APPENDIX TO

“PERCEPTION OF TIMING IS MORE CONTEXT SENSITIVE THAN SENSORIMOTOR SYNCHRONIZATION”

EFFECTS OF ABSENCE OF INTENSITY VARIATION
IN THE MUSICAL TEST EXCERPTS OF EXPERIMENTS 1 AND 2

Introduction

In contrast to remote temporally varying context, which was hypothesized to increase the quasi-random variability of an internal timekeeper or of higher-level perceptual processes, local intensity variation was expected to affect the pattern and/or magnitude of systematic timing variability, which is specific to music and other structured sequences.

In a series of earlier perceptual studies, participants’ ability to detect local deviations from temporal regularity in musical excerpts has been assessed (Repp, 1992a, 1992b, 1995, 1998b, 1998c, 1998d, 1999b, 1999c). The detectability of timing deviations of constant magnitude varies greatly with their position in the music. This pattern of systematic variation of detection scores, which primarily reflects the rhythmic grouping structure of the music, has been termed the “detection accuracy profile” (Repp, 1998d). It is believed to be due to specific temporal expectations induced by the musical structure—that is, to systematic modulations of the period of the perceptual timekeeper (or attentional oscillator) that predicts future tone onsets in the music (Repp, 1999b). The musical excerpts used in those earlier experiments always contained expressive dynamics, and it is possible that the detection accuracy profile was influenced, if not entirely
caused, by this variation. The reason why expressive dynamics was always present in the earlier materials is that music sounds rather crude and unnatural without it. Nevertheless, an examination of its effect on time perception seemed overdue.

There is evidence that the perception of timing is influenced by intensity differences between successive sounds. Following up on early research by Woodrow (1909), Tekman (1995, 1997) demonstrated that temporal intervals preceding louder sounds tend to be perceived as longer than other intervals in a sequence. This finding may be related to the observation that the corresponding interval is shortened in accentuated finger tapping, as long as the tempo is not very fast (Billon & Semjen, 1995; Billon, Semjen, & Stelmach, 1996; Piek, Glencross, Barrett, & Love, 1993; Semjen & Garcia-Colera, 1986). Thus there may be a connection between perception and action with regard to interactions between force (intensity) and timing. Therefore, removal of intensity variation from musical materials may affect the detection accuracy profile for deviations from regularity, especially in positions where the critical temporal interval previously had preceded a loud tone.

If intensity variation affects time perception by modulating an internal timekeeper, and if exactly the same timekeeping processes are involved in time perception and action timing, then removal of intensity variation should also have effects on the timing of synchronized finger taps. Here, too, quasi-random variability (due to inherent timekeeper and motor variance) must be distinguished from systematic variability induced by musical structure. Several recent studies have established that finger taps synchronized with mechanically timed (i.e., entirely isochronous) music exhibit a systematic pattern of deviations
from regularity (Repp, 1999a, 1999b: Exp. 3, 1999c). This “tap timing profile”, usually reported in terms of variation in inter-tap intervals, is specific to a particular musical structure, but it is only weakly related to the detection accuracy profile for the same music, which indicates that it is sensitive to different aspects of the musical structure. Nevertheless, here, too, the question arises to what extent these systematic variations may have been caused by the intensity variation in the music, which was always present in earlier studies and which may have affected the period of the timekeeper that controlled the timing of the taps.

Comparison of the results of Experiments 1 and 2, in which intensity variations were absent, with those of earlier similar experiments in which intensity variations were present (Repp, 1999b, 2000a) was expected to shed some light on these questions. As with the effect of remote temporal variation, it was hypothesized that local intensity variation would have a greater impact on time perception than on the timing of synchronized action.

**Experiment 1**

The absence of local intensity variation in the test excerpts was expected not only to result in a general facilitation of detection of deviations from temporal regularity (as already observed in Repp, 1998d) but also to have an effect on the shape of the detection accuracy profile. However, significant variation in detection accuracy was expected to remain; in other words, it seemed unlikely that the detection accuracy profiles obtained in earlier experiments were solely due to intensity variation. This expectation was
confirmed by the significant main effect of position obtained in the ANOVA (see main text).

Since there were no significant interactions involving position, it seemed reasonable to calculate an average detection accuracy profile across the four precursor conditions and to compare it to the average detection accuracy profile obtained in a previous experiment of similar design (Repp, 1999b: Exp. 1) that used the same musical excerpt, but with a typical pattern of expressive dynamics similar to that shown in Figure 1b. That experiment employed a large number of participants (n = 41), resulting in a very reliable estimate of the typical detection accuracy profile. It included 10 test blocks, without precursors, and the Δt value was changed adaptively from block to block for each participant, as in the present pretest, to achieve an average hit percentage of about 50%. The two average detection accuracy profiles are shown superimposed in Figure A1a, and their difference (“without dynamics” minus “with dynamics”) is plotted in Figure A1b.

Detection scores were somewhat higher overall in Experiment 1 than in the earlier experiment, but this difference may be due to different criteria for selecting Δt values. What is of interest here is any difference in profile shape. In a two-way ANOVA with the grouping variable of experiment (2) and the within-participant variable of position (36), the Experiment x Position interaction, which reflects differences in profile shape, was clearly significant, F(35,1785) = 4.9, p < .0001. Figure A1b shows that elimination of dynamic variation led to a relative improvement of detection scores in the fifth IOI of each bar. This is precisely where, in the earlier study, the top-line intensity contour dropped precipitously because a loud sustained melody tone was followed by a soft accompaniment
tone (see the musical score in Fig. 1). Thus it was relatively harder to detect an increment in IOI duration when the IOI was initiated by a loud tone and terminated by a soft tone than when it was framed by tones of equal intensity. By contrast, there seemed to be no effect of removing intensity variation on hit percentages in positions where previously the intensity increased abruptly from the accompaniment to the melody (position 1 of bars 1, 2, 3, and 5, and position 7 of bar 3; see Fig. 1). Thus the effect of intensity differences on detectability was asymmetric, with only an abrupt intensity decrease creating interference. A role of dynamic variation in temporal discrimination was also suggested by a moderate but significant correlation of .50, $d.f. = 34$, $p < .01$, between the difference of the two detection accuracy profiles (Fig. A1b) and the difference in MIDI key-press velocities between successive top-line tones in the musical excerpt used by Repp (1999b).

There were some other local differences between the two detection accuracy profiles (Fig. A1a), but these were less systematic and therefore difficult to interpret. It is clear, however, that they were not due to temporal expectations induced by the atypical timing profile of the expressively timed precursors. In that case, for example, the greatly lengthened initial IOIs in bars 2, 3, and 5 (Fig. 1a) should have led to expectations of lengthening and hence to selectively reduced detection scores for lengthening of these IOIs, which was not the case. Thus the results are consistent with Repp’s (1998b) failure to find any effect of a specific precursor timing pattern on the detection accuracy profile. The variations in the detection accuracy profile seem to be caused by local structural factors that affect timing perception, not by cognitive temporal expectations based on a temporal pattern memory.
Even though the average detection accuracy profile of Experiment 1 exhibited reduced variation across positions in the music relative to the profile obtained previously with dynamic variation present (Fig. A1a), it still showed a significant negative correlation \( r = -0.73, \text{ d.f.} = 34, p < .001 \) with a typical expressive timing profile for the music (Repp, 1999b: Fig. 1). In Repp’s (1999b) experiment, the analogous correlation had been –.86. Thus it can still be argued that the present detection accuracy profile reflects structurally based expectancies of a typical expressive timing pattern (cf. Repp, 1992b, 1998b, 1998d), albeit perhaps expectancies for the typical timing pattern of an atypical (hypothetical) performance that is deprived of dynamic variation.

**Experiment 2**

By eliminating all intensity variation from the test excerpts, Experiment 2 addressed the question of whether the previously observed systematic deviations from regularity in taps synchronized with isochronous music (Repp, 1999b, 1999c, in press) were partially or entirely due to the presence of expressive dynamics in the music. If so, that would suggest that intensity differences affect not only the perception of sequence timing but also the timing of action, perhaps via a perception-action coupling. The expectation, however, was that intensity variation would have little effect on synchronization.

Additional data were available from a separate brief experiment in which participants tapped in synchrony with strictly isochronous versions of the Chopin Etude excerpt containing neither hesitations nor intensity variation. The excerpt was presented 30 times, in three blocks of 10 trials each, without precursors. This experiment was conducted at the end of the second session of
Experiment 1 and thus involved the participants of that experiment. The data of one participant were rejected because of phase drift, so that the results were based on data from 11 participants.

Two comparisons were conducted across experiments, one involving isochronous musical excerpts and the other involving excerpts containing hesitations, from which the data point triplets containing the phase correction had been removed. First, the average ITI profile from the ancillary experiment just mentioned (without expressive dynamics) was compared with the ITI profile obtained in an earlier experiment in which participants tapped in synchrony with the same isochronous musical excerpt, the only difference being that the music had expressive dynamics (Repp, 1999b: Exp. 3). These two profiles are shown in Figure A2a. Second, the average ITI profile from the no-precursor condition of Experiment 2 (without expressive dynamics) was compared with an average ITI profile for the music containing expressive dynamics (Repp, 1999b: Exp. 2). This second pair of profiles is plotted in Figure A2b.

The two ITI profiles in each panel of the figure were quite similar to each other; their correlations were .79 and .80, respectively (d.f. = 31, \( p < .0001 \); first 3 data points excluded). Across the two panels, the profiles were also similar. A two-way ANOVA with the variables of experiment (2) and position (33, first 3 omitted) on the data shown in Figure A2a showed the position main effect to be reliable, \( F(32,704) = 11.3, p < .0001 \), but the Experiment \( \times \) Position interaction was nonsignificant, \( F(32,704) = 1.3, p < .12 \). In a similar ANOVA on the data shown in Figure A2b, the main effect of position was again reliable, \( F(32,704) = 13.3, p < .0001 \), and the Experiment \( \times \) Position interaction also reached significance, \( F(32,704) = 1.5, p < .05 \). However, the interaction was weak and did not seem to
follow a systematic pattern. In particular, there were no systematic changes at position 5 in each bar, where the removal of intensity variation had a systematic effect on the detectability of hesitations (Fig. A1). Therefore, it may be concluded that expressive dynamics had a negligible effect on the systematic deviations from regularity of the ITIs.

It may be asked whether the systematic variation in tap timing is a consequence of systematic variations in tap force. The ITI profiles in Figure A2 consistently show lengthened ITIs preceding metrically strong positions. If participants tapped more forcefully in metrically strong positions, the results would run counter to the finding that deliberately accented taps at a moderate tempo are preceded by shortened and followed by lengthened ITIs (Billon & Semjen, 1995; Billon et al., 1996; Piek et al., 1993). To further investigate this issue, an analysis of average key depression velocities, key release velocities, and key dwell times was conducted on the tapping data collected in the ancillary experiment conducted at the end of Experiment 1. These analyses showed that none of these kinematic quantities varied systematically as a function of position in the music or correlated significantly with either the asynchronies or the ITIs. Thus the systematic timing variation seems to originate in timing control, not in systematic kinematic variation. As suggested earlier (Repp, 1999b), it seems to reflect a modulation of the period of an internal timekeeper in response to structural features of the music.
FOOTNOTES

1 A caveat must be added with regard to dynamic variation in music. In this study, the nominal intensities (MIDI keypress velocities) of individual tones were equalized. However, this may not have completely eliminated perceived loudness variation. First, the number of nominally simultaneous tone onsets varied from position to position, and this may have resulted in corresponding variations in perceived dynamics because the different voices were not completely segregated perceptually. Second, there may have been interactions between pitch and perceived dynamics, such that lower piano tones (which are richer in harmonics) sounded louder than higher piano tones of equal intensity. This would also not be surprising in view of a common positive covariation of pitch and intensity in music, which may lead listeners to expect high tones to be louder than low tones, and hence to be perceived as softer when intensities are equal. The complete elimination of perceived dynamic variation in complex music is not trivial and was not attempted in the present study.

2 The ITI profile for the isochronous precursors in Experiment 2 was also similar in shape to these profiles.
REFERENCES


FIGURE CAPTIONS

Fig. A1. (a) Average detection accuracy profiles (hit percentages) in Experiment 1 (“without dynamics”) and in Repp (1999b: Exp. 1; “with dynamics”), with standard error bars. (b) The difference between the two profiles (“without dynamics” minus “with dynamics”). Filled triangles indicate inter-onset intervals (IOIs) initiated by melody tones, as in Figure 1.

Fig. A2. Two sets of inter-tap interval (ITI) profiles, with standard error bars, for music without and with expressive dynamics, respectively. (a) Music without hesitations. (b) Music with hesitations, but with compensatory responses excluded. (See text for origin of data.)
Fig. A1
Fig. A2