Music Perception
Spring 1998, Vol. 15, No. 3, 265–289

The Detectability of Local Deviations from a Typical Expressive Timing Pattern

BRUNO H. REPP
Haskins Laboratories

In three experiments, musically trained listeners tried to detect local deviations from a typical pattern of expressive timing in the initial measures of Chopin's Etude in E Major. The goal of the experiments was to neutralize structurally based perceptual biases ("timing expectations") that had been revealed previously in the detectability of local deviations from metronomic timing in the same music. In Experiment 1, which presented deviations from full-scale expressive timing, the perceptual biases were reversed, indicating that listeners' timing expectations were smaller than the typical timing variation in performance. Experiments 2 and 3 therefore presented deviations from a down-scaled timing pattern (25% or 10% of its original magnitude). Although neither experiment succeeded in completely neutralizing the perceptual biases, correlational evidence suggested that listeners' timing expectations were about five times smaller than typical expressive timing variations. The small, essentially subliminal size of the listeners' timing expectations is consistent with recent findings of related unintentional timing variation in deliberately metronomic performance and in finger-tap synchronization with a metronomic rendition of the same music.

Virtually all music is played with expression—that is, with systematic variation in its kinematic and acoustic parameters that can be measured on a continuous scale but cannot be captured by a discrete notational system. One of the most important expressive parameters in performances of the Western standard repertoire is timing. Although musical scores give a picture of metronomic regularity, this is an abstraction because the music is never performed with metronomic precision. Expressive timing is integral to this music, even though there are many possible ways of timing a performance.

Experimental evidence in support of this statement has been obtained in a series of experiments (Repp, 1992, 1995b, in press-a, in press-b) in which listeners had to detect small local increments (lengthenings) or decrements

Address correspondence to Bruno H. Repp, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695. (e-mail: repp@haskins.yale.edu)

(shortenings) in the otherwise equal tone interonset intervals (IOIs) of metronomically paced piano music, generated under computer control. The detectability of these perturbations varied greatly with position in the music, and the "detection accuracy profiles"—graphs showing the variation in percent-correct scores across the different positions—were closely related to a typical expressive timing profile for the same music, obtained by measuring and averaging the IOIs of a number of performances by different pianists. In places where pianists typically lingered, an IOI increment was hard to detect whereas a decrement was usually easy to detect, and the opposite was true in places where pianists typically moved ahead. This perception-performance correlation suggests that listeners expect to hear music played with expressive timing and compensate perceptually for its absence.

Other authors, too, have noted that a particular musical passage may "demand" a particular expressive timing pattern that delineates structural units. Commenting on performance practice, musicologist Roy Howat observed that "judicious tempo adjustment is less intrusive or discernible than the *sagging and cramping* (italics added) that results from imposing a uniform tempo on local material which does not suit it" (Howat, 1995, p. 16). Penel and Drake (in press) have argued that elementary grouping processes lead to distortions of perceived timing in music, distortions that performers themselves need to compensate for in their playing. Repp (in press–b) likewise attributed listeners' "timing expectations" primarily to melodic-rhythmic grouping processes operating on a segmented melody, although other aspects of musical structure (e.g., harmony, meter) may well give rise to timing expectations also.

The present study was not concerned with unraveling the relationship between musical structure and timing in perception or performance (even though this is an important project), for it used only a single musical excerpt of rather unambiguous structure. Rather, its main purpose was to determine the *magnitude* of the timing expectations induced by a particular musical passage, relative to the magnitude of typical expressive timing modulations observed in performance. By introducing expressive timing into the music and having listeners detect deviations from this modulated baseline pattern, the present experiments attempted to neutralize the perceptual biases that, in a metronomic baseline stimulus, were thought to arise from unfulfilled timing expectations. The magnitude of the expressive timing variation necessary to neutralize the biases was expected to provide a measure of the average magnitude of listeners' timing expectations.

The present experiments were very similar in design and method to an earlier experiment that used a metronomic baseline (Repp, in press–b, experiment 1). In that experiment, local IOI increments (lengthenings, hesitations) were significantly easier to detect than IOI decrements (shortenings, accelerations) of the same absolute magnitude, although the detectability of both varied enormously as a function of position. As already indicated, the detection accuracy profile (DAP) for increments showed a high negative correlation with a typical expressive timing profile for the music (i.e., increments were difficult to detect where expressive lengthening was expected), whereas the DAP for decrements showed a smaller positive correlation with the timing profile. From these two DAPs, two additional profiles were derived: A "sensitivity profile" was obtained by averaging the two DAPs, and a "bias profile" was obtained by subtracting the increment DAP from the decrement DAP and dividing by two. In other words, the sensitivity profile represented variations in detectability with position in the music regardless of the direction of the deviation, whereas the bias profile represented variations in the extent to which decrements were easier to detect than increments.[1] Whereas the sensitivity profile was found to be unrelated to the typical expressive timing profile, the bias profile was positively correlated with the timing profile (a positive bias reflects an expectation of lengthening) and therefore was considered an even better measure of listeners' timing expectations than either of the DAPs alone. It was this bias profile that the present experiments tried to neutralize.

The initial hypothesis was that the average listener expects to hear the music with average expressive timing, that is with variations in IOI duration having both the pattern and the magnitude typically observed in performance. Experiment 1, therefore, required listeners to detect deviations from such a typical timing pattern. If this pattern fulfills listeners' timing expectations, as it were, then there should no longer be any significant variation in perceptual bias, and the bias profile should be uncorrelated with the expressive timing profile. As will be seen, this naive prediction proved to be quite wrong. Experiments 2 and 3 then used smaller magnitudes of expressive timing variation, in a search for the magnitude that would neutralize the bias profile.

Another purpose of Experiment 1 was simply to assess the detectability of deviations from an expressive timing pattern. Obviously, such a task would be extremely difficult if the timing pattern (consisting of 36 successive, unequal IOIs extending over nearly 20 s) were arbitrary or carried by a sequence of nonmusical sounds. However, the typical timing pattern of a musical excerpt is not arbitrary, and this should enable musically trained listeners to remember and anticipate it without much practice. It was expected, however, that the deviations, to be detectable, would have to be made larger than in a metronomically timed baseline stimulus. In a previ-

1. The terms "sensitivity" and "bias" bring to mind signal detection theory (see Macmillan & Creelman, 1991), but they are used more informally here. For various reasons, the rigorous methods of signal detection theory could not be applied in the present paradigm (see Repp, in press–b).

ous investigation of this very issue, Clarke (1989) found that, to be equally detectable, timing deviations in an expressively timed musical excerpt had to be about twice as large as deviations from metronomic timing. The present musical materials were more complex than those of Clarke, so that the task was somewhat closer to real music listening. In fact, the task of detecting deviations from typical expressive timing is implicit in performance judgment: Not only do different performances of the same music often differ only by small nuances, but the timing of a single performance may sound locally hesitant or irregular (e.g., due to a technical problem or an expressive mannerism or an imperfect splice in a recording), which implies that the listener expected to hear a different, more typical timing pattern. In places where deviations from an expressively timed baseline are difficult to detect, performers (and recording studio editors) presumably have more leeway than in places where deviations are highly detectable. Positional variations in the detectability of deviations from typical timing thus may provide a measure of the local temporal flexibility of a musical structure.

Experiment 1

METHOD

Materials and Design

The musical excerpt was the beginning of Chopin's Etude in E Major, op. 10, no. 3. A basic computer-generated score is shown in Figure 1a. To provide a clear ending, a chord was substituted for the original notes in the second half of bar 5. Following the initial eighth-note upbeat, the music contains 36 sixteenth-note IOIs and thus 36 possible "target" positions in the detection task. Longer note values are irrelevant to the definition of target IOIs, because there is at least one note onset at every sixteenth-note subdivision of every beat.

The MIDI instructions for the stimuli were created as text in a spreadsheet program and were played back on a Roland RD-250s digital piano with "Piano 1" sound under the control of MAX software running on a Macintosh Quadra 660AV computer.[2] Nominally simultaneous note onsets were synchronous within the accuracy limits of MIDI transmission, and each note ended nominally (i.e., with a "note off" MIDI command) when another note began, which ensured perceived legato articulation without use of the pedal. The IOIs and MIDI velocities (relative tone intensities) were derived from performances of the excerpt on the digital piano by nine skilled pianists (Repp, 1997, in press–b). Each pianist had played the excerpt three times from the score (as shown in Figure 1a), so that there were 27 individual renditions. In each, the IOIs were calculated from the recorded MIDI note onsets, using the onset time of the highest (soprano or alto) note in each cluster of nominally simultaneous onsets. The IOIs and MIDI velocities were subsequently averaged across the 27 performances. Because the pianists' timing profiles were fairly similar to each other, their

2. Measurements revealed that, for reasons internal to the software, the tempo of the sound output was uniformly accelerated by a small amount (about 2.4%). All absolute durations are reported here as they were specified in the MIDI instructions.

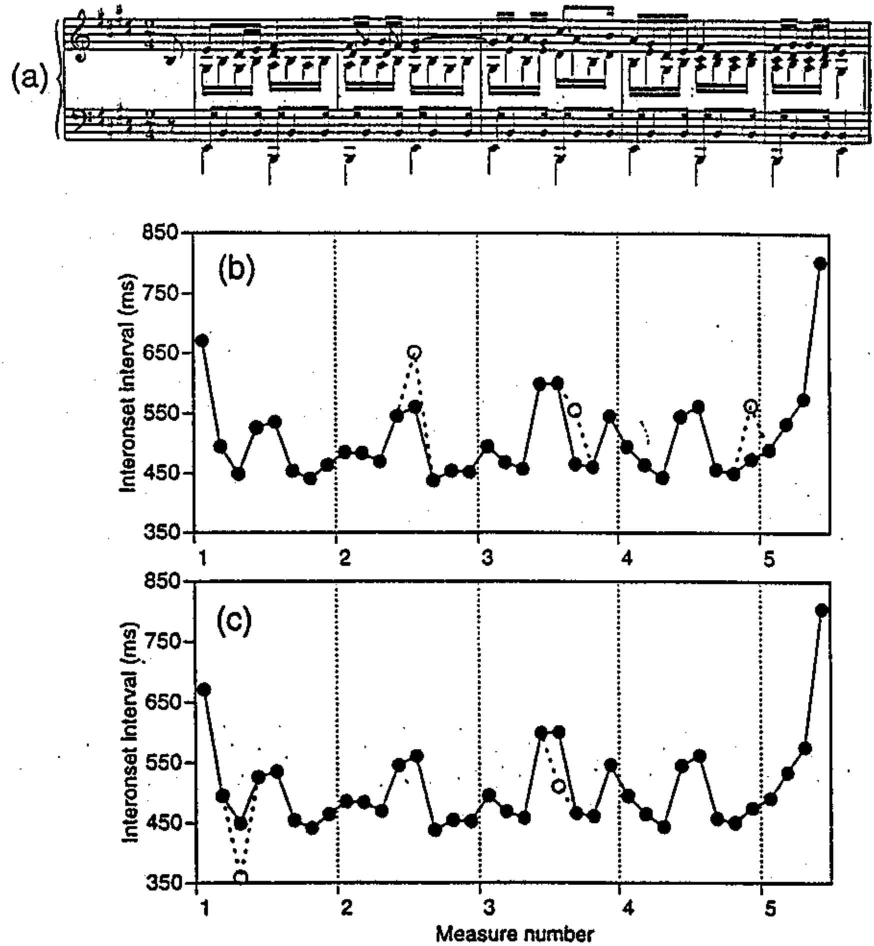

Fig. 1. (a) The beginning of Chopin's Etude in E Major, op. 10, no. 3, without slurs and expression marks and with a final chord added. (b) Example of an Experiment 1 trial containing three 90-ms increments. (c) Example of a trial containing two 90-ms decrements. The baseline timing profile is shown as a solid line connecting filled circles; the changes are shown as dotted lines connecting open circles.

average timing profile was representative and was considered a typical timing profile for this music. The average timing profile was also maximally smooth, because uncontrolled timing variation, which is present to some extent in any individual performance, was averaged out. This average timing profile, then, was considered the best estimate of the average listener's timing expectations. (Note that the perceptual data, too, were going to be averaged across a number of individuals.) The expressive baseline profile is shown as the filled circles in Figures 1b and 1c. It can be seen that IOI durations varied between about 440 and 600 ms, except for the initial and final IOIs, which were longer. In each bar, the fourth and fifth IOIs, which surround the onsets of long melody notes, were lengthened, as was the final IOI in bar 3.[3] The IOI of the initial eighth-note upbeat is not shown because it was never altered in the detection task; its duration was 896 ms.

Two parallel tests were prepared, one for the detection of IOI increments, 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"). Each block had two trials with 4 targets, four with 3, six with 2, four with 1, and two trials with no target. Within each block, a target occurred once in each position, and the 36 target positions were allocated randomly to trials, with successive targets in the same trial separated by at least four unchanged IOIs, usually more. 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 100 ms in the five familiarization trials, and 90, 70, and 50 ms, respectively, in the three test blocks. Figure 1b illustrates a trial with three 90-ms increments, and Figure 1c shows a trial with two 90-ms decrements. 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. At the same time, the onsets of all following notes were delayed or advanced by the same amount (i.e., there was no compensatory change in the immediately following IOI). The analog output of the digital piano was recorded onto digital tape.

The design and materials were identical to those of the earlier experiment with a metronomic baseline, except for (a) the presence of expressive timing in the baseline stimulus, (b) larger timing changes to be detected, and (c) some minor differences in stimulus construction and playback control.

Listeners

Twelve musically trained Yale undergraduates participated as paid volunteers. They were members of the Yale Symphony Orchestra or Concert Band and thus were skilled players of various instruments, although usually not of the piano.

Procedure

The test tape was played back on a digital audiotape recorder. The participants listened at a comfortable intensity over Sennheiser HD 540 reference II earphones. Half of the participants received the increment detection test first, and half received the decrement detection test first. Written instructions explained the experiment in detail, including the facts that there were between zero and four changes in IOI duration to be detected on each trial and that these changes occurred with respect to a fixed expressive timing pattern. This pattern was illustrated by playing the baseline stimulus three times at the beginning of the experiment. 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 note(s) occupying lengthened or shortened IOI(s), *not* the following note(s), should be circled. (Earlier studies had shown a strong tendency toward "late" responses, despite similar instructions.) For the five familiarization trials, the correct responses were indicated 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, participants received a detailed written explanation to take home.

RESULTS AND DISCUSSION

3. In the context of the present study, it is not necessary to discuss the possible structural reasons for these lengthenings. Repp (in press–b) emphasized the grouping of melody notes into segments or gestures, but harmony and meter may play a role as well. This issue is addressed most thoroughly in Repp (1998c).

Overall Accuracy

As in previous experiments in this series, a lenient scoring criterion was adopted, according to which responses adjacent to a target position were considered hits. Of all correct responses, 6% were "early" and 28% "late" in increment detection, whereas 7% were early and 17% were late in decrement detection. The high incidence of late responses and the difference in their frequency between the two conditions replicate results obtained previously; for a detailed discussion, see Repp (in press–b).

Figure 2 shows average percent-correct scores (hits) and chance levels estimated from the false-alarm rates. Following Repp (in press–b), the chance level estimates were calculated as 1/6 (the approximate average guessing probability) times the average number of false-alarm responses per trial times 6/5 (a correction factor for scoring 1/6 of all false alarms as hits) times 100 (to obtain percentages). The hit rates were in the intended midrange (optimal for avoiding ceiling and floor effects in the DAPs) and, naturally, declined across blocks. Decrement detection was easier than increment detection; the difference was significant in a two-way (condition by block) repeated-measures analysis of variance on the hit percentages [$F(1,11) = 23.10, p < .001$], not even taking into account the corresponding difference in false-alarm rates (see Figure 2). The difference in hit rates between the two conditions also increased with the difficulty of the task [$F(2,22) = 5.28, p < .02$, for the condition by block interaction]. The advantage for decrement detection contrasts with previous experiments with metronomically timed baselines, all of which had shown either an advan-

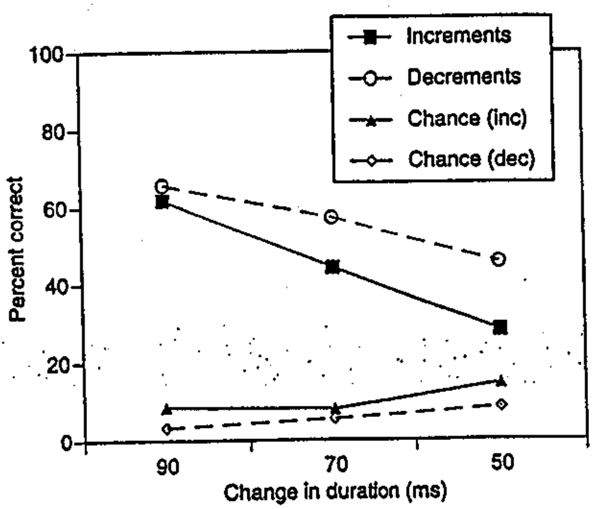

Fig. 2. Overall percent-correct scores (hits) and estimated chance levels in Experiment 1 as a function of condition and test block (amount of duration change).

tage for increment detection or no difference (Repp, in press–a, in press–b). It is also noteworthy that the false-alarm rates were only slightly higher than in earlier studies, and then only for increments. This indicates that listeners rarely mistook the expressive timing modulations of the baseline stimulus for detection targets.

Detection Accuracy Profiles

Figure 3a shows the detection accuracy (hit) profiles (DAPs) for increments and decrements as well as, at the bottom of the panel, the distribution of false alarms across positions—the false-alarm profiles (FAPs). It can be seen that the DAPs were extremely varied, particularly for decrement detection. No significant relationship was found between the two DAPs [$r(34) = -.07$]. Although a positive correlation between the DAP and FAP is expected because the same perceptual biases are thought to underlie both, there was little similarity between the DAP and FAP for increments [$r(34) = .19$, *ns*], whereas the analogous correlation for decrements reached significance [$r(34) = .47$, $p < .01$]. However, the FAP for increments was related to the baseline timing profile (Fig. 1) [$r(33) = .52$, $p < .01$],[4] indicating a tendency for listeners to false-alarm to long IOIs. The FAP for decrements showed a weaker, negative correlation with the timing profile [$r(33) = -.40$, $p < .05$], which indicates a tendency *not* to false-alarm to long IOIs. In general, however, the FAPs are not very reliable because they are based on sparse data, deriving mainly from a few individuals with exceptionally high false-alarm rates.

Even though the DAP for increments was highly varied, it was unrelated to the baseline timing profile [$r(33) = .12$, *ns*]. By contrast, the DAP for decrements showed a very high negative correlation with baseline IOI duration [$r(33) = -.89$, $p < .001$]: Decrements were easy to detect in short IOIs but very difficult to detect in long IOIs. (Note the dips in positions 4 and 5 in each bar, and in position 8 of bar 3; cf. Figure 1.) In the earlier experiment in which deviations from a metronomic baseline were to be detected, the DAP for increments showed a high negative correlation with the typical timing profile, whereas the DAP for decrements showed a *positive* correlation. This was interpreted in terms of perceptual biases due to a perceptual compensation for the absence of expressive timing. These biases were expected to be neutralized in the present experiment. However, it seems that there were instead biases in the opposite direction. This finding was examined further by computing sensitivity and bias profiles from the two DAPs.

4. Because of the exceptionally long duration of the final IOI (see Figure 1), this position was omitted in all correlations with the baseline timing profile. This is evident in the degrees of freedom (33 rather than 34).

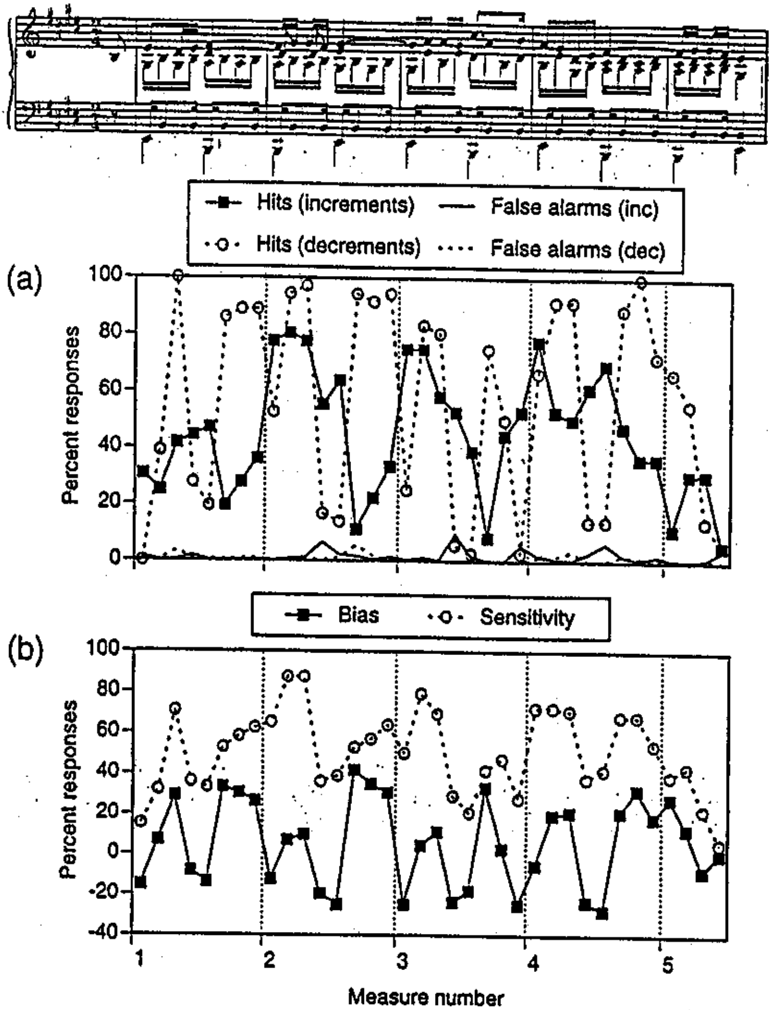

Fig. 3. (a) Detection accuracy profiles and false-alarm profiles in Experiment 1. (b) Sensitivity and bias profiles. The musical score above the figures serves as a reference for the abscissa.

Sensitivity and Bias Profiles

The derived sensitivity and bias profiles are shown in Figure 3b. The sensitivity profile is the average of the two DAPs and thus shows the average detectability of a timing change regardless of its direction. The bias profile is the difference of the two DAPs (decrements minus increments) divided by two and thus shows the extent to which shortenings are easier to detect than lengthenings. It is evident that, contrary to predictions, the bias profile is far from flat. In fact, the sensitivity and bias profiles are about equally varied, and they also show a positive correlation with each other [$r(34) = .48$, $p < .01$], owing to the greater variation in the DAP for decrements than in the DAP for increments (Figure 3a).[5]

The sensitivity profile showed a substantial negative correlation with the baseline timing profile [$r(33) = -.77$, $p < .001$]. Thus sensitivity declined with increasing IOI duration, which is consistent with psychophysical data on duration discrimination (e.g., Drake & Botte, 1993; Friberg & Sundberg, 1995). However, even though the sensitivity profile obtained with a metronomic baseline did not correlate with the typical timing profile, it exhibited considerable similarity to the present sensitivity profile [$r(34) = .61$, $p < .001$]. This suggests that baseline IOI duration may not have been the primary cause of the variations in sensitivity (see the General Discussion later).

The bias profile was effectively the inverse of that obtained with a metronomic baseline [$r(34) = -.72$, $p < .001$], and it correlated negatively with the baseline timing profile [$r(33) = -.78$, $p < .001$], in contrast to the bias profile for a metronomic baseline. The most likely reason for this reversal of biases is that the variations in the baseline timing profile were larger than the listeners' perceptual timing expectations. Whereas, with a metronomic baseline, IOIs expected to be long were perceived as too short, so that shortenings were easier to detect than lengthenings in these IOIs, in the present stimuli they may have been perceived as too long, with the consequence that lengthenings were easier to detect than shortenings. The opposite was true for IOIs expected to be short. This suggests that a similar experiment with reduced baseline timing variations might succeed in neutralizing the perceptual biases. Experiments 2 and 3 explored this possibility. Although it may seem surprising at first that listeners do not expect what is typical in performance, the present experiments do not concern cognitive expectations (which may well match the typical timing profile) but perceptual expectations (or auditory distortions; Penel & Drake, in press) that are automatically elicited by the musical input. It seems quite reasonable, at least in hindsight, that the magnitude of listeners' perceptual timing expectations would be much smaller than the magnitude of typical expressive timing modulations.

Detectability of Deviations from a Typical Timing Profile

The results of Experiment 1 show that musically trained listeners can detect deviations from a typical timing profile without extensive training, provided that the deviations are not too small. This confirms that the typical (average) timing profile has a privileged relationship to the musical structure, which makes it possible to remember and anticipate its pattern. How-

5. The sensitivity and bias profiles would be uncorrelated if both DAPs exhibited the same variance, regardless of their shapes.

ever, the fact that detection accuracy varied greatly with position suggests that deviations were perceived not so much with respect to a remembered temporal shape as in terms of the local continuity and structural appropriateness of each temporal pattern as it unfolded. This is presumably what an ordinary music listener does in judging the expressive timing of any individual live or recorded performance, where there is no explicit "baseline" to compare it with. Listeners were particularly sensitive to shortenings of short IOIs, perhaps as sensitive as they would be in a metronomic baseline. This finding indicates that, in this particular musical excerpt, the short IOIs defined a relatively stable basic or maximal tempo, local increases in which were perceived as irregularities suggesting lack of timing control. By contrast, listeners were very insensitive to shortenings of long IOIs, which amount to a reduction of expressive lengthening. Such reductions do not seriously violate either the local temporal continuity or the musical structure; in fact, performances by different artists often exhibit different degrees of expressive lengthening within the same qualitative timing pattern (Repp, 1998c).

On the average, lengthenings were about equally detectable in long and short IOIs. Lengthening of a long IOI results in an exaggeration of expressive lengthening, whereas lengthening of a short IOI is structurally inappropriate. The average degree of acceptability of these two kinds of distortions may be similar. However, this is not the whole story because the DAP for increments was far from flat: There was strong positional variation in the detectability of increments, even though this variation was not correlated with IOI duration. Its discussion is postponed until the General Discussion.

Experiment 2

Experiment 2 used a baseline stimulus in which the expressive timing variations were reduced to 25% of their original size. This degree of reduction was chosen arbitrarily, as the magnitude of listeners' timing expectations was not known a priori. One consequence of the reduction was that the timing variations were now rather inconspicuous (at least as long as specific attention was not drawn to them), so that the task could be presented to the listeners as one of detecting deviations from even timing.

Unlike Experiment 1, Experiment 2 used IOI increments and decrements that were fixed percentages of baseline IOI duration. Together with the substantial reduction in the range of baseline IOI durations, this was intended to eliminate any decrease in overall sensitivity with increasing IOI duration.

METHOD

Materials and Design

The baseline expressive timing variations were reduced to 25% of their magnitude in Experiment 1, assuming a hypothetical minimal IOI duration of 450 ms: $IOI_{new} = 450 + (IOI_{old} - 450)/4$. The underlying assumption here was that timing expectations are essentially expectations of local slowing (expressive emphasis) relative to an expressively neutral tempo, defined in the present musical excerpt by the accompaniment during sustained melody notes (see Figure 1). Consequently, the range of variation in IOI durations, apart from the initial and final IOIs, was less than 50 ms (about 440–490 ms). The timing changes to be detected were 12% of the baseline IOI duration in the familiarization trials, and 10%, 8%, and 6%, respectively, in the three test blocks (all rounded to the nearest millisecond). In all other respects, the materials and design were the same as in Experiment 1.

Listeners

Ten musically trained Yale undergraduates from the same pool participated as paid volunteers.

Procedure

The procedure was the same as in Experiment 1, except that the three introductory presentations of the baseline stimulus were omitted. The instructions no longer made any reference to expressive timing and mentioned only that deviations from even timing (local hesitations or accelerations) were to be detected.

RESULTS AND DISCUSSION

Overall Accuracy

Of all correct responses, 8% were "early" and 29% "late" in increment detection, whereas 6% were early and 19% were late in decrement detection. Once again, these percentages are in good agreement with those obtained in earlier experiments. Figure 4 shows the average percent-correct scores (hits) and the chance levels estimated from the false-alarm rates. Naturally, there was a decline in accuracy across blocks. However, in contrast to Experiment 1, decrement detection was not easier than increment detection [$F(1,11) = 0.26$]. Because increment detection had been easier than decrement detection in the earlier experiment using a metronomic baseline, the reduced timing variations in the present baseline stimulus evidently succeeded in neutralizing this overall difference. Moreover, overall accuracy was similar to that with a metronomic baseline; no significant difference was found between comparable conditions in the two experiments. The false-alarm rate was even slightly lower in the present experiment.

Detection Accuracy Profiles

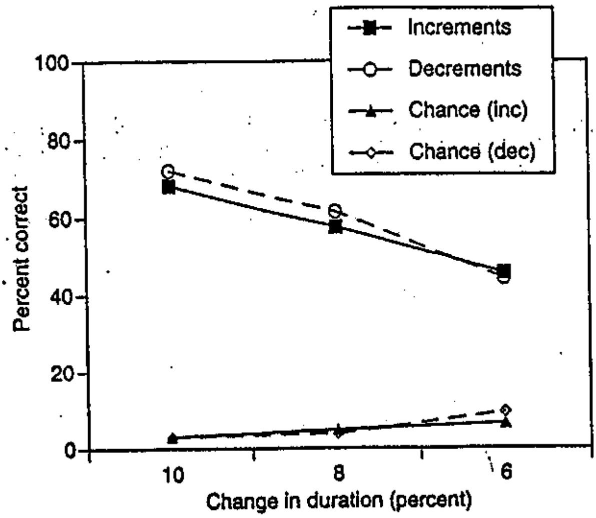

Fig. 4. Overall percent-correct scores (hits) and estimated chance levels in Experiment 2 as a function of condition and test block (amount of duration change).

Figure 5a shows the DAPs and FAPs. Contrary to expectations, the two DAPs were not at all flat, although they were less varied than in Experiment 1. Again, the DAPs were statistically unrelated [$r(34) = .27$, *ns*]. As in Experiment 1, the DAP and FAP for increments were unrelated [$r(34) = .15$, *ns*] whereas the DAP and FAP for decrements showed a significant similarity [$r(34) = .67$, $p < .001$]. The FAP for increments showed a weak correlation with the baseline timing profile [$r(33) = .42$, $p < .05$], which reflects a tendency to false-alarm to long IOIs, as in Experiment 1. The FAP for decrements showed no such relationship.

Also as in Experiment 1, the DAP for increments was unrelated to the baseline timing profile [$r(33) = -.04$], whereas the DAP for decrements still showed a negative correlation with baseline IOI duration. That correlation was smaller than in Experiment 1 but highly significant [$r(33) = -.58$, $p < .001$]. Thus, decrements were still easier to detect in short than in long IOIs.

Sensitivity and Bias Profiles

The derived profiles are shown in Figure 5b. The profiles were not significantly related [$r(34) = -.23$, *ns*]. Contrary to expectations, there was still considerable variation in each profile, although the bias profile was clearly less varied than in Experiment 1. Both profiles again showed negative correlations with baseline IOI duration, but they were much smaller than in Experiment 1 [$r(34) = -.44$ and $-.46$, respectively, $p < .01$]. The

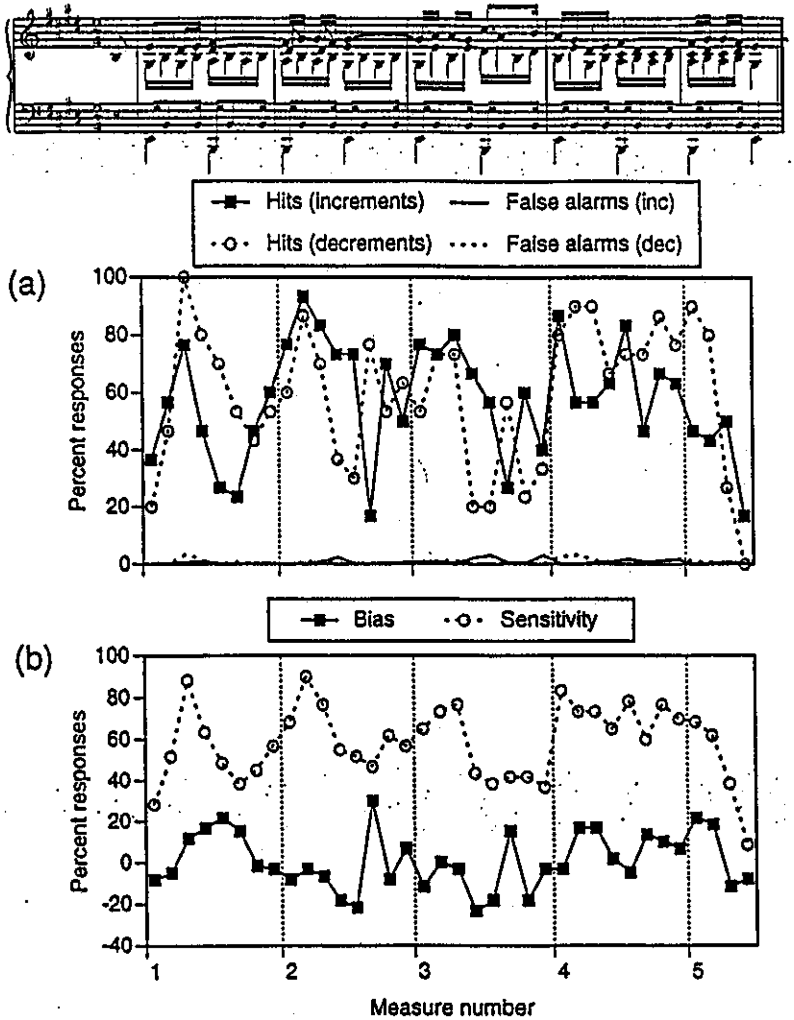

Fig. 5. (a) Detection accuracy profiles and false-alarm profiles in Experiment 2. (b) Sensitivity and bias profiles.

sensitivity profile was quite similar to that in Experiment 1 [$r(34) = .79$, $p < .001$], and the bias profiles of Experiments 1 and 2 also resembled each other [$r(34) = .61$, $p < .001$].

The persistence of a negative correlation between the bias profile and the typical timing profile suggests that the baseline timing variations may still have been larger than listeners' timing expectations. It should be recalled that a high positive correlation between the bias profile and the typical timing profile was obtained with a metronomic baseline. Therefore, a third experiment was conducted in which the baseline timing variations were reduced further.

Experiment 3

METHOD

Materials and Design

The baseline expressive timing variations were reduced to 10% of their magnitude in Experiment 1, again assuming a hypothetical baseline IOI duration of 450 ms: $IOI_{new} = 450 + (IOI_{old} - 450)/10$. The resulting range of variation in IOI durations, apart from the initial and final IOIs, was less than 20 ms.[6] The timing changes to be detected were 11% of the IOI duration in the familiarization trials, and 9%, 7%, and 5%, respectively, in the three test blocks (rounded to the nearest millisecond). In all other respects, the materials and design were the same as in Experiments 1 and 2.

Listeners

Twelve paid volunteers participated. They were mostly Yale undergraduates who had responded to a summer advertisement on campus. On average, they were less sophisticated musically than the participants in Experiments 1 and 2, although all had some musical training and could read music. Repp (1998b) has shown that the shape of the DAP for increment detection in a metronomic rendition of the Chopin Etude excerpt varies but little with musical experience (see also Repp, 1995b).[7]

Procedure

The procedure was the same as in Experiment 2.

RESULTS AND DISCUSSION

Overall Accuracy

Of all correct responses, 5% were "early" and 42% were "late" in increment detection, whereas 10% were early and 37% were late in decrement detection. The percentages of late responses were exceptionally high, evidently due to the listeners' lesser musical experience, and here showed little difference between the two conditions.[8] Figure 6 shows the average percent-correct scores (hits) and the chance levels estimated from the false-alarm rates. Apart from the obvious decline in accuracy across blocks, increment detection was significantly easier than decrement detection [$F(1,11)$

6. It was discovered that the stimuli used in earlier experiments had contained some temporal jitter on the order of a few milliseconds, owing to MIDI transmission delays. Although this was not a serious problem (for details and evidence in support of this statement, see Repp, 1998b), it was corrected in the present stimuli.

7. However, musically inexperienced listeners perform more poorly overall. The data of three additional listeners in Experiment 3 were excluded because of their low detection scores, and several other volunteers did not even attempt the experiment because they could not hear the changes in the practice trials.

8. Two listeners spontaneously commented on their tendency to circle the note following a change, attributing it to the difficulty of keeping track of the music in the score. Because the following note is heard when a change is detected, backtracking is required in marking the response. This and the subsequent catching up may have been a difficult maneuver for some poor music readers, although none of them gave 100% late responses.

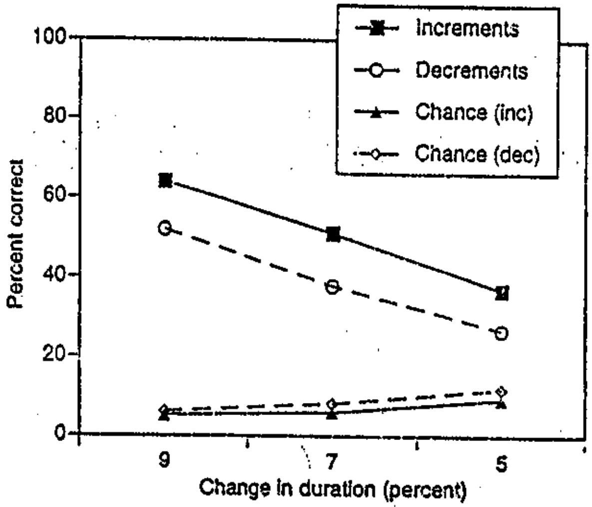

Fig. 6. Overall percent-correct scores (hits) and estimated chance levels in Experiment 3 as a function of condition and test block (amount of duration change).

= 8.8, $p < .02$]. This replicates results obtained with a metronomic baseline and contrasts with the overall results of Experiment 2 and especially Experiment 1.

Detection Accuracy Profiles

Figure 7a shows the DAPs and FAPs. The DAPs were by no means flat. As in Experiment 2, they showed a weak positive relation, which reached significance here [$r(34) = .36$, $p < .05$]. This implies that variations in sensitivity (which affect both profiles similarly) were larger than variations in bias (which affect the profiles in opposite ways). The DAP and FAP for increments were positively correlated here [$r(34) = .51$, $p < .01$], as were the DAP and FAP for decrements [$r(34) = .45$, $p < .01$]. The FAP for increments no longer showed a significant positive correlation with the baseline timing profile [$r(33) = -.09$], nor did the FAP for decrements [$r(33) = .27$, *ns*].

A negative relationship between the DAP for increments and the typical timing profile, which is characteristic of experiments with metronomic baselines, began to emerge [$r(33) = -.35$, $p < .05$], while the DAP for decrements no longer showed a negative correlation with baseline IOI durations [$r(33) = .15$, *ns*].

Sensitivity and Bias Profiles

The derived profiles (Figure 7b) were not significantly related [$r(34) = -.12$, *ns*]. As already mentioned, the bias profile was less varied than the

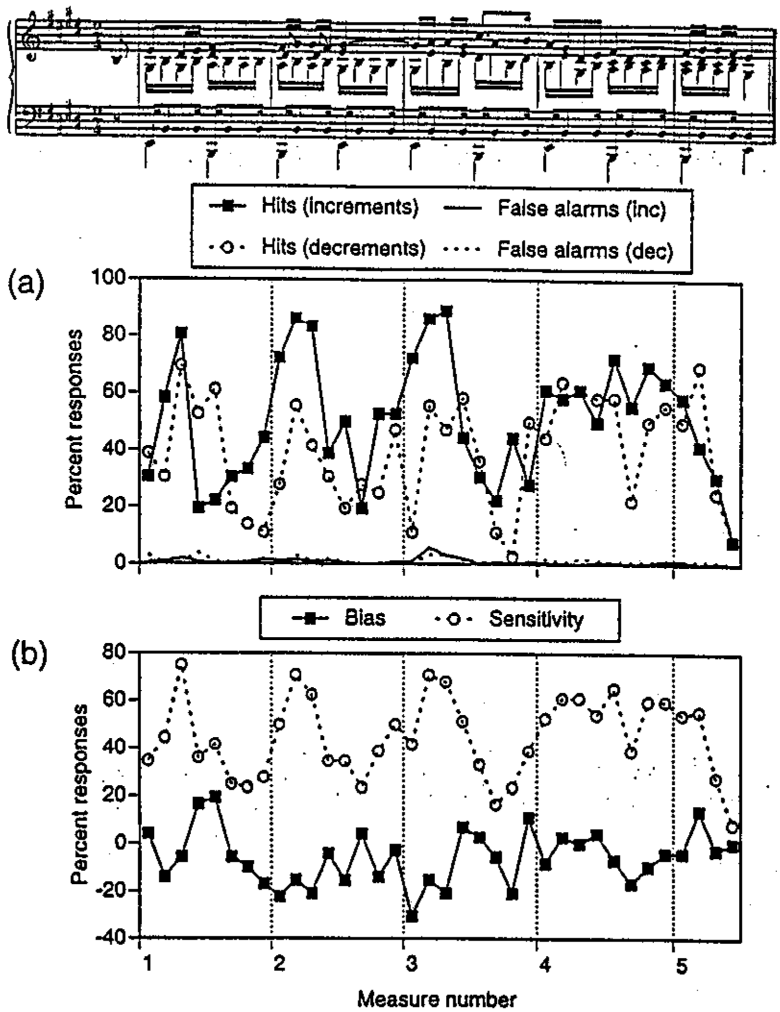

Fig. 7. (a) Detection accuracy profiles and false-alarm profiles in Experiment 3. (b) Sensitivity and bias profiles.

sensitivity profile. The sensitivity profile no longer showed a significant negative correlation with baseline IOI duration [$r(33) = -.14$], whereas the bias profile began to exhibit the positive correlation with the typical timing profile that is characteristically obtained in experiments with metronomic baselines [$r(33) = .43$, $p < .02$]. The sensitivity profile was moderately similar to that in Experiment 1 [$r(34) = .60$, $p < .001$] and highly similar to that in Experiment 2 [$r(34) = .86$, $p < .001$]. The bias profile changed much more dramatically, being unrelated to that in Experiment 1 [$r(34) = -.13$] and only weakly related to that in Experiment 2 [$r(34) = .40$, $p < .05$]. On the whole, these results indicate that the baseline timing variations in Experiment 3 were smaller than the listeners' timing expectations.

General Discussion

It was hypothesized at the outset that using typical expressive timing variations as a baseline might neutralize the perceptual biases that had been observed with a metronomic baseline. Experiment 1 showed this prediction to be wrong: Expressive timing of full magnitude reversed rather than neutralized the bias profile. Experiment 3, on the other hand, yielded results that resembled those for a metronomic baseline, which suggests that 10% timing variations were insufficient for bias neutralization. Experiment 2, using 25% timing variations, probably came closest to the goal.

This conclusion is supported by a correlational comparison of the response profiles of Experiments 1–3 with those obtained for a metronomic baseline, presented in Table 1. It is evident that the DAPs for increments and decrements were differentially affected by the introduction of expressive timing variations. Whereas the DAP for increments lost its similarity to the metronomic DAP only when 100% expressive timing was present, the DAP for decrements lost this similarity with 25% expressive timing and reversed itself with 100% expressive timing. This is due to a change in perceptual bias: The sensitivity profile retained more or less the same shape throughout, whereas the bias profile reversed itself. If a zero correlation of the nonmetronomic bias profile with the metronomic bias profile is taken as the criterion for bias neutralization, then it may be concluded that this neutralization was nearly achieved in Experiment 2. Complete neutralization of bias (a zero correlation) would perhaps have occurred with 20% expressive timing.

Alternatively, a zero correlation between the nonmetronomic bias profile and the typical timing (baseline) profile may be taken as the criterion for bias neutralization. The relevant correlations are shown in the bottom row of Table 2. Those correlations suggest that baseline timing variation

TABLE 1
Correlations of Response Profiles Obtained in Experiments 1–3 with Response Profiles for a Metronomic Baseline

	Experiment		
	1 (100%)	2 (25%)	3 (10%)
DAP (inc)	.18	.42*	.63***
DAP (dec)	–.59***	–.03	.57***
Sensitivity	.61***	.77***	.73***
Bias	–.72***	–.21	.62***

Note. DAP = detection accuracy profile.
*$p < .05$. **$p < .01$. ***$p < .001$.

TABLE 2
Correlations of Response Profiles Obtained in Experiments 1–3 and in the Experiment with a Metronomic Baseline with the Typical Timing Profile (Figure 1)

	Experiment			
	1 (100%)	2 (25%)	3 (10%)	Metronomic
DAP (inc)	.12	−.04	−.35*	−.82***
DAP (dec)	−.89***	−.58***	.15	.70***
Sensitivity	−.77***	−.44**	−.14	−.25
Bias	−.78***	−.46**	.43*	.85***

Note. DAP = detection accuracy profile; inc = increment; dec = decrement.
*$p < .05$. **$p < .01$. ***$p < .001$.

about halfway between 10% and 25% might have led to a zero correlation. Table 2 also summarizes other correlations showing that, as increasing amounts of expressive timing were introduced, the DAP for increments lost its negative relationship with the typical timing profile, the DAP for decrements, like the bias profile, moved from high positive to high negative correlations with the timing profile, and the sensitivity profile, although seemingly stable according to the correlations in Table 1, actually developed a negative relationship with the timing profile. However, this negative correlation was almost certainly enhanced in Experiment 1 by the use of fixed-size duration changes.

From the foregoing, it may be concluded that the average magnitude of listeners' perceptual timing expectations is about one fifth to one sixth of the average magnitude of expressive timing variations in performances of the Chopin Etude excerpt. Considering that the 25% baseline timing variations in Experiment 2 were no longer perceptually obvious—the listeners accepted the baseline stimuli as evenly timed and committed relatively few false alarms—the perceptual biases then are effectively of subliminal magnitude, with a range of no more than 30 ms, except for initial and final IOIs. This finding is in agreement with the informal observation that a truly metronomic performance does not sound subjectively uneven: Its subjective unevenness is subliminal and becomes evident only indirectly when timing perturbations are to be detected. The conclusion is also in agreement with the finding that pianists who intend to play music metronomically still show small timing variations that are related to their expressive timing profiles (Palmer, 1989; Penel & Drake, in press; Seashore, 1938/1967), a result that was recently replicated with the Chopin Etude excerpt (Repp, 1998a). Finally, the conclusion is also consistent with the recent finding of small but systematic variations in the asynchronies between finger taps and musical note onsets when participants were required to tap in synchrony with a metronomic version of the Chopin Etude excerpt; again, these variations were related to the typical expressive timing profile (Repp, 1998b). These studies provide converging evidence that subliminal timing expectations can be evoked by a musical structure.

As long as only correlational evidence is considered, the results of the present study are fairly clear. However, they are not so straightforward at a more detailed level. Although it was possible to minimize the correlation between the bias profile and the typical timing profile, the bias profile never became really flat. In other words, there were always local asymmetries between increment and decrement detection. A closer comparison of the profiles in the four experiments (including the one with the metronomic baseline) may help unravel these asymmetries.

Figure 8 shows the DAPs for increments (Figure 8a) and decrements (Figure 8b) in the four experiments superimposed and aligned by adjusting the percentages of each DAP to yield an average score of 50%. Data point symbols have been omitted, so as not to clutter the graphs too much. The baseline timing profiles of the stimuli are shown in Figure 8c. Consider first the DAPs for decrements (Figure 8b). The changes with respect to the metronomic condition are rather straightforward: Where IOIs were lengthened in the baseline (Figure 8c), detection scores plummeted; where IOIs stayed roughly the same, detection scores increased, relatively speaking. In absolute terms, of course, overall detection performance decreased as expressive timing in the baseline stimulus increased, and so it should not be inferred that any decrements were easier to detect in a modulated than in a metronomic stimulus. Although absolute performance levels are difficult to compare across experiments, it seems that the introduction of expressive timing affected mainly the detection of decrements in expressively lengthened IOIs. This makes sense: As already pointed out, reduction or even negation of typical (average!) expressive lengthening is common in performance and does not affect the underlying basic tempo, which in this excerpt is assumed to be defined by the short IOIs in the accompaniment passages between melodic gestures. The shortening of short IOIs, however, accelerates the basic tempo and therefore is highly detectable even in a modulated performance. In the metronomic condition, however, decrements are easier to detect where lengthening is expected because these IOIs seem subjectively too short to begin with.

Consider now the DAPs for increments (Figure 8a). In the metronomic condition, lengthenings were more difficult to detect where they were expected (i.e., where they were long in the typical timing profile; cf. Figure 8c). Introduction of expressive timing in the baseline increased detection scores in these places, again relatively speaking. (In absolute terms, there was probably no increase anywhere.) However, where there was high detectability in the metronomic condition (i.e., in IOIs that were short in the

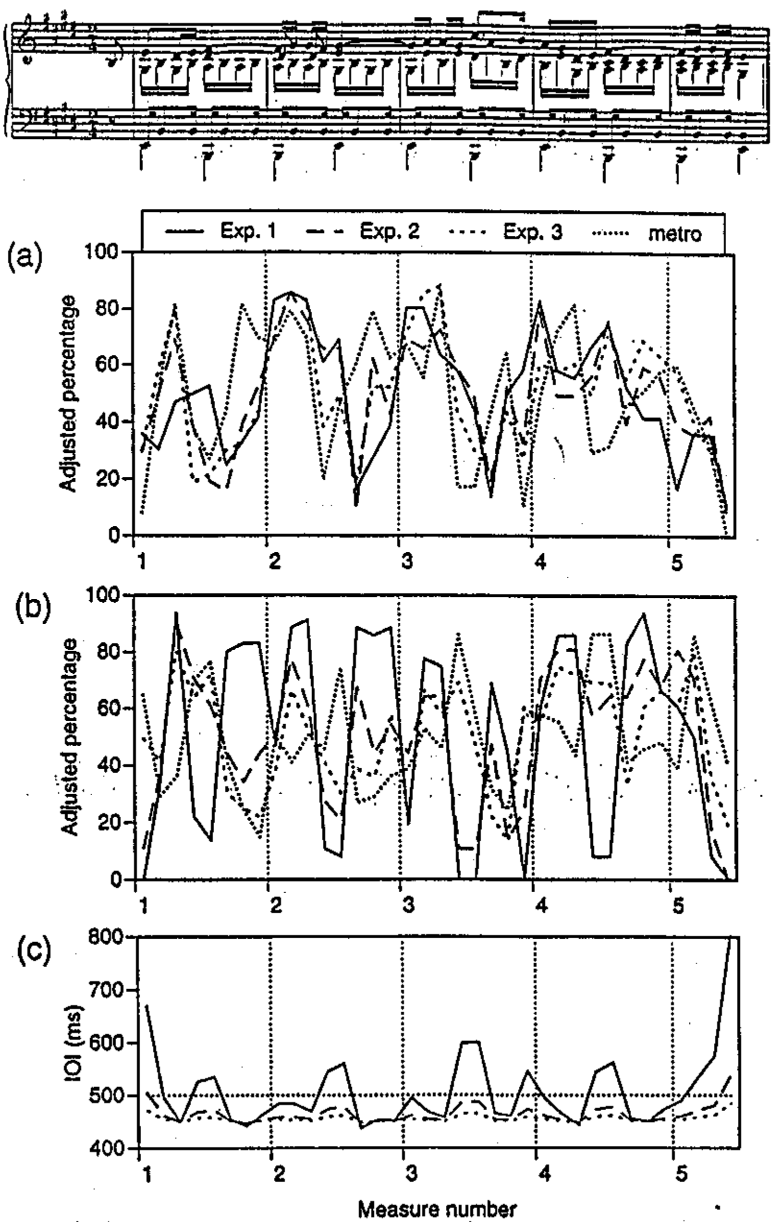

Fig. 8. Detection accuracy profiles for (a) increments and (b) decrements in four experiments (metro = data from Repp, in press–b, experiment 1). All profiles have been centered on 50%. The baseline timing profiles of the stimuli are shown in panel (c).

typical timing profile; cf. Figure 8c), introduction of expressive lengthening elsewhere had an asymmetric effect: Detection scores tended to remain relatively high in short IOIs preceding lengthened IOIs (especially in the initial IOIs of bars 2 and 3) but dropped in short IOIs following lengthened IOIs (especially in the final three IOIs of bars 1 and 2, and in the sixth and seventh IOIs of bar 3). This suggests a contextual factor not considered so far: Expressively lengthened IOIs may automatically create a perceptual expectation of similar lengthening in immediately following IOIs (i.e., of a continuation of the local slowing in tempo), so that increments in these following IOIs are difficult to detect. If one or two short IOIs intervene, this expectation evaporates, resulting in a rapid increase in the detectability of subsequent IOI increment targets, as indicated by the steep increases in the DAPs at the ends of bars 1 and 2 and also in position 7 of bar 3. At the end of bar 4, this effect may have been attenuated by the approaching end of the excerpt, which seems to introduce an expectation of a final ritard that reduces increment detection across the board.

This context effect in the DAPs for increments may be largely responsible for the persistence of systematic variations in bias in Experiment 2, the condition that came closest to neutralizing these variations. Note especially the peaks in positive bias (i.e., expectations of lengthening) in the sixth IOIs of bars 2 and 3 (Figure 5b), which immediately followed two expressively lengthened IOIs, associated with negative bias. The bias profiles of the four experiments, centered around zero, are superimposed in Figure 9b. Representing difference scores, these functions may be compared in an absolute sense, and the asymmetry between the IOIs preceding and following expressively lengthened IOIs should be noted, especially in the first three bars: The increase in negative bias as expressive timing was introduced into the baseline was much more pronounced after expressively lengthened IOIs (position 6 in each bar) than before them (position 3). This difference is a consequence of the asymmetry in the DAPs for increments.

What remains to be explained is the systematic variation in sensitivity. Figure 9a shows the sensitivity profiles of the four experiments aligned and superimposed. They show relatively little change in shape across experiments, although absolute sensitivity to temporal change certainly decreased as expressive timing was introduced. The Experiment 1 profile differs most from the others, which may be attributed to the use of fixed amounts of duration change in that study. According to a generalized Weber's law for duration discrimination (Friberg & Sundberg, 1995; Getty, 1975; Ivry & Hazeltine, 1995), changes of the same absolute size are more difficult to detect in longer IOIs than in shorter IOIs within the range of IOIs used in the present experiments, and the range of baseline IOI durations was wide enough in Experiment 1 for such an effect to emerge. As can be seen, sensitivity in Experiment 1 tended to be lower (relatively speaking) where baseline IOIs were long (cf. Figure 9c) and higher for the surrounding short IOIs. The differences among the sensitivity profiles for the other three experiments, which used proportional duration changes, were fairly small; if anything, the systematic variation observed in the metronomic condition was

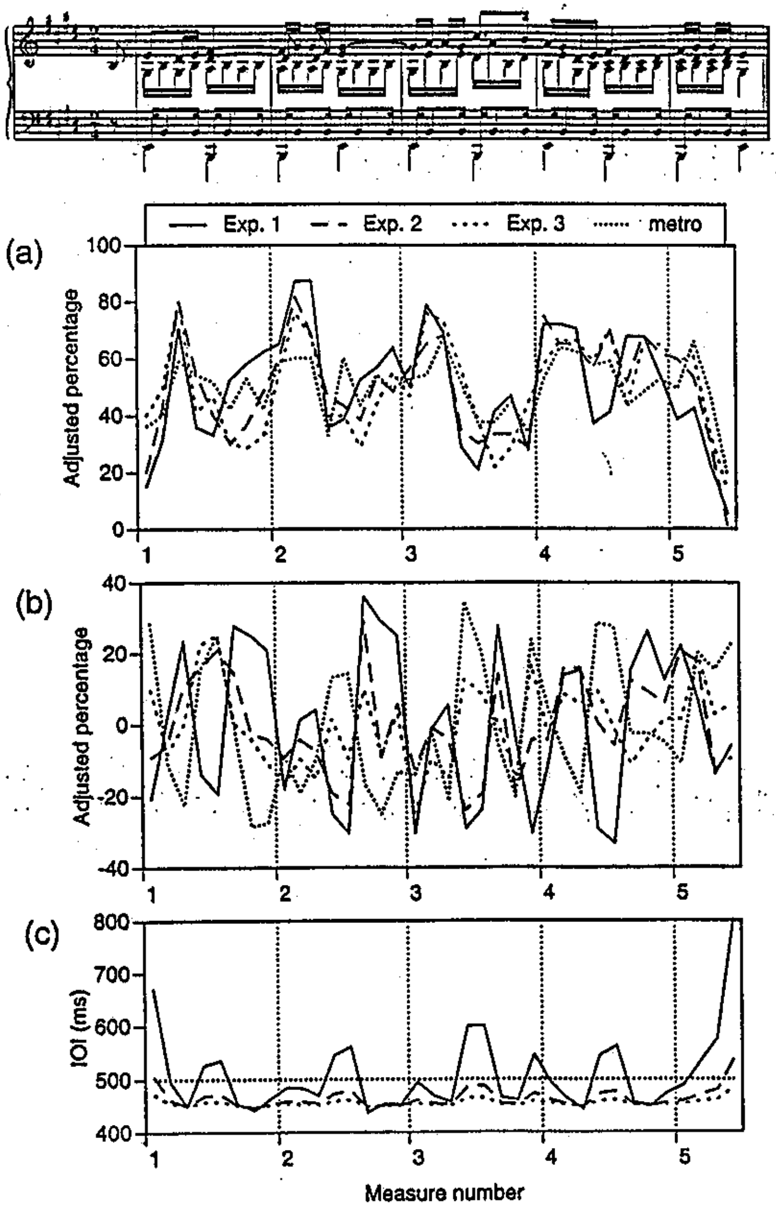

Fig. 9: (a) Sensitivity and (b) bias profiles of four experiments (metro = data from Repp, in press–b, experiment 1). Sensitivity profiles have been centered on 50%, bias profiles centered on zero. The baseline timing profiles of the stimuli are shown in panel (c).

somewhat enhanced in Experiments 2 and 3. This variation, then, must have been caused by a stimulus parameter unrelated to timing.

In the metronomic condition, no significant relationship was found between the sensitivity profile and either the absolute or relative pitch or the absolute or relative intensity of the tones delimiting target IOIs (Repp, in press–b, experiment 1). In the present experiments, there were marginally significant correlations with some of these variables in Experiments 1 and 2, and a moderate correlation with the average intensity of the IOI boundary tones in Experiment 3 [$r(34) = .55, p < .001$]. Indeed, the variations in sensitivity seem to be related to the alternation between melody tones (higher pitched, louder) and accompaniment tones (lower pitched, softer), but not in a simple way. Sensitivity was high at the beginnings of melodic segments (with the notable exception of the final IOI in bar 3) but then decreased, reached a minimum at or shortly after the end of a segment (marked by the onset of a long note), and then increased again until the beginning of the next segment, especially in bars 1–3. Thus, detection accuracy decreased substantially during melodic gestures, whereas it increased during the intervening accompaniment. That increase could be explained as being due to the progressive decay of the long melody tone, which perhaps masked the onsets of the accompaniment tones and thus created temporal uncertainty. During melodic segments, masking of tone onsets may have occurred also, due to the acoustic postrelease decay of the preceding loud tone (Repp, 1995a). However, this does not explain why there was a progressive decrease in sensitivity for within-melody IOIs. This finding, which has been replicated in a new detection experiment using different musical materials (Repp, in preparation), requires further investigation.

In conclusion, the present results suggest that the perceptual timing expectations revealed previously in detection experiments with metronomic baseline stimuli are considerably smaller than the expressive timing variations typically encountered in performance. Introduction of down-scaled expressive timing variation in the baseline stimuli eliminated the correlation between the perceptual bias profile and the typical expressive timing profile, but it did not remove all systematic variation in perceptual bias, because of certain asymmetries in the detectability of IOI increments and decrements. Timing deviations can be detected without much difficulty in stimuli containing full-scale expressive timing, provided that the deviations are sufficiently large and (perhaps) that the listeners are musically trained. However, reductions of expressive lengthening are particularly difficult to detect, whereas reductions of short IOIs are very noticeable. This is reasonable in terms of performance practice: Reduction of typical expressive lengthening is a common strategy, but accelerations beyond a maximal local tempo serve no expressive purpose and suggest poor timing control.[9]

9. This research was supported by National Institutes of Health grant MH-51230. I am grateful to Lisa Robinson for extensive assistance, and to Eric Clarke, Carolyn Drake, and Richard Parncutt for helpful comments on earlier versions of the manuscript.

References

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

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

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

Getty, D. J. (1975). Discrimination of short temporal intervals: A comparison of two models. *Perception & Psychophysics, 18*, 1–8.

Howat, R. (1995). What do we perform? In J. Rink (Ed.), *The practice of performance* (pp. 3–20). Cambridge, U.K.: Cambridge University Press.

Ivry, R. B., & Hazeltine, R. E. (1995). Perception and production of temporal intervals across a range of durations: Evidence for a common timing mechanism. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 3–18.

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

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

Penel, A., & Drake, C. (in press). Sources of timing variations in music performance: A psychological segmentation model. *Psychological Research.*

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

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

Repp, B. H. (in press–a). Obligatory "expectations" of expressive timing induced by perception of musical structure. *Psychological Research.*

Repp, B. H. (in press–b). Variations on a theme of Chopin: Relations between perception and production of timing in music. *Journal of Experimental Psychology: Human Perception and Performance.*

Repp, B. H. (1998a). *Control of expressive and metronomic timing in pianists.* Manuscript submitted for publication.

Repp, B. H. (1998b). *Detecting deviations from metronomic timing in music: Effects of perceptual structure on the mental timekeeper.* Manuscript submitted for publication.

Repp, B. H. (1998c). *Quantitative analysis of expressive variation in pianists' performances of the initial measures of Chopin's Etude in E major: I. Timing.* Manuscript submitted for publication.

Seashore, C. E. (1967). *Psychology of music.* New York: Dover Publications. (Original work published 1938)