

Detecting deviations from metronomic timing in music: Effects of perceptual structure on the mental timekeeper

BRUNO H. REPP

Haskins Laboratories, New Haven, Connecticut

The detectability of a deviation from metronomic timing—of a small local increment in interonset interval (IOI) duration—in a musical excerpt is subject to positional biases, or “timing expectations,” that are closely related to the expressive timing (sequence of IOI durations) typically produced by musicians in performance (Repp, 1992b, 1998c, 1998d). Experiment 1 replicated this finding with some changes in procedure and showed that the perception–performance correlation is not the result of formal musical training or availability of a musical score. Experiments 2 and 3 used a synchronization task to examine the hypothesis that participants’ perceptual timing expectations are due to systematic modulations in the period of a mental timekeeper that also controls perceptual-motor coordination. Indeed, there was systematic variation in the asynchronies between taps and metronomically timed musical event onsets, and this variation was correlated both with the variations in IOI increment detectability (Experiment 1) and with the typical expressive timing pattern in performance. When the music contained local IOI increments (Experiment 2), they were almost perfectly compensated for on the next tap, regardless of their detectability in Experiment 1, which suggests a perceptual-motor feedback mechanism that is sensitive to subthreshold timing deviations. Overall, the results suggest that aspects of perceived musical structure influence the predictions of mental timekeeping mechanisms, thereby creating a subliminal warping of experienced time.

This research is concerned with the perception of music that is normally played with flexible timing. This style of execution is most appropriate for Western art music composed in the 19th century, a period that emphasized the artistic expression of emotion and individuality. Although the printed musical scores show note values that are in integral relationships, their actual durations in performance deviate substantially from integral ratios (see, e.g., Gabrielsson, 1987; Palmer, 1989; Repp, 1992a; Seashore, 1938). This is due to *expressive timing*—continuous modulations of the local tempo that are particularly pronounced in slow, “expressive” passages. These modulations make a performance interesting and engaging, and they also highlight aspects of the musical structure, particularly its division into temporal segments or groups. Group boundaries tend to be marked by a slowing of the tempo in their vicinity (see, e.g., Todd, 1985). Local slowing may also be used to prolong a salient event, such as a dissonance, and, occasionally, to increase the salience of a following note by delaying its onset (Clarke, 1985).

When 19th century piano music is played on an electronic instrument under computer control, so that its tim-

ing is metronomically exact, it sounds rather inexpressive and mechanical, as one might expect. If the music contains note onsets at nominally (i.e., notationally) equal intervals, then its tone interonset intervals (IOIs) are isochronous. In a series of detection experiments using such computer-controlled performances, Repp (1992b, 1995b, 1998c, 1998d) introduced small temporal deviations by lengthening or shortening one or more (well-separated) IOIs by a few percent, which could be perceived as momentary hesitations or accelerations in the isochronous rhythm. It was found that these changes are much more difficult to hear in some structural positions than in others. Although contextually conditioned variation in sensitivity to temporal change may have contributed to some of this variation in detectability, it was largely caused by perceptual biases that made IOI decrements easy to hear where IOI increments were hard to hear and vice versa (Repp, 1998d). It seemed as if listeners expected certain IOIs to be long and others to be short and compensated in perception when these expectations were not met. The IOIs that were expected to be long—and in which increments were much harder to detect than decrements—were precisely those that musicians tended to lengthen in expressive performances of the same music.

This perception–performance parallelism was quantified in terms of the correlation between two empirically determined profiles: a *detection accuracy profile* (DAP) representing the average percent-correct scores of a group

This research was supported by NIH Grant MH-51230. The author is grateful to Lisa Robinson for assistance and to Carolyn Drake and Richard Ivry for helpful comments on the manuscript. Correspondence should be addressed to B. H. Repp, Haskins Laboratories, 270 Crown St., New Haven, CT 06511-6695 (e-mail: repp@haskins.yale.edu).

of listeners for detection of equal-size IOI increments (or decrements) in all possible positions of a musical excerpt, and a *typical timing profile* representing the average IOI durations of a number of expressive performances of the same excerpt. The absolute perception–performance correlation was found to be higher for increment detection (a negative correlation) than for decrement detection (a positive correlation), and it was even higher for a derived *bias profile* representing the difference between increment and decrement DAPs (Repp, 1998d). However, the present study was concerned with increment detection only, since the increment DAP is quite similar to the bias profile and thus is representative of the perceptual biases of interest.

Where do these perceptual biases come from? And why is there a relationship between accuracy in a simple detection task and artistic music performance? One possibility (Repp, 1992b) is that musically trained listeners have developed cognitive expectations of expressive timing through performing on an instrument and listening to many expressive performances. These expectations would then be elicited by the musical structure and bias the perception of actual timing. According to this *cognitive expectation* hypothesis, musically untrained listeners should show much weaker biases than should musically trained listeners, and the biases of the latter also should be malleable to some extent. The findings to date, however, suggest that the perceptual biases are rather inflexible (Repp, 1998c) and independent of musical training (Repp, 1992b, 1995b). The results so far are more compatible with a *psychoacoustic* hypothesis (e.g., Crowder & Neath, 1994; Monahan & Hirsh, 1990) according to which the perceptual biases arise from local constellations of acoustic parameters (pitch, intensity, duration) that influence timing perception. The problems with this hypothesis are that known psychoacoustic phenomena do not seem to sufficiently explain the perceptual results (Repp, 1998d) and that the perception–performance correlation would have to be attributed to performers' compensation for auditory distortions in timing perception (Drake, 1993), which is at variance with the often large size of expressive timing variations.

Therefore, Repp (1998d) proposed that the perceptual biases arise from perception of (and memory for) musical structure. This *structural* hypothesis differs from the cognitive expectation hypothesis in that it assumes that some aspects of musical structure—melodic-rhythmic groups in particular—have an obligatory influence on the experience of time. Since grouping is a basic process of auditory scene analysis (Bregman, 1990) and not specific to music, no specific experience with expressive music performances is required. It is generally agreed that the timing of musicians' movements in expressive performance is closely tied to musical structure, especially rhythmic grouping (Clarke, 1985; Repp, 1992b; Todd, 1985), and this is particularly clear when typical (average) timing patterns are considered (Repp, 1997, 1998b). By assuming that perception is influenced (in part, at least) by

the same structural variables that guide musicians' expressive timing, the structural hypothesis provides a straightforward account of the observed perception–performance correlation.

Drake (1993) proposed a "perceptual hypothesis" that seems to contain features of both the psychoacoustic hypothesis and the structural hypothesis. This hypothesis has been clarified and elaborated in a recent study of timing in both expressive and deliberately mechanical piano performance (Penel & Drake, 1998). Penel and Drake distinguish between lower level, local grouping processes that lead to unintended timing variation even in mechanical playing and additional higher level grouping processes that are under cognitive control and are reflected only in intentional expressive timing. Presumably, only lower level, obligatory perceptual processes play a role when the task is to detect deviations from isochrony, and a close correlation between the perceptual biases and the typical timing profile of intentionally mechanical performances might be expected. To account for the high correlation between the bias profile and the typical expressive timing profile for the same music, it may then be assumed (1) that lower level, obligatory processes represent what is common to different performances, whereas higher level, optional processes are more diverse and serve to differentiate performances, and (2) that the obligatory timing effects are somehow amplified in expressive performance.

The purpose of the present study was to learn more about the origin of the perceptual "timing expectations" uncovered in previous research. Experiment 1 was conducted to demonstrate the perceptual effects to be accounted for and to reexamine effects of musical training with modified methods and a larger group of listeners than was employed in previous studies. Only one previous experiment (Repp, 1995b: Experiment 1) systematically compared listeners with different amounts of musical training in a task that required the detection of IOI increments in monophonic tunes. There was no significant effect of musical training on the shape of the DAP, although musically trained listeners were more accurate overall. In fact, a number of musically less sophisticated participants had to be excluded because they performed at chance level or did not even complete the task. It is possible that these individuals were precisely the ones that would have shown significantly different DAPs. This problem of attrition was avoided in Experiment 1 by adapting the task's level of difficulty to participants' capabilities. The experiment used a musical excerpt that was known from previous studies (Repp, 1998c, 1998d) to give rise to intricately shaped and highly correlated DAPs and timing profiles. Thus, it provided a more stringent test of effects of musical training than the previous study (Repp, 1995b, Experiment 1) and was expected to reveal the extent to which the perceptual biases induced by musical structure are obligatory or mediated by specific experience.

The results of Experiment 1 served as a launching pad for Experiments 2 and 3, which probed into the origin of

the perceptual biases, using the method of perceptual-motor synchronization. These experiments will be introduced after the results of Experiment 1 have been presented.

EXPERIMENT 1

Experiment 1 used the same musical excerpt as an earlier experiment (Repp, 1998d, Experiment 1) that had served as a baseline for several follow-up studies (Repp, 1998a, 1998c, 1998d). Whereas the earlier experiment had included both increment and decrement detection tasks, Experiment 1 included only an increment detection task but tested a much larger group of listeners representing a wider range of musical training. By means of a modified procedure that made it possible to adjust the level of difficulty, sufficient data were collected from each participant to be able to examine and compare individual DAPs.¹ The procedure also eliminated answer sheets with musical notation (used in all previous experiments in this series except for Repp, 1998c, Experiment 3), so that musical literacy was no longer a requirement for participation, and any effects of score reading on time perception (unlikely as they seem) could be ruled out.

Method

Participants. Forty-one individuals participated. All but 2 were paid volunteers recruited through advertisements on the Yale campus. Most of them were students ranging in age from 18 to 25 years, but 4 were Yale employees between 31 and 51 years of age. In addition, the author's daughter (age 16) and research assistant (age 32) participated without pay. All filled out a questionnaire about their musical training and activities.

Materials. The musical excerpt was the beginning of Chopin's *Etude in E major*, op. 10, No. 3. A computer-generated score is shown in Figure 1, with slurs and expression marks omitted. The final chord (not in the original music) gave maximal closure to the excerpt. The initial eighth-note upbeat is followed by continuous motion in sixteenth-note intervals—that is, there are note onsets at regular intervals, eight per bar. The nominal IOIs are indicated below the score. The music is divided into four voices: the soprano melody, which is segmented into rhythmic groups, each ending with a long note, as indicated above the score; the continuous alto accompaniment; the syncopated tenor voice; and the bass, which marks the harmonic alternation between tonic and dominant on the downbeats. The grouping structure of the melody, reinforced by the harmonic changes (which coincide with group-final notes except at the beginning of bar 4), is an important correlate of expressive timing in performance. The graph below the score shows a typical expressive timing profile representing the average IOI durations of performances by 18 graduate student or advanced amateur pianists, each of whom played the excerpt 3 or 10 times on a digital piano (combined data from Repp, 1997, 1998d, in press-b). The duration of the eighth-note upbeat is not included in the graph. It can be seen that all IOIs preceding and following group-final note onsets are lengthened, as is the group-initial IOI at the end of bar 3. The initial and final IOIs of the whole excerpt show extra lengthening.

The music was synthesized under computer (MIDI) control on a Roland RD-250s digital piano. The MIDI instructions were created in text format in a spreadsheet program and were executed by a program written in MAX. All tones were started and stopped in accordance with the nominal note values in the score. The resulting contiguity of consecutive tones (which included overlap due to damped decay; see Repp, 1995a) was sufficient to ensure perceived legato

articulation, so that use of the sustaining pedal (which is essential in human performance) was not necessary. The relative intensities (MIDI velocities) of the individual tones represented the average values of 27 performances by 9 graduate student or advanced amateur pianists (Repp, 1998d); they appropriately emphasized the melody over the accompanying voices and gave the melody a dynamic arch that roughly followed the pitch contour. All IOIs were initially set to 500 msec in duration, except for the initial upbeat, which was 1,000 msec in duration.²

This metronomically timed version served as the basis for the experimental stimuli containing timing deviations. A timing deviation was defined as an increment in the duration of a single IOI by a certain amount. It was implemented by lengthening the durations of all tones that were sounding during that IOI and delaying the onsets of all following tones, so that legato articulation was preserved. Each trial (i.e., presentation of the complete musical excerpt) contained exactly four such IOI increments, separated by at least four (usually more) unchanged IOIs. In the course of a block of nine trials, each of the 36 sixteenth-note IOIs was incremented once, according to a quasi-random schedule. (The eighth-note upbeat IOI was never incremented.) In the practice block, all increments were 100 msec in duration (20% of the baseline IOI). Eight test blocks were created, with increment sizes ranging from 90 msec (18%) to 20 msec (4%) in steps of 10 msec (2%). Additional versions of each test block were prepared by changing the order of the nine trials.

Procedure. The participants sat in front of a Macintosh Quadra 660AV computer, listened to the output of the digital piano over Sennheiser HD 540 II earphones, and pressed the space bar of the computer keyboard whenever they detected a hesitation in the temporal flow of the music. They were told that they had to respond within 1 sec for the response to count as correct. A MAX program controlled stimulus presentation, collected responses and response times, and provided feedback. Three counters were visible on the computer monitor: one showing the trial number, the second counting hits (maximum = 36 in a block), and the third counting false alarms. Any response that did not occur within a 900-msec interval starting 100 msec after the end of a lengthened IOI was considered a false alarm.³ In addition, a yellow light flashed up whenever a correct response was made. The participants were urged to make few false alarms. Successive trials were separated by 4 sec. One block of nine trials took about 4 min.

The practice block (100-msec increments) was presented first. It was followed by 10 test blocks. Using a rough staircase procedure, the increment size in each test block was chosen contingent on the participant's score on the preceding block, so as to converge on a performance level of about 50% correct by the third or fourth block. This performance level was intended to minimize floor and ceiling effects in the DAP. Increment sizes were always different in successive blocks, so that the detection scores in the remaining blocks oscillated around 50% correct. When a particular increment size was presented again, a different order of the trials was used. The participants were free to take a break after the fifth block, and, at the end of the experiment, they were given a detailed explanation.

Results and Discussion

Overall accuracy and musical training. The overall accuracy of each participant was determined by estimating the increment size associated with a hit rate of 50%, using linear regression on the data from Test Blocks 2–10.⁴ The hit rates were usually close to a linear function of increment size and yielded reasonable 50% estimates in all cases. As a measure of the participants' musical training, their combined years of training on all musical instruments studied (including voice) were calculated, which ranged from 0 to 28. Figure 2a shows the 50% detection

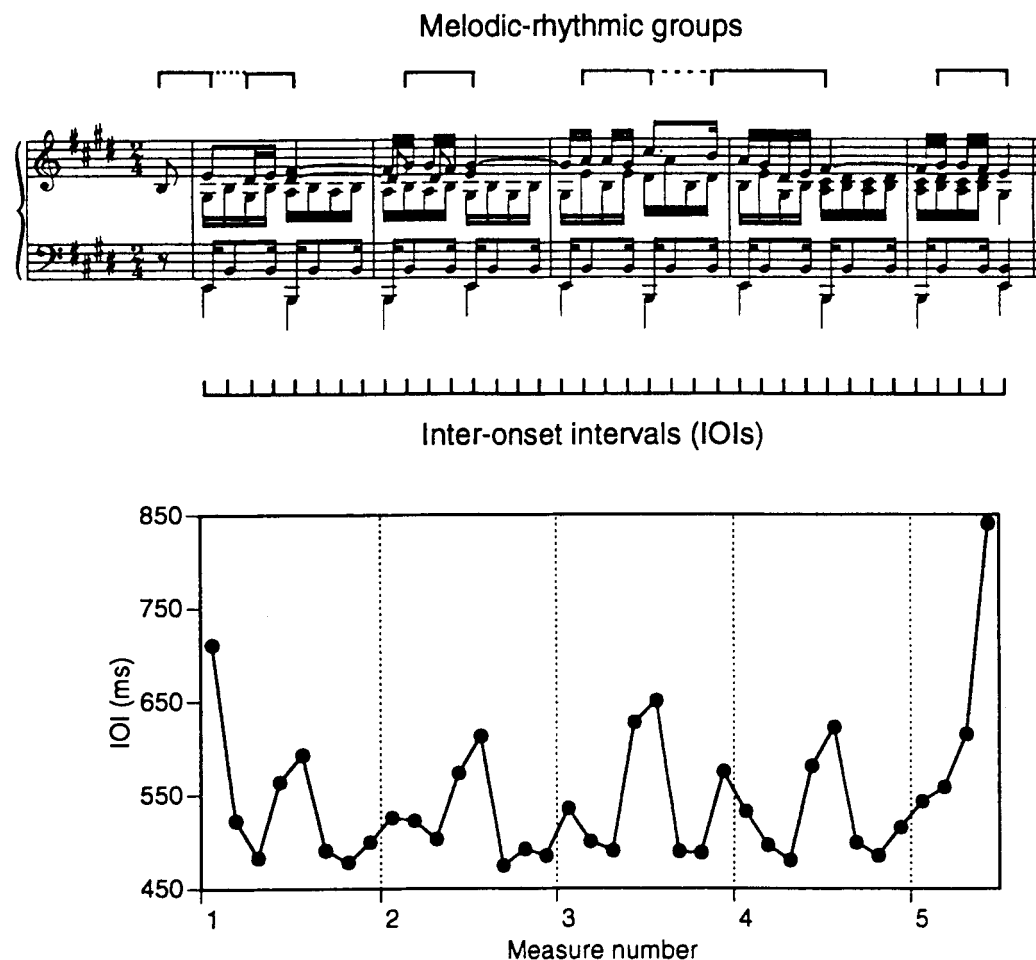


Figure 1. Top: A computer-generated score of the beginning of Chopin's Etude in E major, op. 10, No. 3 (final chord extended), with melodic-rhythmic groups and inter-onset intervals indicated. Bottom: A typical timing profile for this music.

thresholds as a function of years of musical training. The linear correlation between the two variables was $-.36$ ($p < .05$). With two exceptions, the participants with detection thresholds above 35 msec (7%) had little or no musical training, whereas the 5 participants with thresholds below 20 msec (4%) all had at least 10 years of training. Thus, there was some relationship between musical training and sensitivity to timing deviations, but not a very strong one.

The dotted vertical lines in Figure 2a show how, in later analyses, the participants were divided into four groups according to musical training: (I) no training at all ($n = 9$); (II) 1–5 years ($n = 15$); (III) 6–12 years ($n = 10$); and (IV) 16–28 years ($n = 7$). All of those in Group IV had studied more than one instrument.

A second measure of auditory sensitivity whose relationship to musical training could be examined was the slope of the function relating hit rates to increment size, representing the central linear portion of the (in theory) sigmoid psychometric function. The slopes were nega-

tively correlated with the 50% thresholds ($r = -.53$, $p < .001$): The lower the threshold, the steeper the slope. The individual slopes are shown as a function of years of musical training in Figure 2b. The significant correlation of $.47$ ($p < .01$) was largely due to the participants in Group IV, who showed steep slopes of about $.5$ (i.e., 5 additional hits for each 10-msec increase in increment size). Some of the participants in Groups I–III had similarly steep psychometric functions, but the majority had shallower slopes, without any clear relation to musical training in that range.

Since this was not a forced-choice detection task, false-alarm responses could not easily be combined with hits into a single measure of sensitivity such as d' . However, false-alarm rates could be examined as a function of the number of hits across blocks, which again approximated a linear relationship in most cases: False alarms increased as hits decreased. From this linear regression, the number of false alarms associated with a hit rate of 50% was estimated for each participant. The correlation of these

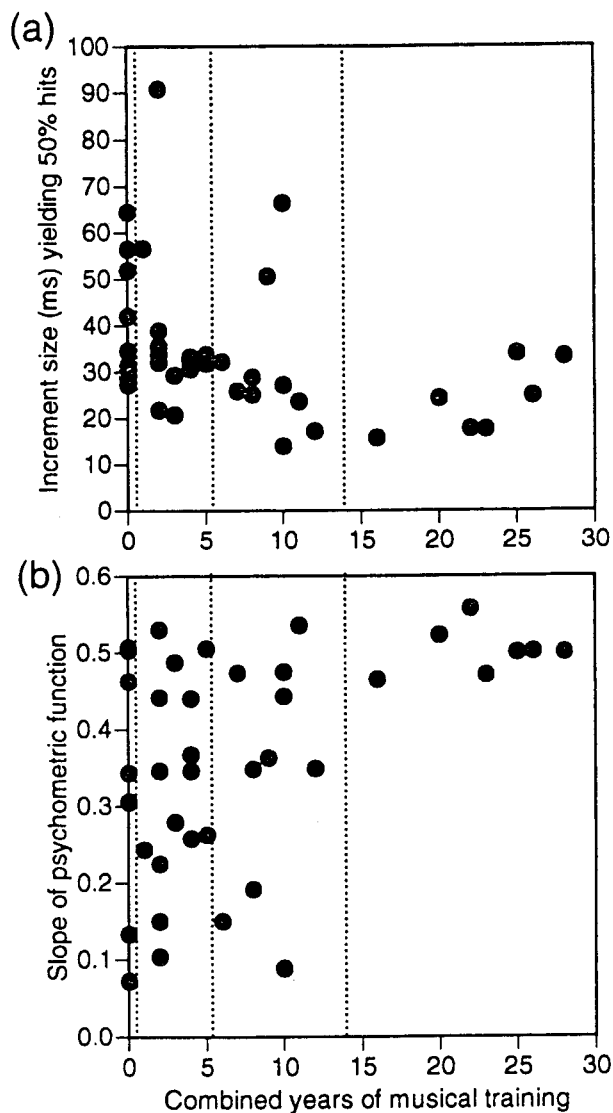


Figure 2. (a) The 50% (hits only) detection threshold as a function of musical training. (b) The slope of the psychometric function as a function of musical training.

estimates with musical training was nonsignificant ($r = .04$). Five participants with conspicuously high false-alarm rates (15–23 per block) were from Groups I–III; the rest showed rates between 3 and 13 per block or between 0.3 and 1.4 per trial. Within each trial, false alarms were distributed over the 36 possible positions, so that the probability of a false alarm being mistakenly accepted as a correct response was fairly low. (At the 50% correct level, it was about 7%.)⁵ There was no significant correlation between the 50% increment thresholds and the associated false-alarm rates ($r = -.20$). The slope of the function relating hits and false alarms varied widely and was clearly related to the false-alarm rate at the 50% threshold ($r = -.70$, $p < .001$): The steeper the (negative) slope, the more false alarms were made. However,

the slope, like the false-alarm rate, was unrelated to musical training ($r = .09$).

The DAP and musical training. For each participant, a DAP was obtained by computing the percentage of hits for each of the 36 positions in the musical excerpt across the 10 test blocks.⁶ A grand average DAP was calculated by averaging all the individual DAPs. Figure 3a shows this grand average DAP with local standard error bars; the musical score is reproduced above the figure for guidance. The DAP was quite similar ($r = .84$, $p < .0001$) to the average DAP for IOI increments obtained in a previous study with similar stimuli (Repp, 1998d, Experiment 1), in which 14 musically trained listeners (comparable to those in the present Groups III and IV) received only three test blocks and responded by circling notes in the musical score.⁷ This demonstrates, (1) that the shape of the DAP is not a consequence of reading a musical score while listening (a possibility already refuted by Repp, 1998c, Experiment 3), (2) that the present paradigm with four increments per trial is just as effective as (and far more economical than) the earlier design in which there were between 0 and 4 increments per trial, and (3) that the lesser musical training of many of the present participants cannot have played a very significant role.

Figure 3b shows the typical timing profile for the Chopin excerpt (Figure 1) in inverted form (i.e., as a local tempo profile, in terms of sixteenth notes per minute), which highlights its striking correlation with the DAP ($r = .90$, $p < .0001$): A small local hesitation in an otherwise metronomically timed performance is difficult to detect precisely where pianists tend to slacken their tempo. The correlation of the DAP with the typical timing profile (Figure 1) was $-.86$.

The DAP, like the typical timing or tempo profile, reflects the melodic-rhythmic grouping structure of the music. One interesting detail is that detection scores were always higher in position 5 than in position 4 of each bar, contrary to the tendency in local tempo. The IOI in position 4 preceded the final long note of a melodic group, whereas the IOI in position 5 was between melody and accompaniment. Also, position 4 was metrically weak, whereas position 5 was metrically strong. Thus, the IOIs in these two positions, although typically lengthened together, may be lengthened for different reasons: in position 4 because of grouping, but in position 5 because of voice alternation and/or metrical salience.⁸ If so, then grouping seems to be somewhat more important than the latter factors in timing perception.

To determine whether musical training affected the shapes of individual DAPs, a two-way analysis of variance (ANOVA) was conducted with the between-group factor of musical training (4 levels) and the within-group repeated-measures factor of position (36 levels). The two-way interaction was small but significant [$F(105, 1295) = 1.64$, $p < .0001$], due to the large number of degrees of freedom. This suggests subtle differences in profile shape as a function of musical training. The average DAPs of the four groups of participants are shown superim-

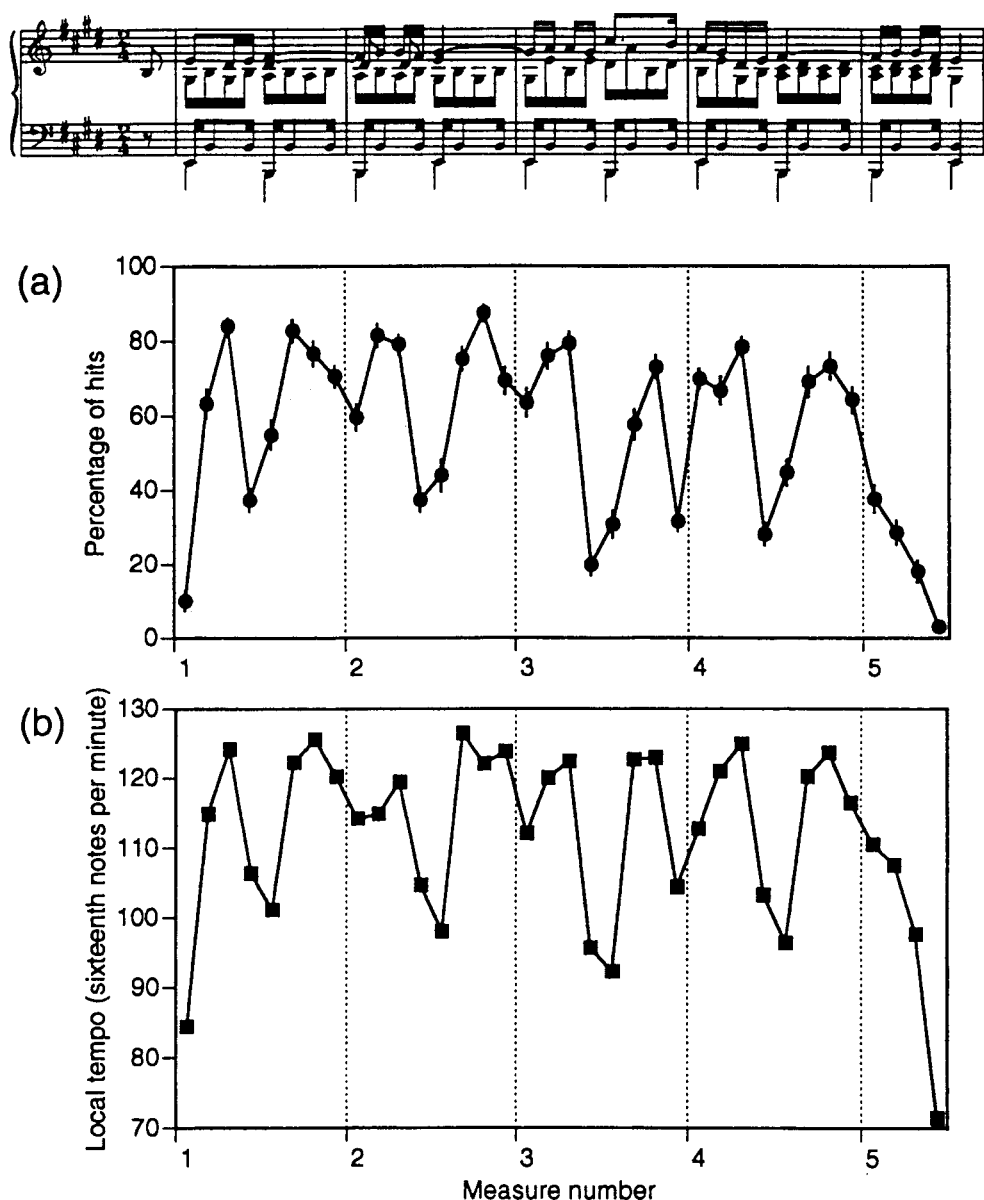


Figure 3. (a) The grand average DAP: Percentage of hits as a function of position in the music. Error bars represent ± 1 standard error. (The musical score is shown above for guidance.) (b) The typical average tempo profile: Local tempo (15,000/IOI) as a function of position in the music.

posed in Figure 4. It appears that the main difference was between Group I (highlighted in the figure) and the others. Group I had a somewhat less varied average DAP than did the other groups, which could have been due to larger individual differences or simply more noise in the data. The DAP of Group I also was less highly correlated with the typical local tempo profile in Figure 3b ($r = .79, p < .001$) than were the DAPs of the other groups (II, .86; III, .89; IV, .92), but the correlation was still quite high and not significantly lower than the other correlations.

In summary, the results of this experiment replicate previous findings of a close correlation between timing

perception and typical performance timing in the same musical excerpt, and they show that this result is not due to participants' degree of musical training, availability of a musical score, or other details of methodology. The perceptual results alone do not reveal whether the peaks and valleys in the DAP reflect variations in auditory sensitivity or in perceptual bias. Earlier experiments (Repp, 1998c, 1998d), however, have demonstrated that the DAP for IOI increments largely reflects variations in bias. While (nondirectional) variations in auditory sensitivity across positions in the music may exist, they are unrelated to the (inherently directional) timing pattern of music perfor-

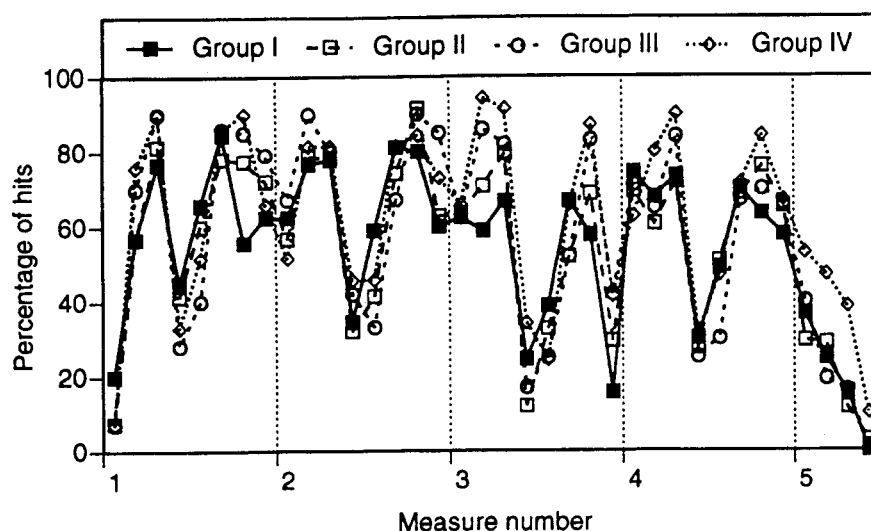


Figure 4. The average DAPs of four groups of participants with increasing musical sophistication.

mance (see Repp, 1998d) and, hence, of little interest in the present context. In fact, the high correlation between the DAP and the typical tempo profile leaves little room for such additional variation in nondirectional sensitivity. The variations in perceptual bias may be interpreted as timing expectations, specifically of expressive lengthening: In the context of metronomic timing, IOIs that are expected to be long may seem too short subjectively, so that an actual lengthening merely restores perception of even timing and therefore is difficult to detect. Although these expectations have a very strong effect on the detectability of IOI increments, the concomitant distortions of time perception must be quite small. This is suggested by the informal observation that a metronomically timed music performance does not sound subjectively irregular and by the corresponding finding that false-alarm responses are not particularly frequent in the detection task. Results of a recent series of detection experiments using temporally modulated baseline stimuli (Repp, 1998a) suggest that the distortions caused by timing expectations are only about one fifth the magnitude of the variations in the typical timing profile, which means that they are largely below the conscious detection threshold.

EXPERIMENT 2

Experiments 2 and 3 explored whether these timing expectations are reflected in the operation of mental timekeeping mechanisms. The rationale was as follows: In order to detect deviations from even timing, a listener must employ a mental timekeeper (i.e., some kind of neural oscillator or memory for interval duration) to track event onsets and predict the onset of the next event from the timing of the immediately preceding events (see, e.g., Desain, 1992; Jones & Boltz, 1989; Large & Kolen, 1994). A sufficiently large discrepancy between the predicted time of onset of an event and its actual time of onset

(or between the predicted event IOI and the actual IOI) will result in the conscious detection of a timing deviation. Timing expectations may be reflected in the predictions of the mental timekeeper: If an IOI is expected to be long, the event terminating it will be predicted to occur slightly later than is suggested by the timing of the immediately preceding context, and the opposite will occur if an IOI is expected to be short. In other words, the period of the timekeeper may be systematically modulated in accordance with the participants' expectations. There is an alternative possibility: The mental timekeeper may be autonomous and purely input-driven, in which case its period should vary randomly around a constant value when it is tracking an isochronous stimulus. In that case, timing expectations would have to have their effect at a subsequent but still relatively early stage in perceptual processing, since they govern whether or not a small timing deviation is consciously detected.

In Experiments 2 and 3, a perceptual-motor synchronization task was used as a window on the operation of the mental timekeeper. There are reasons for believing that the same central timekeeping mechanisms are employed in perception and in motor control, and especially in perceptual-motor coordination (see, e.g., Ivry & Hazeltine, 1995; Keele, Pokorny, Corcos, & Ivry, 1985; Treisman, Faulkner, & Naish, 1992). If so, and if perceptual biases have their effect by modulating the period of a central timekeeper, then synchronization errors and IOIs between successive taps should also reflect these systematic modulations. In other words, taps synchronized with metronomically timed music should be delayed at the ends of IOIs expected to be long, relative to the ends of IOIs expected to be short. Across the different positions in the music, there should thus be an "asynchrony profile" (systematic variation in asynchronies) and a "tap-timing profile" (systematic variation in tap IOIs), both of which should be negatively related to the

DAP of Experiment 1 (Figure 3a) and positively related to the typical performance timing profile (Figure 1). If perceptual biases have their effect at a later stage, then these systematic variations and relationships should be absent.

Experiment 2 used stimuli (taken from Experiment 1) that contained local IOI increments, in order to also investigate how rapidly participants would adjust their tapping to these timing perturbations and whether the speed of this adjustment (also referred to as a *phase correction* or *compensation*) would depend on the conscious detectability of the IOI increments, as assessed in Experiment 1. (Purely isochronous stimuli, without any IOI increments, were used in Experiment 3.) The hypothesis of systematic timekeeper period modulation predicts that the synchronization error (the difference between event onset and tap onset times, which is usually negative) will be smaller for IOI increments that are hard to detect (because these IOIs are expected to be long) than for IOI increments that are easy to detect. Therefore, there may also be a less effective phase correction in the former case than in the latter. However, if the timekeeper that controls perceptual-motor synchronization is autonomous and uninfluenced by perceptual biases, then the speed and accuracy of the phase correction should not depend on the conscious detectability of IOI increments as a function of position in the music. An underlying assumption here was that there is a threshold for perceptual-motor asynchrony, similar to the perceptual detection threshold for IOI increments, below which there is little or no effect on motor performance (Michon, 1967). To test this assumption, Experiment 2 also used two absolute magnitudes of IOI increments. The small IOI increments, which were difficult to detect in Experiment 1, were expected to be compensated for in tapping less effectively than large increments, regardless of whether the timekeeper period is modulated or not.

The role of musical training was of interest in this experiment as well. However, because of the laboriousness of the data analysis, a smaller group of participants was enlisted than in Experiment 1. It was expected that musically trained individuals would be less variable in their motor responses and perhaps more likely than untrained individuals to exhibit evidence of timekeeper period modulation in the synchronization task.

Method

Participants. Twelve individuals participated, 10 of whom had been participants in Experiment 1. They were divided roughly into three groups of 4 according to their musical experience. Group I included 2 undergraduate students (B.T., K.L.), a research assistant (L.R., age 32), and the author (B.R., age 51), all active amateur or semi-professional musicians with at least 10 years of musical training. (K.L. and B.R. had not participated in Experiment 1.) Group II included four undergraduates (H.K., A.L., H.T., A.K.) who had had a few years of musical instruction in childhood but no longer played an instrument, except for H.K. who occasionally played the saxophone. Group III included 2 undergraduates (C.B., J.Z.) and 2 Yale employees (N.F., age 36, and S.S., age 43) who had had no musical training at all, except for 6 months of clarinet instruction in C.B.'s case.⁹ All except L.R. and B.R. were paid for their participation.

H.T., A.K., N.F., C.B., and J.Z. participated in Experiment 2 immediately following Experiment 1; B.T., L.R., A.L., H.K., and S.S. participated in a separate session, a few days to a few weeks after Experiment 1.

Materials and Procedure. The materials were the Experiment 1 blocks with 40- and 20-msec increments. Each of these blocks came in three versions, in which the 9 trials occurred in different orders. Two additional versions, identical to the first two, were created for each increment size. The session started with Block 1 of the 40-msec condition, then Block 1 of the 20-msec condition, and so on, for a total of 10 blocks of 9 trials each. A MAX program played the musical stimuli on the Roland RD-250s digital piano, displayed the current trial number on the computer monitor, and registered the times of key depressions relative to trial onset. The participant sat in front of the Macintosh Quadra 660AV computer, listened to the output of the digital piano over Sennheiser HD 540 II earphones, and tapped with the right index finger on the "enter" key, located in the lower right-hand corner of the computer keyboard. N.F., the only left-handed participant, tapped with the left index finger on the "~" key, located in the upper left-hand corner of the keyboard.¹⁰ Three practice trials were presented first in which the music contained no IOI increments. The participants were instructed to consider the initial upbeat as a "ready" signal and to start tapping with the first downbeat (the second event onset) and then in synchrony with every sixteenth-note onset, at a rate of approximately 2 per second, for a total of 37 taps per trial. They were told to keep the finger resting on the key and to time their key depressions so that the key bottom contact coincided with the note onsets in the music. They were also informed that the music contained some slight hesitations, as in Experiment 1, but that they need not pay attention to them and should simply stay in synchrony with the music. There were short breaks between blocks and a longer break after the 5th block.

Results and Discussion

Average asynchrony profile, tap-timing profile, and variability. The registered key depression times were converted into asynchronies by subtracting the MIDI note onset times from them. Thus, a negative asynchrony means that the tap preceded the note onset, and a positive asynchrony means that the tap lagged behind. Altogether, there were 45 series of asynchronies (5 blocks of 9 trials each) for each participant in each increment size condition. Means and standard deviations were first computed across the 5 replications of each of the 9 trials in each condition. (Each trial had IOI increments in four particular positions.) From these 18 average data vectors, all responses to incremented IOIs were removed. A response was defined as including the tap to the delayed event itself and the two subsequent taps, assuming that the phase correction would be carried out within that time span (see Mates, 1994a; Michon, 1967). With these data triplets removed from the 18 vectors, there remained 12 asynchronies and associated standard deviations per position, representing responses to unperturbed IOIs only. Their averages were subsequently computed, which resulted in a grand average asynchrony profile and an associated variability profile based on $12 \times 5 = 60$ taps per position for each participant.

These grand average profiles are shown in Figure 5a. It can be seen that there was a pronounced tendency to be late with the first tap and then to catch up with the music within the next two or three taps. Therefore, the first three

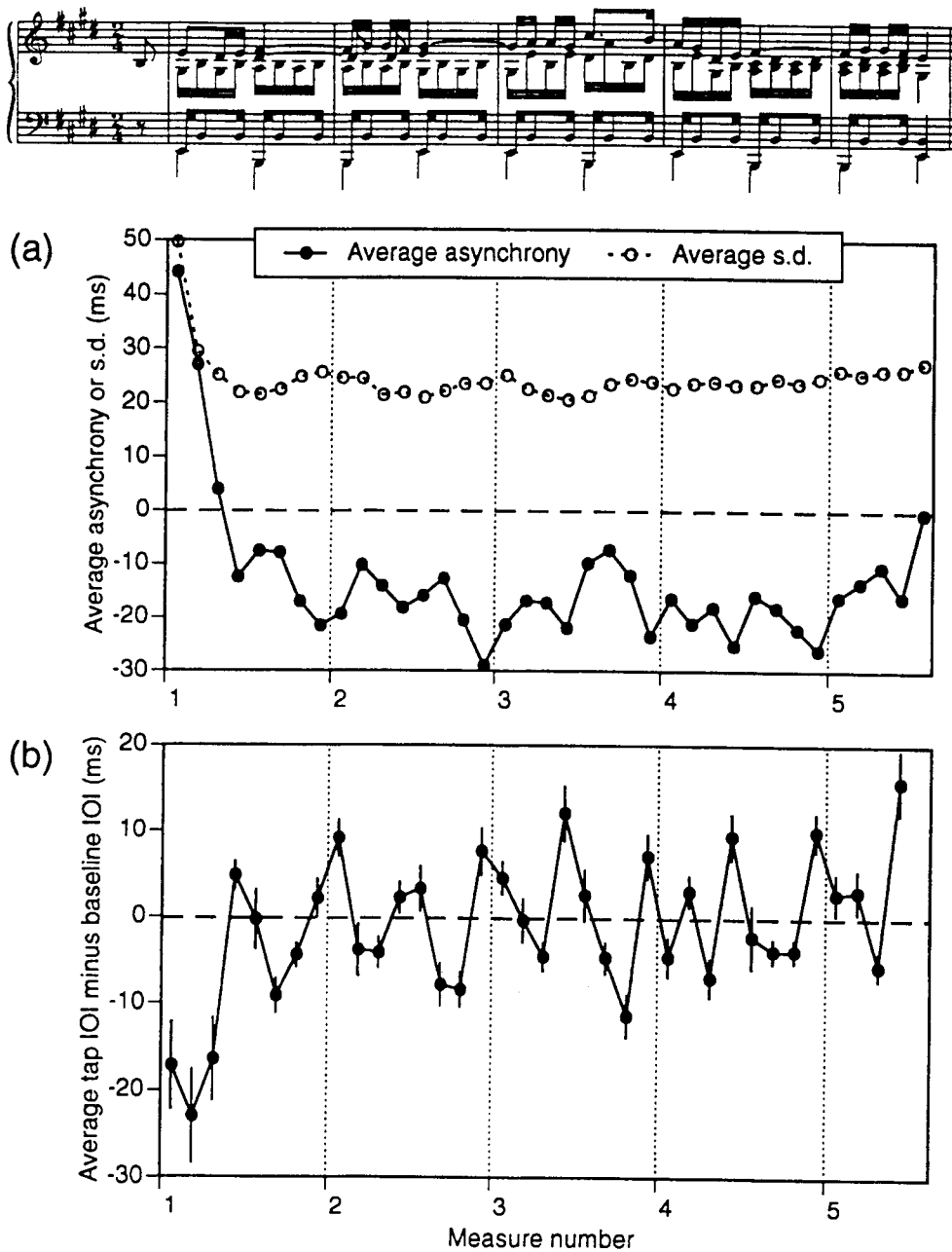


Figure 5. (a) The average asynchrony profile and the average variability profile. (b) The average tap-timing profile with error bars representing ± 1 standard error.

positions were omitted from all following analyses. From the fourth position on, the average asynchronies were negative, which is the anticipation effect usually found in synchronization tasks (see, e.g., Aschersleben & Prinz, 1995). The average within-trial standard deviation was elevated for the first tap but then settled down to a fairly constant value averaging about 24 msec.¹¹ The average asynchronies, however, exhibited systematic variation. This was confirmed by a significant effect of position in a one-way repeated measures ANOVA [$F(33,363) = 5.47$,

$MS_e = 81, p < .001$]. Moreover, as predicted by the time-keeper period modulation hypothesis, the average asynchrony profile was positively related to the typical timing profile (Figure 1) and negatively related to the average DAP of Experiment 1 (Figure 3a). With the asynchrony profile shifted by one position to the left (because the asynchrony with the note onset terminating an IOI was expected to reflect the participants' timing expectations), the correlation was .71 ($p < .001$) with the expressive timing profile and $-.68$ ($p < .001$) with the average

DAP. Thus, taps tended to occur relatively late at the ends of IOIs that tended to be lengthened in performance and in which actual lengthening was difficult to detect (i.e., positions 4 and 5 in each bar, which translates to positions 5 and 6 in the asynchrony profile). However, the relationship was not as striking as that between the DAP and the expressive timing profile.

An average "tap-timing profile" was obtained for each participant by computing the differences between their successive average asynchronies; this is equivalent to the IOIs between taps minus the constant stimulus IOI duration (nominally 500 msec). The grand average tap-timing profile is shown in Figure 5b with standard error bars, representing the variability among the 12 participants, computed separately for each position. The three initial, short tap IOIs were omitted in the following analyses. The nonoverlapping standard error bars suggest that there was significant variation in the tap IOIs, and this was confirmed in a one-way ANOVA [$F(32,352) = 7.96$, $MS_e = 68$, $p < .001$]. The tap-timing profile correlated .60 ($p < .001$) with the typical timing profile and $-.64$ ($p < .001$) with the DAP of Experiment 1. Thus, IOIs that tended to be lengthened in performance also tended to be lengthened in tapping, though again the relationship was not very strong.

The relatively small variations in average tap IOI durations (a range of only about ± 10 msec) and their only moderate correlations with the timing and detection accuracy profiles were almost certainly due to the requirement to stay in synchrony with the metronomically timed music. The nature of the task made it difficult to lengthen two successive IOIs, as was typical in performance. Thus, the IOIs in positions 4 and 5 were typically lengthened in performance (cf. Figure 1), but, in tapping, only position 4 exhibited such a tendency, presumably because the next IOI automatically compensated for the change in the synchronization error. To simulate this constraint, the difference between each IOI and the preceding IOI was computed in the performance timing profile (Figure 1), so that each lengthened IOI caused a shortening of the following IOI and durations oscillated around zero, as in Figure 5b. This IOI difference profile indeed correlated more strongly with the tap-timing profile ($r = .77$, $p < .001$).

Individual differences. Results for each participant, divided into three groups according to musical experience, are summarized in Table 1. The first column shows that there were large individual differences in the anticipation effect. In fact, anticipation was shown by only 5 of the 12 participants: 1 from Group I, 2 from Group II, and 2 from Group III. The others tended to be synchronous with or slightly behind the music, on average. The second column shows the average within-trial standard deviations of the asynchronies. Individual variability was negatively correlated with average asynchrony ($r = -.83$, $p < .001$): Those who tended to anticipate more also showed greater variability (B.R. being an exception). As in Experiment 1, musical experience tended to be associated with good performance in the task (i.e., small standard deviations), but some musically inexperienced participants did quite well.

The last three columns of Table 1 show the correlations between the individual tap-timing profiles and the local tempo profile (Figure 3b), the grand average DAP from Experiment 1 (Figure 3a), and each individual's own DAP (for the 10 participants who had participated in Experiment 1). These correlations were uniformly negative but generally smaller than the correlations between the grand average tap-timing profile and the other two grand average profiles. Moreover, there was no evidence that the participants' tap-timing profiles were more closely related to their own DAPs than to the grand average DAP. Thus, individual differences in Experiments 1 and 2 (which were considerable) seemed to be unrelated. Three participants, 1 in Group II and 2 in Group III, showed nonsignificant correlations. Again, there is a suggestion of a weak effect of musical experience.

To pursue possible effects of musical experience further, a two-way ANOVA with the factors group and position was conducted on the tap-timing profiles (with the three initial IOIs omitted). The group \times position interaction was nonsignificant [$F(64,288) = 1.0$], which indicates that the shape of the average tap-timing profile did not vary significantly as a function of musical experience.

Responses to IOI increments. Whenever a stimulus IOI was lengthened by either 40 or 20 msec, which amounts to a phase shift in the metronomical rhythm, the tap to the next musical event was necessarily out of synchrony by an additional 40 or 20 msec, on average. The questions of interest were how quickly the participants would get back into synchrony with the musical stimulus and whether the speed of this phase correction would depend on the magnitude of the increment and on its rel-

Table 1
Results For Individual Participants:
(A) Average Asynchrony (Without First Three Data Points),
(B) Average Within-Trial Standard Deviation
of Asynchronies (Without First Two Data Points),
(C) Correlation of Tap-Timing Profile With Average
Tempo Profile, (D) Correlation of Tap-Timing
Profile With Average DAP, and (E) Correlation
of Tap-Timing Profile With Individual DAP

Participant	A	B	C	D	E
Group I					
B.T.	8.9	18.4	-.53**	-.49**	-.37*
K.L.	9.0	18.8	-.57***	-.59***	n.a.
L.R.	-.8	21.8	-.37*	-.36*	-.35*
B.R.	-26.4	16.1	-.42*	-.31	n.a.
Group II					
H.K.	3.0	18.4	-.46**	-.50**	-.47**
A.L.	19.6	15.8	-.48**	-.50**	-.42*
H.T.	-105.9	37.6	-.24	-.32	-.29
A.K.	-32.0	35.7	-.71***	-.63***	-.59***
Group III					
C.B.	-0.2	22.7	-.35*	-.41*	-.31
J.Z.	-59.7	31.0	-.16	-.29	-.30
N.F.	16.8	20.2	-.22	-.29	-.13
S.S.	-29.1	26.9	-.39*	-.50**	-.32

* $p < .05$; ** $p < .01$; *** $p < .001$.

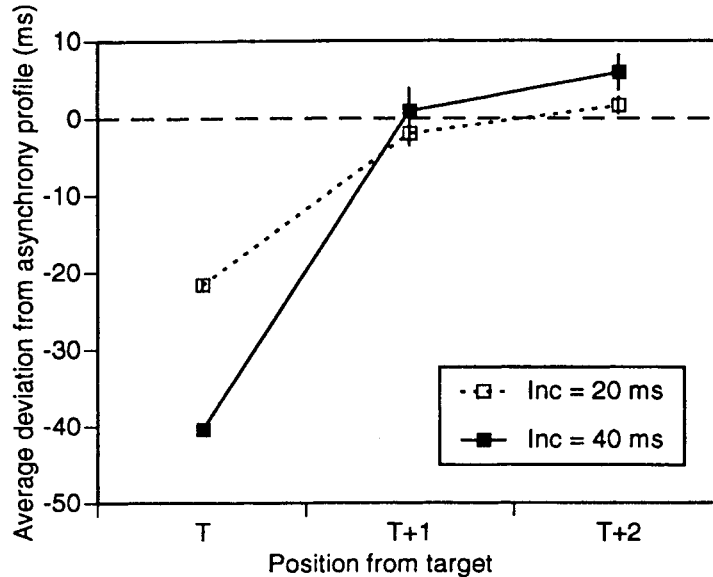


Figure 6. The average deviation from the asynchrony profile in target (T) positions, where IOI increments occurred, and in the two following positions (T+1, T+2), for two increment sizes (20 and 40 msec).

ative detectability in Experiment 1 as a function of position in the music.

To address these questions, the tap asynchronies for the events terminating incremented IOIs and for the two following events were analyzed; these were the response triplets that had been excluded from the preceding analyses. For each participant, there were 360 such triplets, 10 for each position in the music.¹² These triplets were first averaged across the 5 repetitions of each trial for each increment size. Then, each participant's average asynchronies for the corresponding unperturbed events (here computed separately for each increment condition) were subtracted from the phase correction asynchronies, so that the latter were expressed as *deviations from the individual asynchrony profile*. Subsequently, these relative asynchronies were averaged across the 36 positions in the music. Finally, they were averaged across the 12 participants, and the resulting data for the two increment sizes are shown in Figure 6, with standard error bars.

As expected, in the target (T) position (i.e., the position of a delayed musical event) taps occurred 40 msec earlier than usual in the 40-msec increment condition and 20 msec earlier in the 20-msec increment condition. (Since these are relative asynchronies, zero is the baseline.) Already in the following position (T+1), however, the taps were generally back on track. That is, phase shifts were compensated for immediately, regardless of magnitude. Closer inspection of the data revealed some individual differences: Three participants (B.R., L.R., and B.T., all in Group I) did not compensate completely in position T+1, and B.R. did not reach the baseline even in position T+2.¹³ All others compensated rapidly, but some of them tended to overshoot the baseline in the 40-msec increment con-

dition. This was the case in positions T+1 and T+2 for K.L. (Group I), H.K. and A.L. (Group II), and C.B. (Group III) and in position T+2 only for H.T. (Group II) and S.S. (Group III). Apart from this overshoot tendency, however, there seemed to be little difference in the speed of compensation for 40- and 20-msec increments, despite their very different detectability.¹⁴

To determine whether the speed of the phase correction depended on the detectability of the IOI increments as a function of position, the 36 positions in the music were divided into high- and low-detectability positions on the basis of the grand average DAP in Experiment 1. The initial two and final two positions were omitted. The 16 high-scoring positions had an average score of 76.4% correct in Experiment 1, whereas the 16 low-scoring positions had an average score of 44.2%.¹⁵ Figure 7a compares the average responses to high- and low-detectability increments, separately for each increment size. There was very little difference, which suggests that the phase correction did not depend on detectability.

However, detectability of IOI increments presumably depended on the perceived discrepancy between an event onset time and the onset time predicted by an internal timekeeper tracking the musical rhythm. The correlation between the grand average asynchrony profile and the DAP (reported above) implies that taps occurred later in low-detectability target positions than in high-detectability target positions. This is confirmed in Figure 7b, which shows the average *raw* asynchronies (i.e., without subtracting the asynchrony profile). It can be seen that, on average, taps were delayed by about 8 msec in low-detectability positions relative to high-detectability positions; to the extent that this delay reflects directly the

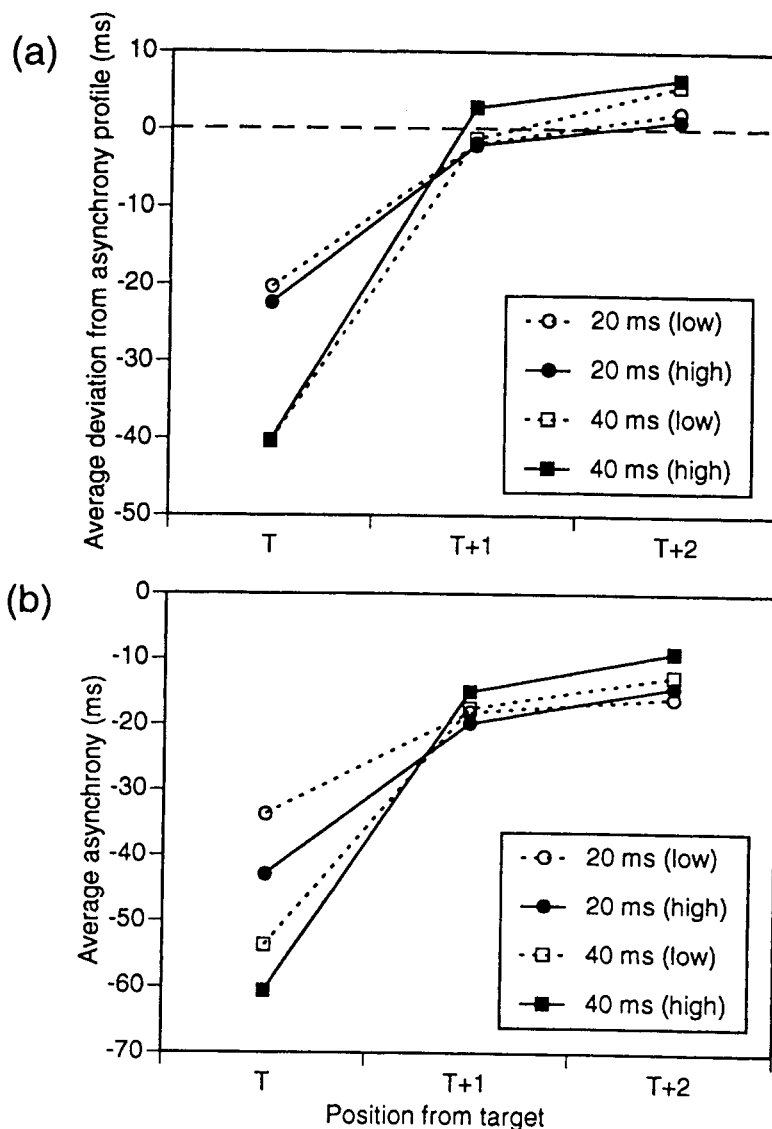


Figure 7. (a) Average deviations from the asynchrony profile in high- and low-detectability target (T) positions and the two following positions (T+1, T+2), for two increment sizes (20 and 40 msec). (b) Actual (raw) asynchronies in the same positions.

activity of a central timekeeper that is also employed in the detection task, it may explain the 32% difference in average detection scores between high- and low-scoring positions. This result is consistent with the timekeeper period modulation hypothesis. However, the rapid motor compensation for timing perturbations, regardless of their detectability, suggests that a different mechanism—perhaps a different timekeeper—is involved in automatic error correction.

EXPERIMENT 3

The main finding of Experiment 2—that synchronization of a motor response with a metronomically regular musical stimulus exhibits systematic deviations from

isochrony—is novel and intriguing. However, the very novelty of the result and the small size of the deviations raise concerns about possible artifacts. Experiment 3 was a partial replication of Experiment 2 with the purpose of ruling out three possible sources of such artifacts.

The first possibility, suggested by both reviewers of this article, is that the presence of IOI increments in the stimulus may somehow have affected the participants' responses to the isochronous portions of the music. Although it is not clear how the random placement of the increments could have resulted in a systematic pattern of asynchronies elsewhere, a similar asynchrony profile should certainly be obtained when there are no increments at all and the music is completely isochronous. Experiment 3 tested the validity of this prediction.

A second possibility, raised by one reviewer, is that the systematic deviations from isochrony in tapping may not reflect modulations in the period of a central timekeeper but rather may be consequences of variations in the force of keypresses. It is known that, when some taps in a sequence are deliberately accented, they occur earlier in time, so that the preceding IOI is shortened and the following one lengthened (Billon & Semjen, 1995; Billon, Semjen, & Stelmach, 1996; Keele, Ivry, & Pokorny, 1987; Semjen & Garcia-Colera, 1986). If the participants in Experiment 2 varied the force of their taps systematically, then the timing variation could have been an epiphenomenon of the force variation, so that no inferences about the flexibility of a central timekeeper are warranted. Small but significant correlations between tap force and tap timing (a negative correlation with the preceding IOI and a positive correlation with the following IOI) were found by Keele et al. (1987) in an unaccentuated tapping task. While the present experimental setup did not allow the measurement of the force or velocity of keypresses, it was easy to register key release times in addition to key depression times. The difference between the two, the "dwell time" or key contact duration, has been shown to be positively correlated with force in accented tapping (Billon & Semjen, 1995). Experiment 3 therefore examined whether there are systematic variations in dwell time and whether they are correlated with the asynchrony profile.

A third possible source of an artifact was discovered by the author: The stimuli used in Experiments 1 and 2, as well as in several previous studies, were found to contain some unintended temporal jitter due to MIDI transmission delays. Although these irregularities were rather small, they constituted a design flaw that needed to be corrected, to make sure that the almost equally small deviations in tap timing did not represent a response to subliminal timing irregularities in the stimulus. Experiment 3 therefore presented participants with corrected, effectively isochronous stimuli. The nature of the stimulus problem is discussed more thoroughly in the Appendix.

Method

Participants. Thirteen summer students were recruited through an advertisement on Yale campus and were paid for their participation. They ranged in age from 19 to 23 years, except for 1 who was only 14 years old, and they all had some musical education ranging from 3 to 21 years of instruction on all instruments combined.

Materials and Procedure. The Chopin Étude excerpt was presented in strictly isochronous form (see Appendix) and did not contain any lengthened IOIs. It was repeated 30 times, in three blocks of 10. Right-handed participants tapped on the "enter" or "=" keys of the computer keyboard; 1 left-handed participant tapped on the "~" key.

Results and Discussion

The average asynchrony profile is shown in Figure 8a (solid line), together with the profile obtained in Experiment 2 (dotted line). The first three positions were omitted in the following analyses. Although the average anticipation effect was slightly larger in Experiment 3 than in Experiment 2, the two profiles were significantly sim-

ilar ($r = .85, p < .001$). A two-way ANOVA showed a nonsignificant main effect of experiment [$F(1,23) = 0.36$] due to large individual differences in the size of the anticipation effect, a highly significant main effect of position [$F(33,759) = 10.32, MS_e = 78, p < .001$], and a nonsignificant interaction [$F(33,759) = 0.86$]. Thus, the results of the two experiments were statistically equivalent. Moreover, the Experiment 3 asynchrony profile correlated $.71 (p < .001)$ with the typical expressive timing profile and $-.72 (p < .001)$ with the average DAP of Experiment 1. These correlations are practically identical with those obtained in Experiment 2.

Figure 8b shows the tap-timing profiles of the two experiments. Again, the first three data points were omitted in the following analyses. These profiles, too, were highly similar ($r = .83, p < .001$). Although nonoverlapping standard error bars suggest significant differences in 7 of the 36 positions, the experiment \times position interaction in the two-way ANOVA was nonsignificant [$F(32,736) = 1.26, p < .16$], as was the main effect of experiment [$F(1,23) = 0.05$]. Only the main effect of position was highly significant [$F(32,736) = 13.04, MS_e = 71, p < .001$]. Thus, these derived profiles, too, were statistically equivalent. Moreover, the Experiment 3 profile correlated $.51 (p < .01)$ with the expressive timing profile and $-.57 (p < .001)$ with the average DAP of Experiment 1. These correlations were only slightly smaller than those obtained in Experiment 2. The findings of Experiment 2 thus were essentially replicated and shown to be due neither to the presence of IOI increments nor to the presence of unintended temporal jitter in the stimuli of Experiment 2.

The average dwell times in Experiment 3 did exhibit significant variation across positions. In particular, the dwell time of the final tap was clearly prolonged, dramatically so in some participants. The dwell time of the initial tap was also slightly longer than the rest. These two positions were omitted from all following analyses. The average dwell times for the remaining positions varied between 186 and 197 msec. This small variation was significant in a one-way ANOVA [$F(34,408) = 4.09, MS_e = 44, p < .001$]. The correlation between the average "dwell-time profile" and the average asynchrony profile, however, was close to zero ($r = .09$). The correlation between average dwell time and average tap IOI duration was likewise nonsignificant ($r = -.21$). There was a significant correlation, however, with the preceding IOI ($r = .47, p < .01$). This correlation is contrary to what would be expected if there were a positive correlation between tap force and dwell time, because tap force correlates negatively with the duration of the preceding IOI (Keele et al., 1987). However, a positive correlation between tap force and dwell time has been demonstrated only in deliberately accented tapping (Billon & Semjen, 1995), and it is possible that this correlation is reversed at the level of unintended variations in force and dwell time. Therefore, this analysis remains inconclusive, and direct measurements of tap force will have to be conducted to clarify the relation between tap force and tap timing.¹⁶

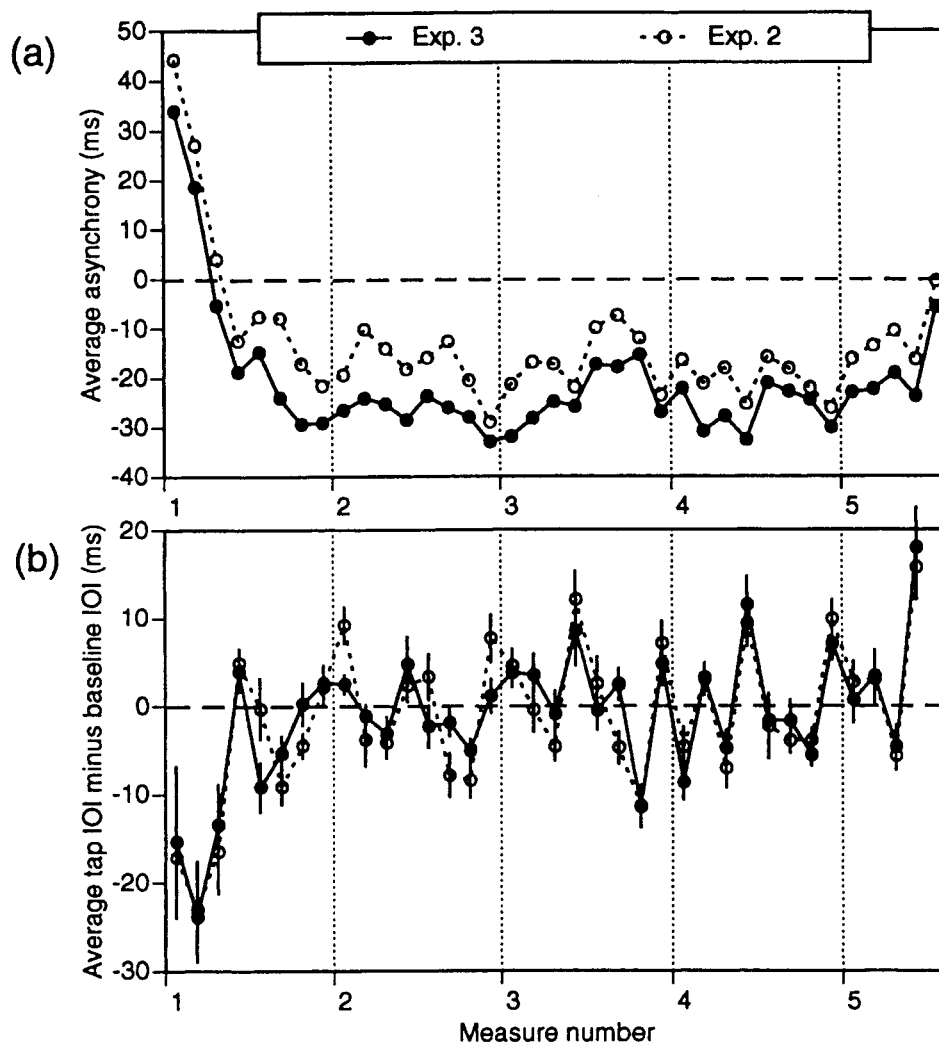


Figure 8. (a) The average asynchrony profiles of Experiments 2 and 3. (b) The average tapping profile of the two experiments, with error bars representing ± 1 standard error.

GENERAL DISCUSSION

Experiment 1 replicated a finding reported in several earlier studies (Repp, 1992b, 1995b, 1998c, 1998d)—namely, that local IOI increments in an otherwise metronomically timed musical excerpt are more difficult to detect in positions where lengthening is typically observed in performance. However, it went beyond previous experiments in that it provided much more reliable data that revealed a remarkably close parallelism between perception and performance. The changes in methodology (individual adjustment of difficulty level, four detection targets per trial, keypress response, immediate feedback) obviously did not reduce the perception–performance correlation. The possibility that the correlation in earlier experiments was due to participants’ reading of printed music on their answer sheets can be safely dismissed (see also Repp, 1998c, Experiment 3).

Experiment 1 also demonstrated that the perception–performance correlation is not due to musical training or specific experience with the performance of classical music. This had already been suggested by earlier results (Repp, 1992b, 1995b), but, again, the present results are more conclusive. To be sure, there was a small effect of musical training: The perception–performance correlation was slightly reduced for completely untrained participants, due to a somewhat less varied DAP, which may have been caused by larger individual differences among members of that group or perhaps just by more noise in their data. However, the correlation was still very high, which suggests that basic musical knowledge acquired from exposure to popular music provides a sufficient basis for the correlation or perhaps that even more basic perceptual skills unrelated to musical acculturation are involved.

As expected, musical training was related to overall accuracy: The participants who performed poorly in the

detection task generally had little musical training. Among the participants who performed well, however, the whole range of musical experience was represented. Thus, musical training is sufficient but not necessary for accurate perception of timing. This is not surprising for two plausible reasons: Musical training may enhance the ability to make fine temporal discriminations, and individuals who are genetically superior in this ability may be more likely to pursue musical activities. Time discrimination ability presumably is a direct reflection of the accuracy of mental time-keeping processes.

The perception–performance correlation demonstrated in Experiment 1 is attributed to an influence of musical structure on both perception and action. Musicians generally pace their performances in a way that is responsive to the structural properties of the music (Clarke, 1985; Todd, 1985). They have considerable freedom in doing so and may choose to ignore some structural features while emphasizing others or may impose alternative cognitive interpretations on the same score. The *average* timing pattern of a group of pianists, however, represents the *timing potentialities* of a given musical structure—the average magnitudes and statistical frequencies of local tempo modulations in a large sample of performances. Similarly, the DAPs of individual participants in the detection task (which also exhibit considerable individual differences) converge onto an average DAP that represents the average perceptual response to the musical structure. These average profiles show a high correlation.

While individual differences in expressive timing largely reflect artists' intentions and interpretations, individual differences in DAPs probably originate at a lower level of perceptual processing. Nevertheless, both average functions may be seen as reflecting the uninterpreted musical structure—that is, the perceptual structure of the sound pattern (Repp, 1998d; see also Penel & Drake, 1998). The division of the sound stream into several perceptually coherent voices is likely to be accomplished by a basic perceptual process akin to auditory scene analysis (Bregman, 1990; Huron, 1989; Scheirer, 1996). A similarly basic process probably recognizes that the soprano melody is segmented into groups terminated by long notes. The approach of a long group-final note and the following shift of onset events from the soprano to the alto voice both seem to have strong timing implications, calling for lengthening in performance and an expectation of lengthening in perception. The small melodic-rhythmic groups are nested in a larger group representing the whole excerpt, which calls for additional lengthening at the beginning and end. Grouping structure accounts for all the major features of the profiles, though additional contributions from harmonic and metrical structure should not be ruled out. Groups are responsible for the gestural quality of music, and a gesture is a movement with a beginning and an end, associated with a characteristic velocity profile that first accelerates and then decelerates (see Todd, 1985). Pianists execute gestures with their hands on the keyboard, whereas perceivers seem to resonate to the gestural implications

of music by warping their subjective perception of time accordingly. This resonance may be enhanced by increasing familiarity with the musical structure: All of the results presented here are from tasks in which the same musical excerpt was repeated many times.

In Experiments 2 and 3, a simple motor synchronization task was used to engage the time-keeping processes that listeners use to track temporal regularity, on the assumption that the same central mechanisms are involved in time perception and in timing of a motor response. The participants' motor responses indeed exhibited timing modulations that were significantly related to the typical expressive timing pattern and to the shape of the average DAP of Experiment 1. This suggests that the predictions of the central timekeeper were affected by the musical structure: Where the structure called for lengthening, the timekeeper's prediction of the next event onset tended to be slightly delayed. This correlation was only of moderate size, which was attributed to the constraints of the synchronization task. It is also possible, however, that structural factors are weighted somewhat differently in the tapping task; for example, metrical structure may play a greater role than in perception and expressive performance. This may account for the presence of additional peaks in the tap-timing profile.

In principle, it could be that the timing modulations did not occur in the timekeeper itself but rather in the delays of the peripheral motor responses (cf. Wing & Kristofferson, 1973). One peripheral explanation, that the timing variation was contingent on unintended variation in the force of the taps, seems unlikely (see note 16). In general, it is unclear how the processing or memory of musical structure could interact directly with peripheral responses, without affecting central control mechanisms first. Such an account would also make it difficult to explain the positional variations in detectability of IOI increments. It would have to be argued that listeners execute covert motor responses and monitor the timing of these responses in relation to physical event onsets. Such a strong motor theory of music perception seems both unwarranted and unparsimonious. It seems more plausible that the grouping structure of music subtly alters the subjective flow of time, comparable to the way in which the curvature of the trajectory of a moving cursor affects its perceived velocity (Viviani & Stucchi, 1992).

A motor theory of rhythm perception was proposed long ago by Stetson (1905). Although he favored explanations in terms of actual changes in the state of participants' musculature, he noted that "such a theory could be easily translated into terms of 'central motor discharges'—if their existence should be demonstrated, and rhythms of that type detected" (Stetson, 1905, p. 258). The "central timekeeper" metaphor of modern theoretical discourse may be equated with the "central motor discharge" envisioned by Stetson—a process that controls the timing of appropriate movements, whether or not they actually occur. In the absence of discrete movements, the process is experienced as a continuous "flow" of conscious time, although underlying it may be a pacemaker issuing discrete pulses at a high rate (Treisman et al., 1992).

The participants in Experiment 2 rapidly adjusted their finger taps whenever an IOI increment occurred, so that they were usually back "on track" with the subsequent tap. Such rapid adjustment to timing perturbations has been observed in previous studies (Michon, 1967; Vos & Helsen, 1992), though they involved an abrupt change in tempo (period) rather than in phase. A surprising finding, however, was that IOI increments that were difficult to detect (either because they were small in size or because they occurred in positions where lengthening was expected) were compensated for just as rapidly as increments that were easy to detect. This suggests that the feedback mechanism that maintains perceptual-motor synchronization is sensitive to discrepancies well below the conscious detection threshold. This is consistent with a synchronization study by Hary and Moore (1987) in which participants were found to track sub-threshold random timing perturbations (see also the Appendix). The hypothesis that the error correction process is independent of the overt detectability of timing perturbations has been investigated further (Repp, in press-a).

Although the negative synchronization error or anticipation effect was not of major interest here, the substantial individual differences found are noteworthy (cf. Mates, Radil, & Pöppel, 1992), as well as the fact that the participants with very large negative asynchronies seemed to compensate for timing perturbations just as effectively as those with small average synchronization errors. Clearly, each participant monitored a subjective criterion of synchrony, but these criteria varied among participants. A number of participants did not show any anticipation effect at all, which may be related to the fact that piano tones have an initial amplitude rise and therefore a perceptual onset (or "P-center") that lags behind their physical onset. According to Vos, Mates, and Kruysbergen (1995), the P-center serves as the basis for synchronization of motor responses. Recent findings (Repp, in press-a), however, cast doubt on this explanation and suggest that properties of the response key may be responsible for the small average anticipation effect.

Finally, it may be added that the biases exhibited in the present detection and synchronization tasks are related not only to the typical timing profile (i.e., intended expressive modulations) of pianists' performances but also to unintended variations in the timing of pianists who try to play the music metronomically. That pianists cannot completely get rid of their habitual timing modulations has been observed repeatedly (Palmer, 1989; Penel & Drake, 1998; Seashore, 1938), but the similarity of expressive and metronomic timing was less striking in these studies than in data obtained recently with the present Chopin Étude excerpt (Repp, in press-b). The conclusion may be anticipated that pianists themselves are subject to automatic influences of musical structure on their central timekeeper. The typical expressive timing pattern obtained from a group of pianists may be thought of as an inten-

tional magnification of the timing pattern induced by the musical structure. In this manner, structural perception may give rise to a norm for the structural cognition that underlies artistic performance.

REFERENCES

- ASCHERSLEBEN, G., & PRINZ, W. (1995). Synchronizing actions with events: The role of sensory information. *Perception & Psychophysics*, 57, 305-317.
- BILLON, M., & SEMJEN, A. (1995). The timing effects of accent production in synchronization and continuation tasks performed by musicians and nonmusicians. *Psychological Research*, 58, 206-217.
- BILLON, M., SEMJEN, A., & STELMACH, G. E. (1996). The timing effects of accent production in periodic finger-tapping sequences. *Journal of Motor Behavior*, 28, 198-210.
- BREGMAN, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- CLARKE, E. F. (1985). Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 209-236). London: Academic Press.
- CROWDER, R. G., & NEATH, I. (1994). The influence of pitch on time perception in short melodies. *Music Perception*, 12, 379-386.
- 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.
- GABRIELSSON, A. (1987). Once again: The theme from Mozart's Piano Sonata in A major (K.331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81-103). Stockholm: Royal Swedish Academy of Music.
- HARY, D., & MOORE, G. P. (1987). Synchronizing human movement with an external clock source. *Biological Cybernetics*, 56, 305-311.
- HURON, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception*, 6, 361-382.
- 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 & Performance*, 21, 3-18.
- JONES, M. R., & BOLTZ, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96, 459-491.
- KEELE, S. W., IVRY, R. B., & POKORNY, R. A. (1987). Force control and its relation to timing. *Journal of Motor Behavior*, 19, 96-114.
- KEELE, S. W., POKORNY, R. A., CORCOS, D. M., & IVRY, R. (1985). Do perception and motor production share common timing mechanisms? A correlational analysis. *Acta Psychologica*, 60, 173-191.
- LARGE, E. W., & KOLEN, J. F. (1994). Resonance and the perception of musical meter. *Connection Science*, 6, 177-208.
- MARANS, M. (1991, December). Timing is everything. *Keyboard*, pp. 95-103.
- 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.
- MATES, J., RADIL, T., & PÖPPEL, E. (1992). Cooperative tapping: Time control under different feedback conditions. *Perception & Psychophysics*, 52, 691-704.
- MICHON, J. A. (1967). *Timing in temporal tracking*. Assen, the Netherlands: van Gorcum.
- MONAHAN, C. B., & HIRSH, I. J. (1990). Studies in auditory timing: 2. Rhythm patterns. *Perception & Psychophysics*, 47, 227-242.
- PALMER, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception & Performance*, 15, 331-346.
- PENEL, A., & DRAKE, C. (1998). Sources of timing variations in music

- performance: A psychological segmentation model. *Psychological Research*, **61**, 12-32.
- REPP, B. H. (1992a). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America*, **92**, 2546-2568.
- REPP, B. H. (1992b). 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 partial 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. (1998a). The detectability of local deviations from a typical expressive timing pattern. *Music Perception*, **15**, 265-289.
- REPP, B. H. (1998b). A microcosm of musical expression: I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major. *Journal of the Acoustical Society of America*, **104**, 1085-1100.
- REPP, B. H. (1998c). Obligatory "expectations" of expressive timing induced by perception of musical structure. *Psychological Research*, **61**, 33-43.
- REPP, B. H. (1998d). Variations on a theme by Chopin: Relations between perception and production of timing in music. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 791-811.
- REPP, B. H. (in press-a). Compensation for subliminal timing perturbations in perceptual-motor synchronization. *Psychological Research*.
- REPP, B. H. (in press-b). Control of expressive and metronomic timing in pianists. *Journal of Motor Behavior*.
- ROADS, C., ET AL. (1996). *The computer music tutorial*. Cambridge, MA: MIT Press.
- SCHERER, E. D. (1996). Bregman's chimerae: Music perception as auditory scene analysis. In B. Pennycook & E. Costa-Gioni (Eds.), *Proceedings of the Fourth International Conference on Music Perception and Cognition* (pp. 317-322). Montreal: McGill University.
- SEASHORE, C. E. (1938). *Psychology of music*. New York: McGraw-Hill.
- SEMEN, A., & GARCIA-COLERA, A. (1986). Planning and timing of finger-tapping sequences with a stressed element. *Journal of Motor Behavior*, **18**, 287-322.
- STETSON, R. H. (1905). A motor theory of rhythm and discrete succession. *Psychological Review*, **12**, 250-270 (Part I), 293-350 (Part II).
- TODD, N. P. (1985). A model of expressive timing in tonal music. *Music Perception*, **3**, 33-58.
- TREISMAN, M., FAULKNER, A., & NAISH, P. L. N. (1992). On the relation between time perception and the timing of motor action: Evidence for a temporal oscillator controlling the timing of movement. *Quarterly Journal of Experimental Psychology*, **45A**, 235-263.
- VIVIANI, P., & STUCCHI, N. (1992). Biological movements look uniform: Evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 603-623.
- VOS, P. G., & HELSPER, E. L. (1992). Tracking simple rhythms: On-beat versus off-beat performance. In F. Macar, V. Pouthas, & W. J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 287-299). Dordrecht: Kluwer.
- VOS, P. G., MATES, J., & KRUYSSBERGEN, N. W. VAN (1995). The perceptual centre of a stimulus as the cue for synchronization to a metronome: Evidence from asynchronies. *Quarterly Journal of Experimental Psychology*, **48A**, 1024-1040.
- WING, A. M., & KRISTOFFERSON, A. B. (1973). Response delays and the timing of discrete motor responses. *Perception & Psychophysics*, **14**, 5-12.
2. Measurements of the output of the digital piano revealed that it was uniformly accelerated by 2.4% due to a peculiarity of the MAX software. The actual IOI duration thus was only 488 msec, and the IOI increments similarly were 2.4% shorter than stated. For additional qualifications regarding the stimuli, see the Appendix.
3. In earlier detection studies of the same type, where listeners responded by circling notes in a musical score, a strong tendency to attribute increments to the following IOI was observed (see Repp, 1998d). The 1-sec limit was intended to capture most of these "late" responses, to the extent that they occurred at all in the absence of a musical score. The average response latency was about 500 msec. Since this was not a speeded task, the latencies were generally too variable to yield reliable information.
4. The scores from Test Block I (i.e., the highest scores) were omitted because they tended to impair the linear fits. They generally fell below the regression line, either because they represented the asymptotic part of the psychometric function or because the participants did relatively poorly because they were still getting used to the task.
5. With an average of two hits per trial (50%), there were 34 "free" positions in a trial, 2 of which contained undetected targets. Since the two positions immediately following a hit are not likely candidates for a false-alarm response, a reasonable estimate of the accidental hit rate is $2/30 = 7\%$.
6. For this analysis, the criterion for hits was tightened by changing the lower cutoff from 100 to 200 msec, which seemed a more reasonable value. Since there were only a few responses with latencies between 100 and 200 msec (probably false alarms to the IOI preceding a target), it did not seem necessary to redo the earlier analyses. False-alarm responses were not scored according to position, in part because they were ambiguous: They could represent either a fast response to the immediately preceding IOI or a slow response to an earlier IOI. In any case, they were so sparsely distributed across the 36 positions that they would have contributed little to a positional measure of accuracy (see Repp, 1998d, for a discussion of average false-alarm profiles in similar experiments).
7. One major difference between the two DAPs occurs toward the end, where the last accuracy peak occurs earlier in the present DAP than in the previous one. The earlier data seem anomalous in that respect, since subsequent experiments in the same series (Repp, 1998c) showed a peak similar to the one observed in the present data. Another major difference of unknown origin occurred in the second half of Bar 1. The present DAP is more reliable because it is based on almost 10 times as much data as the DAP of the earlier study.
8. However, there could also be a contribution of grouping to lengthening in position 5, in that group-final lengthening may extend to the IOI following the final note onset. In addition, harmony may play a role, but it is not clear whether lengthening should be expected before or on a harmonic change.
9. One additional undergraduate without musical training participated but had difficulty tapping in synchrony with the music; his data were excluded.
10. The response keys traveled about 3 mm and made some audible noises, both when they struck the key bed and when they returned to their resting position. If anything, this auditory feedback was expected to increase synchronization accuracy. However, the use of a computer keyboard to collect responses also introduced some "quantization noise" in the data because the computer sampled the state of the keyboard only 60 times per second (i.e., every 16.7 msec). This implied a random delay with a mean of 8.3 msec and a variance of 25.5 msec². The delay was partially counteracted by the fact that key depressions were sensed electronically before the key touched bottom. In any case, the absolute magnitude of the synchronization error (i.e., the size of the anticipation effect commonly found in synchronization tasks; see, e.g., Aschersleben & Prinz, 1995) was not of particular interest here. The quantization variance, too, was considered negligible in view of the large number of observations and in view of the fact that it was much smaller than the total variance of the asynchronies. No correction was applied to the data, as it would have reduced standard deviations only by about 1 msec.
11. The variability among the participants is not displayed in Figure 5a because it largely reflects individual differences in the average magnitude of the anticipation error.

NOTES

1. The analysis of individual differences in perceptual biases was originally a second major purpose of Experiment 1. However, the results turned out to be difficult to interpret and therefore will not be reported.

APPENDIX

12. For increments in the penultimate position, there were only two asynchronies; for increments in the final position, there was only one. Due to a small error in stimulus construction, the 20-msec increment never appeared in the final position.

13. It is interesting to note that B.R. (the author) was the oldest participant by far. It is possible that the efficiency of the error correction process decreases with age.

14. In an earlier study, musically trained individuals detected such increments about 70% and 40% of the time, respectively (Repp, 1998d, Experiment 1).

15. This is for an average detection rate of about 60%, achieved by the adaptive staircase procedure of Experiment 1 (which proceeded from initial high scores to about 50% correct). The values do not reflect the actual detectability of 40- and 20-msec increments in these positions, but they can be estimated from the average percentages given in the preceding note. Accordingly, 40-msec increments may have been detected about 85% of the time in high-scoring positions and 55% in low-scoring positions, whereas 20-msec increments may have been detected 55% and 25% of the time, respectively.

16. Such data will soon be available from recently completed experiments (Repp, *in press-a*). The same musical excerpt was used, and participants tapped on a velocity-sensitive MIDI keyboard. A preliminary analysis was conducted on a subset of the data (averaged across 9 trials). Surprisingly, 7 of the 8 participants showed significant negative correlations between key velocity (monotonically related to tap force) and dwell time: They held down the response key longer when they pressed it more gently, which is the opposite of what has been observed in the context of deliberate accentuation (Billon & Semjen, 1995). Importantly, however, there were no significant correlations between velocity and either preceding or following IOI duration. Three participants showed significant correlations between velocity and asynchrony, but two correlations were positive and one was negative. These preliminary results suggest that variations in average tap force were not responsible for the average tap-timing profile.

The purpose of this methodological Appendix is to document a minor flaw in the stimuli used in Experiments 1 and 2, as well as in earlier studies, and to provide arguments (beyond the empirical evidence of Experiment 3) of why it should not have had any substantial effects on the results. This discussion is not only dictated by scientific honesty but also may be instructive for others who are planning to use complex MIDI stimuli in controlled psychological experiments.

The source of the flaw is painfully obvious in hindsight, but a combination of circumstances (which need not be belabored here) caused the author to ignore it for quite a long time. MIDI instructions are transmitted serially and usually take about 1 msec per instruction (see, e.g., Marans, 1991; Roads et al., 1996). At each sixteenth-note position in the Chopin Etude excerpt (Figure A1), as many as four notes start simultaneously while as many as five preceding notes are terminated. In the MIDI instructions, these events were specified to occur at the same nominal time. Unfortunately, the instructions were arranged so that the note offsets preceded the note onsets, and, moreover, the note onsets were arranged in order of increasing pitch. This meant that the most important notes (the ones with the highest pitch, on which the participants' attention was assumed to be focused) were preceded by a variable number of note offset and onset instructions, ranging from 1 to 8, so that their onsets were delayed relative to their nominal onset times, perhaps by as much as 8 msec. This temporal jitter was not detectable by ear but could have affected the results.

First, the actual magnitude of the jitter was determined. It could not be measured accurately in the recorded sound out-

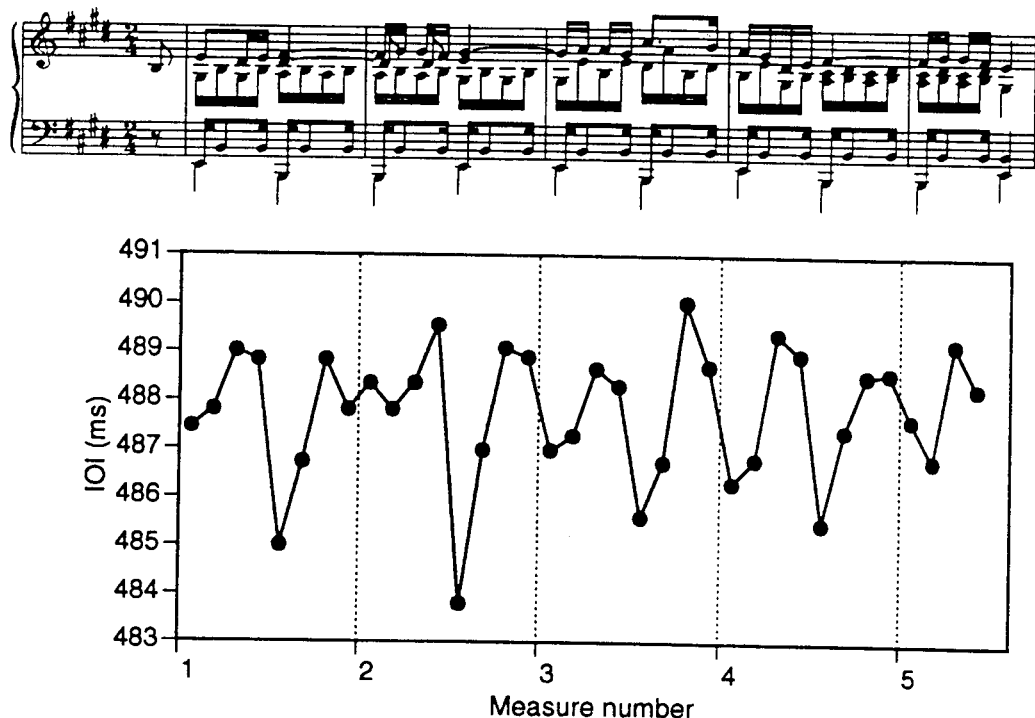


Figure A1. Baseline IOI durations in the stimuli of Experiments 1 and 2, as measured in the acoustic output (see text).

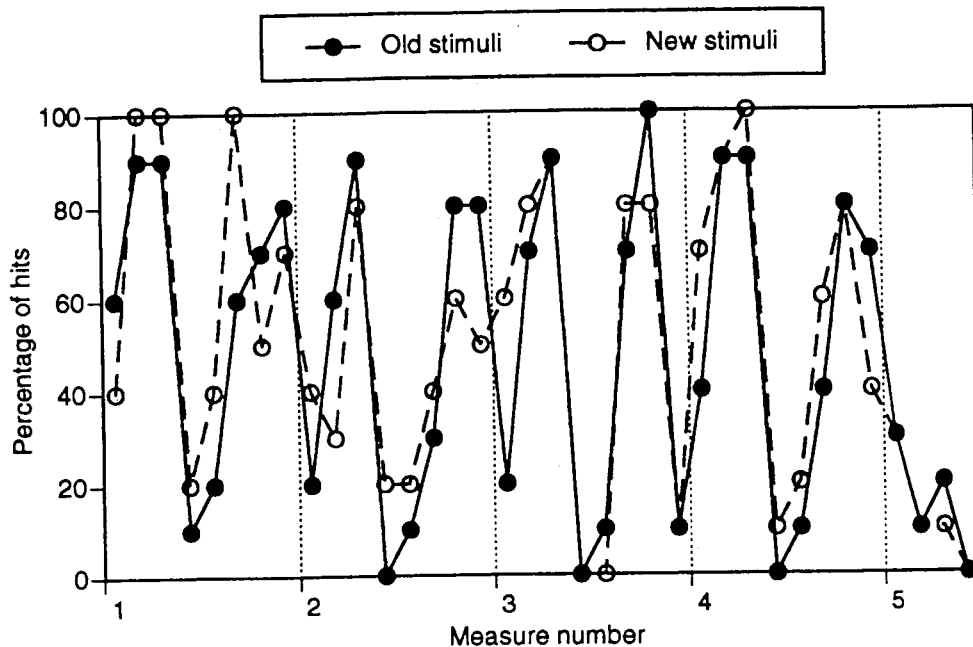


Figure A2. DAPs for the author as a participant in versions of Experiment 1 with the old stimuli (containing jitter) and with new, revised stimuli (without jitter).

put, due to measurement error caused by the overlap of successive tones. (Piano tones have a substantial decay time following their nominal offset; see Repp, 1995a.) To make precise measurement possible, the MIDI instructions of the nominally isochronous baseline stimulus were modified: Additional "note off" instructions were inserted 50 msec after each note onset (to eliminate energy overlap of successive tones), and the MIDI velocities of all note onsets preceding the highest pitched note in a chord were set to zero, so that only the important notes remained. The resulting output (a series of short monophonic tone bursts) was recorded on digital tape, re-input to the computer, and measured using a digital waveform editor. The measurements were carried out three times, and IOIs were computed from the onset measurements. The average standard deviation of the measurements was 0.5 msec, which represents a combination of computer unreliability (if any) and human measurement error. The average measured IOI durations were plotted as a function of the difference in the number of MIDI instructions preceding the notes delimiting the IOI. (These differences ranged from -6 to +4.) The data points were fit very well ($R^2 = .956$) by a regression line with a slope of .61, indicating that MIDI instructions in the author's setup took only 0.6 msec to be transmitted. Thus, the actual range of variation in onset times was only about 4 msec, and that in IOI durations was 6 msec.

Although this variation could not be eliminated, it was effectively nullified in the stimuli of Experiment 3 by (1) putting note offset commands after note onset commands and (2) arranging the note onset commands in order of decreasing pitch, so that the most important note came first. At three points in the music, where a note offset and a note onset for the same pitch coincide, the offset was specified as occurring 1 msec before the onset. The accuracy of the resulting output was verified by carrying out repeated measurements of the sound output using modified MIDI instructions, as described above. The

average IOI durations were all within 1 standard deviation of 488 msec (see note 2).

Despite its small size, the temporal jitter in the earlier stimuli was of concern because it exhibited a regularity related to the musical structure. The "jitter profile" of the IOIs is shown in Figure A1. It can be seen that IOIs were shortest in the fifth position of each bar, precisely where expressive lengthening was maximal (Figure 1) and where IOI increments were difficult to detect (Figure 3a). Otherwise, however, there was little correspondence. The correlation of the jitter profile with the typical expressive timing profile was marginally significant ($r = -.35$, $p < .05$), but that with the average DAP of Experiment 1 was not ($r = .10$).

While the jitter obviously cannot account for the complete pattern of results in Experiment 1, the slightly shorter baseline IOIs in the fifth position of each bar may have made IOI increments somewhat more difficult to detect in these positions. There is no direct evidence for this, however. Although stimuli with the same temporal jitter were used in some earlier experiments (Repp, 1998a, Experiments 1 and 2; 1998c, Experiment 1), Repp (1998d, Experiment 1) used a different MIDI sequencer in which the note onsets were arranged in order of descending pitch, and Repp (1998c, Experiment 3) used stimuli in which a click immediately preceded (by 0.6 msec) the first note onset in each position. Thus, the jitter was reduced to about half its size, deriving only from note offset instructions preceding note onset instructions. In both of these studies, IOI increments were about as difficult to detect in position 5 as in position 4 of each bar. In the present Experiment 1, however, they were easier to detect in position 5, which is contrary to expectations based on the jitter pattern.

As a further check on the validity of the DAPs, the author ran himself through the procedure of Experiment 1 using both the original stimuli and a revised set of stimuli that was free of jitter. The results of this exercise are shown in Figure A2. The re-

sults of the two runs were quite similar ($r = .86, p < .001$), which demonstrates that the jitter did not have any major effect on perception. The difference between the DAPs for old and new stimuli exhibited a weak but significant correlation with the jitter profile ($r = .36, p < .05$), which is consistent with slightly improved increment detection in longer baseline IOIs than in shorter baseline IOIs.

As far as the synchronization task is concerned, Experiment 3 demonstrated that the major pattern of results did not derive from the jitter. If the participants in Experiment 2 had tracked the minute temporal variation in the baseline stimuli, a positive lag-one correlation between the jitter and tap IOIs would be expected (Hary & Moore, 1987; Mates, 1994b; Michon, 1967). This correlation was in fact significant ($r = .67, p < .001$). However, there was also a smaller positive correlation between the average tap IOIs of Experiment 3 and the jitter profile ($r = .36, p < .05$), even though there was no jitter in that experiment, which suggests that the correlation in Experiment 2 was partially due to a (by no means accidental) corre-

lation between the lag-one shifted jitter profile and the typical expressive timing profile ($r = .56, p < .01$). Therefore, the difference between the tap-timing profiles of Experiment 2 and 3 (Figure 8b) was computed. Even though the tap-timing profiles of Experiments 2 and 3 had not been significantly different in the ANOVA, their difference correlated significantly with the jitter profile ($r = .62, p < .001$), which suggests that the participants indeed may have tracked the small jitter in the stimuli (cf. Hary & Moore, 1987). This small variation due to jitter tracking was superimposed on the larger variation of interest in Experiment 2, but it was, of course, absent in Experiment 3. Similarly, in detection tasks such as that used in Experiment 1, there may have been a small overlay of systematic variation due to the jitter, but the gross pattern of the results clearly did not derive from it.

(Manuscript received May 1, 1997;
revision accepted for publication February 7, 1998.)

Forthcoming Articles

The Forthcoming Articles list is not currently printed in this journal. It can be found at the *Perception & Psychophysics* Web site: <http://www.sig.net/~psysoc/prpfth.htm>.