythm and a variety of note values; a bass in the lower voice, which serves as a counterpoint to the melody and marks the harmonic progression; and a middle voice consisting of a steady eighth-note pulse that completes the harmonic structure and serves mainly to mark time.

Vertically, the music can be divided into a four-level binary hierarchy (a *grouping structure*) of nested units. The schematic diagram above the music in Figure 1 is one of several possible ways of representing this structure. Vertical

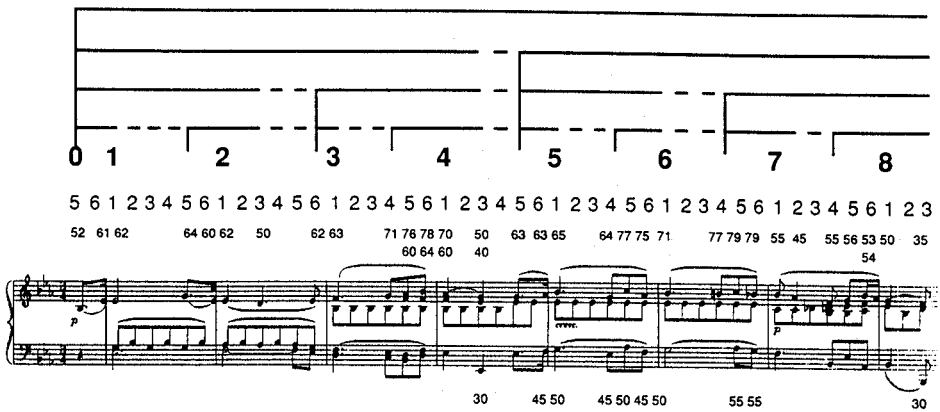

Figure 1. *Score of the musical excerpt used in Experiment 1. The computer-generated score follows the Breitkopf & Härtel Urtext edition, but bar 0 is from the first repeat, while bar 8 is from the second repeat. Above the score are the numbers of bars and eighth-notes, and a schematic representation of the hierarchical grouping structure. The small two-digit numbers above and below the score represent MIDI velocities for melody tones.*

lines represent the onsets of units, whereas continuous horizontal lines show the length of a unit from the onset of the first note to the onset of the last note.[4] The smallest units are conceived here as coherent melodic fragments or *melodic gestures*. Horizontal dashed lines represent time between these gestures. At the lowest level, the music thus comprises eight units, which straddle the bar lines. These melodic gestures are of varying length, which bespeaks Beethoven's rhythmic ingenuity. At the second level, pairs of these short units can be grouped into longer units (phrases). The third level groups these phrases into two sections, and the highest level comprises the whole excerpt (which, in the original music, is followed by another eight-bar section that completes the structure of the minuet).[5]

According to this structural analysis, then, there is a hierarchy of unit boundaries at which final lengthening might be expected to occur in a performance. The "deepest" boundary (where units end at all four levels) is in bar 8

[4]The slurs in the score (which presumably are Beethoven's own) do not indicate the melodic grouping structure; rather, they mark *legato* connections within bars and detached articulations across bar lines. This "articulation structure", which is out of phase with the grouping structure, is ignored here. That the slurs do not represent the grouping structure can be easily proven by the armchair "pause test": it would be much more natural to stop a performance of the music at the ends of melodic gestures than at the ends of slurs.

[5]This performance-oriented representation of the grouping structure differs from the more score-oriented formalism employed by Lerdahl and Jackendoff (1983) in that it does not exhaustively apportion the music to units at the lowest level. For example, the first melodic gesture is assumed to end with the first melody note in bar 1, or more precisely with the onset of the first following note; the following three notes are part of a middle-voice background and hence do not belong to the melodic gesture. In fact, they belong to a different (secondary) grouping structure, that of the middle voice. Lerdahl and Jackendoff (1983) did not deal with the case of several simultaneous grouping structures.

and coincides with the end of the excerpt; the next-deepest boundary is in bar 4; somewhat shallower boundaries are in bars 2 and 6; and the shallowest boundaries are in bars 1, 3, 5, and 7. Accordingly, the most pronounced lengthening (in fact, a *ritardando*) is expected in bar 8, a lesser lengthening in bar 4, and so on. As to the precise location of the expected lengthenings, however, especially at the shallower boundaries where only a single time interval may be affected, the *metric structure* must be taken into account. Final lengthening is generally expected on the last accented tone of a melodic gesture; if a gesture ends with an unaccented tone, it will usually be the preceding accented tone that is lengthened. The gestures ending in bars 2, 4, 7, and 8 have such "weak" endings, in which the penultimate accented tone acts as a harmonic suspense or *appoggiatura*. The last accented tone in each gesture always coincides with the downbeat (first tone) in a bar, so lengthening is generally expected on these metrically strong (and harmonically salient) tones. The metric structure of a piece forms a hierarchy similar to that of the grouping structure (Lerdahl & Jackendoff, 1983). Even though the two structures can be distinguished on theoretical grounds, for our present purpose they lead to the same predictions: final lengthening and emphatic lengthening mostly coincide.

The harmonic structure of the music will not be considered further, as it was not expected to have a major independent influence on the timing microstructure. One special feature of the piece should be noted, however, and that is the prescribed sudden *piano* in bar 7, which follows a *crescendo* through bars 5 and 6. Performers slow down substantially before this dynamic stepdown (Repp, 1990b), presumably to enhance its expressive effect.

An expert performance

Figure 2 presents the timing profile of an expert performance, obtained from a commercial recording by a famous pianist, Murray Perahia (CBS MT 42319). This performance, one of the finest in the set of 19 examined by Repp (1990b), was carefully remeasured for the present study.[6] Tone onsets were determined in visual displays of the digitized acoustic waveform, and the durations of successive eighth-note onset–onset intervals (OOIs) were calculated.[7] These measurements were averaged over four repetitions of the same music (except for bar 1, which

[6]The earlier measurements unfortunately were not at the level of detail required for the present purposes.

[7]Sixteenth-note tones were skipped. Asynchronies among the onsets of nominally simultaneous tones could not be resolved; they were generally small (cf. Palmer, 1989). Chords were thus treated as if they were single tones. A terminological effort is made throughout this paper to distinguish *notes* (printed symbols) from *tones* (single-pitched instrument sounds) and (onset–onset) *intervals* (which include all tones that sound simultaneously). Mixed terms such as "eighth-note interval" and ambiguous terms such as "chord" cannot be avoided, however.

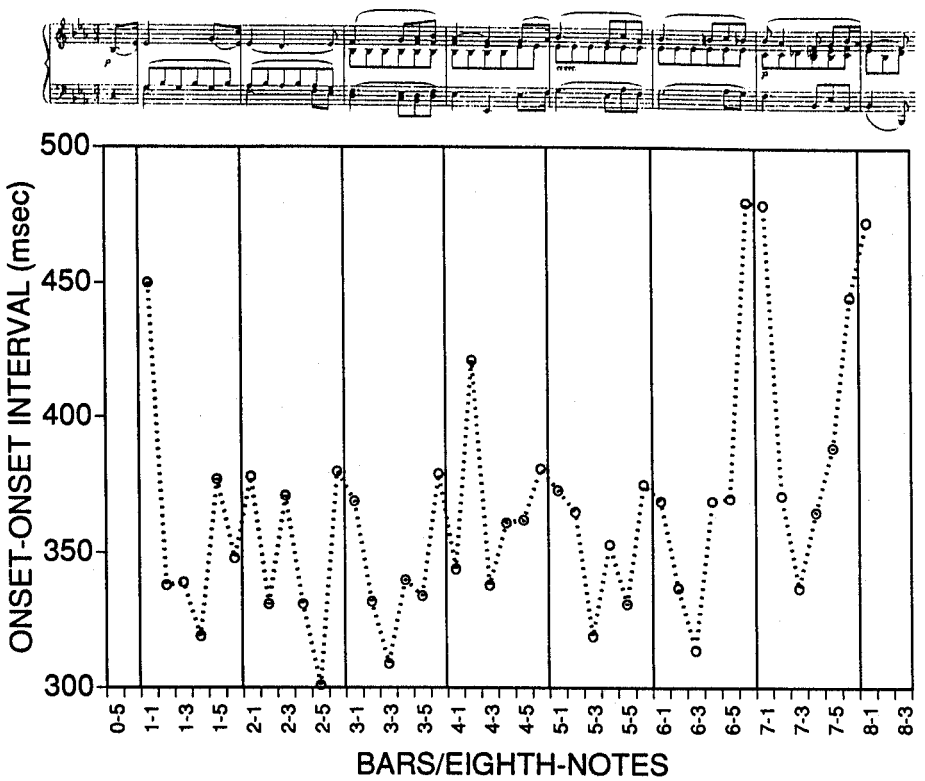

Figure 2. *Timing pattern of an expert performance of the musical excerpt used in Experiment 1.*

occurs in only two literal repetitions in the complete performance). The timing patterns were highly similar across the four repeats, and measurement error is believed to be less than 1%. Data for the initial upbeat (not an eighth-note) and for the final notes (which are not really final in the original music) are omitted in Figure 2.

Each data point in the timing profile represents the interval between the onset of the current eighth-note tone and the onset of the following tone. The pattern can roughly be characterized as a series of ups and downs, with troughs within bars and peaks near bar lines. These peaks are the predicted lengthenings. They generally fall on the first eighth-note interval in each bar, with the preceding interval often lengthened as well. An exception occurs in bar 4, where the second rather than the first eighth-note interval is lengthened; note, however, that both accompany the phrase-final accented tone, which is a quarter-note. Pronounced lengthening (in fact, a gradual *ritardando* spanning four eighth-notes) is found at the end of the excerpt, reflecting the deepest structural boundary there, and the smaller but still prominent peak in bar 4 is associated with a boundary at the next level of depth. The boundaries at the lower levels in the hierarchy are marked

with smaller peaks, with two exceptions: there is substantial lengthening at the beginning of the piece (bar 1) and especially in connection with the sudden dynamic change from bar 6 into bar 7. Finally, it should be noted that the tempo of this performance invariably increased (i.e., OOIs decreased) between melodic gestures (marked by dashes in the diagram in Figure 1), as if to gain momentum for the next gesture.

These observations, even though they are derived from only a single performance, generally confirm the predictions made about the timing microstructure. Below they will be compared to the perceptual results.

Methods

Stimuli

The musical stimuli were generated on a Roland RD-250S digital piano under the control of a microcomputer running a MIDI sequencing program (FORTE). The score was entered manually, such that each eighth-note tone occupied 60 "ticks," the internal time unit of the program. All tones were given their exact notated values (i.e., 30 ticks for a sixteenth-note, 120 ticks for a quarter-note, etc.), except for tones that were immediately repeated, which needed to be separated from the following tone by a brief silence; this silence arbitrarily replaced the last 5 ticks (in eighth-notes) or 2 ticks (in sixteenth-notes) of the nominal tone duration. There were no other temporal modifications in this isochronous performance; thus, for example, the slurs in the score were ignored. Except for repeated tones, therefore, the performance was entirely *legato*, without use of the sustain pedal, and the onsets of simultaneous tones were virtually simultaneous, within the accuracy of the MIDI system. The tempo of the performance was determined by setting the "metronome" in the FORTE program to 88 quarter-notes per minute. Accordingly, one tick corresponded to 5.68 ms, and the duration of a full eighth-note interval was 341 ms. The total performance lasted about 16.4 s.

To increase the musical appeal of the excerpt, the melody tones were assigned relative intensities (coded in the FORTE program as MIDI velocities) derived from a performance on the Roland keyboard by the author, an amateur pianist. These MIDI velocities are shown as the small two-digit numbers above or below the corresponding notes in Figure 1. All other tones (including all in the middle voice) were assigned a constant velocity of 40.[8]

[8]The reason for not varying the intensity of the accompanying tones was that the author's technical skills were not deemed reliable at that very subtle level. The possible influence of the "intensity microstructure" on the perceptual responses will be investigated in the General discussion. Suffice it to note here that the metrically accented downbeats were not usually more intense than neighboring melody tones.

The perfectly isochronous performance was presented to the subjects only as an initial example. In the experimental stimuli, a single eighth-note interval was lengthened by a small amount. The lengthening was achieved by extending all tones occupying that interval by the same number of ticks and by consequently delaying the onset of the following tone(s) and of all subsequent tones. (That is, the lengthening was *not* compensated by shortening the following tone.) Since there were 47 eighth-note intervals in the excerpt, there were 47 possible stimuli with a single lengthened interval. When an interval was bisected by the onset of a sixteenth-note tone, each sixteenth-note interval was lengthened by half the amount.

Four amounts of lengthening were used, based on pilot observations: 10 ticks (56.8 ms or 16.7%) for three initial examples; 8 ticks (45.4 ms or 13.3%) in the first block of trials; 6 ticks (34.1 ms or 10%) in the second block; and 4 ticks (22.7 ms or 6.7%) in the third block. The music was reproduced on the Roland with (synthetic, but fairly realistic) "Piano 1" sound and at an intermediate "brilliance" setting. Recordings were made electronically onto cassette tape. The experimental tape contained three repetitions of the isochronous performance, followed by the three examples of easily detectable lengthening (16.7%). Three blocks of 47 stimuli each followed, in order of increasing difficulty. The order of stimuli within each block was random. The interstimulus interval was 5 s, with longer intervals after each group of 10 and between blocks.

Subjects

Twenty musically literate subjects were paid to participate in the experiment. Most of them had responded to an advertisement in the Yale campus newspaper. They were 10 men and 10 women ranging in age from 15 to 52. A wide variety of musical backgrounds was represented, ranging from very limited musical instruction to professional competence. All subjects had played some musical instrument(s) at some time in their lives (piano, violin, guitar, and others), but only half of them still played regularly.[9]

Procedure

Subjects were mailed a cassette tape with instructions and answer sheets, and they listened at home on their own audio system.[10] They first viewed a sheet with the musical score on it, with bars and eighth-notes numbered. Then they listened to

[9]The author routinely served as a pilot subject, and since his data were quite typical, they were included to replace the data of one subject who performed at chance level in all conditions.

[10]The systematicity of the data proved this "take-home" method to be quite successful for this kind of experiment, in which precise control over volume and sound quality was not essential. Subjects listened to the music in familiar surroundings at a time of their own choosing and thus largely avoided the tense atmosphere that hovers over laboratory experiments.

the examples; on the sheet, the lengthened intervals were indicated numerically as 3–3, 6–2, and 1–5, meaning the third eighth-note interval in bar 3, the second in bar 6, and the fifth in bar 1. After making sure that they heard these lengthenings as slight hesitations in the computer performance (they were allowed to rewind the tape and listen again to the examples, if necessary), the subjects turned to the answer sheets, each of which displayed the score on top. Subjects gave their responses in the numerical notation just explained. They were encouraged to follow along in the score with their pencil tip and to hold it where they perceived a hesitation, but to continue listening until the end of the music before writing down their response. They were asked to write down a question mark if they did not hear any hesitation; wild guesses were discouraged, though a response was welcome if subjects "had a hunch" of where the lengthening might have been. Subjects were asked to take the whole test in a single session, but to take short rests between blocks. Rewinding the tape was strictly forbidden during the test proper, and there was no indication that this instruction was not followed. At the end of the test, subjects completed a questionnaire about their musical experience before returning all materials to the author by mail.

Results and discussion

Overall accuracy

With 47 possible positions of the lengthened time interval, chance performance in this task was about 2%. It was immediately evident that the subjects performed well above chance at all levels of difficulty. However, they were frequently off by one, sometimes two, eighth-notes. Since these near-misses, in view of the low guessing probability, must have reflected positive detection of the lengthened interval in nearly all cases, all responses within two positions of the correct interval were considered correct. (Their distribution will be analyzed below.) By that criterion, chance performance was about 10% correct.

Overall performance was 57% correct. Not surprisingly, performance declined across the three blocks of trials as the amount of lengthening decreased, despite the possible benefits of practice; the scores were 70%, 61%, and 41% correct. This decline was of little interest in itself, and the data were combined across the three blocks in all following analyses. It should be noted that an average performance level of about 50% was optimal for observing variations in detection accuracy as a function of position.

The detection accuracy profile

Average detection accuracy as a function of position is shown in Figure 3. Each data point is based on 60 responses: 3 (blocks) from each of 20 subjects. It can be

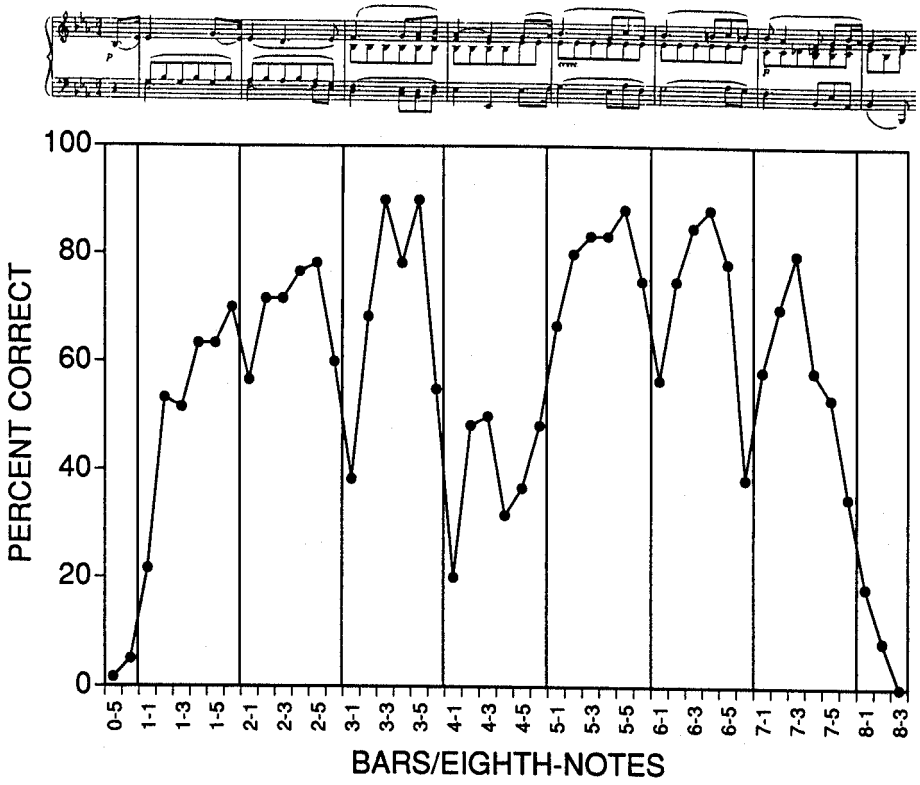

Figure 3. *Percent correct detection as a function of the position of the lengthened eighth-note interval in Experiment 1.*

seen that there were enormous differences in the detectability of lengthening across positions, with scores ranging from 0 to 90% correct.

Lengthening was never detected in the first two intervals and in the last interval. In the case of the initial intervals, this may be explained by the absence of an established beat at the beginning of the piece, and the final interval evidently did not have a clearly demarcated end. The poor score for the penultimate interval (position 8–2) and the steeply declining performance for the preceding intervals cannot be explained on these trivial grounds, however; they appear to be due to subjects' expectation of substantial lengthening (*ritardando*) toward the end of the excerpt.

The remainder of the accuracy profile is characterized by peaks and valleys, the peaks generally in the middle of bars and the valleys near bar lines. This pattern is roughly the inverse of that observed in Murray Perahia's performance (Figure 2). The correlation is -0.59 ($p < .001$).[11] In general, where Perahia slowed down, lengthening was more difficult to detect, and where he speeded up, detection performance improved. The correlation is not perfect, nor should it be expected

[11]The correlation does not include intervals 0–5, 0–6, 8–2, and 8–3.

to be, since both data sets presumably contain other sources of variability and are related only via the structural properties of the music.

In fact, the perceptual data seem to reflect the musical structure more closely than does the performance timing profile. The valleys in the accuracy profile generally coincide with the first note in each bar, which bears the metric as well as gestural accent. This coincidence is evident in bars 2, 3, 4, and 6. In bars 1 and 8, the drop to chance performance for the initial and final intervals, respectively, obscures any local valley on the downbeat, and a similar argument may be made for bar 5, where performance is recovering from an especially large dip in bar 4. The early peak in the last interval of bar 6 is due to the sudden dynamic change on the following note, which normally requires a preparatory lengthening or "micropause" in performance (see Figure 2). Thus the following valley is obscured here, too. Only the pattern in bar 4 remains unexplained: after a momentary rise in performance, there is a second dip in the fourth interval, which is not part of any melodic gesture. This interval, however, marks the end of the first half of the musical excerpt; at this point one could reasonably stop the performance. Thus the dip at position 4–4 may reflect a more abstract structural boundary of the kind represented in the conventional grouping hierarchy of Lerdahl and Jackendoff (1983). Bar 4 also contains a major disagreement with Perahia's performance: Perahia lengthens the second interval in this bar, whereas the dip in the accuracy profile occurs on the first interval. The detection data seem more internally consistent in that respect than Perahia's performance. On the whole, therefore, the perceptual results confirm the predictions based on metric and grouping structure. A final observation is that detection scores invariably increase between melodic gestures, where Perahia's performance just as consistently gained speed.

False alarms

The 43% incorrect responses consisted of question marks (26%) and false alarms (17%, $n = 479$). We now direct our attention to these latter responses, which were clearly unrelated to the intervals actually lengthened. That is, subjects did not hear the lengthened interval but nevertheless believed they had a hunch of where it might have been. Given the low probability of a correct guess, these trials thus served as catch trials, which may reveal the listeners' expectations.

Indeed, the false alarms were far from evenly distributed across positions. Their frequency distribution is shown in Figure 4. It is evident that it followed a pattern not unlike that of the "hits" shown in Figure 3. Here, too, there are peaks within bars and troughs near bar lines, although the peaks are narrower than those for the hits. The correlation between the hit and false-alarm profiles is 0.64 ($p < .001$); the correlation of the false-alarm profile with Perahia's performance timing profile is -0.40 ($p < .01$). This uneven distribution of false alarms signifies

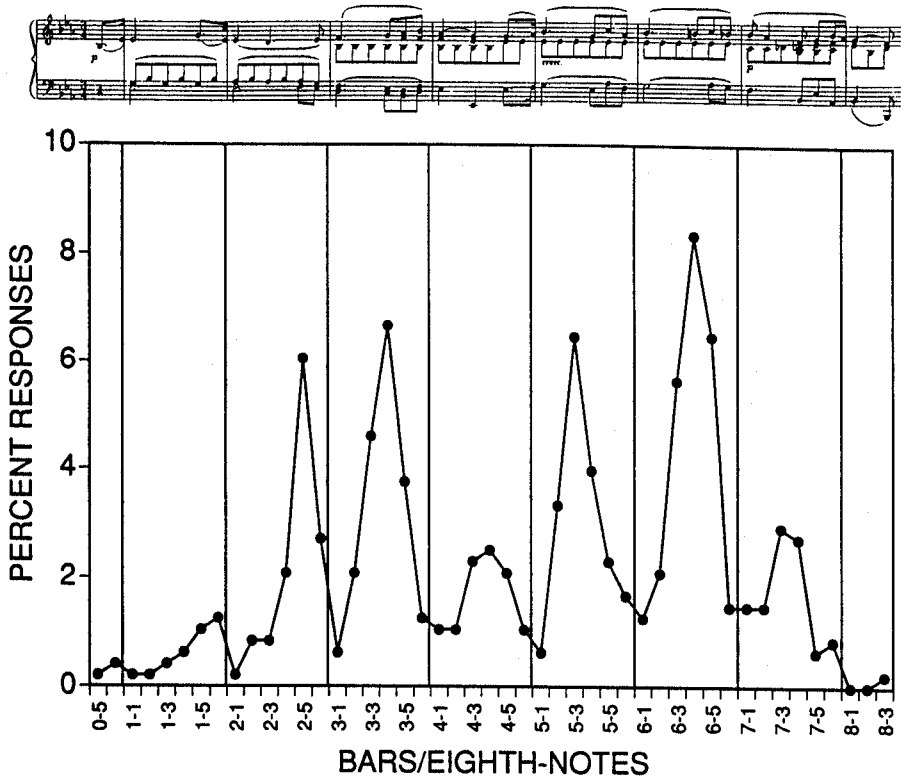

Figure 4. *Distribution of false alarms across eighth-note positions in Experiment 1.*

that there were certain intervals that *a priori* sounded longer than others, and these were ones in which actual lengthening also was easy to detect. These intervals were *not* likely to be lengthened in performance: typically, they either immediately preceded the onsets of melodic gestures or were gesture-initial but unaccented. Thus, under conditions of isochrony, intervals that were expected to be short sounded somewhat long, and intervals that were expected to be long sounded somewhat short.

The variation in hit rates greatly exceeded that in false-alarm probabilities; note that the false-alarm percentages have been magnified ten times in Figure 4 relative to the hit percentages in Figure 3, so as to make their pattern more visible. The comparison is not quite fair, perhaps, because the scoring criterion for correct responses included responses that were off by one or two positions ("near-misses"). However, even when only the percentages of "true hits" are considered (see Figure 5 below), their variation is nearly ten times as large as that of the false alarms. Still, hits and false alarms are probably independent manifestations of the same underlying cause, namely, listeners' structurally determined expectations about timing microstructure. These expectations affect the *criterion* for perceiving a particular interval as lengthened.

Near-misses

Averaged across all 47 positions, correct responses (57% overall) were distributed as follows with respect to the correct position: 1% (−2 positions), 3.6% (−1), 25.7% (on target), 22.6% (+1), and 4.2% (+2). Thus there was a pronounced tendency to locate the lengthened interval in the immediately following position. Other types of near-misses were much less frequent, though (+2) and (−1) responses were about twice as frequent as expected by chance (2%). One possible explanation for this postponement tendency is that subjects necessarily heard the following interval at the time that they realized that an interval had been lengthened; thus, they probably arrested their pencil at that point in the score and forgot to backtrack when writing down their response. Another possibility is that subjects neglected the instructions and attributed the "hesitation" to the tone whose onset was delayed, rather than to the preceding lengthened interval. The delayed tone may also have sounded slightly more intense because of the additional decay of the preceding tone during the lengthened interval (cf. also Clarke, 1988, p. 19, who notes that a delay "heightens the impact" of the delayed note).

If subjects simply had forgotten to backtrack, the profile of (+1) responses should have paralleled that of the true hits; in other words, the relative frequencies of true hits and (+1) responses should have been constant across positions. This was not the case, however. The percentage profiles of true hits, (+1) hits, and (+2) hits are shown in Figure 5.[12] It is evident that each type of response had a distribution with peaks and troughs, but that the peaks in the (+1) and (+2) profiles were generally one and two positions, respectively, to the left of the peaks in the true hit profile (which generally coincide with the peaks in the false-alarm profile–cf. Figure 4). In other words, (+1) responses were most frequent when the lengthened interval occurred immediately before an interval that was *a priori* more likely to be perceived as lengthened, and (+2) responses were most frequent when another interval intervened between these two. Thus, the positional distribution of correct responses was modulated by the same perceptual biases that governed the overall accuracy and false alarm profiles.

These observations receive statistical support: the correlation between the true hit and (+1) profiles aligned as in Figure 5 is 0.24 (n.s.), but it becomes 0.63 ($p < .001$) when the (+1) profile is shifted to the right by one position. Similarly, the correlation between the original true hit and (+2) profiles is −0.26 (n.s.), whereas it is 0.55 ($p < .001$) when the (+2) profile is shifted to the right by two positions.[13] The shifted (+1) and (+2) profiles are also significantly correlated with the false-alarm percentages shown in Figure 4 (0.76 and 0.48, respectively;

[12]For clarity, (−1) responses, which were less systematic, are not included in the figure, nor are (−2) responses, which occurred only by chance.

[13]The analogous operation for (−1) responses did not yield an increase in the correlation, though it reached significance (0.40, $p < .01$), and (−2) responses did not yield any significant correlation at all (0.14).

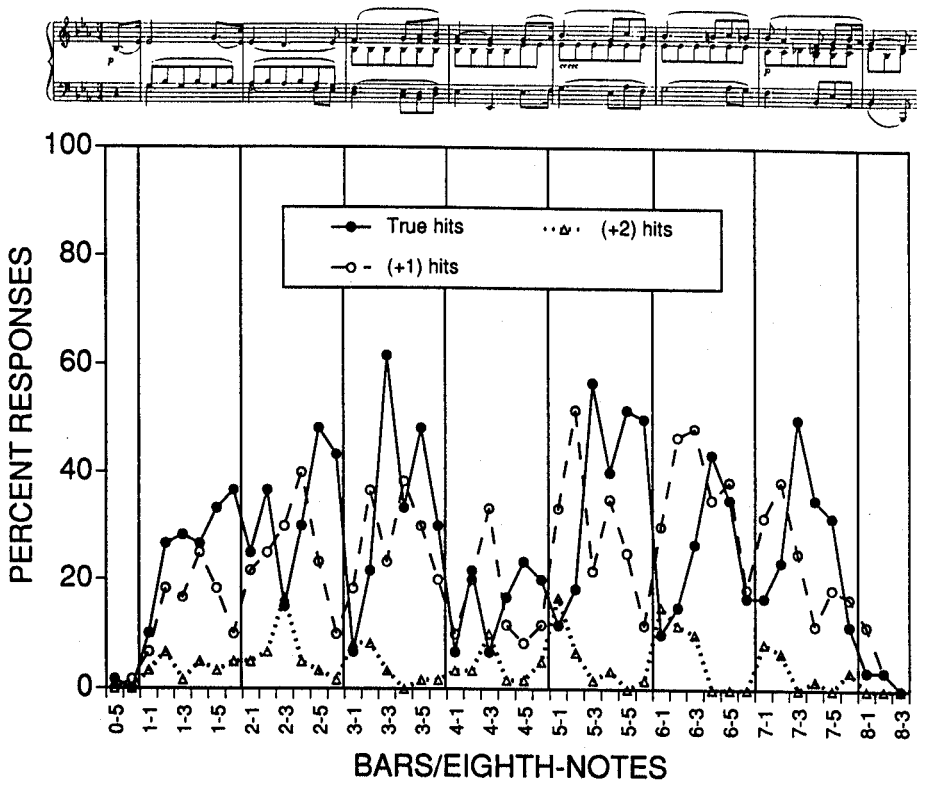

Figure 5. *Percentage profiles of true hits (i.e., on-target correct responses), and of (+1) and (+2) hits (i.e., near-misses) in Experiment 1.*

both $p < .001$), as is the true hit profile (0.60, $p < .001$). Thus, (+1) and (+2) responses, just like true hits, reflect the fact that structurally weak intervals seem *a priori* longer than others.

Yet another way of demonstrating the similarity of the hit and false-alarm profiles is to plot the *frequency distribution* of all correct responses (i.e., the relative frequency with which each position was given as a correct response) and to compare this distribution with that of the false alarms (cf. Figure 4). These two distributions are superimposed in Figure 6. Their similarity is striking ($r = 0.77$, $p < .001$). Virtually all the peaks coincide, and both functions consistently show minima immediately after bar lines. The main difference is that the correct response distribution is flatter, presumably because of the positional constraint imposed by lengthenings that were actually detected, whereas the false-alarm distribution reflects pure perceptual bias. The two distributions evidently reflect the same underlying tendencies (i.e., listeners' expectations).

Individual differences and musical experience

Individual subjects' overall accuracy in the detection task varied considerably,

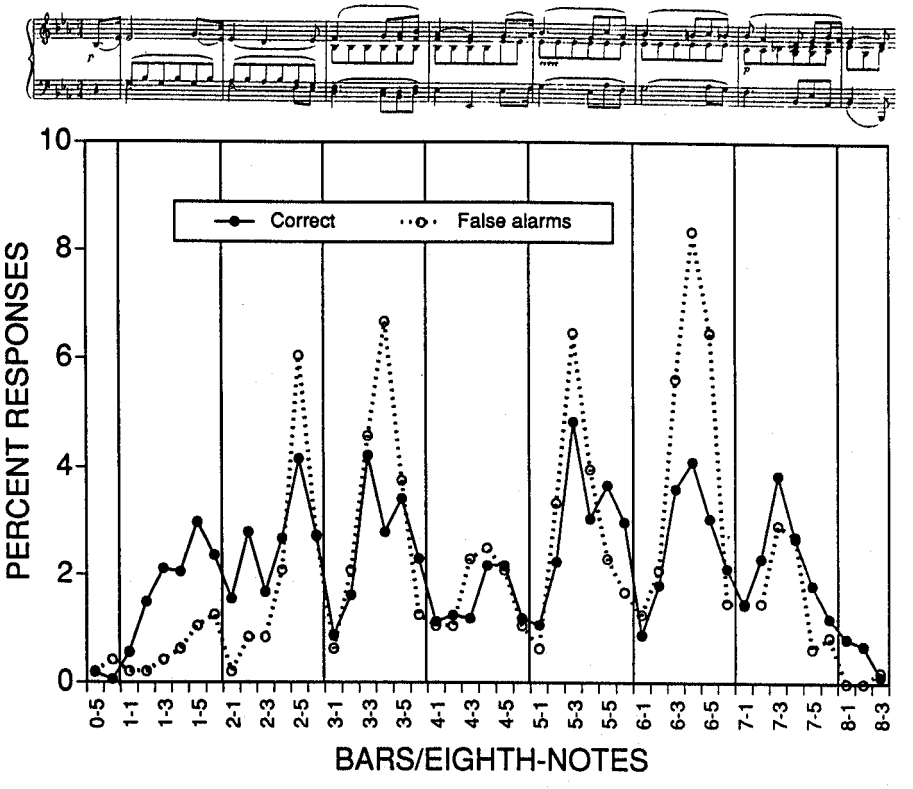

Figure 6. *Frequency distribution of all correct responses combined (according to the response given, not according to the actual location of the lengthened interval) in Experiment 1, compared with the false-alarm distribution (from Figure 4).*

from 26% to 83% correct. One measure of interest was the extent to which each individual subject's accuracy profile (based on only three responses per position, alas) resembled the grand average profile shown in Figure 3. These individual "typicality" correlations ranged from 0.37 to 0.84. The correlation between these correlations and overall percent correct was 0.69 ($p < .001$), indicating that the more accurate subjects also showed the more typical profiles.

Correlations were also computed between the individual accuracy and typicality measures on one hand, and sex, age, and several indices of musical experience derived from the questionnaire responses on the other hand. No striking relationships emerged. There was no sex difference. Correlations with age were negative but nonsignificant. Correlations with years of musical instruction (ranging from 1 to 34, added up across all instruments studied) were positive but nonsignificant. With regard to hours per week of active music making (ranging from 0 to 30), the correlation with the typicality measure, 0.42, reached the $p < .05$ level of significance. In view of the multiple correlations computed, however, this could be a chance finding. Correlations with hours per week spent listening to music were negligible. Finally, pre-experimental familiarity with the music seemed to play no

role. (Twelve subjects indicated that they were totally unfamiliar with the music, 4 subjects were somewhat familiar, and 4 subjects professed greater familiarity.) In the absence of significant correlations, it seems highly unlikely that subjects' timing expectations derived from remembered previous performances of the Beethoven minuet; rather, these expectations must have been induced by the musical structure during the experiment, despite the absence of systematic timing microstructure.

EXPERIMENT 2

Although the results of Experiment 1 demonstrate convincingly the influence of musical structure on listeners' judgments, they were based on only one musical excerpt. Moreover, measurements from a single expert performance were available for comparison. Experiment 2 attempted to replicate the findings using a different musical excerpt, for which detailed timing measurements from 28 different expert performances were available (Repp, 1992). This provided a unique opportunity to relate listeners' performance expectations (as reflected in the detection accuracy profile) to a representative measure of actual performance microstructure.

Musical material

The music used in this experiment constituted the initial eight bars of "Träumerei" ("Rêverie"), a well-known piano piece from the cycle "Kinderszenen", op. 15, by Robert Schumann. The key is F major, and there is no verbal tempo indication. The score, without slurs, is reproduced in Figure 7.[14]

Again, the predominant note value is the eighth-note, which served as the temporal unit in the experiment. The time signature is common time (4/4), so that there are 8 eighth-notes per bar. The excerpt contains 62 eighth-note intervals, not counting the final chord.[15] However, not all of these intervals are marked by tone onsets; 12 of them are unmarked, as indicated by the dashes in

[14]Most editions of the music contain tempo suggestions in the form of metronome values that go back to the composer or his wife Clara, but which are generally considered too fast for contemporary tastes. Slurs were omitted in this experiment in order to avoid visually biasing subjects' auditory perception of the grouping structure. In the Beethoven minuet of Experiment 1, slurs generally ended at bar lines and thus cut across the melodic gestures. The slurs in the Schumann piece (at least in the several editions that were consulted) coincide with the grouping structure, except for a single long slur across bars 6–8. It seemed prudent to eliminate this visual source of grouping information from the answer sheets.

[15]The final chord appears in this form only in the present excerpt, to provide an appropriate conclusion. In the original music, the eighth-note movement continues in the bass voice.

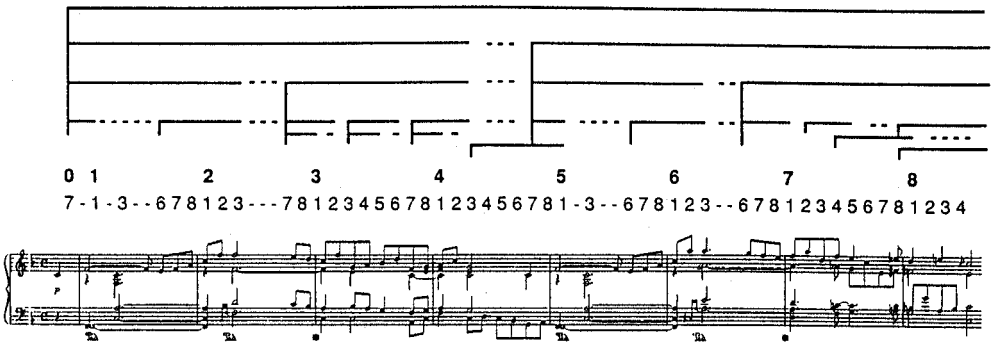

Figure 7. *Score of the musical excerpt used in Experiment 2. The computer-generated score follows the Clara Schumann (Kalmus) edition, but all slurs and expressive markings have been removed. Above the score are the numbers of bars and eighth-notes, and a schematic representation of the hierarchical grouping structure.*

the numbering above the score in Figure 7. Thus there were only 50 OOIs to be measured or manipulated, some of which were multiples of eighth-note intervals.

The music is more complex than that of Experiment 1 in several other respects. Horizontally, four voices (soprano, alto, tenor, bass) can be distinguished, most clearly in bars 3–4 and 7–8. Unlike the Beethoven minuet, where the bass voice merely supported the melody and followed essentially the same grouping structure, the voices in "Träumerei" are more independent of each other and constitute different layers of sometimes staggered melodic gestures.

The vertical structure is represented in Figure 7 by the diagram above the score. At the higher levels, it is similar to that of the Beethoven minuet: one large 8-bar section can be divided into two 4-bar sections, which in turn can be divided into four 2-bar phrases. Even at the lowest level, bars 1–2 and 5–6 resemble those in the minuet, each being composed of two melodic gestures, the second longer than the first. Where the Schumann piece differs from the Beethoven minuet, and becomes considerably more complex, is in bars 3–4 and 7–8.

The phrase spanning bars 3–4 is constituted of four shorter gestures. The first two, in the soprano voice, each comprise 4 eighth-notes and are accompanied by even shorter gestures in the tenor voice. The third gesture in the soprano voice, joined now by tenor and bass, has an additional final note which overlaps the first note of the fourth gesture, now in the bass. That gesture essentially comprises 7 notes, ending on the low F in bar 5, and thus overlaps with the first gesture of the second major section. In fact, the offsets of these two gestures coincide; thus the last gesture of the first major section provides a link with the second major section. Bars 7–8 are even more complex, and quite different from bars 3–4. Instead of being divided between soprano and bass, the melodic gestures descend from soprano to alto to tenor, overlapping each other in the process. The first gesture begins with an extra note and comprises either 8 eighth-notes or two

groups of 4. Anticipating performance data to be discussed below, we assume that it consists of two 4-note gestures. The end of the second gesture overlaps the beginning of the third gesture in the alto (6 eighth-notes), which in turn overlaps the 6-note gesture in the tenor (possibly even extends through it). Although the overall melodic contour seems similar to bars 3–4, the notes are grouped differently. Alternative phrasings are conceivable, though gesture onsets are made unambiguous by the entrances of the different voices.

The metric structure of the piece is irregular. The melodic gestures in bars 1–2 and 5–6 negate the accent on the downbeat and instead postpone it to the second quarter-note beat in bars 2 and 6, respectively. This contradicts the placement of the bar lines in the score; perhaps, if Schumann had lived 100 years later, he would have changed the meter to 5/4 in bars 1 and 5 and to 3/4 in bars 2 and 6. A similar but weaker asymmetry exists in bars 3–4 and 7–8, where there is a strong accent on the second beat of bars 4 and 8, respectively. These metric irregularities are not visible in the score (except for the occurrence of rich chords on the second beats in bars 2, 4, and 6); their existence is derived from musical intuition and knowledge of standard performance practice. Apart from a knowledge of where accents are likely to fall, however, the metric structure as such contributes little to predicting lengthening beyond what can be derived from the grouping structure shown in Figure 7. The same can be said about the harmonic structure, though the salient chromatic transitions in positions 7–4 and 7–8 should be noted; they are likely to receive some emphasis in performance, even though they are metrically weak. Another special feature is the occurrence of grace notes in intervals 2–2 and 6–2, which are likely to result in extra lengthening of the intervals accommodating them.[16]

Performance microstructure

Salient aspects of performance timing are illustrated in Figure 8. This figure shows a timing microstructure extracted from a sample of 28 famous pianists' performances by means of principal components analysis (Repp, 1992). The values represent the factor scores of the first principal component, rescaled into the millisecond domain. Thus, this pattern, while not necessarily corresponding to an outstanding performance, captures what is common to many different performances and contains very little unsystematic variation. The OOIs are plotted in terms of eighth-note intervals; longer OOIs are represented as "plateaus" of multiple eighth-note intervals. The grace notes in bars 2, 6, and 8 are not represented.

[16]These intervals were measured between the onsets of the two F's in the soprano voice; the other tones in the chord in positions 2–3 and 6–3 were sometimes strongly asynchronous (see Repp, 1992). Similarly, the interval 4–8 included a grace note, in addition to being subject to terminal lengthening.

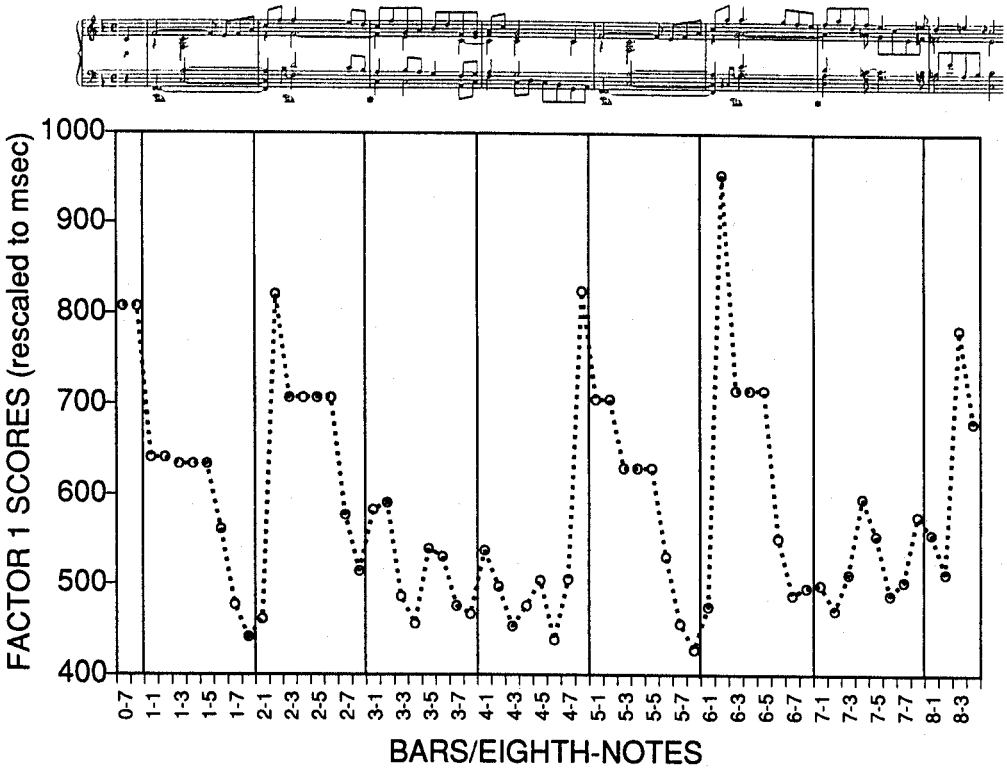

Figure 8. *Generic performance timing microstructure for the musical excerpt of Experiment 2, derived by principal components analysis from the timing patterns of 28 performances by famous pianists (Repp, 1992).*

Again, we can observe a pattern of peaks and valleys, with the peaks reflecting lengthening, mostly at the ends of melodic gestures. Thus, there is a peak at the end of the excerpt, and another at the end of bar 4, which marks the end of the first major section (even though the second section has already begun). Furthermore, the long intervals (in bars 1, 2, 5, and 6), all of which are gesture-final, are relatively extended. The lengthened first upbeat reflects a common start-up effect. The second eighth-note intervals in bars 2 and 6, which are penultimate in their respective gestures, are extra long because they must accommodate the two grace notes (a written-out *arpeggio*) in the bass voice. Finally, a series of small peaks can be seen in bars 3–4 and 7–8. Those in bars 3–4 reflect final lengthening for the eighth-note gestures, while those in bars 7–8 reflect the brief gestures in the bass voice, which are coupled with salient harmonic progressions.

This generic performance pattern may serve as the basis for predictions about the detectability of actual lengthening of individual intervals in an otherwise isochronous performance. Lengthening should be most difficult to detect where there are peaks in the performance timing profile, and easiest where there are valleys.

Methods

Stimuli

The score was entered manually into the FORTE program, with 100 ticks per eighth-note interval. The tempo was set at the equivalent of 60 quarter-notes per minute, so that the duration of an eighth-note interval was 500 ms, and the temporal resolution (1 tick) was 5 ms. The total duration was about 33 s. All tones lasted for their nominal duration, except: (1) those that were immediately repeated, whose last 25 ms were replaced with silence, (2) the left-hand chords tied over into bars 2 and 6, which were shortened by 500 ms, so that they ended with the first eighth-note interval in these bars; and (3) the quarter-note interval preceding the grace note in bar 8, which was shortened by 250 ms, making it effectively a dotted eighth-note interval followed by a sixteenth-note interval (the grace note) lasting 250 ms. The grace notes in bars 2 and 6 started 165 and 330 ms, respectively, into the eighth-note interval; the first of them ended with the onset of the second. Except for repeated tones, then, the performance was strictly *legato*. Sustain pedal was added as indicated, with pedal onset 25 ms after, and pedal offset 25 ms before, the relevant tone onsets.[17]

A fixed intensity microstructure was imposed on the tones (see Table 1). The relative intensities were modeled after acoustic measurements of a recorded performance by a distinguished pianist, Alfred Brendel (Philips 9500 964), and were adapted to the dynamic range of the Roland digital piano. The precise methods of this transfer will not be described and defended here; suffice it to say that the intensity pattern seemed musically appropriate and pleasing to the ear.

With the final chord and the grace notes excluded, there were 50 possible intervals to be lengthened, most of them corresponding to an eighth-note, but some longer. All intervals were lengthened in proportion to their nominal duration; thus, for example, the half-note interval in bar 2 was extended by four times the amount applied to an eighth-note interval. The grace notes in bars 2 and 6 were never lengthened; when the eighth-note interval that contained them was lengthened, the onset of the first grace note (and the corresponding pedal onset) was delayed by that amount. The grace note in bar 8, on the other hand, was lengthened by half the amount of the lengthening applied to the simultaneous eighth-note interval.

For reasons of economy, it was decided to lengthen *two* intervals in each presentation of the excerpt: one in the first half and one in the second half. The second half was defined as beginning with the upbeat on the fourth beat in bar 4; that way, there were 25 possible intervals to be lengthened in each half. The two intervals lengthened in each stimulus were randomly chosen, with the restriction

[17]These were, for pedal onsets, the first tones in bars 1 and 5; for pedal offsets, the first tones in bars 3 and 7; and for both offsets and onsets, the first grace notes in bars 2 and 7.

Table 1. *Intensity microstructure (in MIDI velocities) of the musical excerpt used in Experiment 2. As much as possible, the tones have been assigned to four voices (* = grace notes)*

Bar/eighth-note	Bass		Tenor		Alto		Soprano	
0–7							72	
1–1		42					73	
1–3			26	25	55	61		
1–6							62	
1–7							74	
1–8							70	
2–1							76	
2–2							80	
2–3	48*	40*		59		62	79	
2–7				59			62	
2–8				55			86	
3–1		36		76			73	
3–2							78	
3–3		45		60		67	71	
3–4				76			78	
3–5				59			73	
3–6							88	
3–7		23		41		80	84	
3–8		53		51			74	76
4–1		46		56			76	76
4–2							81	
4–3		48		52		60	69	69
4–4		61						
4–5		68						
4–6		63						
4–7		49					89	
4–8		49						
5–1		25					76	
5–3			30	43	58	69		
5–6							66	
5–7							75	
5–8							74	
6–1							82	
6–2							98	
6–3	55*	65*		57	55	63	91	
6–6							69	
6–7							76	
6–8							75	
7–1		32		65			80	
7–2							90	
7–3							90	
7–4		33		23		36	84	
7–5		48				57	79	
7–6						76		
7–7						77		
7–8		42		57		68	89	
8–1		37		59		62	80	
8–2				58				
8–3				51			55	
8–4							60	54*
8–5		15		61		32	51	

that they never occupied structurally analogous positions in the two halves of the excerpt.

An experimental tape (*Test A*) was recorded, containing three blocks of stimuli preceded by six examples. The first three examples were entirely isochronous; in each of the next three, two intervals were lengthened by 16%. Each of the three successive blocks contained 25 stimuli, arranged in random sequence with ISIs of 5 s, 10 s after each group of 10, and longer intervals between blocks. The amounts of lengthening employed in the three blocks were 14% (70 ms), 12% (60 ms), and 10% (50 ms), in that order.

The degrees of lengthening were chosen on the basis of the author's impressions during stimulus construction; however, as might have been anticipated from Experiment 1, they proved too easy to detect in eighth-note intervals for most listeners, so that there was a ceiling effect in some regions of the accuracy profile. A supplementary test tape (*Test B*) was therefore constructed. It contained three initial examples with 10% lengthening and two additional blocks of stimuli with 8% (40 ms) and 6% (30 ms) lengthening, respectively. Only the eighth-note data from this test were considered.[18]

Subjects

The subjects were musically literate Yale undergraduates who attended a course on the psychology of music. Participation was voluntary, and subjects were paid for their services. Sixteen subjects completed Test A, and 7 of them listened to Test B some weeks later. One additional subject, who had not received Test A, completed a special test comprising two blocks of stimuli with 10% and 8% lengthening, respectively; his 10% data were included in the Test A totals, and his 8% data in the Test B totals.

Procedure

The procedure was the same as in Experiment 1, with two exceptions: first, as already mentioned, subjects gave two responses per presentation rather than just one. Second, rather than giving a numerical response, subjects circled the lengthened note(s) in a copy of the score. (It was sufficient to circle any of several simultaneous notes.) The answer sheets provided a separate miniscore for each presentation in the sequence. The possible responses were first illustrated on a sheet showing also the locations of the lengthened intervals in the initial examples.

[18]Due to an oversight, the longer intervals had been lengthened by the same small amount as eighth-note intervals, which was virtually impossible to detect. These stimuli thus served essentially as catch trials.

Results and discussion

Overall accuracy

With 25 possible choices in each half of the excerpt, the chance level of performance was 4% correct–twice as high as in Experiment 1. Also, subjects seemed to be on target relatively more often than in Experiment 1. Therefore, the criterion for accepting responses as correct was tightened, and only responses that were within one position of the correct interval were considered correct. (Long intervals were counted as a single position.) That way, chance level was 12% correct.

Overall, subjects scored 74% correct on Test A. Average performance declined across the three blocks from 78% to 75% to 70% correct. This surprisingly small effect of increasing the difficulty of the task may have been due in part to a counteracting effect of practice; in part, the reason was that the task was a little too easy, leading to a ceiling effect for most eighth-note intervals. (The scores for eighth-note intervals only were 82%, 79%, and 74%, respectively, in the three blocks.) However, even in Test B, where only eighth-notes were scored, performance was still quite good: 59% and 50% correct, respectively. The amount of lengthening in the last block was 6%, or 30 ms, which evidently was still quite detectable.

Again, since absolute performance levels were not of particular interest, the data were combined across all levels of difficulty in each test.

The accuracy profile

Figure 9 shows percent correct detection as a function of position in the music. The solid line shows the results of Test A, where each data point is based on 49 responses. The dashed line represents the results of Test B (eighth-notes only), where each data point derives from 15 responses. As in the generic performance timing profile (Figure 8), the results for long notes are shown as plateaus; they really represent only a single data point.

Consider the Test A results first. It can be seen that performance varied dramatically across positions, from chance to nearly perfect detection. Lengthening was moderately difficult to detect in all long intervals. It was easy to detect in most eighth-note intervals, but with significant exceptions. Performance was at chance for the last eighth-note interval of bar 4, and rather poor for the preceding interval. These tones represent the end of the melodic gesture in the bass voice which marks the end of the first major section of the excerpt and coincides with the upbeat to the second major section. Lengthening was also difficult to detect in the eighth-note interval preceding the final chord, in the first eighth-note interval of bar 8, and in the second eighth-note intervals in bars 2 and 6, which coincide with grace notes. Most of these intervals also tend to be lengthened in per-

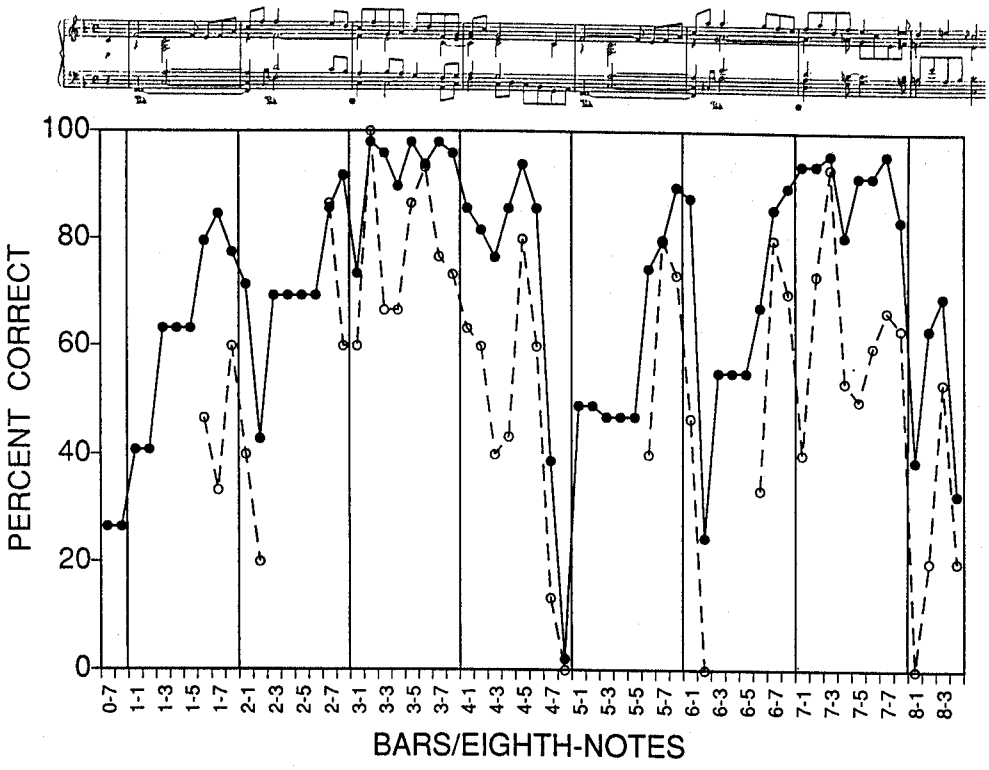

Figure 9. *Detection accuracy profiles obtained in Experiment* 2. *The solid line represents Test A results, the dashed line Test B results (eighth-notes only).*

formance; the accuracy profile is very nearly the inverse of the performance timing profile shown in Figure 8. The correlation between these two profiles is -0.75 ($p < .001$).[19] Thus the results are again consistent with the hypothesis that lengthening is more difficult to detect where it is expected.

Clearly, the gross accuracy profile mirrors the major lengthenings encountered in performance. It is less clear whether there is also any correspondence at the more molecular level of the eighth-note intervals in bars 3–4 and 7–8, because of the paucity of errors in these regions in Test A. Here the Test B results are useful. As can be seen in Figure 9, the Test B results generally confirm but magnify the tendencies observed in the Test A profile. The correlation between the Test A and Test B eighth-note profiles is 0.84 ($p < .001$). The correlations between each of these two profiles and the performance timing profile (eighth-notes only) are -0.70 and -0.49, respectively (both $p < .001$). The second correlation is a good deal lower, suggesting that the match at this more detailed

[19]Each of the long notes was a single data point in any correlation computed. Their representation as multiple data points in the figures is for graphic purposes only.

level is not so close, and that the significant correlations may be largely due to the several eighth-notes that yielded extremely poor scores.

Let us compare, therefore, the exact locations of peaks and valleys in the performance timing and accuracy profiles. The ascending eighth-note gestures straddling bars 1–2 and 5–6 exhibit an acceleration of tempo followed by a dramatic slowing on the penultimate note (Figure 8). This pattern is mirrored rather nicely in the perceptual data (Figure 9). In bars 3 and 4, however, the correspondence breaks down. In the performance profile, there are small peaks in positions 3–1 and 3–2, 3–5 and 3–6, 4–1 and 4–5. These locations do not correspond to valleys in the accuracy profile; on the contrary, there are accuracy peaks in some of these positions. In fact, at this local level (from position 2–6 to position 4–6) there is a *positive* correlation between the performance and Test B accuracy profiles: $r(16) = 0.57 (p < .02)$. In bar 7, there are small performance timing peaks in positions 4–5 and 4–8. The first two locations do have a corresponding valley in the Test B accuracy profile, but the last one does not. Conversely, in the accuracy profile there are pronounced dips in positions 6–6, 7–1, and 8–1 that do not correspond to lengthenings in performance. Nor is there a dip relating to the pronounced *ritardando* at the end of the excerpt. Thus, although there are clear peaks and troughs in the accuracy profile, they do not mirror the performance profile at this detailed level. We will examine later what factors they might reflect (see General discussion).

False alarms

The false-alarm responses were pooled across Tests A and B. There were 300 such responses altogether, or 9.4% of all responses. As in Experiment 1, they were not evenly distributed. Figure 10 shows their frequency histogram. There are some seven intervals that tended to attract false alarms. Two of them are long (in bars 2 and 5); the others are eighth-note intervals. The eighth-note peaks in the false-alarm profile do coincide with peaks in the accuracy profile, especially that of Test B. Otherwise, however, the correspondence is not close. The overall correlations are 0.12 (Test A, all intervals, n.s.) and 0.31 (Test B, eighth-note intervals only, $p < .05$). Thus, again, there were certain intervals that *a priori* seemed longer than others. As in Experiment 1, some of these intervals were located immediately preceding the onset of melodic gestures. The long notes, of course, may have attracted false alarms simply because of their length. The false alarm peaks in positions 7–8 and 8–3, and the corresponding peaks in the accuracy profile, have no obvious explanation at present.

Near-misses

As in Experiment 1, there was a tendency to postpone responses by one position, but it was less pronounced, perhaps due to the slower tempo of the music.

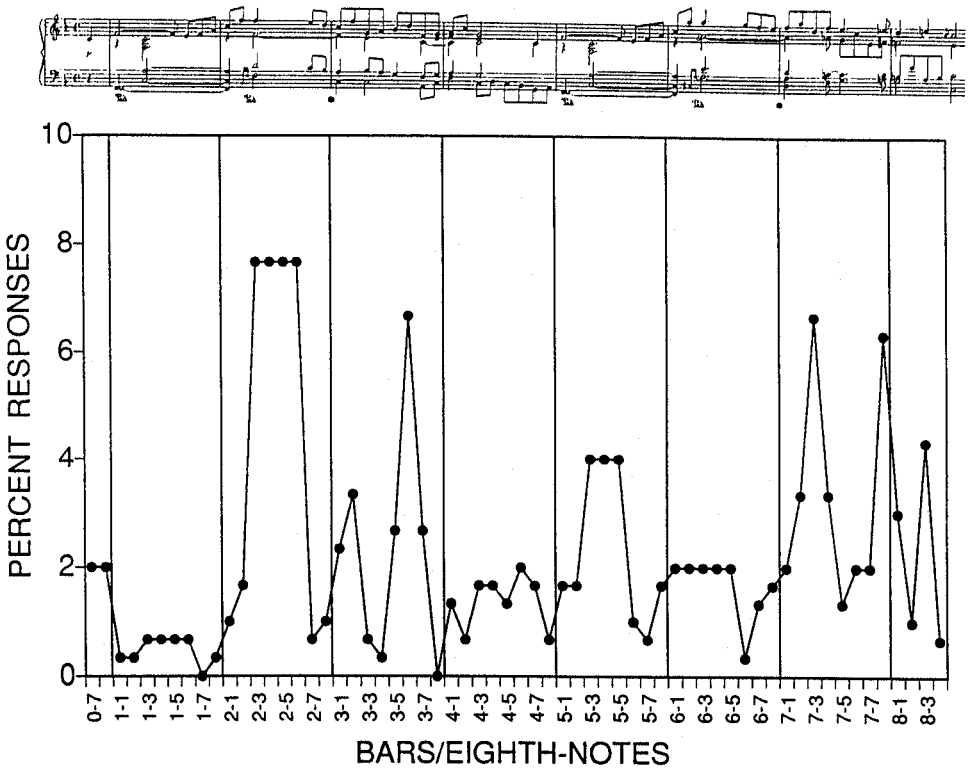

Figure 10. *False-alarm distribution in Experiment 2.*

Mislocations to the preceding interval were rather infrequent. The percentages of these two types of near-misses are shown, together with the "true hit" accuracy profile, in Figure 11. The results of Tests A and B have been combined in this figure; thus the scores for long intervals (which derive from Test A only) appear somewhat elevated. Unlike Experiment 1, there was no convincing tendency for the peaks in the (+1) profile to be shifted one position to the left with respect to the peaks in the true hit profile. The correlation between these two profiles is 0.46 ($p < .001$); it is only slightly higher (0.54) when the (+1) profile is shifted one position to the right. Thus, the influence of perceptual bias on correct responses was less evident than in Experiment 1.

Individual differences and musical experience

Individual overall accuracy scores ranged from 52% to 95% correct in Test A, and from 42% to 80% correct in Test B. None of the (generally positive) correlations with measures of musical experience reached significance. Even though these subjects had, on the whole, received more musical training than the subjects of Experiment 1, only two subjects indicated more than a passing

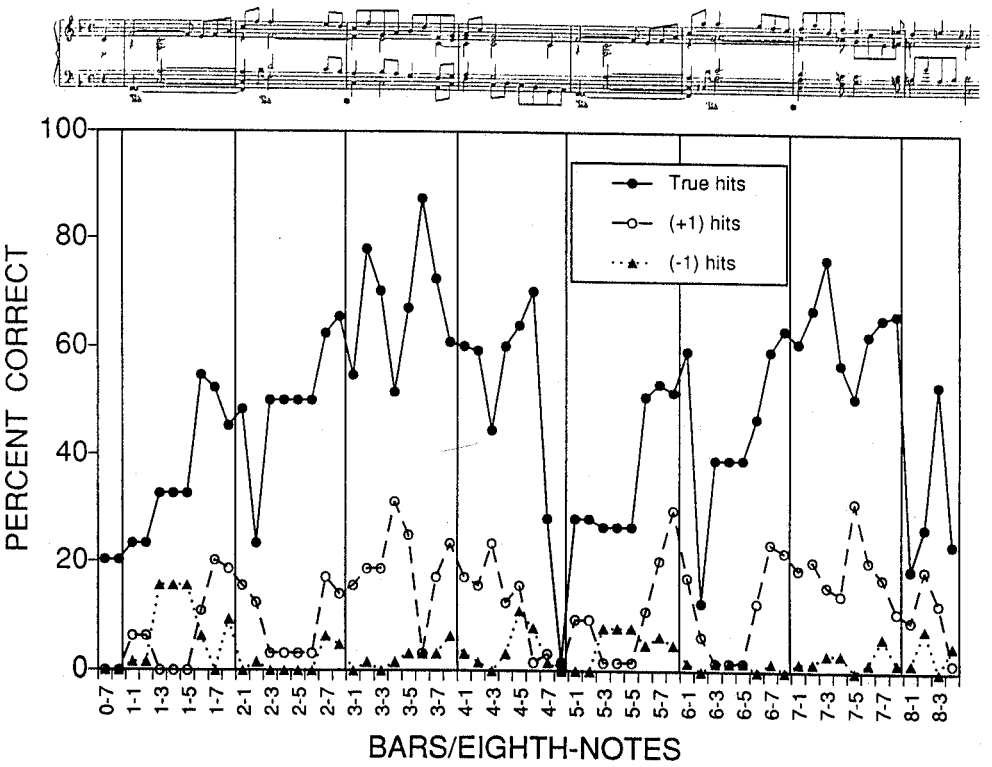

Figure 11. *Accuracy profiles for true hits, and for (+1) and (−1) hits (near-misses) in Experiment 2.*

familiarity with the music.[20] The subject with the highest scores in both tests had received only limited musical instruction and neither played nor listened to classical music. The next highest score was achieved by an accomplished pianist, but the third-highest score belonged to one of the two subjects in this group who had no musical training at all. It is possible that structurally guided expectations inhibit rather than facilitate the detection of lengthening, which would counteract any advantage musicians may have in general rhythmic sensitivity.

GENERAL DISCUSSION

On the whole, the two experiments reported here support the hypothesis that lengthening of musical intervals is more difficult to detect where lengthening is expected. Expectation was defined here mainly with reference to the hierarchical

[20]All subjects in this study had heard the excerpt several times in a lecture given by the author several weeks to several months preceding the experiment. The presentation included an expert performance–the recording by Alfred Brendel.

grouping structure and its reflection in the timing microstructure of expert performances. In both experiments, there was a highly significant negative correlation between a representative performance timing profile and listeners' detection accuracy profile: where performers typically slow down, listeners had trouble detecting lengthened intervals as hesitations. The reason for this is presumably that lengthened intervals in these positions sound regular, while lengthened intervals in other positions are perceived as hesitations that disrupt the rhythm.

It is noteworthy that this result was obtained in the context of a psychophysical detection task. It would be less surprising, for example, if listeners had been asked to provide aesthetic judgments of noticeably lengthened intervals and had been found to be in agreement with expert performance practices. The present listeners, however, were not making aesthetic judgments but listened very carefully for a perturbation in an isochronous sequence of sonic events. Some listeners were remarkably accurate in this task. Nevertheless, they consistently missed certain intervals – usually the ones that were most likely to be lengthened in music performance.

One interpretation of these findings is that, despite the simple task, listeners processed (and learned) the musical structure, which in turn elicited expectations of timing microstructure via some internal representation of performance rules. These top-down expectations *interacted* with subjects' perception of the temporal intervals; they made certain intervals seem longer than others, on which actual lengthening then was especially easy to detect, whereas other intervals seemed relatively short and were difficult to perceive as lengthened. Microstructural expectations thus "warped" the musical time scale in accordance with the perceived grouping structure. This may be a manifestation of the "dynamic attending" discussed by Jones and Boltz (1989).

If this interpretation is correct, several corollaries follow. First, most, if not all, listeners must have grasped the musical grouping structure, at least at the higher levels of the hierarchy, even though a number of them had little music training beyond an ability to read notation. Second, the processing of musical structure and the activation of performance expectations seem to be obligatory. Even though listeners were exposed to many repetitions of mechanically regular performances, there was no indication that their expectations weakened (i.e., that the accuracy profile became flatter) in the course of the experimental session. Third, the origin of these expectations must be in subjects' past experiences with music performance in general, or perhaps in even broader principles of prosody and of the dynamic structure of events. They cannot derive from exposure to performances of the specific music excerpts, for many subjects (especially in Experiment 1) were in fact unfamiliar with these excerpts and never were presented with an expressively timed performance in the course of the experiment.

These conclusions are intriguing and potentially important for our understanding of musical perception and judgment. There is an alternative interpretation of the data, however, which we need to examine closely now: the observed variations in detection accuracy may be direct reflections of the complex *acoustic properties* of the musical stimuli, rather than of expectations elicited by the musical structure as such. The apparently obligatory nature of the phenomenon, its consistency across listeners of widely varying musical backgrounds, and its independence from specific experience with the music would all be consistent with a "bottom-up" account.

It is difficult, of course, to dissociate acoustic from musical structure: music *is* an acoustic structure, and acoustic variations along a number of dimensions in fact define the musical structure for the listener.[21] Although acoustic structure, therefore, is indispensable in music, it may be asked whether there are basic psycho-acoustic principles that make it possible to explain the local variation in detection accuracy without explicit reference to musical structure. If that were the case, it would seem that composers have taken advantage of certain principles of auditory perception in building their musical structures, so as to make recovery of the grouping structure easy and natural. A good example of this kind of argument may be found in David Huron's work (Huron, 1989, 1991; Huron & Fantini, 1989), which demonstrates that J.S. Bach in his polyphonic compositions unwittingly implemented principles of auditory scene analysis (Bregman, 1990).

In another relevant study, Krumhansl and Jusczyk (1990) demonstrated that 4–6-month-old infants are sensitive to phrase boundaries in music: infants preferred listening to Mozart minuets that were interrupted by 1-s pauses at phrase boundaries, rather than to minuets that were interrupted in the middle of phrases. Since it seemed unlikely that infants that young would have had sufficient exposure to music to acquire performance expectations or even a representation of the tonal system (cf. Lynch, Eilers, Oller, & Urbano, 1990), Krumhansl and Jusczyk searched for and identified (in their materials) several acoustic correlates of phrase endings: drops in melody pitch, longer melody notes, and presence of octave intervals. If these factors are operative for infants, they should provide phrase boundary cues for adults as well.

Psycho-acoustic studies with simple sequences of tones have demonstrated that both infants and adults find it more difficult to detect an increment in the duration of a between-group interval than of a within-group interval (Fitzgibbons, Pollatsek, & Thomas, 1974; Thorpe & Trehub, 1989; Thorpe, Trehub, Morrongiello, & Bull, 1988). In these studies, two groups of identical tones were segregated by a change in frequency or timbre. Musical grouping structures are considerably more

[21] Even though musical structure is commonly discussed with reference to the printed score, it must be based on the analyst's sonic image to have any value; purely visual structure in music notation is irrelevant, though it often parallels the auditory/cognitive structure.

complex, of course, and a number of sometimes conflicting factors may be relevant at the same time (cf. Lerdahl & Jackendoff, 1983; Narmour, 1989).

Five such factors were considered in the following analysis. The second and third variables identified by Krumhansl and Jusczyk (1990) were not included because they seem to pertain only to major phrase boundaries, and more to the effects of interruption than of lengthening, which typically is observed before the end of a long tone (if shorter tones accompany it).

Possible psycho-acoustic factors

Pitch change

It is well known from many psychophysical studies that the difficulty of detecting or accurately judging a temporal interval between two tones increases with the pitch distance between the tones (see Bregman, 1990, for a review and many references; also, Hirsh, Monahan, Grant, & Singh, 1990). It could be that, in the present experiments, lengthening was more difficult to detect when the lengthened tone was followed by a large pitch skip. This simple hypothesis is actually difficult to evaluate because, in the musical excerpts used, there were often several simultaneous tones and pitch progressions in several voices. The additional assumption was made, therefore, that listeners paid attention to the principal melody tones, and to other tones only if they did not coincide with a melody tone. This is quite plausible because the melody tones were also louder than the others. This assumption reduces the Beethoven minuet of Experiment 1, for example, to the melody in the upper voice, interspersed with the single notes of the middle voice accompaniment whenever no new melody note occurs (cf. Figure 1). The melody and the middle-voice accompaniment each move in small pitch steps, usually less than 4 semitones, but at points where there is a switch from one voice to the other, larger intervals occur. Large descending pitch skips from the upper to the middle voice occur after the first beat in each bar, which always coincides with a long melody note. This is indeed where dips in the accuracy profile tended to occur in Experiment 1 (Figure 3). Conversely, however, the ascending skips from the middle voice back to the upper voice are associated with *peaks* in detection performance, as well as in the false-alarm profile (Figure 4). Since there is no obvious psycho-acoustic reason why ascending skips should *facilitate* detection performance, the absolute pitch distance hypothesis fails to explain the pattern in the data. Directional pitch change, however, may be a relevant variable. It was expressed in semitones and included in the regression analysis reported below.

Intensity

A second possible hypothesis is that the relative intensities of the tones influenced

the detectability of lengthening. The musical excerpts in both experiments had an intensity microstructure derived from real performances. It could be that louder tones sound longer than softer tones, so that their lengthening is easier to detect. The opposite could also be true, however, if listeners compensate perceptually for a positive correlation between lengthening and intensity in performance. That is, listeners may expect a loud (accented) tone to be longer but, since it has the same duration as the others, it may sound relatively short. To test these hypotheses, the tone intensities (expressed as MIDI velocities) were included in the regression analysis. Whenever several tones coincided, the intensity of the loudest tone was taken.

Intensity change

Another reasonable hypothesis is that the change in intensity from one tone to the next may have played a role. A loud tone may mask the onset of a following soft tone, so that the loud tone sounds longer and the soft tone shorter. This hypothesis was evaluated by including the intensity difference between each tone and the next (taking the loudest tone in a chord) in the regression analysis.

Tone density

Some eighth-note intervals were marked by the onset of a single tone, others by several simultaneous tones. It might be hypothesized that clusters of tones sound inherently longer (or perhaps shorter?) than single tones, so that lengthening then is correspondingly easier (or more difficult) to detect in chords. This hypothesis, which is perhaps the least plausible, was tested by including the number of simultaneous tone onsets as a variable.

Change in tone density

The final variable to be considered was the change in tone density from one eighth-note interval to the next. The relevant hypothesis is similar to that for intensity change.

Regression analysis

Two stepwise linear multiple regression analyses were conducted on the data of each experiment, with either percent correct or false-alarm frequencies as the dependent variable, and the five independent variables listed above. In Experiment 1, the first two and the last two eighth-note intervals were omitted because their extremely low scores were probably due to other causes and would have dominated the correlations. In Experiment 2, only the eighth-note data were analyzed (from Test B only in the case of percent correct), and the three eighth-note intervals that contained grace notes were omitted. That way, the Experiment 1 analyses included 43 intervals, and the Experiment 2 analyses

Table 2. *Intercorrelations among the variables in Experiments 1 and 2*

(a) Experiment 1 ($n = 43$)

	Percent correct	False alarms	PC	I	IC	D
Pitch change (PC)	0.38	0.40				
Intensity (I)	−0.07	0.08	−0.38			
Intensity change (IC)	0.51	0.53	0.85	−0.54		
Density (D)	−0.29	−0.07	−0.47	0.72	−0.53	
Density change (DC)	0.49	0.44	0.85	−0.38	0.79	−0.61

(If $r > 0.47$, $p < .001$; $r > 0.38$, $p < .01$; $r > 0.29$, $p < .05$)

(b) Experiment 2 ($n = 40$)

	Percent correct (Test B)	False alarms (A + B)	PC	I	IC	D
Pitch change (PC)	−0.10	−0.23				
Intensity (I)	0.32	0.45	−0.39			
Intensity change (IC)	0.12	−0.21	0.68	−0.60		
Density (D)	−0.10	0.11	−0.24	0.21	−0.27	
Density change (DC)	0.41	0.21	0.15	0.16	0.03	−0.67

(If $r > 0.49$, $p < .001$; $r > 0.39$, $p < .01$; $r > 0.30$, $p < .05$)

included 40. The intercorrelations among the variables are of greater interest than the regression equations themselves which, with one exception, included only a single independent variable. The intercorrelations are shown in Table 2.

In Experiment 1, three variables (pitch change, intensity change, and density change) correlated significantly with the perceptual results. However, they were also highly intercorrelated, so that only one of them contributed to the regression equation. That was the variable with the highest correlation, intensity change, although density change clearly was an equivalent predictor, at least for percent correct. In other words, in the Beethoven minuet, ascending pitch changes went together with increases in amplitude and with increases in tone density, all of which occurred at the onsets of melodic gestures and facilitated detection of lengthening preceding the change. Conversely, decreases in these three variables went along with poor detection scores. It thus appears that intensity and density decreases, and to a lesser extent pitch decreases, served as natural boundary markers. Of course, this was in large part due to the switching between melody and accompaniment. In the false-alarm analysis, the intensity variable made an additional contribution to the regression equation beyond the effect of intensity change; residual false-alarm rates were positively correlated with intensity. However, the regression equations explain only 26% of the variation in the accuracy profile and 48% of the variation in the false-alarm distribution.

The results of Experiment 2 were different. Although pitch change and

intensity change were correlated with each other, neither of them correlated with either density change or the dependent variables. Density change, however, was a significant, though weak, predictor of detection performance, accounting for 17% of the variance. Intensity, too, showed a weak correlation with percent correct, but did not make a significant additional contribution to the regression equation. Intensity was a stronger correlate of the false-alarm percentages, but again the variance accounted for was small (20%). No additional variables made a significant contribution to the regression equation.

In summary, then, it appears that tone density change was most consistently related to detection performance: when a chord was followed by a thinner texture or a single (accompanying) tone, listeners tended to have difficulty perceiving lengthening of the intervening interval. In other words, these intervals sounded relatively short, to the subjects.[22] This cannot have been due to masking of following tone onsets, which predicts the opposite. Thus it is not clear whether the effect of density change represents an independent psycho-acoustic effect, or whether it is simply a conventional marker of metric strength and/or grouping boundaries that elicited the performance expectations of musical listeners. Experiments employing this variable in nonmusical contexts or with musically inexperienced listeners remain to be done. Also, it must be kept in mind that most of the variation in the accuracy profiles remains unexplained in terms of local acoustic variables. A larger proportion of the variance, at least at a global level, is accounted for by the correlations with the expressive timing profile of expert performances. Whatever the factors are that influence perception, they seem to constrain production as well. These factors are perhaps better captured in an abstract description of the hierarchical metric and grouping structure, which is served by a variety of acoustic variables. A purely reductionistic explanation may not be the most parsimonious one when real music is the subject. Nevertheless, analytical research remains to be done to sort out the many interacting factors that may constrain perception of lengthening.

Parallels with language

The activities of producing and perceiving melodic gestures and phrases are analogous in many ways to those of producing and perceiving prosodic constituents in spoken language (see, for example, Hayes, 1989). The language

[22]This statement is based on the fact that false alarms were rare on these chords. The argument throughout this paper has been that intervals that are expected to be long sound short to listeners under conditions of isochrony. This is in contrast to the "duration illusion" discussed by Thorpe and Trehub (1989) and by earlier authors cited there, according to which a between-group interval sounds longer than a within-group interval. One possible reason for this discrepancy is that the duration illusion concerns silent intervals. If group-final sounds are expected to be lengthened and therefore are perceived as relatively short, a following silence may be perceived as relatively long.

situation most comparable to performing composed music would be reading text aloud or giving a memorized speech (rather than conversational speech, which is more analogous to improvisation in music). Experienced speakers, just like musicians, introduce prosodic variations that make the speech expressive and group words into phrasal units. The closest parallels may be found in timing. The phenomenon of *final lengthening* in constituents of varying sizes is well documented in speech production (e.g., Edwards, Beckman, & Fletcher, 1991; Klatt, 1975; Lehiste, 1973), and its parallelism with final lengthening in music has been pointed out by Lindblom (1978) and Carlson et al. (1989), among others. It has also been shown that pauses in speech, when they occur, tend to occur at phrase boundaries (e.g., Hawkins, 1971; Grosjean & Deschamps, 1975), and their duration tends to be proportional to the structural depth of the boundary (Gee & Grosjean, 1983; Grosjean, Grosjean, & Lane, 1979). Listeners, in turn, use perceived lengthening and pausing to determine the boundaries between prosodic units, as has been shown in experiments using syntactically ambiguous or semantically empty speech (e.g., Scott, 1982; Streeter, 1978). The same cues are likely to aid infants in discovering meaningful units in spoken language (Bernstein Ratner, 1986; Hirsh-Pasek et al., 1987; Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright-Cassidy, 1989); in fact, the infant study by Krumhansl and Jusczyk (1990), cited above, replicated for musical phrases what Hirsh-Pasek et al. (1987) had found for clauses in speech: infants prefer to hear pauses at structural boundaries signalled by prosodic variables.

Final lengthening (sometimes called pre-pausal lengthening) and pausing seem to serve the same function in speech, with pauses taking over wherever lengthening reaches limits of acceptability. In the present research, it was not necessary to distinguish between lengthening and pausing because the *legato* character of the music made silent pauses inappropriate. Because of the equivalent function of lengthening and pausing as structural markers, it seems quite appropriate to subsume them under a single temporal variable (onset–onset time). The contrast between lengthening and pausing in music pertains to the performance dimension of *articulation* (or connectedness, i.e., *legato* versus *nonlegato*), which did not vary in the pieces considered here. This dimension does not have a clear analogue in speech, where "articulation" refers to something quite different.

There does not seem to be a language study in the literature quite analogous to the present experiments, in which all points of potential lengthening (or pausing) were probed perceptually in a sizeable stretch of speech. However, there is little doubt that results parallel to the present findings would be obtained. Klatt and Cooper (1975) showed that phonetic segment lengthening is more difficult to detect in (absolute) phrase-final position, an effect also obtained in simpler sequences of syllables or nonspeech sounds (e.g., Benguerel & D'Arcy, 1986). Boomer and Dittman (1962) found that pauses are more difficult to detect at a terminal juncture (clause boundary) than within a clause, and Butcher (1980)

reported that the pause detection threshold increases as a function of the depth of the structural boundary at which the pause occurred. Duez (1985), who investigated the perception of pauses in continuous speech, concluded that pauses that occurred at unexpected places were *less* often detected than pauses in boundary locations. This finding may have been due to a response bias: Martin and Strange (1968) found that listeners tended to mislocate pauses towards structural boundaries. Still, Duez's findings are at variance with the present results, which showed no bias to mislocate lengthening at phrase boundaries – quite the opposite. The explanation may be that, in speech, lengthening may occur at structural boundaries even when no pause follows. Pauses tend to be heard following lengthened syllables even if no silence is present (Martin, 1970), which confirms the functional equivalence of lengthening and pausing.

It must be kept in mind that the speech in these pause detection tasks was naturally produced, whereas the music excerpts in the present study were mechanically regular. While it would be straightforward to conduct a pause detection experiment with music that contains expressive timing microstructure, it is difficult to envision an experiment in which lengthened segments or syllables would have to be detected in mechanically regular speech. Even the poorest synthetic speech includes many forms of temporal variation that are inherent to consonant and vowel articulation. Nevertheless, even with all this variation present, listeners are quite sensitive to changes in the durations of individual segments, particularly vowels (e.g., Huggins, 1972; Nooteboom, 1973; Nooteboom & Doodeman, 1980; however, see also Klatt & Cooper, 1975). In music, on the other hand, listeners are not particularly accurate in detecting local temporal changes in a temporally modulated performance (Clarke, 1989).

Nooteboom (1973, p. 25) concluded that "The internal representation of how words should sound appears to be governed by rather strict temporal patterns . . .". The norms of a particular language community seem to dictate a rather narrow range of possible temporal realizations for words at a given speaking rate. Nevertheless, these norms do permit modulations of timing and intensity (together with pitch) when words occur in context, not only to demarcate syntactic boundaries, but also to convey focus, emphasis, and emotional attitude. Listeners also seem to have expectations of these modulations (Eefting, 1991), though probably within relatively wide margins. It is these contextual modulations that have close parallels in the expressive microstructure of music, where phrase structure, emphasis, and emotions are conveyed by timing and intensity (as well as notated pitch) variations. Listeners also have expectations about these variations, and although it is difficult to gauge at present how precise these expectations are, they may be quite analogous to those for speech. Specific expectations due to musical training and sophistication are probably superimposed on a substrate of very general principles applying to speech, music, and perhaps auditory events in general. These general principles include group-final

lengthening as well as lengthening and raising of intensity to convey emphasis. Louder and longer events attract attention, so speakers and performers use this device to draw the listener's attention to certain points in the acoustic structure. Final lengthening may be a general reflection of relaxation and of the replenishing of energy (e.g., through breathing) to start a new cycle of activity. Thus the prosodic and microprosodic structures of speech and music may rest on the same general foundation, which is derived from our experience with real-world events and activities.

This common foundation may not only arise from a general desire to imbue music and speech with "living qualities" (Clynes & Nettheim, 1982), but it may be a direct consequence of the hierarchical event structure that music and speech have in common, and of the processing requirements that production and perception of such structures entail. This argument has been pursued by such authors as Cooper and Paccia-Cooper (1980), Gee and Grosjean (1983), and Wijk (1987) for speech production, and by Todd (1985) for music performance. The characteristic slowing down or pausing at structural boundaries enables speakers and performers to plan the next unit, and it in turn enables listeners to group what they hear into meaningful parts by closing off one unit and "laying the foundation" (Gernsbacher, 1990) for the next one. The presence of an appropriate performance (micro)structure may not be absolutely necessary for basic comprehension, but it certainly facilitates "structure building" (Gernsbacher, 1990) for the listener. What the present results seem to demonstrate is that, for music at least, listeners build their mental structures anyway, based on whatever clues are available in the music played. However, the absence of timing microstructure may be perceived as a lack of humanity and consideration on the part of the performer (a computer, in this instance). A poor music performance, like a poor public speech, may be offensive to a sensitive listener, whereas expert performers are loved by the audience.

Implications for musical aesthetics

The evaluation of musical performances is a topic that has barely been touched by psychologists, though it is of vital interest to music critics, performers, and teachers. Perceptual results of the kind reported here, in conjunction with analyses of musical microstructure in representative samples of performances (Repp, 1990b, 1992), begin to provide an objective foundation for judgments of performance quality, particularly in educational contexts. If it is indeed true, as the present findings suggest, that certain performance variations sound more "natural" than others at a perceptual level, then it is presumably part of the performer's task to provide these variations ("to discharge faithfully their aesthetic responsibilities", as Narmour, 1989, p. 318, expresses it). Once a good

model of these natural variations is available (which is still a task for the future; but see Todd, 1985, 1989, and Gee & Grosjean, 1983, for speech), performances may be judged with respect to the degree to which they meet the model's predictions, and these judgments may well be supported by the evaluations of expert listeners and critics.[23] Furthermore, such a normative model then could provide a basis for characterizing the nature of differences among performances that are perceived as equally adequate with respect to the model. Such differences may include the expression of different underlying structures for the same music, as well as different scalings of the magnitude of the microstructural variations. Musical performance (questions of technical accuracy aside) could thus be decomposed into two aspects: (1) expression of structure, which is open to objective evaluation; and (2) choice of scale, which is a matter of individual preference, within broad limits. Similarly, listeners' expectations could be partitioned into a shared structural component (leaving aside instances of structural ambiguity) and individual scale preferences, if any. This is surely an oversimplification of such complex activities as musical performance and judgment, but it is a modest beginning of trying to analyze these processes, which traditionally have been shrouded in mystery.

References

Benguerel, A.-P., & D'Arcy, J. (1986). Time-warping and the perception of rhythm in speech. *Journal of Phonetics*, *14*, 231–246.

Bernstein Ratner, N. (1986). Durational cues which mark clause boundaries in mother-child speech. *Journal of Phonetics*, *14*, 303–309.

Boomer, D.S., & Dittmann, 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.

Carlson, R., Friberg, A., Frydén, L., Granström, B., & Sundberg, J. (1989). Speech and music performance: Parallels and contrasts. *Contemporary Music Review*, *4*, 391–404.

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

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

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

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

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg

[23] I am not trying to suggest here that the most "natural" performances are necessarily the most preferred. They should be perceived as pleasing and, at their best, as beautiful. Truly noteworthy performances will often be deviant in some respect, however, just as fascinating faces, landscapes, or personalities have nonideal properties.

(Ed.), *Studies of music performance* (pp. 76–181). Stockholm: Royal Swedish Academy of Music.

Clynes, M., & Nettheim, N. (1982). The living quality of music: Neurobiologic patterns of communicating feeling. In M. Clynes (Ed.), *Music, mind, and brain* (pp. 47–82). New York: Plenum Press.

Cooper, W.E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Duez, D. (1985). Perception of silent pauses in continuous speech. *Language and Speech, 28*, 377–389.

Edwards, J., Beckman, M.E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society, 89*, 369–382.

Eefting, W. (1991). The effect of "information value" and "accentuation" on the duration of Dutch words, syllables, and segments. *Journal of the Acoustical Society of America, 89*, 412–424.

Fitzgibbons, P.J., Pollatsek, A., & Thomas, I.B. (1974). Detection of temporal gaps within and between tonal groups. *Perception & Psychophysics, 16*, 522–528.

Gabrielsson, A. (1987). Once again: the theme from Mozart's Piano Sonata in A major (K. 331). A comparison of five performances. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81–103). Stockholm: Publications issued by the Royal Swedish Academy of Music, No. 55.

Gee, J.P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*, 411–458.

Gernsbacher, M.A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Grosjean, F., & Collins, M. (1979). Breathing, pausing and reading. *Phonetica, 36*, 98–114.

Grosjean, F., & Deschamps, A. (1975). Analyse contrastive des variables temporelles de l'anglais et du français: Vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica, 31*, 144–184.

Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology, 11*, 58–81.

Hartmann, A. (1932). Untersuchungen über metrisches Verhalten in musikalischen Interpretationsvarianten. *Archiv für die gesamte Psychologie, 84*, 103–192.

Hawkins, P.R. (1971). The syntactic location of hesitation pauses. *Language and Speech, 14*, 277–288.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (Eds.), *Phonetics and phonology, vol. 1: Rhythm and meter* (pp. 201–260). New York: Academic Press.

Henderson, M.T. (1936). Rhythmic organization in artistic piano performance. In C.E. Seashore (ed.), *Objective analysis of music performance* (pp. 281–305). Iowa City, IA: The University Press (University of Iowa Studies in the Psychology of Music, Vol. IV).

Hirsh, I.J., Monahan, C.B., Grant, K.W., & Singh, P.G. (1990). Studies in auditory timing: I. 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.

Huggins, A.W.F. (1972). Just noticeable differences for segment duration in natural speech. *Journal of the Acoustical Society of America, 51*, 1270–1278.

Huron, D. (1989). *Voice segregation in selected polyphonic keyboard works by Johann Sebastian Bach*. Unpublished doctoral dissertation, University of Nottingham, UK.

Huron, D. (1991). The avoidance of part-crossing in polyphonic music: Perceptual evidence and musical practice. *Music Perception, 9*, 93–104.

Huron, D., & Fantini, D. (1989). The avoidance of inner-voice entries: Perceptual evidence and musical practice. *Music Perception, 7*, 43–48.

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

Kemler Nelson, D.G., Hirsh-Pasek, K., Jusczyk, P.W., & Wright-Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16*, 53–68.

Klatt, D.H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics, 3*, 129–140.

Klatt, D.H., & Cooper, W.E. (1975). Perception of segment duration in sentence contexts. In A. Cohen & S. G. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 69–89). New York: Springer-Verlag.

Krumhansl, C.L., & Jusczyk, P.W. (1990). Infants' perception of phrase structure in music. Psychological Science, *1*, 70–73.

Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America*, *54*, 1228–1234.

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

Lindblom, B. (1978). Final lengthening in speech and music. In E. Gårding, G.Bruce, & R. Bannert (Eds.), *Nordic prosody* (pp. 85–101). Lund University: Department of Linguistics.

Lynch, M.P., Eilers, R.E., Oller, D.K., & Urbano, R.C. (1990). Innateness, experience, and music perception. *Psychological Science*, *1*, 272–276.

Martin, J.G. (1970). On judging pauses in spontaneous speech. *Journal of Verbal Learning and Verbal Behavior*, *9*, 75–78.

Martin, J.G., & Strange, W. (1968). The perception of hesitation in spontaneous speech. *Perception & Psychophysics*, *3*, 427–438.

Narmour, E. (1989). On the relationship of analytical theory to performance and interpretation. In E. Narmour & R. Solie (Eds.), *Explorations in music, the arts, and ideas: Essays in honor of Leonard B. Meyer* (pp. 317–340). New York: Pendragon.

Nooteboom, S.G. (1973). The perceptual reality of some prosodic durations. *Journal of Phonetics*, *1*, 25–45.

Nooteboom, S.G., & Doodeman, G.J.N. (1980). Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America*, *67*, 276–287.

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

Povel, D.-J. (1977). Temporal structure of performed music: some preliminary observations. *Acta Psychologica*, *41*, 309–320.

Repp, B.H. (1990a). Further perceptual evaluations of pulse microstructure in computer performances of classical piano music. *Music Perception*, *8*, 1–33.

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

Repp, B.H. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei". Manuscript submitted for publication.

Scott, D.R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America*, *71*, 996–1007.

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

Sloboda, J.A. (1985). Expressive skill in two pianists: Metrical communication in real and simulated performances. *Canadian Journal of Psychology*, *39*, 273–293.

Streeter, L.A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America*, *64*, 1582–1592.

Sundberg, J. (1988). Computer synthesis of music performance. In J.A. Sloboda (Ed.), *Generative processes in music* (pp. 52–69). Oxford: Clarendon 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. (1985).A model of expressive timing in tonal music. *Music Perception*, *3*, 33–58.

Todd, N.P. (1989). A computational model of rubato. *Contemporary Music Review*, *3*, 69–88.

Wijk, C. van (1987). The PSY behind PHI: A psycholinguistic model for performance structures. *Journal of Psycholinguistic Research*, *16*, 185–199.