Variations on a Theme by Chopin: Relations Between Perception and Production of Timing in Music

Bruno H. Repp
Haskins Laboratories

A note interonset interval (IOI) increment in mechanically timed music is more difficult to detect where expressive lengthening typically occurs in artistic performance. Experiment 1 showed this in an excerpt from a Chopin etude and extended the task to IOI decrement detection. A simple measure of variation in perceptual bias was derived that correlated highly with the average timing pattern of pianists' performances, more so than with acoustic surface properties of the music. Similar results, but decreasing correlations, were obtained in each of four subsequent experiments in which the music was simplified in stages. Although local psychoacoustic effects on time perception cannot be ruled out completely, the results suggest that musical structure (melodic-rhythmic grouping in particular) has temporal implications that are reflected not only in musicians' motor behavior but also in listeners' time-keeping abilities.

This research was concerned with the perception of timing in music relative to timing in music performance. It is well-known that music is not played the way it looks in a printed score. Musical notation is an abstraction that does not fully represent the details of musicians' performances, just as printed text does not represent a speakers' tempo, rhythm, and intonation. The most important parameters of what Clynes (1983) has termed the "expressive microstructure" of music are timing, dynamics, and articulation. The focus of this research was on timing; dynamics (i.e., the relative intensity of musical tones) played a secondary role, and articulation (i.e., the manner of attack and connectedness of successive tones) was not varied.

There are now many empirical studies of expressive timing, mostly in piano performance, that demonstrate that performers rarely produce the precise ratios between note interonset intervals (IOIs) that are documented in a score (see, e.g., Drake & Palmer, 1993; Epstein, 1995; Gabrielsson, 1974, 1987; Henderson, 1936; Palmer, 1989; Repp, 1992c, 1995b; Todd, 1985). An expressive performance—particularly of Western art music from the 19th century—is often characterized by continuous tempo modulations. A plot of the (normalized) IOI durations as a function of score position is one way of representing the *timing profile* of a performance. Individual performances differ in their timing profiles, which is one reason why it would be futile to represent this detailed variation in a score. However, expressive timing is far from arbitrary, and there is considerable commonality among the timing profiles of different musi-

cians, notwithstanding the occasional deviant pattern. The commonality is especially striking among less experienced artists such as advanced student pianists, whose individual timing profiles may be considered to be variants of a shared standard or norm—a typical timing profile—that may be approximated by the average timing profile of a group of pianists (Repp, 1995b, 1997a).

The existence of such typical timing patterns has made it possible to formulate rules and models of expressive timing, some of which have been applied in automatic performance systems with varying success. A detailed review of this work is beyond the scope of this article, but it is fair to say that current models are far from complete. Two different approaches may be distinguished:[1] The rules developed by the Stockholm group (Friberg, 1991; Sundberg, Friberg, & Frydén, 1989) are based mostly on local surface features of the music; for example, they include the enhancement of durational contrast between successive tones, the shortening of tones in an ascending melody, the lengthening of tones preceding a pitch jump, and the lengthening of metrically stressed tones in certain contexts. By contrast, the model of Todd (1985, 1995) is based on an analysis of the hierarchical phrase (or grouping) structure of a composition (Lerdahl & Jackendoff, 1983). It postulates temporal acceleration-deceleration shapes within each phrasal unit, and the performance timing profile is assumed to result from the superposition of several hierarchically nested timing shapes. While the Stockholm rules have been evaluated primarily by ear, Todd's model has been shown to capture major features of the timing profiles of actual performances by pianists. However, although Todd's model has the structural basis that the Stockholm model lacks, it does not account for systematic local deviations from the timing shapes. A complete

---

Correspondence concerning this article should be addressed to Bruno H. Repp, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695. Electronic mail may be sent to repp@haskins.yale.edu.

---

[1] A third approach, too complex and idiosyncratic to be considered here, is that of Clynes (1983), which nevertheless has led to the most impressive computer synthesis results to date.

model may have to include both high- and low-level rules (see Oosten, 1993).

An interesting question related to the topic of this article is to what extent deviations from strict timing in music performance are intentional and to what extent they are obligatory. In several studies, researchers have instructed pianists to play without expression but still have found systematic variation in timing. Timing variation due to purely technical difficulties (e.g., large jumps on the keyboard) can usually be ruled out as a factor in these studies. Thus, Gabrielsson (1974) asked pianists and percussionists to tap out simple rhythms to a metronome beat but nevertheless found systematic timing deviations within each beat. For example, in a sequence of two shorter notes preceding a longer note, the second IOI was typically longer than the first. Drake and Palmer (1993) compared pianists' performances of simple rhythms, with instructions to play first "in a normal musical fashion" and then "in a flat, mechanical way." Although the instructions had an effect on dynamics and articulation, they left the timing patterns unchanged. As in Gabrielsson's study, there was a consistent tendency to lengthen the IOI preceding a longer note. Although this could be seen as a purely local effect, Drake and Palmer interpreted it as a lengthening of the final IOI in rhythmic groups because longer notes usually define the end of a rhythmic group.

Indeed, this group- or phrase-final lengthening is a ubiquitous phenomenon in expressive performance that seems to account for much of the timing variation in more complex music (Drake & Palmer, 1993; Repp, 1992c; Todd, 1995). However, when Palmer (1989) asked pianists to play excerpts from pieces by Mozart and Brahms "unmusically" (i.e., without expression), she found a considerable reduction of expressive timing variation, although the remaining timing profiles showed some similarity to the profiles for "musical" performances. Thus, the magnitude of expressive timing variations in realistic performance is under musicians' control, but underlying them there may be smaller, involuntary tendencies that may also have been captured with the simpler materials of Gabrielsson (1974) and Drake and Palmer (1993). It is tempting to align this voluntary–involuntary distinction with that between structure-based and local timing effects: Voluntary timing may be the consequence of cognitive structural analysis ("interpretation") of the music, whereas involuntary timing may have lower level origins, perhaps in local auditory interactions or in the kinematics of keyboard performance. However, it is also possible that the involuntary timing is a residual consequence of structural perception that falls below the perceptual discrimination threshold and is therefore unconscious and uncontrollable. In contrast to the flexibility of cognitively guided expressive timing, which is often highly individual or idiosyncratic, this residual timing variation would presumably be tied closely to the auditory structure of the music (i.e., it would represent a typical timing pattern).

Given that musicians produce systematic timing variations in performance, it is natural to hypothesize that passive listeners will expect to hear such variations when they listen to music, at least if they are familiar with the musical style

and the general principles underlying expressive timing. This can be easily confirmed in terms of aesthetic judgment: Music without expressive timing sounds dull, and music with atypical or random timing generally sounds bad. However, the finding that musicians find it difficult to play with perfectly even timing suggests that there may be a connection not only with aesthetics but with the actual perception of timing. For example, one might predict that a performance that the musician intended to be evenly timed but that in fact contains residual timing variation will be judged by listeners to be more evenly timed than an evenly timed (computer-controlled) performance. To my knowledge, this prediction has not yet been tested directly, but indirect tests of a perception–performance connection with regard to timing have been conducted.

Repp (1992a) presented listeners with piano music played under computer control without expressive timing, so that all IOIs were of equal duration. The listeners' task was to detect small increments in single IOIs in different positions of the music. Over several trials, all positions were "probed" in this way, and, by averaging across several increment sizes and a number of listeners, a *detection accuracy profile* (DAP) was obtained that portrayed the average percentage of correct detections as a function of position in the music. This profile was highly varied and was negatively correlated with a typical expressive timing profile of the music derived from pianists' performances: The lengthening of an IOI was more difficult to detect when that IOI was likely to be lengthened in performance, which naturally occurred especially at the ends of melodic-rhythmic groups and phrases. Such a perception–performance correlation was obtained even with simple monophonic melodies (Repp, 1992b, 1995c), although these materials—and consequently their DAPs and performance timing profiles—had less internal structure, and the correlation rested mainly on a decline in detection accuracy that paralleled the phrase-final *ritardando*.

One interpretation of these findings, favored by Repp (1992a), is that they reflect listeners' cognitive expectations of expressive timing. These expectations may have been acquired from extensive exposure to music performances that exhibit expressive timing, which include not only classical music but various forms of "soft" popular music and film music. In this view, then, the DAP represents listeners' perceptual compensation for the absence of typical expressive timing. The listeners' expectations result from their processing of the musical structure in conjunction with their past musical experience, and these expectations are brought to bear on the music and interact somehow with the perception of musical time. Consequently, local deviations from even timing that are in agreement with expectations are more difficult to detect (i.e., sound more even or natural) than deviations that violate expectations.

An alternative, psychoacoustic explanation of a perception–performance correlation has been proposed by Drake (1993): The complex acoustic structure of music may give rise to local auditory interactions that affect the detectability of timing deviations and the perceived relative durations of IOIs. In this view, it is not so much the listener who compensates for the absence of expected timing variation,

but it is the performer who compensates for these auditory distortions in perceived timing. This may apply especially to situations in which the performer intends to produce even timing, as in the experiments by Gabrielsson (1974) and Drake and Palmer (1993). However, it cannot explain the large temporal variation and individual differences encountered in realistic expressive timing. Although it is conceivable that typical expressive timing represents a deliberate exaggeration of basic compensatory actions, even conventional expression is commonly attributed to the performer's analysis and representation of the musical structure (e.g., Clarke, 1985; Epstein, 1995; Palmer, 1989).

Given that music performance is commonly talked about in structural terms, and given that there seems to be such a close parallelism between the production and perception of timing, it also seems appropriate to talk about perception in structural terms. According to Repp's (1992a) expectation hypothesis, timing perception is influenced by timing production (through musical experience); according to Drake's (1993) psychoacoustic hypothesis, it is the other way around. A third possibility, tentatively proposed by Repp (1995c), is that both are reflections of the musical structure, particularly of its division into rhythmic groups. The natural way to represent a cognitive group in action (if it is represented at all) is an acceleration–deceleration pattern, as described by Todd (1985, 1995): Acceleration increases cohesiveness within the group, whereas deceleration creates or enhances boundaries between groups. The perceptual act of grouping may result in a corresponding warping of the subjective flow of time—an "expectation" of deceleration in the vicinity of group boundaries. Grouping structure thus may have "timing demands" that both performers and listeners respond to in analogous ways, except that performers (who start with a cognitive representation or performance plan) can modify their actions at will, whereas listeners (who are processing real-time input) cannot or only to a limited extent. This may be called a "structural perception hypothesis."

To clarify, I assume here that there are three stages in musical perception and performance: auditory processing, structural perception, and structural cognition. Psychoacoustic interactions among elementary parameters such as frequency (pitch), intensity (loudness), and duration take place at the auditory processing stage. Structural perception goes beyond auditory processing and yields a representation of the perceptual organization of a musical object, at least of its grouping structure.[2] Structural cognition operates on this "primary" structure and elaborates, disambiguates, or modifies ("interprets") it. The psychoacoustic, structural perception, and expectation (i.e., structural cognition) hypotheses are aligned with these three stages.

In this research I did not address the distinction between the second and third hypotheses. Whether listeners' "timing expectations" for music are flexible and under cognitive control, as the expectation hypothesis predicts, or inflexible, as the structural perception hypothesis predicts, was investigated in a separate set of experiments, whose results suggest inflexibility (Repp, 1998). There is also little evidence so far that these perceptual "expectations" vary significantly as a

function of musical experience, although musically trained individuals are more accurate overall in detecting timing deviations in music (Repp, 1995c, in press). However, these results are also consistent with the psychoacoustic account because local auditory processing interactions are likely to be inflexible and independent of musical experience.

This article is concerned with the distinction between the psychoacoustic and structural perception hypotheses. However, they are not mutually exclusive, and therefore a decisive empirical test should not be expected. After all, musical structure is conveyed by acoustic and auditory properties, and to the extent that psychoacoustic interactions underlie or are congruent with structural effects, they cannot be empirically distinguished. Rather, the discussion revolves around parsimony and plausibility. With regard to parsimony, it may be asked whether a set of results is explained more simply and exhaustively with reference to elementary properties such as pitch, intensity, and duration or with reference to musical structure. The structural perception account has an inherent advantage here: It takes for granted that the typical performance timing profile is an embodiment of structural effects on timing, and therefore a high perception–performance correlation is prima facie evidence for a structural basis of perceptual effects. The viability of a psychoacoustic account may be gauged by its ability to explain the perceptual results just as well or even better. With regard to its plausibility, it is necessary to link the observed perceptual effects to phenomena that have been obtained in psychophysical research with more primitive stimuli. An ad hoc appeal to novel psychoacoustic interactions would not be convincing.

What sorts of relevant psychoacoustic interactions have been observed previously? Two kinds of effects on timing perception need to be distinguished: those affecting *sensitivity* and those affecting *perceptual bias* (or constant error). More is known about the former than about the latter. For example, it may be inferred from a study by Monahan and Hirsh (1990) that changes in IOI duration are more difficult to detect next to a longer IOI or in the initial or final position of a sequence (i.e., next to a *very* long IOI). At the beginning of an isochronous sequence, it may take a few IOIs to develop a stable perceptual reference for duration discrimination (Drake & Botte, 1993; Schulze, 1978). Furthermore, variations in sensitivity may be caused by variations in the pitch distance between successive tones because of a generalized tendency to assign widely separated tones to different auditory groups or streams, which results in less precisely defined IOIs between such tones (Bregman, 1990; Drake, 1993; Fitzgibbons, Pollatsek, & Thomas, 1974; Thorpe & Trehub, 1989; Thorpe, Trehub, Morrongiello, & Bull, 1988).

---

[2] Of course, there is more to musical structure than just grouping: meter and harmony, for example. Some of these may require cognitive processing, and they may also have an influence on timing. In the present context, however, it is sufficient to refer to grouping as the major determinant of expressive timing in the musical materials used.

A moment's thought reveals, however, that these variations in sensitivity are irrelevant to the perception–performance relationship that is of interest here. By definition, a local change in sensitivity is not direction specific; it applies equally to increments and decrements in IOI duration. Musicians who, according to Drake (1993), would wish to compensate for such an effect would not know whether to lengthen or shorten the relevant IOI; basically, they would be free to do either. The point of expressive timing, however, is that it is direction specific, with relative lengthening of IOIs in some positions (e.g., at the ends of rhythmic groups) and relative shortening in others. Therefore, any corresponding psychoacoustic effects also must be direction specific (i.e., perceptual biases).

These phenomena have received much less attention in the psychophysical literature. One relevant finding is the "auditory kappa effect" (Crowder & Neath, 1995; Shigeno, 1986). It seems that an IOI between tones separated widely in pitch is perceived as being longer than an IOI between tones close in pitch. However, Repp (1995c) found no evidence of this effect in monophonic melodies, and Tekman (1992, 1995) also failed to find a perceptual interaction between pitch and timing. However, Tekman did find an interaction between intensity and timing, in that an IOI preceding a loud tone was perceived as being longer than one preceding a soft tone (see also Woodrow, 1909). In Drake's (1993) study, moreover, there was an indication that an IOI preceding a long IOI was perceived as being relatively short, as one might predict on the basis of the rhythm production data of Gabrielsson (1974) and Drake and Palmer (1993). Thus, there is some evidence that pitch, intensity, and duration all may interact with perception of timing. These effects were considered in this research.

The purpose of Experiment 1 was to replicate the basic perception–performance correlation with a new musical excerpt and to clarify the nature of the perceptual effects embodied in the DAP by extending the experimental paradigm to include both increment and decrement detection. In previous experiments (Repp, 1992a, 1992b, 1995c), I investigated only the detection of increments in IOI duration. The variation in detection accuracy across positions in the music thus may have been a combination of variations in sensitivity and perceptual bias. To be sure, the high correlation of the DAP with a typical performance timing profile indicated that variation in bias predominated, but it was desirable to confirm this and to separate variations in sensitivity from variations in bias. The inclusion of both increment and decrement detection tasks in Experiment 1 made this possible, at least approximately, according to the following simple rationale: Positional variations in sensitivity should affect the DAPs for increments and decrements in the same way, whereas positional variations in bias should have opposite effects on the two DAPs. To the extent that variations in bias predominate, the DAP for decrements should be inversely related to the DAP for increments. A rough estimate of variations in sensitivity (a sensitivity profile) can then be obtained by averaging the two DAPs, whereas an estimate of variations in perceptual bias (a bias profile) can be obtained by calculating (half) their differ-

ence.[3] The sign of the difference was chosen to yield a positive correlation with the timing profile; the bias profile thus represents "expectations of lengthening." Variations in sensitivity were not expected to be related to typical expressive timing in performance, whereas the bias profile was expected to correlate even more strongly with performance timing than either of the two DAPs.

Experiments 2–5 were replications of Experiment 1 with changes in the musical material. These "variations" of the "theme" used in Experiment 1 took the form of increasing simplification, the intention being to proceed from complex original piano music to a simple monophonic tune, thus bridging the gap between the materials of preceding studies in this series (Repp, 1992a, vs. Repp, 1992b, 1995c). Typical timing profiles were obtained from pianists' performances of these variations, so that the perception–performance correlation could be examined at each stage of the journey. The correlation was expected to decrease because of a weakening of the grouping structure with increasing simplification, resulting in a reduction of systematic variation in both perceptual bias and expressive timing profiles; however, a significant correlation was expected to persist even with the simplest materials. The viability of a psychoacoustic explanation of the bias profile was assessed in each experiment by computing multiple correlations with acoustic surface properties of the music. A psychoacoustic account would be convincing if the multiple correlation exceeds the simple perception–performance correlation and if the specific psychoacoustic effects involved seem plausible. In addition, specific consequences of the changes in the musical materials were predicted, and the accuracy of these predictions was examined by comparing results between experiments.

Experiment 1

In Experiment 1 I examined the relative detectability of IOI duration increments and decrements in the different positions of a new musical excerpt taken from the 19th-century piano literature. The methodology was similar to that used by Repp (1992a), but the decrement detection task and the derivation of a bias profile were new features. The perceptual results were compared with a typical timing profile derived from pianists' performances of the music, and their correlations with acoustic surface properties of the music were examined.

Method

Listeners. Fourteen paid volunteers served as listeners. All were musically trained undergraduates who were members of the Yale Symphony Orchestra. Musically trained individuals were preferred because they are sensitive to timing deviations and could

---

[3] In analysis of variance terms, the two profiles represent the position main effect and the Position × Condition interaction, respectively. Although the concepts of sensitivity and bias are strongly associated with signal-detection theory (SDT; see Macmillan & Creelman, 1991), they are used here without recourse to SDT analysis methods, mainly because of the small number of responses per participant and position.

read the musical notation on the answer sheets.[4] Because their primary instrument was not the piano, they were not overly familiar with the music.

Materials. The music, shown in Figure 1, was the beginning of Chopin's Etude in E major, op. 10, No. 3. To give closure to the excerpt, the final tonic chord was prolonged to the end of the bar. With the exception of this final chord and the initial 8th-note upbeat, note onsets occur at regular intervals corresponding to 16th-note divisions in the score (8 per bar). However, the pitches are divided into four horizontal layers or voices that are traditionally referred to as "soprano," "alto," "tenor," and "bass." Continuous 16th-note motion takes place only in the alto voice, which has an accompanying function. The melody is in the soprano voice, and it is divided into five rhythmic groups or melodic segments that end with longer notes, as indicated by brackets above the score in Figure 1. The tenor and bass voices (played by the left hand) form a repetitive rhythmic pattern that reinforces the harmonic alternation of tonic and dominant. Altogether, the music defines 36 IOIs and thus 36 possible "target" positions in the detection task. If necessary, these positions are referred to in terms of bar and 16th-note numbers (e.g., "3-6" refers to the sixth IOI in Bar 3). Note that IOIs between successive notes are defined regardless of voice membership. The fact that there are longer IOIs within the soprano voice is irrelevant to the definition of target IOIs, although it is, of course, of structural significance. Given the small size of the changes to be detected and the well-known increase in the absolute detection threshold with baseline duration (see, e.g., Friberg & Sundberg, 1995), it seems obvious that listeners could detect an IOI increment or decrement only as a change in the target (16th-note) IOI, not as a change in any longer IOI that may occur within a voice.

The musical stimuli were created under computer Musical Instrument Digital Interface (MIDI) control on a Roland RD-250s digital piano with "Piano 1" sound. The baseline 16th-note IOI duration was 500 ms. Each note in the music ended nominally (i.e., with a "note off" MIDI command) when another note began,

which ensured perceived *legato* articulation.[5] The pedal was not used. The MIDI velocities (relative intensities) of the notes varied, which was necessary to differentiate the voices and give the music some minimal expression. The specific velocity values were obtained by averaging the MIDI velocities of three performances by the author (an amateur pianist) on the digital piano; they are shown graphically below the music in Figure 1.[6] Four velocity units are approximately equal to 1 dB in peak sound level (Repp, 1993a). It can be seen that the soprano melody was louder than the other voices and reached a peak in Bar 3, just before the highest pitch; the alto voice was softer but showed a similarly arched shape; the lower voices and the subsidiary alto voice notes were the softest and showed the smallest dynamic variation. The sudden drop in the intensity of the soprano and alto notes in Position 2-5 was an idiosyncratic feature.

Two parallel tests were prepared, one for the detection of IOI increments and the other for the detection of IOI decrements. Each test began with five familiarization trials, which were followed by three blocks of 18 test trials each. There were a few seconds of silence between trials and longer intervals between blocks. Each trial (a complete presentation of the musical excerpt) contained between 0 and 4 lengthened or shortened IOIs ("targets"). In each block, there were 2 trials with 4 targets, 4 with 3, 6 with 2, 4 with 1, and 2 with no target. Within each block, the 36 target positions were allocated randomly to these trials, although successive targets were always separated by at least four baseline IOIs, usually more. Because each position served as a target once in each block, a listener had a total of three opportunities to detect a change in each IOI. The increment and decrement detection tests used exactly the same randomization.

The size of the duration increment or decrement decreased from block to block in each test. It was approximately 52 ms in the 5 familiarization trials and about 42, 31, and 21 ms, respectively, in the three test blocks. To be exact, these were 50, 40, 30, and 20 "ticks" (480 ticks = 500 ms) in the Performer MIDI sequencing program, which was used for stimulus construction and playback. Whenever an IOI was changed in duration, the durations of all notes that were "on" during the IOI were changed in the MIDI instructions, so as to maintain *legato* articulation everywhere. Thus, for example, when the IOI in Position 2-6 was lengthened, five notes were lengthened by the same absolute amount: the two
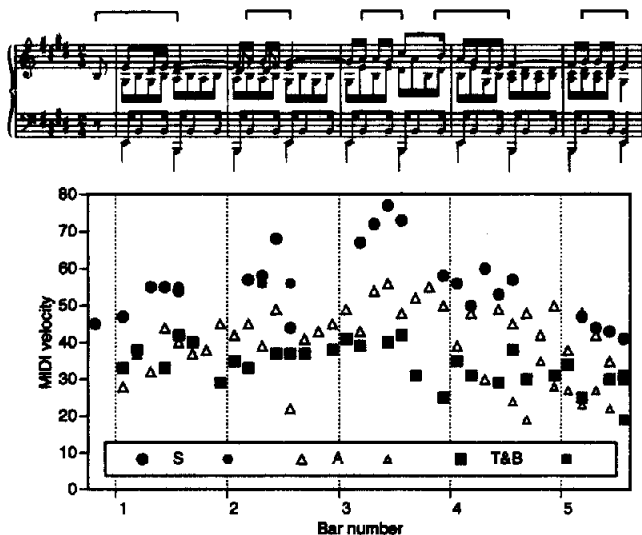

Figure 1. Top: The beginning of Chopin's Etude in E major, op. 10, No. 3, with slurs and expression marks omitted and the final chord extended. Brackets indicate rhythmic groups in the melody. Bottom: The MIDI velocities of the notes, divided into soprano (S), alto (A), and tenor-bass (T&B) voices. Smaller symbols represent subsidiary (lower) notes in each voice.

---

[4] In earlier experiments, musically untrained participants often could not hear small timing deviations, which resulted in a high attrition rate (Repp, 1992b, 1995c). The role of musical training was reinvestigated in a recent study and was found to be small (Repp, in press).

[5] Acoustically, successive tones of different pitch overlapped because piano tones take some time to die away after a key is released; this postrelease decay was present in the synthetic sounds used (see Repp, 1995a).

[6] The performances from which the typical timing profile was derived were not yet available at the time of stimulus construction. If expression is viewed as something that is added to the music by the performer (a score-based perspective), the presence of expressive dynamics may seem undesirable in an experiment on timing. I, however, regard expressive variation as an essential part of the music that can only be changed or removed, not added (a performance-based perspective). From that perspective, the present stimuli were intended to be "music minus expressive timing." Expressive dynamics were removed only in the last stage of simplification in Experiment 5.

sustained quarter notes (E_4, $G\#_4$) in the soprano voice, the 16th note (B_3) in the alto, the 8th note (B_2) in the tenor, and the sustained quarter note (E_2) in the tenor and bass. At the same time, the onsets of all subsequent notes in the music were delayed by the same amount. The change to be detected thus is best described as a phase shift in the isochronous rhythm; there was no compensatory adjustment in the IOI after a target. The analog output of the digital piano was recorded onto digital audiotape (DAT).

Procedure. The test tape was played back on a DAT recorder. The participants listened at a comfortable intensity over Sennheiser HD 540 reference II earphones. Half of them received the increment detection test first and half the decrement detection test. Written instructions explained the experiment in detail, including the fact that there were between 0 and 4 IOI increments or decrements to be detected on each trial. The responses were marked on answer sheets that reproduced the upper staff of the musical score (soprano and alto voices) for each trial. The instructions emphasized that the notes filling the lengthened or shortened IOIs, not the following notes, should be circled with a vertical ellipse. (My earlier studies had shown a strong tendency toward "late" responses.) For the 5 familiarization trials, the correct responses had been filled in on the answer sheet. Random guessing was discouraged; when no change was detected on a trial, a question mark was to be entered on the answer sheet. At the end of the experiment, listeners received a detailed written explanation to take home.

Performance measurements. Nine pianists were recorded playing the Chopin excerpt. Six of them were graduate students of piano at the Yale School of Music, 1 was a talented undergraduate pianist, and 2 were graduate students of musicology with adequate pianistic skills. They were recorded on the Roland RD-250s digital piano, which has a weighted keyboard and is equipped with a simple sustain pedal. Each pianist was given a brief warmup period to get used to the keyboard and rehearse the music. (Some of them had not played the Chopin etude previously, but the initial bars are easy enough to be sightread well after a few attempts.) As soon as the pianist felt ready, he or she played the music three times from the computer score shown in Figure 1 (without the brackets). The data were saved in MIDI format and later imported as text files into a spreadsheet program. The onset of the highest note in each 16th-note position was located and given a label. (These were soprano notes when present and alto notes otherwise.) The differences between the onset times of these notes yielded the IOIs that constituted the timing profile. To reduce uncontrolled variability, I averaged the IOIs over each pianist's three performances. Subsequently, the nine individual average timing profiles were subjected to a principal-components analysis to determine whether they represented more than one underlying timing pattern. The answer was negative: The first principal component, which was equivalent to the grand average timing profile, accounted for 85% of the variance. Thus, it seemed that the student performances represented individual variants of a single normative timing pattern (see Repp, 1995b, 1997a, for related findings and discussion). The grand average timing profile of all 9 pianists then provided a good estimate of a typical timing profile for the Chopin excerpt.[7]

Results and Discussion

Overall accuracy.[8] Responses adjacent to the correct position, both "early" and "late," were scored as correct. This liberal criterion was used to take into account a strong tendency of listeners—despite the explicit instructions—to circle the note(s) after a lengthened IOI, observed also in previous experiments (Repp, 1992a, 1992b, 1995c). Early responses were included because it seemed that they some-
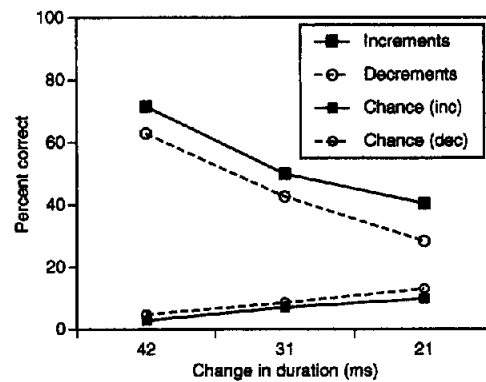

Figure 2. Overall percent correct scores and chance levels for the two tasks of Experiment 1 as a function of change in duration (test block). inc = increments; dec = decrements.

times resulted from uncertainty about the precise position of a change or careless marking of responses, although some of them were probably false alarms. In any case, they were so infrequent that their exclusion would not have changed the data significantly. Early and late responses are discussed in a separate section toward the end of this article.

The overall percent correct detection scores are shown in Figure 2 as a function of condition and test block. They were subjected to a two-way repeated measures analysis of variance (ANOVA). Naturally, scores declined as the difficulty of the task increased, $F(2, 26) = 162.28, p < .0001$. Also, the detection of increments was easier than the detection of decrements, $F(1, 13) = 12.57, p < .004$. The average scores were just right for DAPs because they avoided floor and ceiling effects. Individual grand average scores ranged from 29% to 64% correct.

Figure 2 also shows estimates of chance levels based on the false-alarm responses. These estimates were calculated as follows: If a single random guess were made on a trial, then the probability of being correct, given the lenient scoring criterion, would be approximately $3n/36$ or $n/12$, where n is the number of targets in the trial.[9] Because the average number of targets per trial was 2, the average guessing probability was 1/6. The average rates of false-alarm responses per trial in the three blocks were .167, .424, and .590 in the increment detection condition and .290, .515, and .786 in the decrement detection condition. Because about 1/6 of all true false alarms must have been scored as

---

[7] There is an implicit assumption here that a qualitatively similar, involuntary timing profile would have been obtained if the pianists had been asked to play evenly, without expressive timing. This assumption was supported by recently collected data (Repp, 1997b). Thus, the typical timing profile obtained here was essentially a reflection of "uninterpreted" musical structure.

[8] "Overall" means here "averaged across all positions." Only actual responses ("hits" and "false alarms") were considered in the analysis.

[9] The "approximately" derives from the fact that guessing probabilities are somewhat smaller for targets in the initial and final interonset intervals.

correct responses, a correction was applied by multiplying the false-alarm rates by 6/5. Finally, estimates of chance probabilities were obtained by multiplying the corrected false-alarm rates by the guessing probability of 1/6, and then multiplying by 100 to obtain percentages. It is evident that the average detection scores in the last block of trials were still much better than chance.

Detection accuracy and false-alarm profiles. The two DAPs—the percent correct scores as a function of position in the music averaged over listeners and test blocks—are shown in Figure 3A. First, it can be seen that there was enormous variation in the relative detectability of IOI duration changes across positions in the music. This was true for both increments and decrements. Second, the two profiles were inversely related, as predicted: When the detectability of increments was high, the detectability of decrements was usually low and vice versa. The correlation
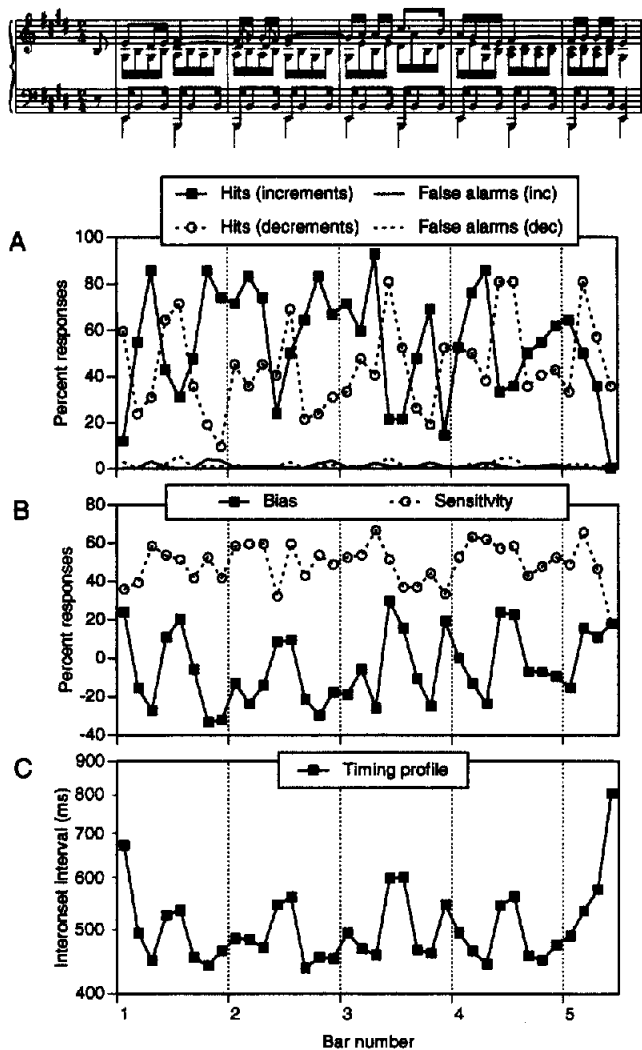

Figure 3. A: Detection accuracy and false-alarm profiles for duration increments (inc) and decrements (dec) in Experiment 1. B: The derived sensitivity and bias profiles. C: A typical timing profile for the Chopin "theme," shown on top.

between the two DAPs was $-.53$ ($p < .001$).[10] Clearly, however, the two functions were not exact mirror images. For example, detection scores decreased over the last three positions in both tasks.

In principle, false-alarm profiles (FAPs) could be a more direct indicator of a perceptual bias than the bias profile discussed later. However, false-alarm frequencies varied radically among listeners, and their distribution over 36 positions resulted in very sparse data. In the course of each block of 18 trials, a listener had 15 opportunities to make a false-alarm response to a given position. (On 3 trials, there was a target in that position or in an adjacent position, so a false-alarm response would have been scored as a correct response.) The false-alarm percentages shown on the bottom of Figure 3A were calculated relative to this theoretical maximum. The response distributions were clearly nonuniform and showed the expected positive relationship to the respective DAPs: The DAP–FAP correlation was .59 ($p < .001$) for increments and .75 for decrements. The two FAPs, like the two DAPs, were negatively correlated with each other ($r = -.58$).

Sensitivity and bias profiles. Parallel tendencies in the two DAPs could be due to positional variations in sensitivity that, by definition, are neutral with regard to the direction of the change to be detected. As proposed in the introduction, a rough estimate of variations in sensitivity may be obtained by calculating the average of the two DAPs (i.e., the position main effect). At the same time, a better estimate of the variations in perceptual bias common to the two DAPs may be obtained by computing their difference (decrements minus increments), divided by 2 (i.e., the Position × Condition interaction).[11] The resulting sensitivity and bias profiles are shown in Figure 3B. Their correlation ($-.25$) was not significant.[12]

The sensitivity profile showed that sensitivity to durational change was reduced in the final position, where there was no following IOI for comparison. Somewhat lowered sensitivity in the initial two positions also was suggested. These effects were expected on psychophysical grounds. There was additional internal variation in this profile, but it did not follow a systematic pattern from bar to bar. The bias profile, by contrast, showed a much more regular pattern, with twin peaks in Positions 4 and 5 in every bar and a peak at the end of the excerpt. These positions coincided with the ends of rhythmic groups in the melody, indicating that listeners "expected" these IOIs to be lengthened (i.e., they found it hard to detect increments but easy to detect

[10] All correlations reported in this article are Pearson product–moment correlations based on the 36 interonset intervals (IOIs), sometimes with the final IOI omitted (as noted), and thus they have 34 or 33 *dfs.*

[11] Of course, both effects were highly significant in an analysis of variance. These trivial statistics are omitted here.

[12] The expected correlation between the sum and the difference of two variables is zero when their variances are equal, regardless of the magnitude of the correlation between the variables. The increment detection accuracy profile (DAP) was somewhat more varied than the decrement DAP, which accounted for the small negative correlation.

decrements in these IOIs). In addition, there were peaks in Position 3-8, where an expressive anacrusis (unique to the fourth melodic segment) occurred in the music, and in Position 1-1, indicating expectations of lengthening at the beginning of a longer rhythmic group or a whole phrase.

Timing profile. The average performance timing profile is shown in Figure 3C. As predicted, it bore a striking similarity to the bias profile ($r = .75$, $p < .001$). The only large discrepancy was in the final position, which may be attributed to a floor effect in the DAP for increments (see Figure 3A) that depressed the bias estimate. When the final position was omitted, the correlation rose to .85. This was about as high a value as could be expected given the relatively small number of responses on which the perceptual data were based. Because of the long final IOI in performance and the corresponding floor effect in increment detection, this IOI was omitted in further correlational analyses pertaining to perception–performance relationships; an asterisk (r^*) is used to signify this. The original DAP for increments was also (negatively) correlated with the timing profile ($r^* = -.82$, $p < .001$). This finding replicates the perception–performance correlations obtained in earlier studies (Repp, 1992a, 1992b, 1995c): Wherever pianists tend to slow down, listeners have difficulty detecting a duration increment. They also seem to find it easier to detect a decrement in those positions, although the correlation between the DAP for decrements and the timing profile was not as impressive ($r^* = .70$, $p < .001$). As predicted, the sensitivity profile was not significantly correlated with the timing profile ($r^* = -.25$).

The shape of the timing profile, like that of the bias profile, can be explained by reference to the melodic-rhythmic grouping structure of the music. In addition to phrase-initial lengthening (Position 1-1) and phrase-final *ritardando* (Positions 5-1 to 5-4), there was a lengthening around each group boundary (Positions 4 and 5 in each bar) and at the beginning of the fourth segment (Position 3-8). The lengthening in Position 4 of each bar is the phenomenon of slowing before the long note that ends a rhythmic group, which was observed by Gabrielsson (1974) and Drake and Palmer (1993). The lengthening in Position 5 of each bar, although of about the same magnitude, presumably has a different explanation. It occurs after a rhythmic group has ended, while the long melody note is sustained but the accompaniment in the alto voice continues. It could be seen as a strategy of segregating the accompaniment from the melody to prevent any misperception of the melody continuing into the accompaniment. However, it is not necessary to give a definitive explanation of pianists' timing strategies here. The important finding here is that listeners seemed to "expect" precisely what artists typically do.

Acoustic variables. According to Drake's (1993) psychoacoustic hypothesis, however, it may be the other way around: Pianists' expressive timing may be a compensatory adjustment to auditory temporal distortions. In that case, it would be inappropriate to call the listeners' biases "expectations." It may be asked, then, whether changes in low-level acoustic variables such as pitch, intensity, and duration could have produced the observed distortions in perceived

timing. Perceptual and cognitive structures such as rhythmic groups, harmonic progressions, and meter are by definition excluded from a psychoacoustic account.

However, even a description of the music in terms of pitch, intensity, and duration is not straightforward. If the music were regarded simply as a series of undifferentiated chordal events, then duration would not be a factor because all IOIs are the same length; however, pitch and intensity would have to be averaged over the notes in each chord, which is problematic. It seems more reasonable to grant the auditory system the ability to carry out a simple scene analysis (Bregman, 1990) that parses the music into its separate voices.[13] Because the melody was perceptually more salient than the accompaniment, both because of its higher pitch (Palmer & Holleran, 1994) and its higher intensity (see Figure 1), it is reasonable to assume that listeners focused on the soprano voice whenever they could and that they switched attention to the alto voice only when there was no note onset in the melody. According to this hypothesis, the pitches and intensities of the most salient notes are the acoustic parameters to consider. However, duration also comes in as a stimulus variable because there are longer IOIs within the soprano voice (and in the lower voices).

Could the duration of melody notes have had a direct effect on the perception of timing? Whenever a long melody note occurred, listeners had to "switch" to the accompanying alto voice to detect deviations from even timing, and a switch from the soprano to the alto voice boils down to a change in pitch and intensity between the tones whose onsets delimit the target IOI. For the same reason, durational contrast or assimilation within the melody is an unlikely mechanism, especially preceding a long note. Slightly more plausible is the hypothesis that a sustained melody note masked the tone onsets in the accompaniment, so that they were less salient than they would otherwise have been. However, this effect (to the extent that it was not already eliminated by auditory segregation of the voices) amounts to a slightly enhanced intensity difference between the voices. As pointed out earlier, there is no reason to believe that listeners responded on the basis of perceived changes in the durations of longer IOIs. Therefore, there seems to be no good reason for including duration as a psychoacoustic variable because its effects can be reduced to those of pitch and intensity. In Experiments 2–4 I examined the consequences of eliminating IOI duration differences in the melody and the lower voices.

To determine whether the bias profile may have been caused by psychoacoustic interactions due to the variation in pitch and intensity among the most salient tones, eight acoustic variables were defined: the pitch (MIDI pitch number) and intensity (MIDI velocity) of the highest pitched tone initiating each IOI (the "target tone"); the average pitch and intensity of this tone and the following highest

---

[13] Actually, this analysis is not primitive because the conditions that normally lead to auditory stream segregation (e.g., high event rates, large pitch separations, silent intertone intervals) were not present in the music. Rather, the proposed "musical scene analysis" is a generalization of auditory scene analysis.

pitched tone; the relative (signed) pitch and intensity difference between these successive tones; and their absolute (unsigned) pitch and intensity difference.[14] The auditory kappa effect would be reflected in a positive correlation between the bias profile and the absolute pitch difference between adjacent tones. This correlation was not significant ($r^* = -.03$). However, bias tended to be high when the pitch of the target tone was high ($r^* = .55, p < .001$) and when the average pitch of that tone and the next was high ($r^* = .48, p < .01$) but low when intensity increased between the two tones ($r^* = -.50, p < .01$). There does not seem to be any evidence from psychoacoustic studies indicating that high tones are perceived as being shorter than low tones. The last effect, however, is consistent with the findings of Tekman (1992, 1995): IOIs preceding relatively loud tones tend to be perceived as being relatively long, which means that they are expected to be short (negative bias). In a stepwise multiple regression analysis, pitch height and intensity difference together yielded an R^* of .62 with the bias profile.[15] This correlation was significantly ($p < .05$) smaller than the simple correlation of the bias profile with the typical performance timing profile ($r^* = .85$).

The relationship between the sensitivity profile and the acoustic surface variables was also investigated. However, none of the eight correlations reached significance. Because some targets were preceded by other targets in the same trial, variation in sensitivity may have been related to the distance of a target from a preceding target. However, although sensitivity was low for the two targets with the shortest lags after a preceding target (Positions 3-8, Lag 4, and 2-4, Lag 6), overall there was no significant relationship.

The timing profile, like the bias profile it resembled, correlated with pitch height ($r^* = .39, p < .05$) and velocity difference ($r^* = -.39, p < .05$) but less strongly. Expressive lengthening thus tended to occur at points of high pitch (Positions 4 and 5 in each bar) and at points of large intensity drops (Position 5 in each bar). These tendencies were well accounted for in terms of the timing implications of rhythmic grouping. The correlations with surface variables resulted from the facts that melody and accompaniment (necessarily) differed in pitch register and intensity and that the accompaniment filled the gaps between melodic segments. However, it is doubtful that these variables as such, rather than the musical structure they convey, drive the musicians' actions. Not only are the correlations small, but there is much evidence that musicians plan their performance on the basis of a structural representation of the music (Clarke, 1985; Palmer, 1989), and auditory perception follows rather than precedes that plan (although auditory imagination and feedback also surely play a role). To the extent, then, that the timing profile is an expression of the typical musician's understanding of the musical structure (which includes but is not limited to rhythmic grouping), its high correlation with the perceptual bias profile suggests that a similar structural representation influences the perception of timing. The psychoacoustic variables merely serve to convey that structure.

Experiment 2

Experiment 1 successfully replicated Repp's (1992a) original findings for increment detection with a new musical excerpt of similar complexity. By including decrement detection and distinguishing variations in bias from variations in sensitivity, it purified the perception-performance correlation and confirmed that it is caused by variations in perceptual bias. Both expressive timing and perceptual bias seem to reflect primarily the division of the music into melodic-rhythmic groups. The most important cue for this grouping structure was the duration of the IOIs in the melody: Long IOIs signaled the ends of rhythmic groups. In Experiment 2, this cue was eliminated by substituting repeated 16th notes for all long melody notes. This enabled listeners to track the timing of the music by focusing on the soprano voice alone without any switching between soprano and alto voices.

From a psychoacoustic perspective, this manipulation should result in a considerable increase in sensitivity and a reduction of bias at (and also between) previous "switch points" because of the reduction of pitch and intensity differences between successive tones. From a structural perspective, some reduction of bias at these points may also be expected because of a weakening of the melodic grouping structure. However, this structure was by no means eradicated: The division of the melody into five segments was still conveyed by pitch contour and intensity. In fact, a division of long melody notes into shorter notes of the same pitch does not significantly alter the musical structure. It only makes the melodic segments somewhat less distinct and also gives the music a more trivial character, which may have effects on both expressive timing and perceptual bias. However, no dramatic change in the bias profile was expected, and the parallelism between perception and performance was expected to persist. The crucial question was whether the variations in bias could be explained by the psychoacoustic properties of the modified melody, which presumably now was the sole basis for listeners' temporal tracking.

Method

Listeners. Twelve new paid volunteers served as listeners. As in Experiment 1, they were musically trained undergraduates who were members of either the Yale Symphony Orchestra or the Yale Concert Band.

Materials and procedure. The score of "Variation 1" is shown on top of Figure 5.[16] The music was synthesized on the digital piano after making the appropriate modifications in the MIDI

---

[14] Strictly speaking, these were not acoustic variables but acoustic and perceptual properties inferred from MIDI parameters.

[15] The criterion for entering a variable into the regression model was that its (partial) correlation had to have a probability level of less than .15, which is conservative.

[16] The music retained the supplementary soprano notes starting in Positions 1-5 (D#), 2-3 (D#), and 2-5 (E) at their original values (see Figure 1), but they are not shown in Figure 5 because their notation looked awkward. They were also omitted on the participants' answer sheets.

instructions for the theme (see Figure 1). The MIDI velocity of each repeated 16th note was the same as that of the original long note in the theme. There were no gaps between repeated notes. Apart from this modification, the stimuli, test tapes, randomization, and procedure were exactly the same as in Experiment 1. The listeners' answer sheets showed the upper staff of the score of the modified excerpt.

Performance measurements. The same 9 student pianists who had played the theme for Experiment 1 subsequently played Variation 1 three times. They were asked to play with the expression that seemed artistically appropriate for the music. An average timing profile was obtained from the IOIs between melody note onsets.

Results and Discussion

Overall accuracy. As can be seen in Figure 4, overall accuracy was similar to that found in Experiment 1 (see Figure 2); the difference between experiments was not significant in a three-way ANOVA with the factors of experiments (2), conditions (2), and blocks (3). Rather than facilitating the detection of timing changes overall, the addition of melody notes reduced the difference between increment and decrement detection accuracy, which was no longer significant. This was due mainly to an increase in decrement detection scores. However, the Experiment × Condition interaction did not reach significance.

DAPs and FAPs. These profiles are shown in Figure 5A. Although the two DAPs show an inverse relationship in the initial bars, as they did in Experiment 1, they are more nearly parallel in Bars 3–5 here. Their correlation was −.06. Nevertheless, both DAPs were similar to those in Experiment 1 ($rs = .74$ and .70, respectively, $ps < .001$), but they were not identical, as indicated by the Experiment × Position interactions in separate two-way ANOVAs on the DAPs for increments, $F(35, 840) = 2.67, p < .0001$, and for decrements, $F(35, 840) = 1.81, p < .004$. The FAPs, shown at the bottom of Figure 5A, correlated positively with their respective DAPs ($rs = .42$ and .66, $ps < .01$ and .001, for increments and decrements, respectively).

Sensitivity and bias profiles. The absence of a correlation between the two DAPs could mean that increment and

Figure 5. A: Detection accuracy and false-alarm profiles for duration increments (inc) and decrements (dec) in Experiment 2. B: The derived sensitivity and bias profiles, with the corresponding profiles of Experiment 1 shown for comparison. C: A typical timing profile for Variation 1, shown on top, with the timing profile of the "theme" shown for comparison.

decrement detection are unrelated processes, which seems implausible. Alternatively, it may be the result of a reduction of variations in bias relative to those in sensitivity. The sensitivity and bias profiles are shown in Figure 5B, together with the corresponding profiles of Experiment 1 (dotted lines without data points). Again, the profiles were similar across experiments ($rs = .81$, respectively, $ps < .001$), but they also exhibited significant differences, $Fs(35, 840) = 1.96$ and $2.52, ps < .001$ and .0001, respectively, indicating that the stimulus manipulation had some reliable effects on the shapes of these profiles.

The nature of these effects can be seen by comparing the profiles of the two experiments in Figure 5B, especially at the previous "switch points" between voices. In Positions 1-1, 1-5, 2-5, 3-5, and 4-5, a "down" switch from soprano to alto voice was made unnecessary by the introduction of
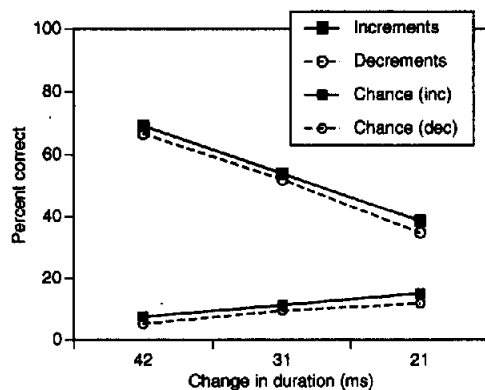

Figure 4. Overall percent correct scores and chance levels for the two tasks of Experiment 2 as a function of change in duration (test block). inc = increments; dec = decrements.
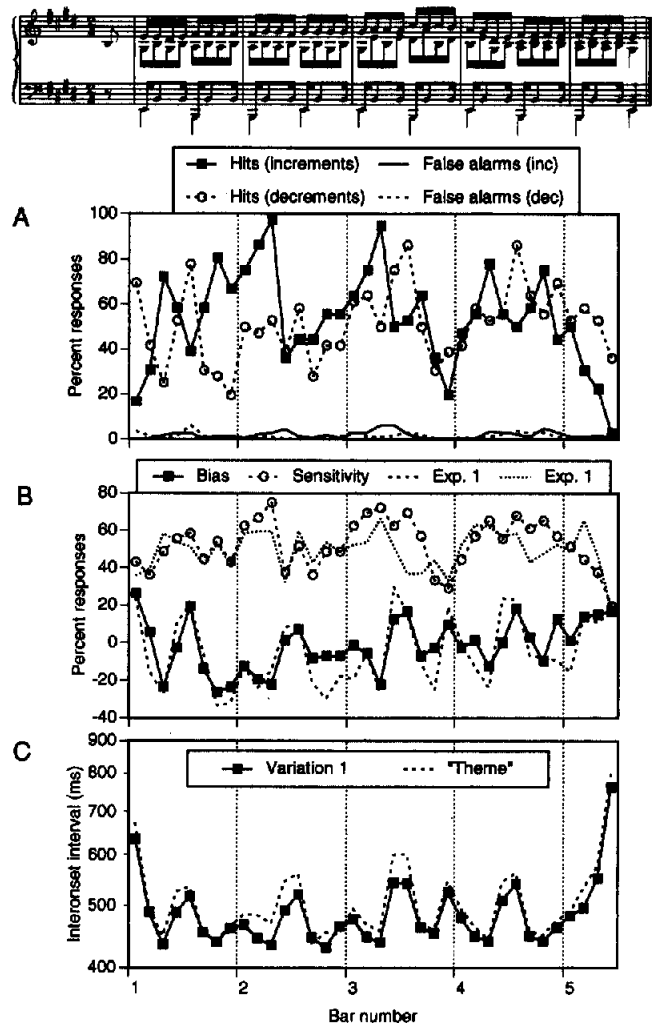
repeated melody notes. There was a large increase in sensitivity at 3-5, where the added melody notes were the most intense, and smaller benefits at 1-1, 1-5, and 4-5, but not at 2-5. However, there was no effect at all on the magnitude of the positive bias at any of these points. In Positions 1-2, 2-1, 3-1, 3-7, and 5-1, an "up" switch from alto to soprano was eliminated. Here, there were no systematic effects on sensitivity, but in four of the five positions (with the exception being 2-1) there was a substantial reduction in negative bias. The significance of these specific effects was tested in an ANOVA with experiments (2) as a between-subject factor and condition (2), switchpoint (2), and bar (5) as within-subjects factors. The increase in sensitivity at down but not at up switches (the Experiment × Switchpoint interaction) fell short of significance, $F(1, 24) = 3.53, p < .08$, but the triple interaction with bar was significant, $F(4, 96) = 5.83, p < .001$, which indicates significant changes in some bars but not in others. The change in bias at up but not at down switches (the Experiment × Condition × Switchpoint interaction) was significant, $F(1, 24) = 8.61, p < .008$, and did not interact with bar. It is consistent with both a psychoacoustic effect of intensity difference (Tekman, 1992, 1995) and a weakening of the grouping structure, although it is not clear why it was restricted to up switches.

Another systematic effect on bias can be seen in Figure 5B: In the fourth position of each of Bars 1–4 (the IOI immediately preceding a long melody note and a down switch in Experiment 1), there was a considerable reduction in positive bias. This effect (the Experiment × Condition interaction) was also significant in a separate ANOVA on these four positions, $F(1, 24) = 9.52, p < .006$. It represents a reduction in the expectation of lengthening preceding a long melody note that marks a melodic segment boundary (Drake & Palmer, 1993; Gabrielsson, 1974). These changes are consistent with both psychoacoustic and structural perception accounts.

Timing profile. Figure 5C shows the average timing profile of Variation 1, together with the timing profile of the theme from Experiment 1. It is apparent that the addition of repeated melody notes reduced expressive tempo modulations somewhat. In particular, the expressive lengthening at the end of each rhythmic group (Position 4 in each bar) was reduced, sometimes also in the following IOI. This makes sense in terms of the greater continuity between the melodic segments and the more trivial character of the variation, which invited less expression. However, group-final lengthening was still clearly present despite the absence of long melody notes. The shape difference between the two timing profiles (the Experiment × Position interaction) was highly significant in an ANOVA, $F(35, 280) = 2.65, p < .0001$, despite great qualitative similarity ($r = .98$). The main effect of experiment also reached significance due to the slight increase in average tempo. The timing profile of Variation 1 (without the final data point) was still highly similar to the perceptual bias profile in Figure 5B ($r^* = .81, p < .001$), and it also correlated with the DAPs in Figure 5A ($rs^* = -.69$ and $.52, ps < .001$ and $.01$, respectively, for increments and decrements).

Acoustic variables. The added melody notes changed the most salient pitches and intensities, which now resided exclusively in the soprano voice. The sensitivity profile correlated positively with the MIDI velocity of the highest tone occupying the target IOI ($r^* = .39, p < .05$), with the average velocity of the two boundary tones ($r^* = .58, p < .001$) as well as with the velocity difference between the two tones ($r^* = .40, p < .05$): a higher intensity and/or an increase in intensity facilitated detection. However, the bias profile was not significantly related to any pitch or velocity properties of the soprano voice, in contrast to the correlations with pitch height and velocity difference obtained in Experiment 1. Because the bias profile was similar in shape to that in Experiment 1 ($r = .81, p < .001$), it may be concluded that the pitch or intensity variation of the most salient tones was not the cause of the major variations in bias in either experiment.

This conclusion rests on the assumption that the listeners in this experiment were tracking the timing of the melody notes. However, it is possible that they still shifted their attention back and forth between the soprano and alto voices despite the presence of soprano notes throughout. This, however, would seem to be a perceptual strategy based on structural processing of the music; it is beyond a psychoacoustic account, as it is understood here. Moreover, such a strategy would actually be counterproductive in the detection task. It is more likely that listeners indeed focused on the soprano voice throughout but were nevertheless influenced by the timing implications of the melodic-rhythmic grouping structure, as conveyed by the remaining cues (mainly pitch contour). The results of Experiment 2, then, support the hypothesis that it is the perceptual grouping of the musical events into successive units, rather than psychoacoustic perturbations of time perception, that creates the variations in perceptual bias and in pianists' expressive timing.

In the remaining three experiments, I investigated the effects of additional changes in the music on timing perception and production, with the aim of showing that a perception–performance correlation would persist even for stripped-down musical structures. The contrast between psychoacoustic and structural perception accounts played a lesser role, however, because they generally made similar predictions in Experiments 3–5.

Experiment 3

There is another way of removing durational variation in the theme's melody and of bridging the gaps between its melodic segments: to remove the accompanying alto voice when it coincides with soprano notes, to simultaneously reduce each long melody note to a 16th note, and to remove any supplementary soprano notes. The result, "Variation 2," is shown in musical notation at the top of Figure 7. It consisted of a rhythmically uniform melody composed of original soprano and alto notes (i.e., a compound melody), still accompanied by the rhythmic pattern of the lower voices.

Any variations in bias caused by the large pitch and intensity jumps between the original soprano and alto voices should be equally present or even enhanced in Experiment 3.

However, effects due to the durational variation in the melody tones of the theme should be reduced. They may not be eliminated completely because perceptual segregation of melody and accompaniment in the compound melody may cause "rests" to be perceived after group-final melody notes. However, because the conditions for true auditory streaming (Bregman, 1990) are not present, this must be considered a form of structural listening. From a structural perception perspective, there is not much difference between the theme and Variation 2. The compound melody of Experiment 3 is certainly more continuous than the separate voices in Experiment 1, and therefore a reduction of expressive timing and of perceptual bias may be expected at group boundaries. This may also reduce the perception–performance correlation somewhat, as in Experiment 2, but it should still be substantial.

Method

Listeners. Twelve new paid volunteers served as listeners. As in Experiments 1 and 2, they were musically trained undergraduate students.

Materials and procedure. Variation 2 (see the top of Figure 7) was created by deleting notes from the theme's original MIDI instructions and by reducing the durations of long melody notes to 500 ms. The remaining notes had the same MIDI velocities as in Experiments 1 and 2; thus, the soprano and alto parts of the compound melody were distinguished both by pitch and intensity (see Figure 1). The test sequences and the procedure were exactly the same as in Experiments 1 and 2. The listeners' answer sheets showed the upper staff of the score of the modified excerpt.

Performance measurements. The same 9 student pianists who had played the earlier excerpts went on to play Variation 2 three times. They were asked to play with the expression that seemed artistically appropriate for the music. An average timing profile was obtained from the IOIs of the compound melody notes.

Results and Discussion

Overall accuracy. As can be seen in Figure 6, overall accuracy was similar to that found in Experiment 1; the difference was not significant, nor was any interaction
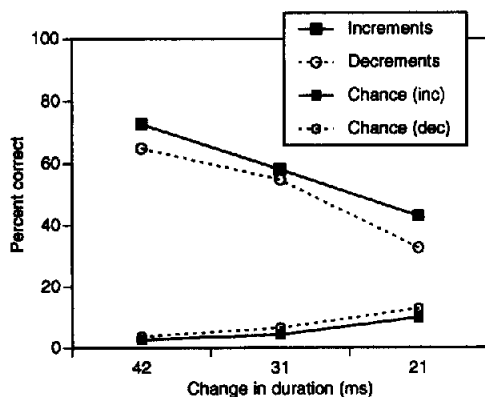

Figure 6. Overall percent correct scores and chance levels for the two tasks of Experiment 2 as a function of change in duration (test block). inc = increments; dec = decrements.
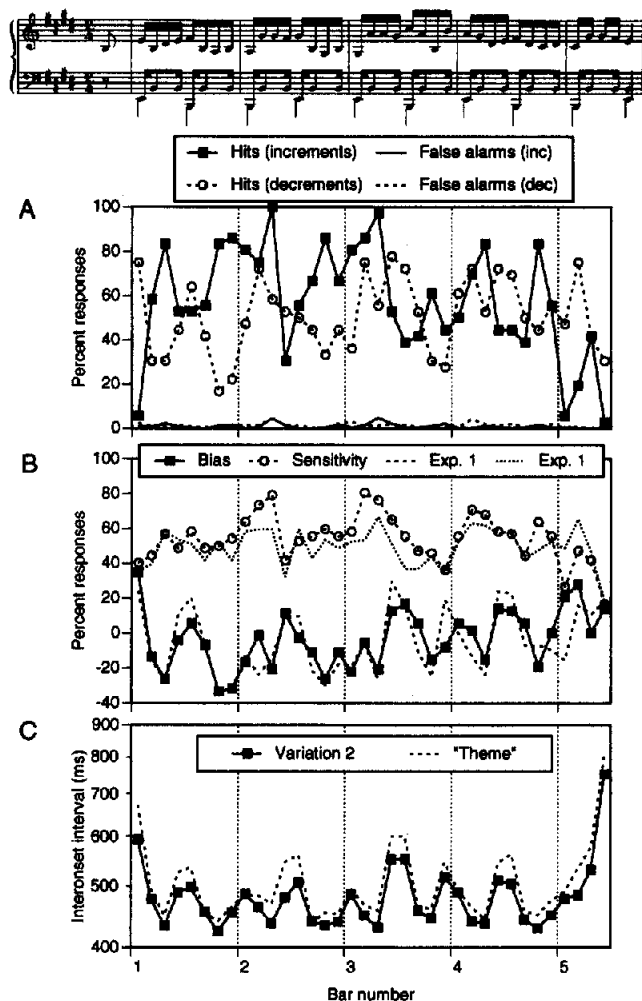

Figure 7. A: Detection accuracy and false-alarm profiles for duration increments (inc) and decrements (dec) in Experiment 3. B: The derived sensitivity and bias profiles, with the corresponding profiles of Experiment 1 shown for comparison. C: A typical timing profile for Variation 2, shown on top, with the timing profile of the "theme" shown for comparison.

involving experiments. The advantage for increment over decrement detection fell short of significance here.

DAPs and FAPs. Figure 7A shows that the two DAPs were again inversely related in Bar 1, with a less clear relationship thereafter. The correlation between the two profiles was nonsignificant ($r = -.22$). Both DAPs were similar to those in Experiment 1 ($rs = .77$ and $.70, ps < .001$) as well as to those in Experiment 2 ($rs = .79$ and $.74$). Nevertheless, their shapes were significantly different from those of the DAPs in Experiment 1, $Fs(35, 840) = 2.89$ and $1.87, ps < .0001$ and $.002$, respectively. The FAPs correlated weakly but positively with the respective DAPs ($rs = .46$ and $.43, ps < .01$).

Sensitivity and bias profiles. These derived profiles are shown in Figure 7B, together with the corresponding profiles from Experiment 1, to which they were significantly

similar ($rs = .64$ and $.81$, $ps < .001$, respectively). ANO-VAs indicated that the shapes of both sensitivity and bias profiles were affected by the change in the music, $Fs(35, 840) = 1.70$ and 3.07, $ps < .008$ and $.0001$, respectively. It appears that the removal of accompanying notes increased sensitivity in Bars 2 and 3, although there was a reduction in sensitivity in Bar 5. (The overall increase was not significant, as already noted.) The reduction was caused by a dramatic dip in the DAP for increments at 5-1 and neighboring positions (see Figure 7A). This unexplained dip was also reflected in an increase of positive bias in these positions. Otherwise, the changes in the bias profile consisted mainly of a lowering of peaks of positive bias (1-4, 1-5, 2-5, 3-4, 3-8, 4-4, and 4-5). The reductions in Positions 1-4, 3-4, and 4-4 may be attributed to the absence of long melody notes (as in Experiment 2). Those in Positions 1-5, 2-5, and 4-5 (down switches) may be due to the increased coherence of soprano and alto voices at this point.

Timing profile. Figure 7C shows the average timing profile of Variation 2, together with the timing profile of the theme from Experiment 1. The effect of the musical manipulation on performance was very similar to that found in Experiment 2 (see Figure 5C): Again, the expressive lengthening toward the end of each melodic segment was reduced, probably due to the increased continuity between soprano and alto voices. These changes were similar to some of the changes observed in the perceptual bias profile (see Figure 7B). The shape difference between the two timing profiles in Figure 7C (the Experiment × Position interaction) was highly significant, $F(35, 280) = 3.12$, $p < .0001$, despite great qualitative similarity ($r = .98$). The main effect of version was also significant because of the increase in average tempo. The timing profile of Variation 2 was significantly similar to the perceptual bias profile ($r^* = .68$, $p < .001$), and it was also correlated with the original DAPs ($rs = -.65$ and $.42$, $ps < .001$ and $.01$, respectively).[17] The perception–performance correlation was smaller here than in the preceding experiments, partly because of the unexplained increase in bias (due to low detectability of increments) around Position 5-1.

Acoustic variables. The elimination of accompanying notes did not change the salient pitch and velocity properties of the music, which were defined in terms of the compound melody notes. The sensitivity profile correlated significantly with the velocity of the target tone ($r^* = .43$, $p < .01$) and with the average velocity of the two boundary tones ($r^* = .57$, $p < .001$), which replicated correlations obtained in Experiment 2, but not in Experiment 1. There was no correlation between sensitivity and the velocity difference between successive tones. However, the bias profile did correlate with that variable ($r^* = -.49$, $p < .01$) as well as with the pitch of the target tone ($r^* = .51$, $p < .01$) and with the average pitch of both boundary tones ($r^* = .42$, $p < .01$). These correlations replicated almost exactly the findings of Experiment 1. However, here the R^* in a stepwise regression analysis ($.61$) was not significantly lower than the perception-performance correlation ($.68$).

Considered by themselves, these results would seem to be equally compatible with psychoacoustic and structural per-

ception accounts. Given the results of Experiments 1 and 2, however, the preferred interpretation is that the variations in perceptual bias once again reflected the (somewhat weakened) timing implications of melodic-rhythmic grouping, not auditory distortions caused by the primitive acoustic correlates of the melodic groups.

Experiment 4

In Experiment 4 I focused on the role of the lower voices, which had been ignored so far. Although clearly less salient than the melody, they nevertheless provided an important harmonic and rhythmic foundation. Their uneven rhythm could have affected detection performance: Low-pitch notes were sustained through Positions 3 and 7 in each bar, which were exactly the positions in which there was the strongest negative bias and in which increments were particularly easy to detect. To examine the role played by the lower voices, I omitted them from the simplified excerpt used in Experiment 3, leaving only the monophonic compound melody of the right-hand part ("Variation 3"). The effect of this manipulation could be gauged by comparing the results of Experiment 4 with those of Experiment 3.

If the repetitive rhythm and pitch pattern of the lower voices had any psychoacoustic effects on the detectability of duration changes, then these effects should exhibit a pattern that, like the rhythm itself, is replicated from one beat to the next. The structural perception hypothesis does not make such a specific prediction. Elimination of the lower voices removes vertical harmony, leaving only the harmonic implications of the monophonic compound melody, which is unchanged from Experiment 3. It also removes another cue to grouping, namely the harmonic change from tonic to dominant or from dominant to tonic that occurs at the end of each melodic group. This harmonic and textural impoverishment may lead to a further reduction in expressive timing and perceptual bias, but no qualitative changes in these profiles were expected, and the perception–performance correlation was expected to be smaller but still significant.

Method

Listeners. Twelve new paid volunteers served as listeners. Again, they were musically trained undergraduates.

Materials and procedure. Variation 3 (see the top of Figure 9) resulted from deleting all tenor and bass notes from the MIDI instructions of Variation 2. The test sequences and the procedure were exactly the same as those used in Experiment 3.

Performance measurements. The same 9 student pianists who had played the earlier excerpts went on to play Variation 3 three times. They were asked to play the monophonic melody with expression, but without pedal and "as if it were a single melody."

---

[17] There was a small negative correlation between the sensitivity profile and the timing profile ($r^* = -.35$, $p < .05$) and a similar correlation between the sensitivity and bias profiles ($r = -.39$). Both were artifacts attributable to greater variation in the detection accuracy profile (DAP) for increments than in the DAP for decrements (see Figure 7A).

An average timing profile was obtained by the same procedures as previously.

Results and Discussion

Overall accuracy. As shown in Figure 8, overall accuracy was similar to that found in the earlier experiments, but there was no difference here between increment and decrement detection. There was no significant change in accuracy relative to Experiment 3.

DAPs and FAPs. The DAPs in Figure 9A were not significantly related to each other ($r = .11$), but they were similar to those of Experiment 3 ($rs = .79$ and $.67, ps < .001$) as well as to those of Experiment 1 ($rs = .70$ and $.39$, $ps < .05$). Note that the DAP for increments remained more stable across experiments than the DAP for decrements. A comparison of these DAPs with those of Experiment 3 (see Figure 7A) suggests that removal of the low notes caused a considerable reduction in the variability of both profiles. However, only the change in the DAP for increments was significant, $F(35, 770) = 2.09$, $p < .0004$. The FAPs correlated weakly but positively with the respective DAPs ($rs = .46$ and $.37, ps < .05$).

Sensitivity and bias profiles. The derived profiles are shown in Figure 9B, together with the corresponding profiles from Experiment 3. The shapes of both profiles were affected by the change in the music, although the F values were small, $Fs(35, 770) = 1.60$ and $1.76, ps < .02$ and $.005$, respectively. To determine whether removal of the lower voices had a cyclically repeating effect, I subdivided the 36 positions into 9 groups of 4 (corresponding to quarter-note beats). The individual positions within these groups were compared in two-way ANOVAs, with the factors of experiment (2), within-beats position (4), and beat (9). The Experiment × Within-Beats Position interaction was nonsignificant for sensitivity but reached significance for bias, $F(3, 66) = 3.47$, $p < .03$, due to a relative decrease in negative bias in within-beats Positions 3 and 4. This effect can be seen most clearly in Bar 1, especially in its second half (see Figure 9B), but it was not consistently present
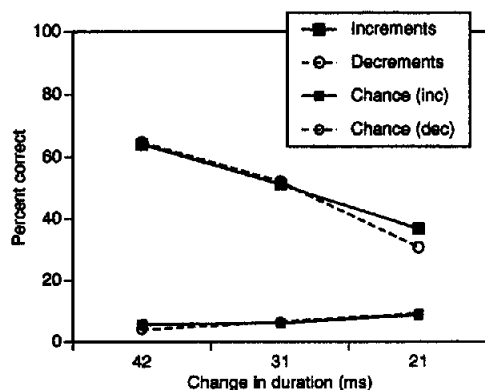

Figure 9. A: Detection accuracy and false-alarm profiles for duration increments (inc) and decrements (dec) in Experiment 4. B: The derived sensitivity and bias profiles, with the corresponding profiles of Experiment 3 shown for comparison. C: A typical timing profile for Variation 3, shown on top, with the timing profile of Variation 2 shown for comparison.

Figure 8. Overall percent correct scores and chance levels for the two tasks of Experiment 4 as a function of change in duration (test block). inc = increments; dec = decrements.
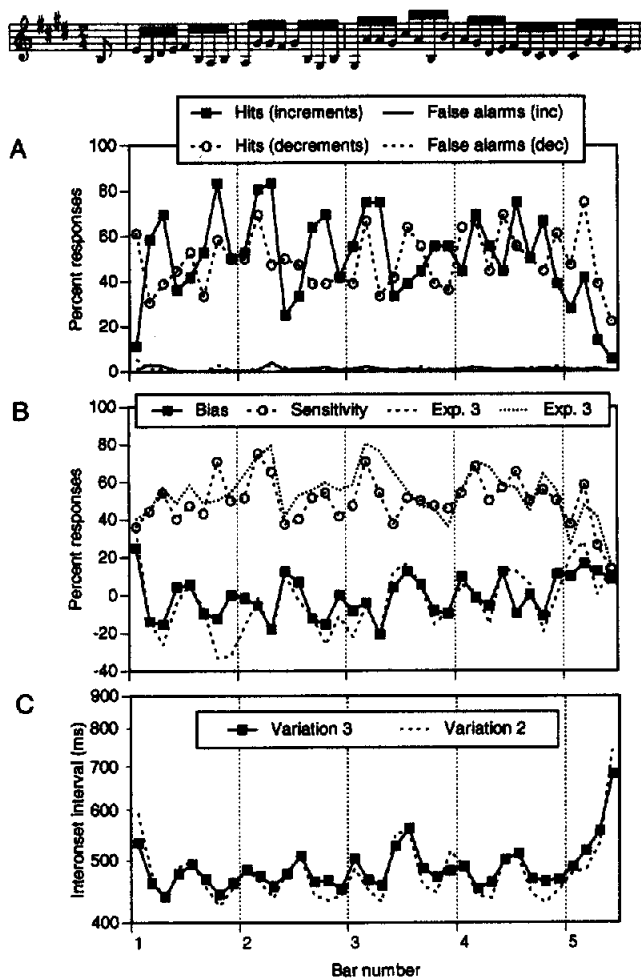
elsewhere. The Experiment × Beat × Within-Beats Position interaction was significant, $F(24, 528) = 1.63, p < .03$. Both the sensitivity and bias profiles were similar to the corresponding profiles in Experiment 3 ($rs = .72$ and $.80$, respectively) and Experiment 1 ($rs = .62$ and $.63$, respectively).

Timing profile. Figure 9C shows the average timing profile of Variation 3, together with the timing profile of Variation 2. A further reduction in expressive tempo modulations can be seen. The shape difference between the two timing profiles (the Experiment × Position interaction) was highly significant in a two-way ANOVA, $F(35, 280) = 6.04$, $p < .0001$, despite great qualitative similarity ($r = .95$). An analysis of within-beats positions, similar to the one described earlier, revealed a significant Experiment × Within-Beats Position interaction, $F(3, 24) = 15.34, p < .0001$, due to a relative lengthening of the short IOIs in Positions 2 and 3. This tendency, which indicated an effect of the (removal

of the) cyclic tenor–bass rhythm on expressive timing, can be seen fairly clearly in Figure 9C as a raising of the "valleys" in the timing profile, although this, too, varied from beat to beat, $F(24, 192) = 4.12$, $p < .0001$, for the triple interaction. The cyclic effect observed in performance matched the weak effect found in perception only partially (in within-beats Position 3). Nevertheless, the timing profile of Variation 3 was significantly similar to the perceptual bias profile ($r^* = .66$, $p < .001$), and it also was correlated with the DAP for increments in Figure 9A ($r^* = -.64$, $p < .001$), but not longer with the DAP for decrements ($r^* = .24$, ns).[18]

Acoustic variables. The sensitivity profile of Experiment 4 did not correlate significantly with any of the pitch or velocity properties of the compound melody. Thus, the correlations with velocity obtained in Experiment 3 were not replicated, even though they might have been expected to emerge more clearly without the lower voices. The bias profile still correlated with the velocity difference between the boundary tones ($r^* = -.51$, $p < .005$) and with the pitch of the target tone ($r^* = .34$, $p < .05$), which replicated the findings of Experiments 3 and 1. The R^* (based on velocity difference and absolute pitch difference) was .56, which was not significantly lower than the perception–performance correlation (.66).

The results of Experiment 4, then, do indicate some influence of the rhythmic pattern of the lower voices on both perception and performance, but these effects were not exactly parallel, for unknown reasons. Neither a psychoacoustic nor a structural perception account is particularly furthered by this finding, but the significant perception–performance correlation still tends to encourage the latter.

Experiment 5

The tones of the excerpt used in Experiment 4 varied in only two parameters, pitch and intensity, which were positively correlated ($r = .64$, $p < .001$), as they typically are in performance (Repp, 1996). The MIDI velocity values provided an important cue to grouping structure, particularly in the compound melody of Variations 2 and 3. To assess the contribution of intensity differences, I made the MIDI velocities of all notes equal in Experiment 5, leaving only pitch variation.

Some psychoacoustic studies of isochronous rhythm perception have failed to indicate any effect of relative intensity (Schütte, 1978; Terhardt & Schütte, 1976). However, a correlation of sensitivity with MIDI velocity was found in Experiments 2 and 3. More important, Tekman (1992, 1995) demonstrated that IOIs preceding loud tones were perceived as being longer than IOIs preceding soft tones, which seems to be consistent with the negative correlation of bias with MIDI velocity difference obtained in Experiments 1, 3, and 4. Thus, a reduction of variations in both sensitivity and bias might be expected when all MIDI velocities are made equal. From a structural perception perspective, removal of expressive dynamics reduces the expressive quality of the music and deprives the listener of a perceptual cue to rhythmic grouping structure, and therefore a reduction of variations in perceptual bias was likewise
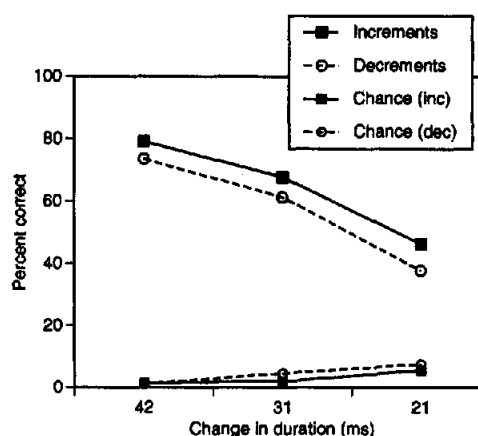

Figure 10. Overall percent correct scores and chance levels for the two tasks of Experiment 5 as a function of change in duration (test block). inc = increments; dec = decrements.

predicted. The question was whether the pitch structure alone would be sufficient to support a bias profile that is still related to performance timing. Repp (1992b, Experiment 1; Repp, 1995c, Experiment 1) found significant perception–performance correlations with melodies composed of equal-intensity tones, although they rested primarily on the final few IOIs, where a large *ritardando* in performance corresponded to a gradual decrease in increment detection scores.

Method

Listeners. Ten new paid volunteers served as listeners. Again, they were musically trained undergraduates.

Materials and procedure. Variation 3 was used again, but all MIDI velocities were set to the same arbitrary value of 60. The test sequences and the procedure were exactly the same as those used in Experiment 4.

Performance measurements. The performances of Variation 3 were used for comparison because, to determine the perception–performance correlation, it was necessary to have an expressively modulated timing profile. (If pianists had been asked to play without variations in dynamics, the expressive timing profile might have become unrepresentative of artistic performance.)

Results and Discussion

Overall accuracy. Figure 10 shows that overall accuracy was higher than in Experiment 4 (see Figure 8), even though the difference did not reach significance, $F(1, 20) = 3.56$, $p < .08$, because of large between-listener variability. The trend was consistent, however, with Repp's (1992b) finding that dynamic variation interferes somewhat with the detection of timing deviations. A significant advantage of

[18] As in Experiment 3, there was a negative correlation ($r^* = -.42$, $p < .01$) between the sensitivity profile and the timing profile as well as a similar correlation between the bias and sensitivity profiles ($r = -.44$, $p < .05$), again due to greater variation in the detection accuracy profile (DAP) for increments than in the DAP for decrements (see Figure 9A).

increment over decrement detection emerged again, $F(1, 9) = 7.72, p < .03$.

DAPs and FAPs. The DAPs in Figure 11A show a small positive correlation ($r = .35$, $p < .05$), mainly because of the lower scores at the beginning and end of the excerpt. Systematic variation in the profiles seemed further reduced, but they were still related to those of Experiment 4 ($rs = .73$ and .50, $ps < .001$ and .01, respectively), Experiment 3 ($rs = .69$ and .47, $ps < .001$ and .01, respectively), Experiment 2 ($rs = .62$ and .35, $ps < .001$ and .05, respectively), and Experiment 1 ($rs = .53$ and .22, $ps < .01$, ns, respectively). Again, the DAP for increments remained more stable across experiments than the DAP for decrements. The change in the DAP for increments between Experiments 4 and 5 was significant, $F(35, 770) = 1.63, p < .02$, whereas that in the DAP for decrements fell short of significance, $F(35, 770) = 1.43, p < .06$. The FAPs correlated weakly with the respective DAPs ($rs = .27$ and .47, $ps = ns$ and $<.01$, respectively).
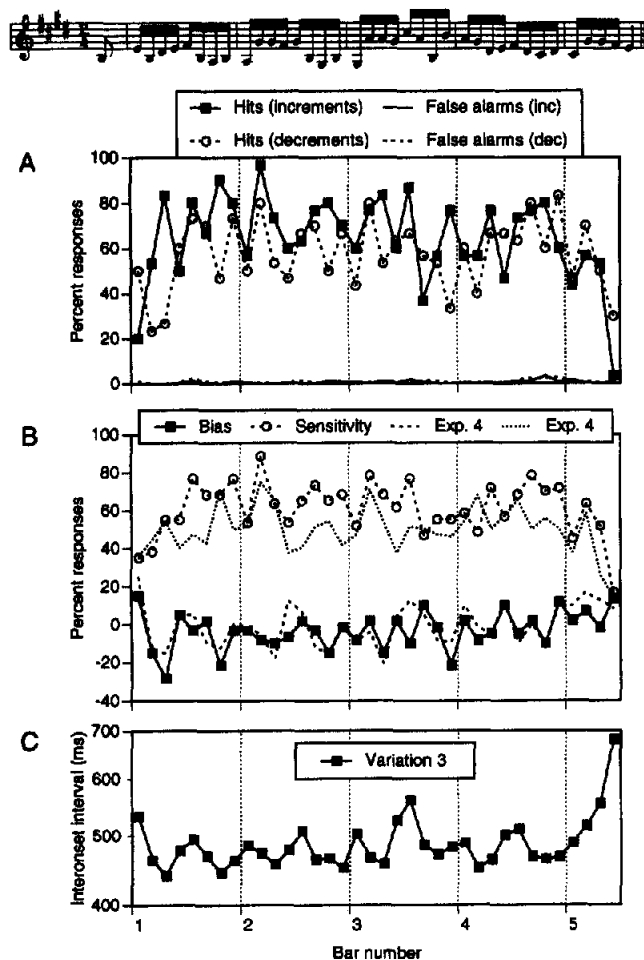

Figure 11. A: Detection accuracy and false-alarm profiles for duration increments (inc) and decrements (dec) in Experiment 5. B: The derived sensitivity and bias profiles, with the corresponding profiles of Experiment 4 shown for comparison. C: The typical timing profile for Variation 3 (from Figure 9C).

Sensitivity and bias profiles. These profiles are shown in Figure 11B, together with the corresponding profiles from Experiment 4. The shape of the sensitivity profile was changed significantly by the elimination of expressive dynamics, $F(35, 770) = 2.15, p < .0003$, but there was no significant effect on the bias profile. The change in the sensitivity profile was caused by increases in sensitivity in various positions, even though the overall increase fell short of significance. Both profiles remained significant similar to the corresponding profiles in Experiment 4 ($rs = .60$ and .73), Experiment 3 ($rs = .62$ and .66), Experiment 2 ($rs = .56$ and .48), and Experiment 1 ($rs = .44$ and .46, $ps < .01$).

Timing profile. Figure 11C shows the timing profile of Variation 3 (from Figure 9C). It was still significantly related to the bias profile ($r^* = .41$, $p < .05$) and the DAP for increments ($r^* = -.37$, $p < .05$), but the correlations were considerably lower than in Experiment 4.

Acoustic variables. The sensitivity profile correlated weakly and negatively with absolute pitch difference ($r^* = -.42$, $p < .05$), which is the first and only hint of an effect of pitch separation in this study. However, the bias profile did not correlate significantly with any pitch variables, even though pitch effects should have emerged most clearly in this experiment. This suggests that previous correlations of bias with pitch height (Experiments 1, 3, and 4) were due to structural factors rather than pitch as such. A similar conclusion may be drawn about previous effects of intensity difference on bias because removal of intensity variation had no significant effect on bias.

"Late" Responses

The focus of this research was on external correlates of the variations in perceptual bias, not on the internal processes causing these variations. One aspect of the data bears on models of temporal processing, however. As in my earlier studies of IOI increment detection (Repp, 1992a, 1992b, 1995c), listeners frequently attributed an increment to the IOI *following* the target IOI by circling the following note(s) on the answer sheet, even though they were specifically instructed not to do so. There were large individual differences in this tendency, but no listener ever gave close to 100% of such "late" responses, as should be the case if the instructions had been simply misunderstood. One possible explanation is that listeners occasionally forgot to backtrack when marking their responses, for (as was always pointed out in the instructions) the lengthening of an IOI or target tone could be detected only at the time that the following tone was heard. If this were the complete explanation, then the tendency to give late responses should be equally pronounced in increment and decrement detection tasks. In two unpublished experiments with simple melodies, however, I observed that the average percentage of late responses in decrement detection was much lower than that in increment detection. Although it was still higher than chance level, the results suggested that failure to backtrack does not account for the majority of late responses in the increment detection task and that these responses may have a perceptual origin.

This research replicated this finding, as can be seen in Figure 12. The figure shows the percentages of early and late responses (out of all correct responses) in the five experiments, separately for increment and decrement detection. In every experiment, the percentage of late responses was lower in the decrement than in the increment detection task. The difference was highly significant in an overall two-way ANOVA on late responses, $F(1, 55) = 25.74, p < .0001$, whereas the interaction with experiments was nonsignificant. The variation in the overall percentage of late responses across experiments was not significant because of large between-listeners variability. Late responses in the decrement detection task were still much more frequent than early responses, as determined in another two-way ANOVA, $F(1, 55) = 18.51, p < .0001$. This suggests that late responses have two independent causes, one being a failure to backtrack on the answer sheet and the other being perceptual and applying only to increment detection. Alternatively, there may be perceptual causes of different strength in increment and decrement detection.

Late responses were also examined as a function of position in the music. A "late response profile" represents the percentage of correct responses that were late as a function of position. Figure 13 shows these profiles for Experiment 1. It is evident that late responses were more frequent in some positions than in others. In the increment detection task, there was a systematic pattern: In each bar, there were peaks in Positions 2 and 6. (The only exception was Position 1-2, where the peak was shifted to 1-1.) The DAP and the FAP for increments (see Figure 3A) generally showed peaks in Positions 3 and 7. Although the correlation of the late response profile with the DAP ($r = .34, p < .05$) did not increase dramatically when the former was shifted to the right by one position ($r = .40, p < .05$), because of the different relative heights of the peaks, the data nevertheless indicate that positions in which increments were easy to detect (i.e., that showed a negative bias) attracted responses away from a preceding increment target. The same tendency was noted by Repp (1992a), and it was also found in the present Experiments 2 and 3. By contrast, the late response
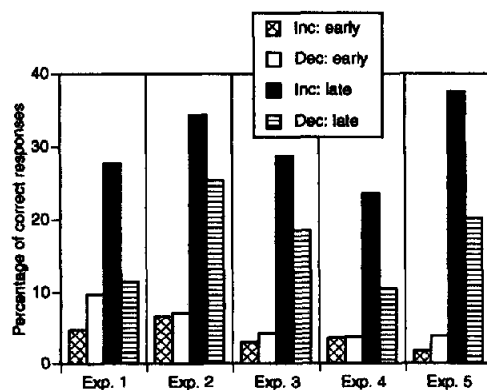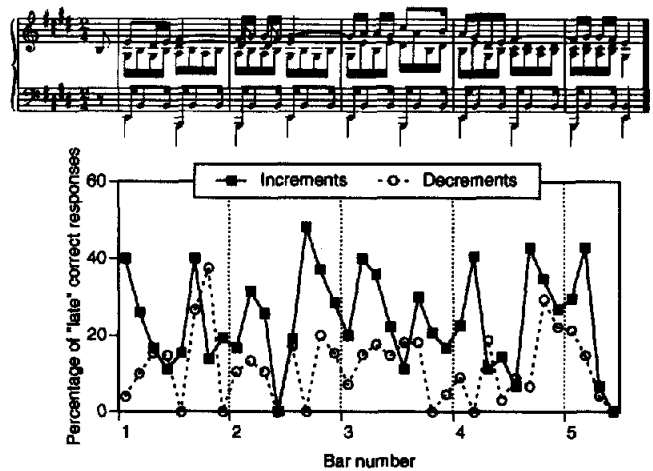

Figure 13. The late response profiles of Experiment 1.

profile for the decrement detection task, shown for Experiment 1 in Figure 13, was unsystematic and unrelated to the DAP for decrements, and this was also true in the other experiments. These findings may hold some valuable clues about the perceptual processes underlying the detection task.

General Discussion

The present series of experiments and the distinction between the psychoacoustic and structural perception hypotheses are symbolic of different theoretical and methodological traditions in psychology. In the experiments I used materials that ranged from complex, original piano music (Experiment 1) to a simple sequence of tones more characteristic of psychophysical experiments (Experiment 5). In that respect, they spanned the gap between psychomusicology and psychoacoustics, two research areas that are also instantiated by two of my earlier studies using the present detection paradigm (Repp, 1992a, 1995c). These areas are associated with two contrasting explanatory approaches to perceptual phenomena, which have been termed here *structural* versus *psychoacoustic*. The choice between them is in part a philosophical matter relating to the preferred level of description for the domain studied. Music can be described either in terms of its acoustic properties or in terms of the structural concepts provided by music theory. These two levels of description are not mutually exclusive, but their relative appropriateness depends on the complexity of the musical materials: For a simple sequence of tones, an acoustic characterization may be adequate, but for an extended, polyphonic piece of music this becomes cumbersome, and a structural description in terms of rhythmic groups, hierarchical phrases, harmonic prolongation, and so on is to be preferred.

The issue is not entirely philosophical, however. Empirical evidence is relevant also, and in this research I have tried to provide some. The phenomenon to be explained is the consistent perception–performance correlation: the parallelism between the average perceptual bias profile derived

Figure 12. Percentages of early and late responses, out of all correct responses, in the two tasks of Experiments 1–5. Inc = increments; Dec = decrements; Exp. = experiment.

from a detection task and the average expressive timing profile produced by musicians. This correlation was obtained in each of the five experiments reported here, although it decreased in magnitude as the music was increasingly simplified. This decline was attributed to a progressive weakening of the melodic-rhythmic grouping structure, which resulted in reductions of expressive timing variation and of variation in perceptual bias, thus lowering the correlation but not changing the qualitative aspects of the respective profiles. The results thus are very consistent with a structural perception account. The perceptual segmentation of the music into melodic-rhythmic groups, it is claimed, has temporal implications (acceleration–deceleration patterns) that are realized both in performance and perception. Musicians carry out a cognitive analysis of the music before performance and then act out the temporal implications on their instrument (Todd, 1985). Listeners carry out a perceptual analysis on-line (familiarity with the music may help) and "expect" to hear the temporal implications of the groups. If these implications are not realized, perceptual compensation (i.e., perceptual bias) will be the result. Of course, listeners may also carry out a cognitive analysis of what they hear, but this is not likely to influence timing perception.

Admittedly, the data are also partially consistent with a psychoacoustic account. However, the combined effects of the variables considered were not always sufficient to account for the full perception–performance correlation. This was most clearly the case in Experiment 2, but also in Experiment 1. Moreover, the effects obtained do not always correspond to what is known from psychophysical studies. Thus, there was little evidence for the auditory kappa effect (i.e., an increase in negative bias with absolute pitch distance). It is possible that this effect is obtained only when successive tones are separated by silence, as they were in the studies of Shigeno (1986) and Crowder and Neath (1995). The positive correlation between pitch height and bias (obtained in Experiments 1, 3, and 4, but not 5), which suggests that higher tones are perceived as being shorter than lower tones, does not seem to correspond to any known interaction between pitch and duration. Only the association between intensity decrease and bias resembled an effect described by Tekman (1992, 1995). It is possible, however, that a more sophisticated model of auditory processing than has been presented here will provide a more complete explanation of the perceptual data.

The most serious problem of a psychoacoustic account, as already noted in the introduction, is in explaining expressive timing in performance and the parallelism between perception and performance. Expressive timing is generally considered to be the result of the musician's conscious or subconscious structural analysis of the score (see, e.g., Clarke, 1985; Palmer, 1989; Repp, 1992c; Todd, 1985).[19] The deviations from even timing are usually large enough to be perceived as such. To compensate for auditory interactions and distortions, however, the musician's goal would have to be to play as evenly as possible (Drake, 1993). Now, it is indeed possible that the typical intended expressive timing profile is merely a scaled-up version of the typical unin-

tended timing profile.[20] However, what purpose does the change in scale serve? The original motivation of compensation is thereby lost. Presumably, a performer wants to clarify the musical structure by enhancing its timing implications, and even involuntary timing variations are probably a residue of this tendency. It is easier to extend the structural concepts of music-theoretical discourse to music perception than to apply the reductionistic concepts of psychoacoustics to music performance. Using a common vocabulary for both draws attention to the close parallelism between music perception and performance with regard to timing.

The present results raise several interesting questions to which the remainder of the discussion is devoted. The computation of sensitivity and bias profiles, which seemed reasonable in Experiment 1, became more problematic in the later experiments because of an asymmetry between the DAPs for increments and decrements: The latter were less varied, less reliable, and less clearly related to performance timing than the former. Also, the negative correlation between the two DAPs vanished, so that it was not generally true that decrements were difficult to detect where increments were easy to detect or vice versa. The analyses nevertheless focused on the bias profile because it provided a simpler representation of the data and because it correlated more strongly with the timing profile than did either DAP. Alternatively, it would have to be concluded that increment and decrement detection are unrelated processes, at least in music. All the findings obtained for bias profiles also hold for increment DAPs, so the conclusions of this study do not rest on the bias profiles. However, they do not hold as strongly for decrement DAPs. There could be a reason for this: Expressive timing deviations in performance are usually lengthenings rather than shortenings (Repp, 1992c). Although the baseline duration to which timing deviations should be compared is not always clear, the distribution of IOIs in typical expressive performances is decidedly asymmetrical, and lengthening lends emphasis and enhances boundaries, whereas shortening merely increases continuity. Therefore, listeners may have stronger "expectations" of lengthening than of shortening and thus will compensate more strongly for the absence of expressive lengthening in an isochronous musical excerpt. In other words, the biases underlying increment detection may be stronger than any biases underlying decrement detection.

How does the musical structure cause perceptual biases in increment detection? The present study did not address this interesting issue directly, but some speculations are possible. There is compelling evidence from studies of visual motion

---

[19] The reader may wonder what effect the availability of a (partial) score to the listeners may have had on the bias profile. Did their timing expectations perhaps derive from an analysis of the printed music? Certainly not. Repp (1998, Experiment 3; Repp, in press, Experiment 1) found essentially the same detection accuracy profiles and bias profiles when the participants pressed a response key without ever seeing a score.

[20] Data from a recent experiment (Repp, 1997b) tend to confirm this for the present Chopin etude excerpt.

perception for a coupling between motor and perceptual processes: Lawful variation in movement velocity as a function of trajectory shape is perceived as constant velocity, whereas constant velocity is perceived as varying (Derwort, 1938; Viviani & Stucchi, 1989, 1992a, 1992b). Something analogous may happen in audition if listeners are sensitive to the kinematic implications of musical structure: The music may seem to "move" at a constant velocity when played with expressive timing (although considerable attenuation of the timing variations is probably required to achieve this impression), but at an uneven velocity when played isochronously. Music may have *temporal curvature*, so to speak, and when it is absent, listeners exhibit perceptual compensation. This sounds not unlike a psychoacoustic account, and indeed it shares the same problem in accounting for full-scale expressive timing, with its audible deviations from evenness. However, the point here is that the implications of musical structures are primarily kinematic and that listeners essentially "move as they listen," even if they are not musicians—a motor theory of musical time perception (see also Clynes & Walker, 1982; Repp, 1993b; Todd, 1995).

Although the perceptual biases found here seem to be obligatory (Repp, 1998), even in musically untrained listeners (Repp, 1995c, in press), performers can move in ways contrary to the "timing demands" of the music, as analyses of expert performances have shown (Repp, 1992c, 1997a). Nevertheless, the demands are likely to exist, perhaps independent of style and period (which only affect the musical structure), and performers' decisions to violate them presumably are cognitive in nature, often resting on alternative structural interpretations (Palmer, 1989, 1992). One important issue for future research is whether there are corresponding individual differences in listeners. It has been assumed so far that all musically trained listeners exhibit more or less the same bias profile, and there is no evidence to the contrary as yet. A thorough investigation of individual differences will require the collection of larger amounts of data per listener. It would be fascinating to discover that some individuals—perhaps precisely those who produce atypically timed performances—have atypically shaped bias profiles. However, because detection of timing deviations is a relatively low-level perceptual process, it may turn out that individual differences in the shape of the bias profile are small (as also would be predicted by a psychoacoustic account). Another important question is whether individuals without musical training or with training in styles different from Western classical music exhibit the same biases. In other words, are the temporal implications of Western classical music specific to its style and culture, or do they reflect more general properties of complex sound structures that transcend stylistic and cultural domains?

More specific questions may be asked about the perceptual processes underlying the detection of timing deviations. Listeners obviously must use some internal time keeper—a mental "clock" or oscillator (Desain, 1992; Jones & Boltz, 1989; Large & Kolen, 1994; Povel & Essens, 1985)—to detect deviations from isochrony. However, there may be an asymmetry between increment and decrement detection: An event that occurs earlier than predicted may precipitate the

next "tick" of the internal clock and thus may cause more rapid adjustment to a phase shift in isochronous timing than an event that occurs after the predictive tick has been issued. Four findings seem to suggest such a processing asymmetry.

First, decrement detection was somewhat more difficult overall than increment detection, even though it represented a larger Weber fraction (i.e., if the duration of the shorter interval is the denominator in both cases). Thus, listeners may be less conscious of a deviation that leads to a more rapid phase resetting of the mental clock. Second, the DAP for decrements was generally somewhat less varied than the DAP for increments and less highly correlated with the timing profile. This, too, may be attributable to a rapid adjustment of the internal time keeper to early-occurring events, which would reduce the effect of any biases induced by the musical structure. Third, a reduction in the frequency of late responses was observed in the decrement detection task relative to the increment detection task. If late responses have a perceptual basis, the finding indicates that listeners are more prone to perceive an IOI following a lengthened IOI as long than they are to perceive an IOI following a shortened IOI as short. This is again consistent with a rapid adjustment of the internal time keeper to early-occurring events: Early-occurring events advance the tick about to be issued while maintaining the period of the clock cycle, so that the duration of the next cycle will be accurately predicted. Late-occurring events can affect only the next tick, which occurs at a considerable temporal distance and therefore may not be delayed sufficiently, so that the following event may also occur late with respect to it. It may take several clock cycles to adjust to a delayed phase shift (Large & Kolen, 1994). Fourth, late responses in the increment detection task were more frequent to targets preceding IOIs that were perceived as being relatively long (i.e., had a negative bias). This suggests that the bias enhanced the tendency to perceive the IOI following a lengthened IOI as long, which supports the hypothesis that the tendency is perceptual in nature. In the decrement detection task, on the other hand, there was no systematic variation in the incidence of late responses across positions in the music, which suggests that there was no perceptual tendency to be enhanced by a bias.

An important question for future research is whether the perceptual bias reflects a modulation of the time-keeping function itself or an external bias that operates at a subsequent stage in perception. If the ticks of the mental clock are issued at unequal intervals, these perturbations must be quite small to be consistent with the listeners' high accuracy in the task. Shaffer (1981) postulated a "flexible clock" to account for expressive timing in performance, but expressive timing variation is much larger than the modulation envisioned here. Alternatively, the time keeper may maintain its mechanical regularity (except when it needs to adjust to phase shifts), and the bias may represent a variable criterion that decides whether a registered timing deviation reaches consciousness. This possibility seems more consistent with theories of oscillatory timing mechanisms (Large & Kolen, 1994), but the issue needs to be investigated empirically, and relevant research is in progress (Repp, in press).

References

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

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

Clynes, M. (1983). Expressive microstructure in music, linked to living qualities. In J. Sundberg (Ed.), *Studies of music performance* (pp. 76–186). Stockholm, Sweden: Royal Swedish Academy of Music.

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

Crowder, R. G., & Neath, I. (1995). The influence of pitch on time perception in short melodies. *Music Perception, 12,* 379–386.

Derwort, A. (1938). Untersuchungen über den Zeitablauf figurierter Bewegungen beim Menschen [Research on the time course of figural movements in humans]. *Pflügers Archiv für die gesamte Physiologie des Menschen und der Tiere, 240,* 661–675.

Desain, P. (1992). A (de)composable theory of rhythm perception. *Music Perception, 9,* 439–454.

Drake, C. (1993). Perceptual and performed accents in musical sequences. *Bulletin of the Psychonomic Society, 31,* 107–110.

Drake, C., & Botte, M.-C. (1993). Tempo sensitivity in auditory sequences: Evidence for a multiple-look model. *Perception & Psychophysics, 54,* 277–286.

Drake, C., & Palmer, C. (1993). Accent structures in music performance. *Music Perception, 10,* 343–378.

Epstein, D. (1995). *Shaping time: Music, the brain, and performance.* New York: Schirmer.

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

Friberg, A. (1991). Generative rules for music performance: A formal description of a rule system. *Computer Music Journal, 15,* 56–71.

Friberg, A., & Sundberg, J. (1995). Time discrimination in a monotonic, isochronous sequence. *Journal of the Acoustical Society of America, 98,* 2524–2531.

Gabrielsson, A. (1974). Performance of rhythm patterns. *Scandinavian Journal of Psychology, 15,* 63–72.

Gabrielsson, A. (1987). Once again: The theme from Mozart's Piano Sonata in A major (K.331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81–103). Stockholm, Sweden: Royal Swedish Academy of Music.

Henderson, M. T. (1936). Rhythmic organization in artistic piano performance. In C. E. Seashore (Ed.), *Objective analysis of musical performance* (pp. 281–305). Iowa City: Iowa University Press.

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

Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science, 6,* 177–208.

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

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, England: Cambridge University Press.

Monahan, C. B., & Hirsh, I. J. (1990). Studies in auditory timing: 2. Rhythm patterns. *Perception & Psychophysics, 47,* 227–242.

Oosten, P. van (1993). Critical study of Sundberg's rules for expression in the performance of melodies. *Contemporary Music Review, 9,* 267–274.

Palmer, C. (1989). Mapping musical thought to musical perfor-

mance. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 331–346.

Palmer, C. (1992). The role of interpretive preferences in music performance. In M. R. Jones & S. Holleran (Eds.), *Cognitive bases of musical communication* (pp. 249–262). Washington, DC: American Psychological Association.

Palmer, C., & Holleran, S. (1994). Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Perception & Psychophysics, 56,* 301–312.

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

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

Repp, B. H. (1992b). Detectability of rhythmic perturbations in musical contexts: Bottom-up versus top-down factors. In C. Auxiette, C. Drake, & C. Gérard (Eds.), *Proceedings of the Fourth Rhythm Workshop: Rhythm perception and production* (pp. 111–116). Bourges, France: Imprimérie Municipale.

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

Repp, B. H. (1993a). Some empirical observations on sound level properties of recorded piano tones. *Journal of the Acoustical Society of America, 93,* 1136–1144.

Repp, B. H. (1993b). Music as motion: A synopsis of Alexander Truslit's "Gestaltung und Bewegung in der Musik" (1938). *Psychology of Music, 21,* 48–72.

Repp, B. H. (1995a). Acoustics, perception, and production of legato articulation on a digital piano. *Journal of the Acoustical Society of America, 97,* 3862–3874.

Repp, B. H. (1995b). Expressive timing in Schumann's "Träumerei": An analysis of performances by graduate student pianists. *Journal of the Acoustical Society of America, 98,* 2413–2427.

Repp, B. H. (1995c). Detectability of duration and intensity increments in melody tones: A possible connection between music perception and performance. *Perception & Psychophysics, 57,* 1217–1232.

Repp, B. H. (1996). The dynamics of expressive piano performance: Schumann's "Träumerei" revisited. *Journal of the Acoustical Society of America, 100,* 641–650.

Repp, B. H. (1997a). The aesthetic quality of a quantitatively average music performance: Two preliminary experiments. *Music Perception, 14,* 419–444.

Repp, B. H. (1997b). *Control of expressive and metronomic timing in pianists.* Manuscript submitted for publication.

Repp, B. H. (1998). Obligatory "expectations" of expressive timing induced by perception of musical structure. *Psychological Research, 61,* 33–43.

Repp, B. H. (in press). Detecting deviations from metronomic timing in music: Effects of perceptual structure on the mental timekeeper. *Perception & Psychophysics.*

Schulze, H.-H. (1978). The detectability of local and global displacements in regular rhythmic patterns. *Psychological Research, 40,* 173–181.

Schütte, H. (1978). Subjektiv gleichmäßiger Rhythmus: Ein Beitrag zur zeitlichen Wahrnehmung von Schallereignissen [Subjectively even rhythm: A contribution to the temporal perception of sound events]. *Acustica, 41,* 197–206.

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

Shigeno, S. (1986). The auditory tau and kappa effects for speech and nonspeech stimuli. *Perception & Psychophysics, 40,* 9–19.

Sundberg, J., Friberg, A., & Frydén, L. (1989). Rules for automated performance of ensemble music. *Contemporary Music Review, 3,* 89–109.

Tekman, H. G. (1992). *Perception of rhythmic structure in tone sequences: Interactions of relative timing, intensity, and pitch.* Unpublished doctoral dissertation, Dartmouth College, Hanover, NH.

Tekman, H. G. (1995). Cue trading in the perception of rhythmic structure. *Music Perception, 13,* 17–38.

Terhardt, E., & Schütte, H. (1976). Akustische Rhythmus-Wahrnehmung: Subjektive Gleichmäßigkeit [Acoustic rhythm perception: Subjective regularity]. *Acustica, 35,* 122–126.

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

Viviani, P., & Stucchi, N. (1989). The effect of movement velocity on form perception: Geometric illusions in dynamic displays. *Perception & Psychophysics, 46,* 266–274.

Viviani, P., & Stucchi, N. (1992a). Biological movements look uniform: Evidence of motor–perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 603–623.

Viviani, P., & Stucchi, N. (1992b). Motor-perceptual interactions. In G. E. Stelmach & J. Requin (Eds.), *Tutorials in motor behavior II* (pp. 229–248). Amsterdam: Elsevier.

Woodrow, H. (1909). A quantitative study of rhythm. *Archives of Psychology, 14,* 1–68.