Psychol Res (1998) 61: 33–43

1076

ORIGINAL ARTICLE

Bruno H. Repp

Obligatory "expectations" of expressive timing induced by perception of musical structure

Received: 5 November 1996 / Accepted: 26 February 1997

Abstract Previous research has shown that, in a task requiring the detection of local deviations from mechanically precise timing in music, the relative detectability of deviations in different positions is closely related to the typical expressive timing pattern musicians produce when playing the music. This result suggests that listeners expect to hear music expressively timed and compensate for the absence of expressive timing. Three new detection experiments shed additional light on the nature of these timing expectations in musically trained listeners. Experiment 1 shows that repeated exposure to an atypically (but not unmusically) timed performance leaves listeners' timing expectations unaffected. Experiment 2 demonstrates that the expectations do not manifest themselves when listeners merely imagine the music in synchrony with a click track. Experiment 3, however, shows that the timing expectations are fully operational when the click track is superimposed on the music. These results reveal timing "expectations" to be an obligatory consequence of the ongoing auditory perception of musical structure.

Introduction

This study is concerned with the relationship between music performance and the perception of timing in music. It is well known that music – particularly music in the Western Romantic tradition, which concerns us here – must be played expressively in order to give joy to performers and listeners. An important aspect of this expression is the almost continuous modulation of the local tempo, or *expressive timing*. In large part, expressive timing seems to be a function of the music's division into successive temporal segments, variously called rhythmic groups, melodic gestures, phraselets, or motives. Each such segment seems to be associated with an acceleration-deceleration pattern, with the final deceleration usually being more pronounced than the initial acceleration and being especially notable at the ends of larger units such as whole phrases. The performer's (often subconscious) intention seems to be to "act out" the music's hierarchical grouping structure and thereby to communicate it to listeners (Povel, 1977; Clarke, 1985; Palmer, 1989; Todd, 1985, 1995).

Listeners attuned to classical music evidently appreciate this, though exactly why they do is a matter of speculation. Expressive timing is not necessary to perceive the grouping structure of music, expect in cases of high ambiguity. Usually there are other, stronger cues to grouping, such as the notated rhythmic structure (the pattern of unequal note interonset intervals, or IOIs). A long IOI usually signals the end of a group, and an additional lengthening of the preceding short IOI(s) is merely icing on the cake. This icing is important to listeners, however, perhaps because it enables them to participate empathically in the musician's actions. Perception of the unadorned structure usually gives pleasure only to music theorists (Cook, 1990).

When music is played with mechanical exactitude under computer control, the absence of expressive timing has not only aesthetic but also perceptual consequences. This is not immediately evident to introspection, for a mechanical performance simply sounds mechanical. However, in a task that requires listeners to detect small local deviations from the perfectly even timing of a mechanical performance, detection accuracy varies greatly as a function of position in the music. It turns out that a deviation is difficult to detect when it is in the same direction as the deviation that would typically be found in musicians' expressive timing at that same point in the music (Repp, 1992, 1995a, in press). In particular, the lengthening of an IOI is hard to detect when musicians would normally lengthen that IOI; the analogous relationship with regard to detection of shortening is less pronounced,

B. H. Repp
Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695, USA
e-mail: repp@haskins.yale.edu

perhaps because in performance, relative lengthening is expressively more significant than relative shortening. It seems, then, as if listeners *expected* to hear expressive timing and compensated perceptually for its absence in a mechanical performance. That is, an IOI expected to be long is perceived as relatively too short, and this is the reason why an actual small lengthening of that IOI is difficult to detect. In other words, listeners' timing expectations may interfere with the mechanism that keeps track of a steady beat, such as a mental clock or oscillator (e.g., Desain, 1991; Large & Kolen, 1994; Povel & Essens, 1985). Alternatively, listeners' timing expectations may leave the mental timekeeper unaffected but may govern some subsequent cognitive criterion that decides whether or not a deviation from even timing becomes conscious.

The aim of the present experiments was to learn more about the origin and nature of these timing expectations. Are they true expectations in that they exist prior to the perception of the musical object and are brought to bear on it, as it were? Or are they rather an automatic consequence of the on-line auditory and cognitive processing of musical structure, in which case they probably should not be called expectations at all? (Hence the quotation marks in the title of this paper). In the former case, they should be relatively flexible and under cognitive control. For example, it should be possible to have different timing expectations for the same music, to apply similar expectations to an unstructured auditory stimulus, or to suspend expectations when not attending to music. In the latter case, the "expectations" should be inflexible and obligatory, elicited by structured materials only, and unaffected by attention.

Experiment 1 examined whether listeners' timing expectations can be influenced by repeated exposure to an unusually (but not unmusically) timed performance of the same music. If timing expectations are flexible and based on past experience with expressive performances, they should be at least partially assimilated to the performance that serves as a model. If timing "expectations" are a function of the musical structure, however, they should not change.

Experiment 2 tested whether timing expectations can be applied to a nonmusical stimulus (a click sequence) by imagining music in synchrony with it. Again, if the expectations are applied to the auditory input, there may be an effect, perhaps an attenuated one. However, if they derive from on-line auditory and cognitive processing, they should not be in evidence in this task.

Experiment 3 superimposed clicks on music and asked listeners to attend to the clicks and ignore the music. If timing expectations are under cognitive control, they should be reduced or absent in this task. If they are obligatory consequences of auditory structural processing, however, they should be fully present.

All three experiments used the same music and employed very nearly the same design as Exp. 1 in Repp (in press), whose results provide baseline data for comparison with the present findings. While it would be inappropriate to re-describe this experiment in detail here, it is necessary to summarize its results and describe its methodological features that are shared with the present experiments. For details omitted here, the reader is referred to Repp (in press).

The baseline experiment

Method

Materials. The music, shown in Fig. 1a, was the beginning of Chopin's Etude in E major, op. 10, No. 3, with the final chord extended to the end of the fifth bar. Between the initial downbeat and the final chord, the music defines 36 sixteenth-note IOIs which served as "target" positions in the detection task. The melody, in the soprano voice, is divided into five rhythmic groups or melodic segments which end with longer notes on the second beat of each measure while the accompaniment in the other voices continues. While this melodic grouping structure is important for expressive timing and timing perception, it is irrelevant to the definition of target IOIs, due to the increase in the detection threshold with IOI duration (see, e.g., Friberg & Sundberg, 1995): It was assumed that

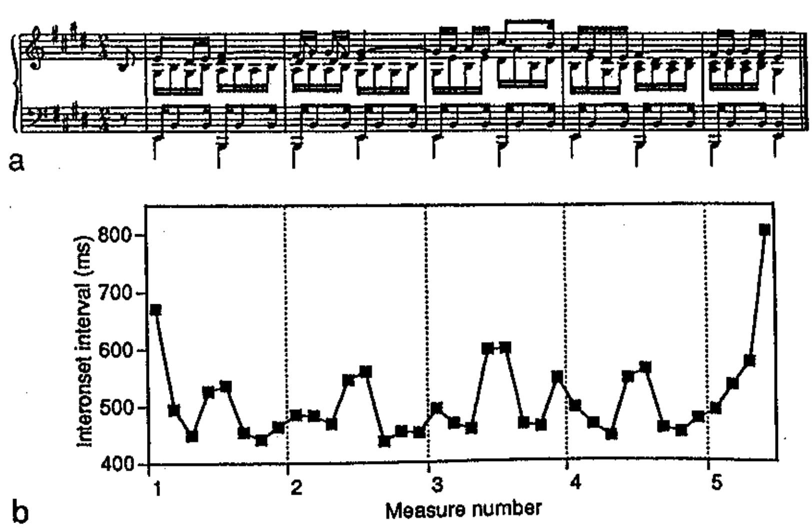

Fig. 1 **(a)** The musical score of the initial measures of Chopin's Etude in E major, op. 10, No. 3. Slurs and expression marks have been deleted, and the final chord has been extended. **(b)** A typical expressive timing profile, representing the average timing of 9 pianists' performances (Repp, in press: Exp. 1)

small timing changes would be detectable only with reference to the shortest IOI in which they occurred.

The music was synthesized under computer (MIDI) control on a Roland RD-250s digital piano, with a baseline sixteenth-note IOI duration of 500 ms. Although there was no expressive timing, expressive dynamics was present; the MIDI velocities (relative intensities) of the notes were taken from a human performance.

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. Each trial (a complete presentation of the musical excerpt) contained 0–4 lengthened or shortened IOIs ("targets"). In each block, there were 2 trials with 4 targets, 4 trials with 3, 6 trials with 2, 4 trials with 1, and 2 trials with no target. Two targets in close succession were avoided. In the course of a block, each of the 36 IOIs served as a target once.

The size of the duration increment or decrement was 52 ms in the five familiarization trials, and 42, 31, and 21 ms, respectively, in the three test blocks. Whenever an IOI was changed in duration, the durations of all notes that were "on" during the IOI were changed in the MIDI instructions and the onsets of all following notes were delayed by the same amount, so as to maintain *legato* articulation everywhere. The change to be detected thus was a phase shift in the isochronous rhythm underlying the performance.

Listeners and procedure. Fourteen musically trained undergraduate students participated. The test tape was played back on a DAT recorder, and participants listened at a comfortable intensity over Sennheiser HD 540 II earphones. Half of them received the increment detection test first, and half the decrement detection test. Participants circled the note(s) representing the lengthened or shortened IOI(s) on answer sheets that reproduced the upper staff of the musical score for each trial. For the 5 familiarization trials, the correct responses had already been filled in. Random guessing was discouraged; when no change was detected on a trial, a question mark was to be entered on the answer sheet.

Performance measurements. Nine skilled pianists, mostly graduate students at a major music school, were MIDI-recorded playing the Chopin excerpt three times on the digital piano after brief practice. The IOIs were determined from the onset times of the highest notes in chords.[1] An average *timing profile* was obtained by linearly averaging the IOIs over all pianists' performances, after ascertaining that the individual profiles were fairly similar to each other. This average profile is shown in Fig. 1b, where it is evident that pianists tended to slow down at the ends of melodic-rhythmic groups (around the second beat of each bar) and especially at the end of the excerpt, as well as at the beginning.

Results

Overall accuracy

Responses adjacent to a target position were scored as correct because there was a strong tendency to circle the note following a target position, especially in the increment detection task. Naturally, average percent correct scores declined as increment sizes decreased across the three blocks, from about 70% to about 40% (well above chance level, which was about 10%). Increment detection was significantly better than decrement detection.

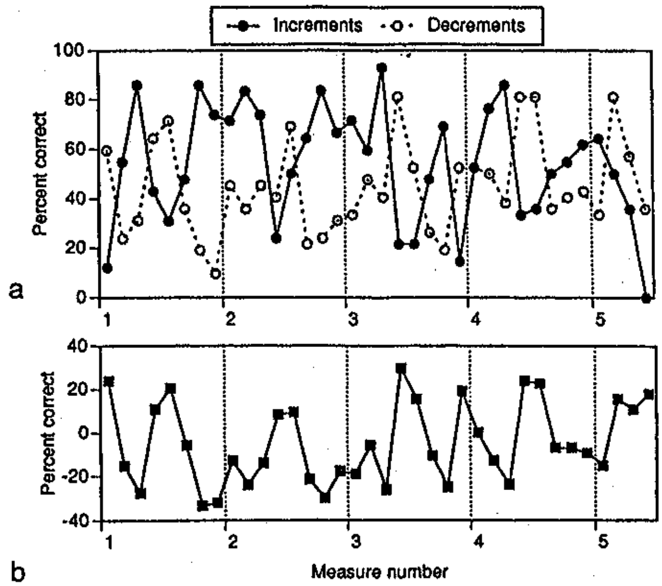

Fig. 2 **(a)** Detection accuracy profiles for IOI increments and decrements in the baseline study. **(b)** The bias profile derived from these data

Detection accuracy profiles

By combining all listeners' responses across all three blocks, an average detection accuracy profile (DAP), representing average percent correct scores (hits) as a function of position in the music, was obtained for each condition. As can be seen in Fig. 2a, detection accuracy varied greatly from position to position, and the DAPs for increment and decrement detection were negatively correlated [$r(34) = -0.53$, $p < .001$]: In general, decrements were easy to detect where increments were hard to detect, and vice versa. The DAP for increments was negatively correlated with the average performance timing profile shown in Fig. 1b [$r(33) = -0.82$, $p < .001$], whereas the DAP for decrements showed a positive correlation with the timing profile [$r(33) = 0.70$, $p < .001$].[2] Thus, an increment was difficult to detect where pianists would typically slow down, whereas it was easy to detect where pianists would typically move ahead; and the reverse for decrements. The higher perception-performance correlation for increments than for decrements, which was replicated in four subsequent experiments (Repp, in press), suggests that listeners' expectations concern primarily lengthenings, in agreement with the greater importance of lengthenings than of shortenings in expressive performance. A *bias profile* was subsequently derived by subtracting the increment DAP from the decrement DAP and dividing by two, under the assumption that these functions exhibited underlying

---

[1] There are slight asynchronies between nominally simultaneous note onsets in human performance. The notes with the highest pitch are usually the most salient and the most carefully timed.

[2] The final IOI was omitted in these correlations because of a floor effect in increment detection and extreme lengthening of the final IOI in performance.

perceptual biases of opposite polarity. This bias profile, shown in Fig. 2b, was even more highly correlated with the performance timing profile [$r(33) = 0.85$, $p < .001$]. In a way, it represents the listeners' timing expectations: A high bias indicates an expectation of relative lengthening.

Discussion

Repp (in press) contrasted three possible explanations of the perception-performance correlation. The first is that listeners have formed timing expectations from experience with expressively timed performances and that these expectations are applied to the evenly timed music, which results in perceptual compensation effects (Repp, 1992). A second possibility is that the complex acoustic structure of music causes auditory distortions in the perceived rhythm which are reflected in the detection accuracy results; musicians' expressive timing is seen as a form of (over-)compensation for these auditory distortions (Drake, 1993). A third proposal is that the perceptual and cognitive processing of musical grouping structure has strong timing implications that are reflected both in listeners' perceptual compensation and in musicians' kinematics (Repp, in press). The present experiments are relevant to the distinction between the first of these alternatives and the other two. The first ("expectation") hypothesis assumes that timing expectations are cognitive, experience-based, and at least somewhat flexible. The psychoacoustic hypothesis clearly implies the opposite – an obligatory relationship between acoustic structure and perceptual bias in timing judgments. The third ("structural") hypothesis, too, assumes a rigid connection between structure and timing but describes it in musicological rather than psychoacoustic terms, in order to provide a more palatable explanation of the perception-performance parallel. The distinction between the psychoacoustic and structural hypotheses thus is more epistemological than empirical. The structural hypothesis can be extended, however, to incorporate a higher level of structural interpretation which has the flexibility assumed by the expectation hypothesis. In that case, the empirical issue addressed here can be framed entirely within the structural hypothesis: Are timing expectations purely bottom-up or top-down, perceptual or cognitive, inflexible or flexible?

It is important to note that musicians *are* flexible in their expressive timing. Experienced artists, in particular, sometimes deviate considerably from the typical pattern instantiated by an average timing profile (Fig. 1b), sometimes but not always as a consequence of alternative structural conceptions of the music. Clearly, expressive timing is under cognitive control. In a principal components analysis on 15 famous pianists' timing profiles for the initial phrase of Chopin's E-major Etude (Fig. 1a), Repp (1997) found three very different underlying timing patterns, two of which are shown in

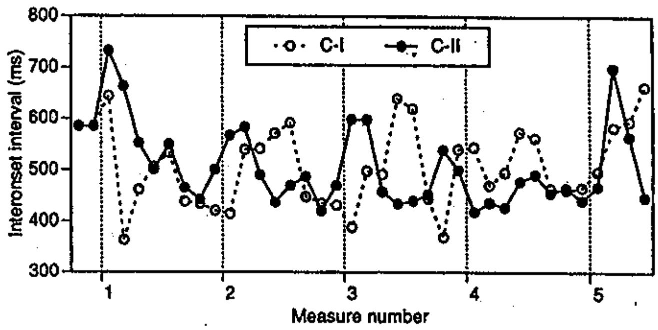

Fig. 3 The first two principal components (Varimax-rotated) of 15 famous pianists' timing profiles, rescaled into milliseconds (Repp, 1997)

Fig. 3. The first component, representative of 11 pianists, is similar to the average student profile shown in Fig. 1b. The second component, which is highlighted in Fig. 3, was representative of only 3 pianists and seems to reflect an alternative interpretation of the melodic grouping structure: The slowing down just before the beginning of each melodic segment suggests that the previous segment was viewed as including the whole duration of the final long note. In the standard interpretation, melodic segments seem to end with the onset of their final note or perhaps with the onset of the first subsequent note in the accompaniment, so that there are between-segment "gaps" filled with accompaniment only. These observations lead directly to the present Exp. 1.

Experiment 1

If listeners were exposed repeatedly to a performance with an atypical timing profile, such as the second component pattern in Fig. 3, would their perceptual biases in the detection task change? A significant effect of exposure would indicate that listeners' timing expectations are malleable and depend on specific experience; thus it would support the "expectation" hypothesis or top-down structural explanation outlined above. It would be incompatible with a psychoacoustic or bottom-up structural account. A negative finding, on the other hand, would be most compatible with a bottom-up account, though – as negative findings go – it would not necessarily contradict the other hypotheses.

Method

Materials and design. Experiment 1 differed from the baseline experiment in three ways:
(1) Only IOI increment detection was tested.
(2) Each trial was preceded by an expressively timed performance of the same music. That precursor performance always had the

atypical timing profile of the second principal component of the previous analysis of 15 famous pianists' timing profiles (Fig. 3).[3] The interval between a the precursor and the following trial was 2 s; that between a trial and the next precursor was 4 s.
(3) The increments to be detected were larger. As a pilot subject, the author found the task more difficult than previously and decided to use increments of 80, 60, and 40 ms, respectively, in the three test blocks. In the five familiarization trials, the increment was 100 ms.

In other respects, Exp. 1 was identical with the baseline experiment, including the quasi-random schedule that assigned targets to trials. One technical difference was that the MIDI instructions were created using arithmetic operations in a spreadsheet program, and the resulting stimulus sequences were played back and recorded on digital audio tape under control of a program written in MAX.[4]

Listeners and procedure. The listeners were 12 musically trained Yale undergraduates who were paid for their services. They were members of the Yale Symphony Orchestra or Concert Band and thus were skilled players of various instruments, but few of them were also pianists. The Chopin Etude thus was not very familiar to them, if at all. The procedure was the same as in the baseline experiment, except that the instructions pointed out that the precursors required no response, only passive listening.

Results and discussion

Overall accuracy

Of all responses scored as correct, 25% were "late" (i.e., attributed to the IOI following the target IOI). This percentage is similar to that observed in earlier increment detection studies; for a discussion, see Repp (in press). Figure 4 shows the average percentage of correct responses (hits) in the three test blocks, as well as the results of the baseline experiment for comparison. On the bottom of the figure, estimates of chance level performance derived from the false-alarm rates are shown. (See Repp, in press, for their derivation). Overall, accuracy was higher in Exp. 1 than in the baseline study, which indicates that the difficulty of the present task had been somewhat overestimated by the author when he chose the larger increment sizes. Nevertheless, it is clear that the task was more difficult than in the baseline study: Scores in Block 1 are almost equal, even though the increment was almost twice as large in Exp. 1. Also, the baseline scores in Block 1 (42-ms increments) were

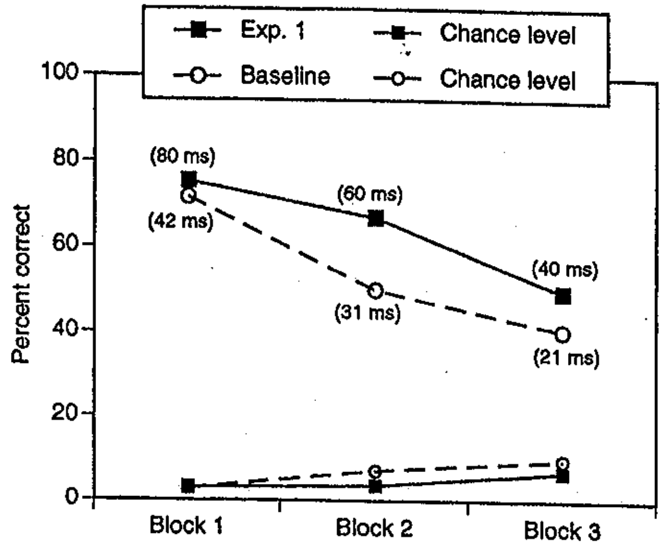

Fig. 4 Overall accuracy and chance level estimates in Exp. 1 and in the baseline study

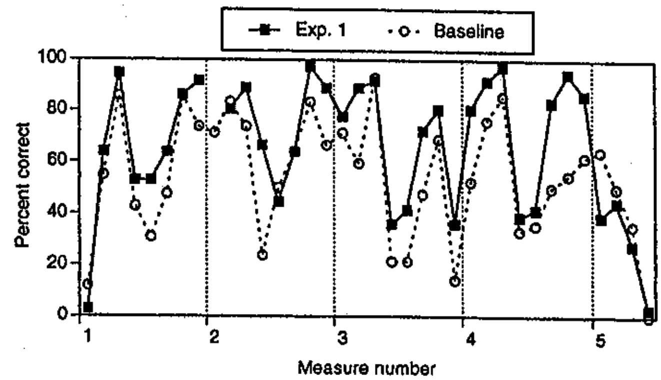

Fig. 5 Detection accuracy profiles for IOI increments in Exp. 1 and in the baseline study

significantly higher than those in the present Block 3 (40-ms increments) $[F(1,24) = 15.57,\ p < .0007]$, even though the present subjects had had more practice by that time. This is a fortuitous but interesting result: It appears that a temporally modulated precursor interferes with the detectability of deviations from isochrony, and this interference seems to persist throughout the test trial (see Fig. 5), even though 18 s would seem plenty of time to adapt to the isochronous rhythm. Large and Jones (submitted) have reported a similar long-term effect of exposure to temporal irregularity and have explained it in terms of a widening of the temporal "expectancy window" of a mental oscillator.[5]

[3] The IOIs in Fig. 3 were obtained by multiplying the standardized component scores with the average standard deviation of the original performances and adding this product to the grand average IOI of all performances (Repp, 1997). The duration of the initial upbeat, which had not been included in the principal components analysis, was the average of all pianists' upbeat durations. The MIDI velocities were similar to, but not identical with, those of the music in the test trials; they represented the average velocities of the 9 student performances.

[4] In the baseline study, a MIDI sequencer (PERFORMER) had been used for both stimulus generation and playback, which explains why the increments had not been exact multiples of 10 ms: The durations had to be specified in "ticks", with 1 tick = 1.0417 ms. Note that MAX was not employed for stimulus generation, only for playback of complete MIDI files.

[5] It might be asked whether it could have been the atypicality of the precursor's expressive timing rather than its temporal modulation that affected detection performance. The author has run himself in a condition with typically timed precursors and obtained the same amount of interference relative to the no-precursor baseline as with atypically timed precursors. This is consistent with an explanation in terms of exposure to temporal irregularity.

Detection accuracy profile

The DAP is shown in Fig. 5, together with that of the baseline study.[6] It is clear from the graph that the DAPs were extremely similar in shape [$r(33) = 0.84$, $p < .001$]. The only substantial difference, apart from the difference in overall level of accuracy, was in the second half of bar 4, where the present DAP showed a sharp peak while the baseline DAP did not. There is reason to believe, however, that the baseline DAP was somewhat atypical in this respect; subsequent experiments with similar materials (Repp, in press) did show a peak there, which also makes more sense in terms of the musical structure. Moreover, the difference cannot be rationalized in terms of the precursor profile, which was similar to the typical profile in that region (see Fig. 3). The present DAP was significantly correlated with the typical expressive timing profile shown in Figure 1b [$r(32) = -0.82$, $p < .001$], and this perception-performance correlation is of the same magnitude as that in the baseline study. The DAP was not significantly related to the atypical timing profile instantiated in the precursor [$r(32) = -0.29$, n.s.]. Finally, the distribution of false-alarm responses over positions (not shown in Fig. 5) was also positively correlated with that in the baseline study [$r(33) = 0.59$, $p < .001$].[7]

Conclusion

It may be concluded from these data that the timing expectations underlying the DAP are resistant to modification through repeated exposure to a specific (atypical) performance. Almost certainly, then, they do not derive from experience with specific *typical* performances of the Chopin Etude in the past; anyway, most of the participants were not very familiar with the music. The principles underlying the timing expectations must be more general, be they tacit knowledge of the rules governing expressive music performance or psychoacoustic interactions or the timing implications of melodic-rhythmic grouping. These three alternatives correspond to the three explanations outlined above, and while a positive finding in Exp. 1 would have ruled out a psychoacoustic account, the negative finding obtained is less informative. Although it seems least consistent with the expectation hypothesis, several reasons could be imagined why the priming technique was not effective:

(a) Musically trained listeners' expectations may be far too ingrained to be changed in a brief experiment.
(b) The listeners may have rejected the precursor as a cognitive model because they recognized its atypicality.
(c) An atypical timing pattern may be difficult to remember (cf. Clarke, 1993).

Certainly, however, the results are consistent with explanations that assume a rather rigid connection between acoustic or musical structure and the perception of timing.

Experiment 2

In this experiment, the question was whether the processing of the musical sound structure is necessary to elicit timing expectations, or whether it would be sufficient just to imagine the music. If the expectations can be invoked through imagination, this would favor the expectation hypothesis, whereas a negative result would indicate that the "expectations" are a by-product of bottom-up perceptual processing and therefore should be placed between quotation marks.

Method

Materials and design. The design differed from that of the baseline experiment in the following ways:
(1) As in Exp. 1, only increment detection was tested.
(2) Each trial consisted not of music but of a series of 38 clicks, which were actually brief, high-pitched tones produced on the digital piano. They had a fundamental frequency of 4186 Hz (C_8) and a nominal duration of 20 ms.[8] The unmodified IOI duration was 500 ms; only the initial IOI, which was never a detection target, was twice as long, simulating the initial upbeat in the music. The increment sizes used were 50 ms in the five familiarization trials and 40, 30, and 20 ms in the three test blocks, similar to the values used in the baseline study.
(3) As in Exp. 1, each trial was preceded by a fixed precursor, but here it was the Chopin excerpt played with *typical* expressive timing. The timing profile represented the average IOIs of 11 expert performances that had loaded most highly on the first principal component in Repp's (1997) analysis.[9]
(4) The five familiarization trials were preceded by three presentations of the precursor alone, to familiarize the listeners with the music.

Listeners and procedure. The listeners were 12 musically trained Yale undergraduates from the same pool as in previous experiments. They were paid for their services. The instructions asked the participants to listen to the precursor music without responding and then to imagine it as vividly as possible in synchrony with the click sequence while trying to detect lengthened IOIs. Listeners

[6] Due to an error in stimulus construction, position 2-1 did not occur as a target in Exp. 1.

[7] False-alarm rates were subject to large individual differences, and the responses were distributed over 36 positions, which resulted in very sparse and unreliable data. For this reason, the methods of signal detection theory could not be applied in this paradigm. Nevertheless, false alarms presumably reflect the same perceptual biases that shape the DAP.

[8] As in all (simulated) piano tones, the energy did not stop abruptly but decayed following the nominal offset (see Repp, 1995b). This decay was very rapid, however, for such a high-pitched tone. There were also lower spectral components in the sound which simulated the thud of key-bed contact.

[9] This average profile was used in preference to the statistically extracted first-component profile (Fig. 3) or the average profile of student performances (Fig. 1b) because it received the highest average rating in an aesthetic judgment task that included performances with all three profiles (Repp, 1997). In any case, these three profiles were quite similar to each other.

marked their responses in the musical score, on the same answer sheets as previously. In order to enter responses correctly and rapidly, they had to read along while listening, and this was expected to aid musical imagery as well. Participants were discouraged from using techniques that might interfere with imagining the music, such as counting or tapping with the pen on successive notes in the score. They were made to believe that the study was about musical imagery.

Results and discussion

Overall accuracy

Of all the correct responses, 33.5% were "late". Thus, the strong tendency to attribute increments to the following IOI persisted even with click sequences. The overall percent correct scores are shown in Fig. 6, together with the baseline data. Detection scores were somewhat higher in the present experiment, even though the increments were slightly smaller, but the difference was not significant.[10]

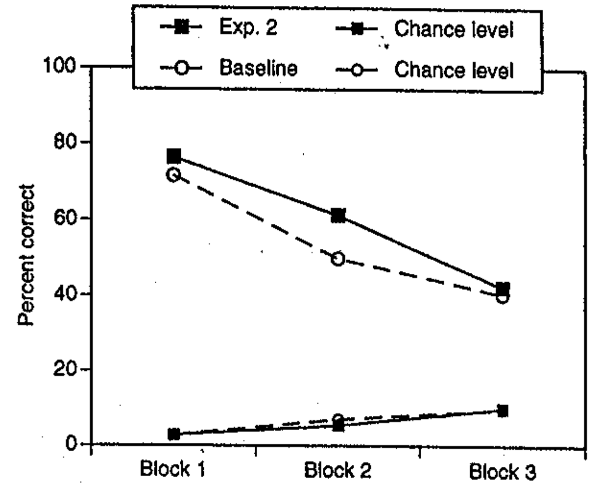

Fig. 6 Overall accuracy and chance level estimates in Exp. 2 and in the baseline study

Detection accuracy profile

The DAPs for the clicks and the music baseline are shown in Fig. 7. The DAP for the clicks seems rather irregular and bears little resemblance to the baseline DAP, except for low scores in the initial and final positions, which are easily explained by the availability of only one adjacent unmodified IOI for comparison. Without these two extreme data points, the correlation between the two DAPs is 0.2, which is nonsignificant. Likewise, there was no correlation between the click DAP and the typical timing profile instantiated in the music precursor [$r(32) = -0.07$, n.s.].

Although some random variability was to be expected with only 36 observations per data point, the click DAP was more variable than expected. In order to determine whether any of this variability was systematic, a one-way repeated-measures ANOVA was carried out, with the initial and final data points omitted. The F ratio was not large but significant, due to the large number of degrees of freedom [$F(33, 363) = 2.85$, $p < .001$]. It was still significant when the four initial data points were omitted, suggesting that there was systematic variability beyond a possible initial tuning-in to the periodicity of the sequence (Drake & Botte, 1993). What could have been the cause of this variability?

Many of the trials contained multiple targets. Although these targets were separated by at least four and usually more unmodified IOIs, it is possible that the phase shift caused by an early target had repercussions for later targets, in proportion to the distance between them. The oscillatory mechanism that keeps track of isochrony may need some time to adjust to the phase shift (Large & Kolen, 1994). To test this hypothesis, targets that were preceded by another target on the same trial were identified in each block, and their distance from the nearest preceding target was determined. Distances larger than 12 positions were excluded. For comparison, the present correct scores of targets not preceded by another target were examined as a function of their position in the music, up to the 12th position. (Targets preceded by another target were all beyond the 12th position.) The initial and final positions were excluded. The results of this exercise were suggestive, but not conclusive. The clearest pattern was obtained in Block 3, which also yielded the most varied DAP: Detection accuracy tended to increase both with the distance from a preceding target [$r(9) = 0.57, p < .07$] and, for targets not preceded by another target, with distance from onset [$r(9) = 0.60, p < .06$], though the

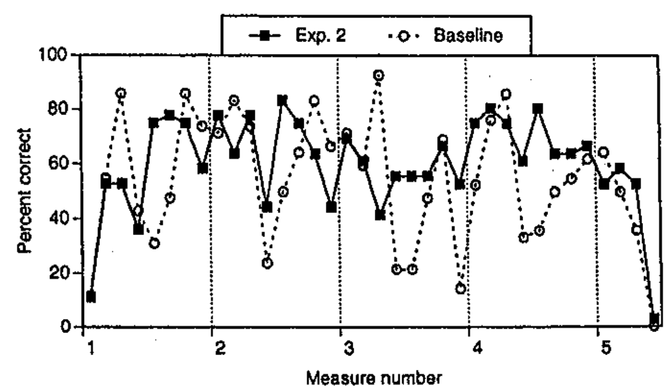

Fig. 7 Detection accuracy profiles for IOI increments in Exp. 2 and in the baseline study

---

[10] The reader may wonder whether the temporally modulated music precursor interfered with the detection of IOI increments, as it did in Exp. 1. Being equally curious, the author tested himself both in Exp. 2 and in a condition without precursors and indeed achieved higher detection scores in the latter. This result suggests that IOI increment detection is easier in a click sequence than in music and that the absence of a significant difference in Fig. 6 was due to the interference caused by the precursor in Exp. 2.

correlations fell short of significance. Moreover, the average accuracy for the latter targets was some 20% higher than for the former. This suggests that preceding targets did indeed interfere with increment detection. In Blocks 1 and 2, however, only small trends in the same direction could be discerned. Moreover, these trends were not sufficient to cause a correlation with the DAP of the baseline study, in which exactly the same assignment of targets to trials had been used.

Conclusion

The results show that merely imagining music does not produce perceptual bias effects comparable to those observed when music is heard. There is no reason to doubt that the listeners tried to imagine the music to the best of their ability. It is possible, however, that their auditory imaging abilities were limited; perhaps, only individuals with exceptional musical imagination would show perceptual bias effects. Nevertheless, for ordinary listeners with musical training, perception of the musical sound structure seems to be essential for the biases to emerge. This argues against the expectation hypothesis.

Experiment 3

It remains possible, however, that the clicks in Exp. 2 somehow neutralized listeners' timing expectations. Perhaps listening to a nonmusical sequence of sounds is incompatible with having expectations based on music, or the internal time-keeper that is used to track clicks is impervious to these expectations. In Exp. 3, the clicks were superimposed on the music, but the listeners were asked to direct their attention to the click sequence and ignore the music. Three outcomes were possible:
(1) If the results look like those of Exp. 2, they would suggest that the perception of click timing is indeed immune to music-based timing expectations or that such expectations are generated only when the music is attended to.
(2) If the results look like those of the baseline study, they would indicate that perceptual biases are an automatic consequence of hearing music, even if it is unattended.
(3) If the perceptual biases are present but weaker than usual, then an additional experiment may be necessary in which listeners are instructed to pay attention to the music, to see whether the presence of clicks attenuates timing expectations.

In order to focus the listeners' attention on the clicks as much as possible, the response sheets showing musical notation were eliminated, and listeners pressed instead a response key whenever they detected a deviation from isochrony. Since there were no precursors in this experiment, there was time to include both increment and decrement detection tasks in a single session, as in the baseline study.

Method

Materials and design. The materials differed from those of the baseline experiment only in that each note or chord onset in the music was accompanied by a click, so that any change in IOI duration occurred simultaneously in the click sequence and in the music.[11] The clicks were the same as in Exp. 2 and were incorporated into the MIDI instructions for the music trials. They had a fixed MIDI velocity of 70, whereas all but the five loudest melody notes in the music had velocities below 60 (see Repp, in press: Fig. 1). The clicks were perceptually prominent and clearly distinct from the music. The IOI increments and decrements were 50 ms during familiarization and 40, 30, and 20 ms in the three blocks, the same as in Exp. 2 and similar to the baseline study.

Listeners and procedure. The participants were 12 musically trained Yale undergraduates from the same pool as in the previous experiments. Only 1 of them had participated in a previous experiment in this series that required attention to the music. Half the listeners received the increment detection task first, and half the decrement detection task. A Macintosh Quadra 660AV computer running a MAX program controlled the sequence of trials and played the pre-assembled MIDI files on the digital piano. Individual participants listened on-line over earphones plugged into the piano. Whenever they heard a deviation from isochrony, they pressed the space bar of the computer keyboard. The response and its time from the onset of the trial were registered and saved by the computer. The subjects were asked to focus on the clicks and ignore the music, which was described as irrelevant to the task. They were also asked to respond quickly.

Results and discussion

Overall accuracy

A response was scored as correct if it fell within 1 s of the end of a modified IOI (i.e., within the two following unmodified IOIs).[12] Figure 8 shows the overall percent correct scores and estimated chance levels. The baseline data are not included in this figure, but they were quite similar (see Repp, in press: Fig. 2). The average accuracy in the two experiments was very nearly the same, which suggests that the listeners in Exp. 3 did not derive any benefit from the presence of the clicks. The difference between increment and decrement detection scores fell short of significance here [$F(1,11) = 3.94$, $p < .08$], but a joint ANOVA of the two experiments revealed a significant difference [$F(1,24) = 13.96$, $p < .001$] and no significant interaction with experiment.

[11] While it would have been methodologically desirable to have the changes occur only in the click sequence and not in the music, this was not feasible because asynchronies between the clicks and the music would have provided undesirable cues to timing changes.

[12] "Late" responses did seem to occur but could not be identified unambiguously in this experiment. The arbitrary 1-s criterion included most but not all of these responses.

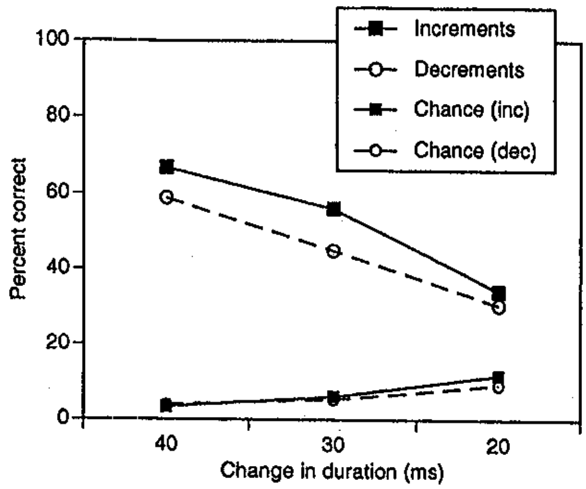

Fig. 8 Overall accuracy and chance level estimates in Exp. 3

Detection accuracy and bias profiles

The DAPs are shown in Fig. 9a. They show a mild negative correlation [$r(34) = -0.36$, $p < .05$]. The corresponding baseline data are shown in Fig. 2a. The present DAP for increments is highly similar to the baseline DAP for increments [$r(34) = 0.86$, $p < .001$], whereas it shows no relationship at all to the DAP for clicks in Exp. 2 [$r(32) = 0.09$, n.s., initial and final data points omitted]. The DAP for decrements is also similar to that in the baseline study, though less strikingly so [$r(34) = 0.55$, $p < .001$]. Repp (in press) found the DAPs for decrements to be generally less stable across experiments than those for increments.

The bias profile derived from the DAPs is shown in Fig. 9b. It, too, shows a close similarity to the corresponding profile in the baseline study (Fig. 2b) [$r(34) = 0.78$, $p < .001$]. Moreover, the bias profile is positively correlated with the typical timing profile shown in Fig. 1b [$r(33) = 0.77$, $p < .001$ final data point omitted].

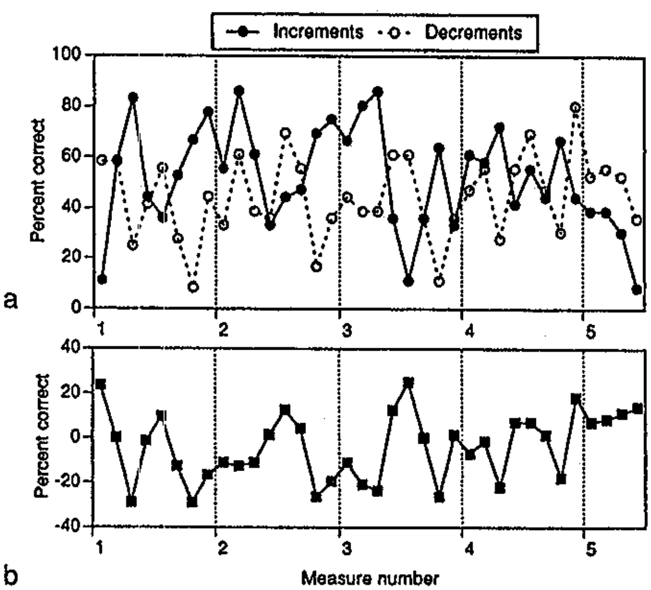

Fig. 9 (a) Detection accuracy profiles for IOI increments and decrements in Exp. 3. (b) The bias profile derived from these data

Conclusions

The results of this experiment are very similar to those of the baseline study, and they bear no similarity to the results for clicks alone (Exp. 2). Moreover, the perceptual biases were just as strong as in the baseline study. The absence of such biases in Exp. 2 thus was not due to the presence of clicks but rather to the ineffectiveness of auditory imagery. Evidently the biases are contingent on the auditory perception of a concrete musical structure, but they do not require focused attention.

It might be argued that the listeners did not follow the instructions and attended to the music rather than the clicks. This is difficult to rule out, but an intention to follow instructions should be granted to participants in an experiment. The results then demonstrate either that unattended music creates timing expectations just as attended music does and that these expectations affect the perception of click timing, or that clicks cannot be selectively attended to in the presence of music. Additional research would be necessary to disentangle these two possibilities, but in either case the perceptual biases seem to be an obligatory consequence of hearing music.

General discussion

The present experiments are part of a long and continuing series of studies on the nature and origin of timing expectations in music. Repp (in press) discussed three hypotheses – (1) the expectation, (2) the psychoacoustic, and (3) the structural – and addressed primarily the distinction between the second of these and the other two, arguing for (1) and (3), in part on epistemological grounds. The present experiments are relevant to the distinction between the first hypothesis and the other two. Although none of the three experiments is very conclusive by itself, taken together they make a fairly strong case against the expectation hypothesis. That is, it does not seem to be the case that listeners' timing expectations are cognitive, flexible, and applied to music or other stimuli in a top-down fashion. They are insensitive to experience with a model performance (Exp. 1), they do not emerge when music is imagined (Exp. 2), and they are back in full force when the music is (intended to be) unattended (Exp. 3). These findings are all consistent with the other two explanations, one based on the acoustic surface characteristics of the music and the other based on the processing of the musical structure, particularly the segmentation of the melody into rhythmic groups. The present results cannot distinguish between these two alternatives.

There are several important questions that have not been addressed in the present study but that are currently under investigation. One concerns the role of musical experience. Perhaps only musically trained individuals have such ingrained knowledge of the principles of expressive timing that its elicitation by musical stimuli cannot be avoided. Perhaps musically untrained listeners would prove to be more flexible. Most experiments in the present series have used musically trained listeners because they perform well in the detection task. Those experiments that included musically less experienced listeners have found only small differences in the DAP for increments (Repp, 1995a). Following Exp. 3, the author has recently designed an on-line version of the detection task in which the size of the temporal changes to be detected can be adapted to the listener's perceptual capabilities. This will make it possible to test a wider range of listeners in the future. It would be interesting to repeat the present Exp. 3 with musically untrained listeners, to see whether they are more successful in ignoring music and attending to clicks.

Another interesting question is whether the perceptual biases elicited by music interact directly with the time-keeping mechanism that enables a listener to detect deviations from isochrony. This issue is now under investigation. The present experiments do not bear on this topic, save for one small observation: The "late" response tendency was just as pronounced with clicks (Exp. 2) as with music (Exp. 1) and thus is not specific to music. Repp (in press) suggested that this tendency is at least partially perceptual in origin: The IOI following a lengthened IOI tends to be perceived as lengthened, also. This would be expected if the internal time-keeper does not immediately reset its phase when a phase shift is encountered in the isochronous rhythm (see Large & Jones, 1997; Large & Kolen, 1994; Mates, 1994a, b). Since the tendency is much reduced in decrement detection (Repp, in press), it may be that the internal time-keeper adjusts more rapidly to negative than positive phase shifts because its next prediction or "tick" is precipitated by an early auditory event onset.

In part, the "late" response tendency may be due to listeners' failure to backtrack on their answer sheets, since a change in an IOI obviously can be detected only during the following IOI. Experiment 3 did away with answer sheets, but on the basis of the response latencies alone it was impossible to distinguish between slow responses to the target IOI and fast responses to the following IOI. However, "late" responses did seem to be present, judging from the number of responses that occurred during the second and even third IOI following a target IOI. This would be consistent with a perceptual explanation.

Experiment 3 lays to rest another issue that was raised by a reviewer of Repp's (in press) manuscript. It was suggested that reading the score while hearing the music might create the perceptual biases. While this seemed unlikely to begin with, Exp. 3 now demonstrates that this is not the case.

It may seem surprising from a structural perspective that the precursors in Exp. 1 had no effect. The atypical timing pattern presented to the listeners was believed to illustrate an alternative interpretation of the grouping structure of the music, and since the listeners were not overly familiar with the music to begin with, it is not clear why they did not accept the cognitive parsing suggested by the precursor. One possibility is that the atypical timing pattern did not really represent a valid structural alternative, but merely a denial of the typical timing pattern – a performance strategy or mannerism. This issue needs to be pursued further with materials that exhibits true structural ambiguities. If it turns out that listeners' timing expectations are really unaffected by different structural interpretations, this would confirm the tentative conclusion reached here – that they are a cognitively impenetrable effect caused by the uninterpreted musical structure.

Acknowledgements This research was supported by NIH grant MH-51230. I am grateful to Lisa Robinson for extensive assistance, and to Henkjan Honing and Dirk-Jan Povel for helpful comments on the manuscript.

References

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

Clarke, E. F. (1993). Imitating and evaluating real and transformed musical performances. *Music Perception*, 10, 317–341.

Cook, N. (1990), *Music, imagination, and culture*. Oxford, UK: Clarendon Press.

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.

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

Large, E. W., & Jones, M. R. (1997). The dynamics of attending: How we track time-varying events. Manuscript submitted for publication.

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

Mates, J. (1994a). A model of synchronization of motor acts to a stimulus sequence: I. Timing and error corrections. *Biological Cybernetics*, *70*, 463–473.

Mates, J. (1994b). A model of synchronization of motor acts to a stimulus sequence: II. Stability analysis, error estimation, and simulations. *Biological Cybernetics*, *70*, 475–484.

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

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

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

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

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). Variations on a theme of Chopin: Relations between perception and production of timing in music. *Journal of Experimental Psychology: Human Perception and Performance.*

Todd, N. P. [McA.] (1985). A model of expressive timing in tonal music. *Music Perception, 3*, 33–58.

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