THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2009, 62 (12), 2306–2312

Psychology Press
Taylor & Francis Group

Short article

Rhythmic sensorimotor coordination is resistant but not immune to auditory stream segregation

Bruno H. Repp
Haskins Laboratories, New Haven, CT, USA

In a recent study of musicians' sensorimotor synchronization with auditory sequences composed either of beat and subdivision tones differing in pitch or of beat tones only, Repp (2009) found that the phase correction response (PCR) to perturbed beats was inhibited by the presence of subdivisions regardless of whether beats and subdivisions formed integrated or segregated perceptual streams. The present study used a different paradigm in which perturbed subdivisions triggered the PCR. At the slower of two sequence tempi, the PCR was equally large in integrated and segregated conditions, but at the faster tempo stream segregation reduced the PCR substantially. This new finding indicates that although the PCR is strongly resistant to auditory stream segregation, it is not totally immune to it.

Keywords: Auditory scene analysis; Stream segregation; Synchronization; Perception–action dissociation; Phase correction.

In a recent study, Repp (2009) reported a striking dissociation between auditory perception and control of action timing, the latter being assessed by the automatic phase correction response (PCR) in sensorimotor synchronization. The PCR is the shift of the tap following an unexpectedly shifted tone. Musically trained participants were presented with not-quite-isochronous ABBABB ... sequences (where A and B, respectively, signify beat and subdivision tones differing in pitch) and with baseline sequences in which the B tones were omitted. From time to time, one of the A tones occurred early or late. There were two tasks: In the perception task, participants pressed a key whenever they detected a shifted A tone; in the synchronization task, they tapped in synchrony with the A tones. The main independent variables were the pitch separation between the A and B tones and the sequence tempo. The perceptual results showed that, relative to baseline, the presence of B tones facilitated the detection of shifted A tones when the pitch separation was small (2 semitones, st) but not when it was moderate (10 st) or large (48 st), indicating that stream segregation occurred already with the 10-st separation and certainly with the 48-st separation,

Correspondence should be addressed to Bruno H. Repp, Haskins Laboratories, 300 George Street, New Haven, CT 06511–6624, USA. E-mail: repp@haskins.yale.edu

This research was supported by National Science Foundation Grant BCS-0642506.

http://www.psypress.com/qjep
DOI:10.1080/17470210903118107

regardless of sequence tempo. The synchronization results showed that, relative to baseline, PCRs to shifted A tones were reduced by the presence of B tones, more so at the slower sequence tempo (as previously shown by Repp, 2008), but these reductions were equally large at all pitch separations. These results suggested to Repp (2009) that the PCR is immune to stream segregation and that perceptually segregated streams are functionally integrated in sensorimotor coordination, at least by musicians.

When A and B tones can be integrated perceptually into a single rhythm, the B tones serve as temporal references for the perception of shifted A tones: Shifts can be detected as changes in the short B–A or A–B interval rather than in the long A–A interval. This explains the facilitation of detection at small pitch separations (cf. Jones, Jagacinski, Yee, Floyd, & Klapp, 1995). The reduction of the PCR by B tones is likewise believed to be due to the B tones serving as temporal references, together with the preceding A tone, in this case for the temporal placement of the next tap (Repp, 2008). Although participants are instructed to synchronize with the beat and ignore the subdivisions, they nevertheless seem to rely on both in timing their actions. Because the subdivisions are never shifted, they serve as stable references and thus attenuate the PCR to a shifted beat tone. Perceptual stream segregation evidently prevents the B tones from serving as perceptual references but not from serving as action references.

The present study attempted to replicate Repp's (2009) findings by using the complementary paradigm in which the A tones remain stable, and the B tones are shifted from time to time. Repp (2008) showed that shifted subdivision tones do elicit a PCR, and that this PCR is larger when the sequence tempo is slower. (Beat cycle durations were 540 or 720 ms.) Will this type of PCR also be immune to auditory stream segregation?

Method

The methods were similar to those of Repp (2009) but differed in some important (and some unimportant) details. Only a small and a very wide pitch separation were used. Baseline sequences, which here consisted of sequences with A tones omitted, were included only in the perception task. In the synchronization task, such sequences were not appropriate in the present paradigm because the task was to synchronize with A tones, not to tap in the gap between pairs of B tones (a much more difficult task; see Repp, 2005b).

Participants

The participants included 9 graduate students from the Yale School of Music (6 women, ages 22–28 years), who were paid for their services, and the author (age 64 years). All were regular participants in perception and synchronization experiments. Three of the students and the author had been participants in the previous study (Repp, 2009).

Materials and equipment

ABBABB . . . sequences were generated online by a program written in MAX 4.0.9, running on an Intel iMac computer. The tones (piano timbre) were produced by a Roland RD-250s digital piano according to musical-instrument-digital-interface (MIDI) instructions from the MAX program and presented over Sennheiser HD540 II headphones. A and B tones had equal nominal durations (40 ms) and intensities (MIDI velocities). The sequence tempo was either fast (beat cycle duration = 450 ms; interonset intervals, IOIs, of 150 ms between successive tones) or slow (beat cycle duration = 600 ms; IOIs of 200 ms between successive tones). The A tones had either a higher or a lower pitch than the B tones, and the pitch separation between the tones was either narrow (2 st) or wide (46 st). At the narrow separation, the pitches of the tones were C3 (131 Hz) and D3 (147 Hz); at the wide separation, they were C3 and A#6 (1865 Hz). The factorial combination of two tempi, two pitch separations, and two pitch assignments resulted in eight sequences that constituted one block of trials in the synchronization task. In the perception task, each block included in addition four baseline sequences (two fast, two slow) in which A tones were omitted (.BB.BB.BB . . .).

The pitch of the tones in those sequences was either C3 or A#6.

Each sequence contained eight shifted pairs of B tones. The positions of these pairs were determined randomly, with the earliest possible position of the first shifted pair being the 8th cycle, and between 4 and 7 unperturbed beat cycles intervening between successive perturbations. Consequently, sequences were of variable length, containing between 47 and 70 cycles. The magnitudes of the shifts were $\pm 2\%$, $\pm 4\%$, $\pm 6\%$, and $\pm 8\%$ of the cycle duration. Thus, shifts ranged from ± 9 ms to ± 36 ms at the fast tempo and from ± 12 ms to ± 48 ms at the slow tempo.

Procedure

Participants came for two sessions, typically one week apart. The first session was always the synchronization task. Participants were instructed to tap in synchrony with the beats (A tones) and ignore the subdivisions (B tones). They were told that the beats were perfectly regular but that some small temporal irregularities might occur in the subdivisions. Participants started each trial by pressing the space bar of the computer keyboard and started tapping with the third A tone. They tapped with the index or middle finger of their preferred hand on a Roland SPD-6 percussion pad that was held on the lap. Eight blocks of 8 trials each were presented, with short breaks between blocks.

In the second session, the perception task, participants were instructed to press the down-arrow key on the computer keyboard as quickly as possible whenever they heard the slightest temporal irregularity in a sequence. They were told that some sequences would not contain any beat tones. Participants started each sequence by pressing the space bar and kept their finger on the response key throughout. Five blocks of 12 trials each were presented, with short breaks between blocks.

Results

Perception task

Reaction times (RTs) were measured from the onset of the first shifted B tone in each pair. Responses with RTs between 200 and 1,200 ms were considered correct (hits); others were considered false alarms. The average number of false alarms per session was 21, which translates into 0.35 per sequence. Hit percentages naturally increased with the magnitude of the shift to be detected.

The average hit percentages in each condition are shown in Panels A and B of Figure 1. The horizontal dashed lines represent the results for the baseline sequences, collapsed over the pitch variable, which clearly did not have any effect. (Standard errors are not shown for the baseline conditions but were similar to those in the other conditions.) Detection performance clearly exceeded baseline levels at the 2-st pitch separation, at both tempi. At the slow tempo, detection scores dropped to baseline levels at the 46-st pitch separation. At the fast tempo, scores dropped below baseline levels, indicating interference from the separate stream of A tones.

A $2 \times 2 \times 2$ repeated measures analysis of variance (ANOVA; baseline condition not included) yielded significant main effects of pitch separation, $F(1, 9) = 96.06$, $p < .001$, and of tempo, $F(1, 9) = 36.18$, $p < .001$, as well as a significant interaction between these two variables, $F(1, 9) = 23.64$, $p < .001$. In addition, the main effect of pitch assignment also reached significance, $F(1, 9) = 10.39$, $p = .01$: Detection performance was slightly better when the B tones were lower than the A tones. Separate 2×2 ANOVAs on each tempo condition confirmed highly reliable ($p < .001$) main effects of pitch separation for each. The main effect of pitch assignment reached significance only at the fast tempo, $F(1, 9) = 5.78$, $p = .04$.

Synchronization task

The PCR to each shifted pair of B tones was measured by subtracting the asynchrony (with the A tone) of the immediately following tap from the asynchrony of the preceding tap. The PCRs for the same shift magnitudes were averaged across blocks, and these averages were then linearly regressed onto shift magnitude. Because the regression line passes through the origin (or nearly so), the slope of the regression line times

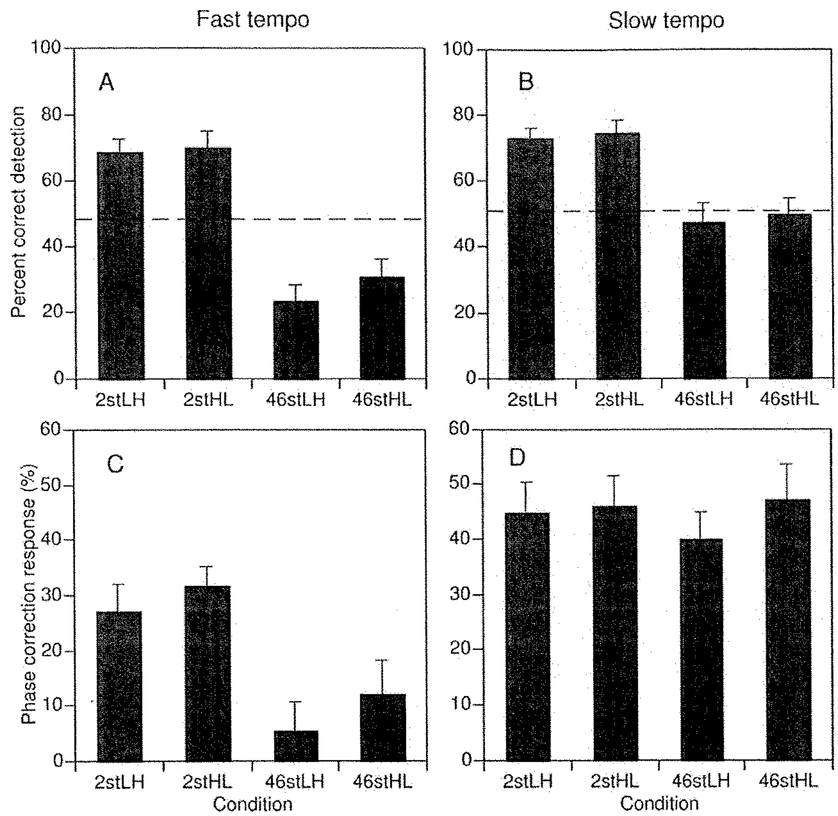

Figure 1. *Percentage of correct detection responses (Panels A, B) and PCR as a percentage of shift magnitude (Panels C, D) in four conditions at fast and slow tempi. Dashed horizontal lines (Panels A, B) represent baseline conditions (averaged across pitch). 2st, 46st = pitch separation (i.e., 2 and 46 semitones). LH = low A tones, high B tones; HL = high A tones, low B tones. Error bars are standard errors.*

100 essentially expresses the mean PCR as a percentage of shift magnitude.

These percentages are shown in Panels C and D of Figure 1. As expected, PCRs were larger at the slow than at the fast tempo, $F(1, 9) = 24.86$, $p < .001$. Pitch separation also had a significant main effect, $F(1, 9) = 14.80, p = .004$, but it interacted significantly with tempo, $F(1, 9) = 16.38$, $p = .003$. Separate 2 × 2 ANOVAs on the two tempo conditions revealed a significant reduction in PCRs at the 46-st pitch separation at the fast tempo, $F(1, 9) = 19.33$, $p = .002$, but not at the slow tempo, $F(1, 9) = 0.68$, $p = .430$. In the three-way ANOVA, the main effect of pitch also reached significance, $F(1, 9) = 5.31$, $p = .047$: PCRs tended to be larger when the B tones were low than when they were high. This effect did not reach significance, however, in the separate analyses of the tempo conditions.

Discussion

The perception results clearly demonstrate that auditory stream segregation occurred at the wide pitch separation, as one should have expected given earlier results in the literature (e.g., Sussman, Wong, Horváth, Winkler, & Wang, 2007; van Noorden, 1975). Relative to baseline sequences in which changes in the long (300 or 400 ms) intervals between pairs of B tones had to be detected, the presence of A tones improved detection of temporal shifts when the pitch separation was narrow by enabling perceptual integration of A and B tones into a single rhythm, so that changes in short (150 or 200 ms) B–A or A–B intervals could be detected. No such improvement occurred when the pitch separation was wide, which indicates perceptual segregation of A and B tones, even though the task encouraged perceptual integration.

The PCR results at the slow tempo replicate Repp's (2009) findings of immunity to stream segregation: Even though the perception results show unambiguous evidence of segregation at the wide pitch separation, the PCR was not affected at all. The results at the fast tempo, however, diverge from the previous findings: Here stream segregation did reduce the PCR significantly; in fact, the PCR practically disappeared. Apparently, a PCR triggered directly by shifted subdivisions is more sensitive to stream segregation than is the inhibition of the PCR to a shifted beat tone by intervening subdivisions. The reason for this difference is not entirely clear. In each case, beat and subdivision tones are assumed to serve as temporal references for tap placement. One possible explanation is that the PCR triggered by shifted B tones depends on the precise timing of those tones whereas a reduction of the PCR triggered by shifted A tones requires only the presence of intervening B tones. It is likely that stream segregation is not categorical but varies in strength. At the wide pitch separation at the fast tempo, it may be argued, segregation may have been strong enough to obliterate any subconscious sensitivity to the relative timing of A and B tones, whereas the presence of B tones could still be registered. Although sensitivity to the relative timing of A and B tones is not necessary for a PCR—these tones could serve as independent temporal references—perhaps the results indicate that relative timing plays a subconscious role, after all. In a similar vein, it may be that the PCR reflects subconscious registration of asynchronies between taps and tones. Conscious detection of sensorimotor asynchronies is not likely to play a role because it is quite poor (see Repp, 2000; Repp & Knoblich, 2007), and the linearity of the PCR regression function through the origin argues against any role of a detection threshold in phase correction. However, if subconscious registration played a role, it could plausibly be argued that asynchronies of taps with A tones (the synchronization targets) are easier to register than increases or decreases in the large asynchrony of taps with B tones. That might then explain why strong stream segregation erases the PCR to shifted B tones while it preserves the PCR to shifted A tones. It does not explain, however, why intervening B tones continue to reduce the PCR to shifted A tones. Moreover, the author has long argued on the basis of various findings (see Repp, 2005a) that the PCR is not based on subconscious detection of asynchronies or of changes in sequence timing but reflects phase resetting with reference to time points defined by pacing events. This hypothesis may yet be proven to be incorrect, but the present findings are not sufficient to reject it. The hypothesis also attributes the resistance of the PCR to stream segregation to the reliance of phase correction on different temporal information (time points) than conscious perception relies on (intervals between time points). It is not necessary to attribute the finding to different processing streams for action and perception, and Repp's (2009) initial reference to Milner and Goodale (1995) may have been misleading in that regard. For a similar argument in the visual–spatial domain, see Smeets, Brenner, de Grave, and Cuijpers (2002).

One difference between the experiment of Repp (2009) and the present one that has not been mentioned so far is that the wide pitch separation was 48 semitones (four octaves) previously, whereas it was only 46 semitones here. This change was made because someone might argue that an octave relation between the A and B pitches reduces stream segregation. However, such an effect seems quite implausible considering the very wide pitch separation, at which the octave relationship is not perceptually salient. To the author's knowledge, there is no evidence that musical consonance between successive complex tones has any effect on stream segregation.

Admittedly, stream segregation in the present study and in Repp (2009) was not as strong as that in classic studies of the phenomenon.

To create strong stream segregation, Van Noorden (1975) and many subsequent perceptual studies employed ABA.ABA.ABA ... sequences composed of pure tones with IOIs of about 100 ms. However, such sequences are problematic in a synchronization task because it is difficult to keep synchrony with the B tones if the IOI between A and B tones is less than 200 ms, and the tones are similar in pitch (Repp, 2005b). This prevents a comparison between integrated and segregated conditions unless the tempo is made very slow. (Whether stream segregation facilitates synchronization with ABA.ABA.ABA ... sequences remains to be investigated.) Tapping instead in synchrony with the A tones, which occur at intervals of only twice the IOI, at a very fast tempo would be too strenuous and close to the rate limit of unimanual tapping (Peters, 1980). The A tones in ABBABB ... sequences occur at a more comfortable pace, and musicians are able to synchronize with them as long as the IOI between A and B tones is longer than about 130 ms, even if the pitches of the tones are identical (Repp, 2007). This rate limit prevents the use of ABBABB ... sequences with very short IOIs. Moreover, shifted B tones may not generate a reliable PCR when the sequence tempo is too fast. In the present study, PCRs were clearly smaller at IOIs of 150 ms than at IOIs of 200 ms, and it is likely that they would have vanished at even shorter IOIs, thus causing a floor effect. Given these considerations, the present IOIs of 150 and 200 ms seemed a reasonable choice, and stream segregation was ensured by using a very large pitch separation and by confirming that participants were unable to gain any perceptual benefit from the presence of such widely separated A and B tones in a sequence. The use of piano tones rather than pure tones was motivated by a desire to make the results more relevant to music perception. The perceptual results of Repp (2009) and of the present study confirm that the principles of auditory stream segregation extend to complex harmonic tones with gradual offsets (damping).

The present results also revealed somewhat better detection performance and larger PCRs when the B tones were low than when they were high. When the pitch separation was large, Repp (2009) found somewhat better detection performance and larger PCRs when the A tones were low than when they were high; note that in the earlier paradigm the A tones triggered the PCRs because they contained the timing perturbations. Both results thus indicate that low tones were perceptually more salient than high tones. However, the reason for this may not have been pitch as such but the fact that the low piano tones had a longer decay of energy following their nominal offset than did the high piano tones, which may have increased their perceived loudness through temporal integration (Zwislocki, 1960).

One unexpected finding was that, at the fast tempo, tones in the irrelevant stream of tones (B tones in Repp, 2009; A tones here) interfered with conscious perception of timing changes. This does not invalidate the absence of perceptual facilitation as a criterion for the occurrence of stream segregation, and it may be taken as evidence for the increasing strength of stream segregation. It is paradoxical, however, that segregated streams begin to interact again when stream segregation increases. This interference effect requires further research.

It should be remembered that the present results and those of Repp (2009) were obtained with highly trained musicians. The author has tested a group of participants without any musical training in the task of Repp's Experiment 2 but obtained highly variable and therefore inconclusive results. It appears that the synchronization task is too difficult for most people who do not have music training.

In conclusion, by providing the first demonstration of an effect of auditory stream segregation on temporal action control in sensorimotor coordination, the present findings refute Repp's (2009) premature claim that sensorimotor coordination bypasses auditory scene analysis entirely. At the same time, however, the results confirm that the PCR in sensorimotor synchronization is strongly resistant to stream segregation: It remains unchanged at parameter settings that are more than sufficient to prevent perceptual integration of sequence elements into a single rhythm.

Original manuscript received 12 January 2009
Accepted revision received 10 June 2009
First published online 15 July 2009

REFERENCES

Jones, M. R., Jagacinski, R. J., Yee, W., Floyd, R. L., & Klapp, S. T. (1995). Tests of attentional flexibility in listening to polyrhythmic patterns. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 293–307.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford, UK: Oxford University Press.

Peters, M. (1980). Why the preferred hand taps more quickly than the non-preferred hand: Three experiments on handedness. *Canadian Journal of Psychology*, *34*, 62–71.

Repp, B. H. (2000). Compensation for subliminal timing perturbations in perceptual-motor synchronization. *Psychological Research*, *63*, 106–128.

Repp, B. H. (2005a). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review*, *12*, 969–992.

Repp, B. H. (2005b). Rate limits of on-beat and off-beat tapping with simple auditory rhythms: 2. The role of different kinds of accent. *Music Perception*, *23*, 167–189.

Repp, B. H. (2007). Perceiving the numerosity of rapidly occurring auditory events in metrical and non-metrical contexts. *Perception & Psychophysics*, *69*, 529–543.

Repp, B. H. (2008). Multiple temporal references in sensorimotor synchronization with metrical auditory sequences. *Psychological Research*, *72*, 79–98.

Repp, B. H. (2009). Segregated in perception, integrated for action: Immunity of rhythmic sensorimotor coordination to auditory stream segregation. *Quarterly Journal of Experimental Psychology*, *62*, 426–434.

Repp, B. H., & Knoblich, G. (2007). Toward a psychophysics of agency: Detecting gain and loss of control over auditory action effects. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 469–482.

Smeets, J. B. J., Brenner, E., de Grave, D. D. J., & Cuijpers, R. H. (2002). Illusions in action: Consequences of inconsistent processing of spatial attributes. *Experimental Brain Research*, *147*, 135–144.

Sussman, E., Wong, R., Horváth, J., Winkler, I., & Wang, W. (2007). The development of the perceptual organization of sound by frequency separation in 5–11-year-old children. *Hearing Research*, *225*, 117–127.

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven, The Netherlands.

Zwislocki, J. J. (1960). Theory of temporal auditory summation. *Journal of the Acoustical Society of America*, *32*, 1046–1060.