Ψ Psychology Press
Taylor & Francis Group

Short article

Segregated in perception, integrated for action: Immunity of rhythmic sensorimotor coordination to auditory stream segregation

Bruno H. Repp

Haskins Laboratories, New Haven, CT, USA

Auditory stream segregation can occur when tones of different pitch (A, B) are repeated cyclically: The larger the pitch separation and the faster the tempo, the more likely perception of two separate streams is to occur. The present study assessed stream segregation in perceptual and sensorimotor tasks, using identical ABBABB ... sequences. The perceptual task required detection of single phase-shifted A tones; this was expected to be facilitated by the presence of B tones unless segregation occurred. The sensorimotor task required tapping in synchrony with the A tones; here the phase correction response (PCR) to shifted A tones was expected to be inhibited by B tones unless segregation occurred. Two sequence tempi and three pitch separations (2, 10, and 48 semitones) were used with musically trained participants. Facilitation of perception occurred only at the smallest pitch separation, whereas the PCR was reduced equally at all separations. These results indicate that auditory action control is immune to perceptual stream segregation, at least in musicians. This may help musicians coordinate with diverse instruments in ensemble playing.

Keywords: Auditory scene analysis; Stream segregation; Streaming; Sensorimotor synchronization; Rhythm.

Dissociations between perception and action have been an important research topic since the seminal work of Milner and Goodale (1995) on the dorsal and ventral processing streams in vision. Conscious visual perception and visual guidance of action often rely on different kinds of sensory information, especially in the context of visual illusions (e.g., Aglioti, DeSouza, & Goodale, 1995; Carey, 2001; Ganel, Tanzer, & Goodale, 2008; Smeets & Brenner, 2006). There is less research on such perception–action dissociations in the auditory modality (though see Hafke, 2008; Loui, Guenther, Mathys, & Schlaug, 2008; Repp, 2000, 2001, 2006). Here, a new dissociation with regard to auditory scene analysis is described.

Auditory scene analysis

The general principles of auditory scene analysis are now well known, thanks to the pioneering research of van Noorden (1975) and Bregman (1990). In some of his experiments, van Noorden interleaved A–A–A–A–A–A– ... and B—B—B— ... sequences of tones such that they created a "galloping" rhythm (ABA–ABA–ABA– ...). He varied the pitch distance between the A and B tones as well as their interonset interval (IOI). When the pitch distance was only a few semitones (below the "fission boundary"), participants reported hearing a single stream of tones, regardless of IOI duration. When the pitch distance was larger, IOI and pitch distance jointly determined the likelihood of hearing one or two streams. The longer the IOI, the larger the pitch distance had to be in order for participants to report being unable to hear a single integrated stream. Between this "temporal coherence boundary" and the fission boundary, either one or two streams could be heard, with varying likelihoods.

Van Noorden (1975) also observed that stream segregation affects temporal discrimination. For example, he showed that a temporal displacement of all B tones relative to the A tones was increasingly difficult to detect as the pitch distance between the A and B tones was increased. More recently, Jones, Jagacinski, Yee, Floyd, and Klapp (1995) found that the temporal displacement of a single tone in a 3 : 2 polyrhythm was more difficult to detect when the two component sequences were widely separated in pitch. When tones are integrated into a single stream, the temporal displacement of a single tone can be perceived as a change in a short IOI between successive tones, whereas in segregated streams, a change in a longer within-stream IOI must be discerned. Because temporal discrimination approximately obeys Weber's law, according to which the just noticeable difference increases in proportion to interval duration, the latter discrimination is more difficult.

Winkler, Takegata, and Sussman (2005) used a similar logic to investigate the mismatch negativity (MMN) in auditory event-related brain potentials to omitted tones. They presented long sequences that could be perceived as either one or two streams and asked participants to report their percepts. Two distinct MMN components were observed: an early one that occurred regardless of the reported percept, and a later one that occurred only when an integrated percept was reported. On the basis of these results, they argued that there are two stages in auditory organization: an early integrated bottom-up stage and a later stage during which a conscious (potentially segregated) percept is formed under top-down influences.

Phase correction in sensorimotor synchronization

The present study compared a perceptual task (detecting timing deviations) with a sensorimotor task (synchronized tapping), using the same auditory sequences. To maintain synchrony of discrete movements such as tapping with an event sequence, phase correction is necessary (Repp, 2005b; Vorberg & Schulze, 2002). Typically, phase correction occurs automatically and without awareness. Introduction of a local phase perturbation in an otherwise isochronous sequence elicits a *phase correction response* (PCR) of the immediately following tap, a shift of the tap in the same direction as the perturbation but typically smaller. The PCR increases if the sequence IOI is increased (Repp, 2008a, 2008b).

Repp (2008a) showed that the PCR to a shifted tone is reduced when one or more subdivision tones of similar pitch are interpolated between the sequence tones. This occurs presumably because the subdivision tones (which are not shifted) serve as additional temporal references for the next tap. This effect was exploited in the present study.

In several previous studies the PCR, unlike perceptual judgements of sequence timing, has proven to be insensitive to local changes of pitch in a tone sequence (Repp, 2000, 2002, 2003). In a paradigm resembling the scenarios of auditory stream segregation, Repp (2003) interleaved two isochronous tone sequences whose pitches were either 3 or 20 semitones (st) apart. Participants had to

synchronize their taps with one sequence (the target) while ignoring the other (the distractor). The distractor caused systematic phase shifts or modulations in the sequence of taps, regardless of pitch separation. This was attributed to automatic phase correction, which appeared to be based on an integrated representation of the target and distractor sequences. Although it is remarkable that a 20-st separation had no effect at all in a task that encouraged perceptual segregation, the sequence tempi were slower than those typically employed in stream segregation studies; thus, the temporal coherence boundary (van Noorden, 1975) may not have been crossed.

The present study

The present study used ABBABB ... sequences (used previously by Sussman, Wong, Horváth, Winkler, & Wang, 2007, to investigate stream segregation) with three pitch separations and two IOI durations. Small temporal displacements of single A tones (*event onset shifts*; see Repp, 2002) were introduced into these otherwise isochronous sequences. In the perceptual task, participants had to detect these deviations, and the percentage of correct detections (hits) was the dependent variable. In the sensorimotor task, participants

had to tap in synchrony with the A tones, and the mean PCR to shifted tones was the dependent variable. Sequences in which the B tones were omitted were included as a baseline in both tasks.

It was expected that, at a small pitch separation of A and B tones, detection of temporal irregularities would be facilitated by the presence of B tones whereas the PCR in synchronized tapping would be reduced. At a larger pitch separation, sufficient for stream segregation, detection scores were expected to return to baseline. The question of interest was whether the PCR would do the same. Figure 1 gives a schematic illustration of the paradigm (see caption for explanation).

EXPERIMENT 1

Method

Participants
Nine paid volunteers (5 women, ages 22–28 years) and the author (age 63 years) participated. The paid participants were graduate students at the Yale School of Music who had studied their instrument(s) for 13 to 24 years. The author plays the piano at an advanced amateur level.
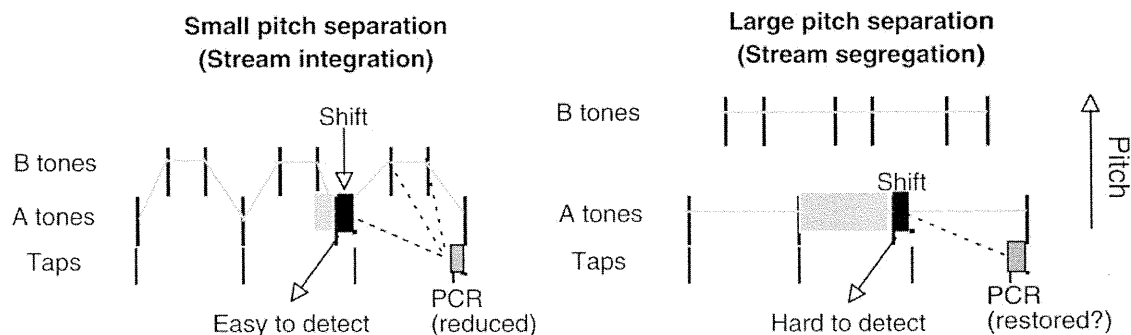
Figure 1. *Schematic diagram of the experimental paradigm. Vertical lines represent tones or taps, with time running from left to right. Vertical distance between A and B tones indicates pitch separation. Grey lines connect tones within the same perceptual stream. Black rectangle indicates (forward) shift of a tone. Dark grey rectangle indicates phase correction response (PCR) of the following tap. Dot marks position of an event before its shift. Light-grey rectangle indicates the complement of the temporal interval within which a shifted tone must be detected. Dotted lines indicate temporal references for tap placement. The PCR is "reduced" or "restored" relative to a baseline condition containing A tones only.*

Materials and equipment

Baseline sequences (A—A—A ...) consisted of 91 identical digital piano tones whose pitch was either C7 (2,093 Hz) or A#7 (3,729 Hz), with an IOI of either 450 ms ("fast") or 600 ms ("slow"). Composite (ABBABB ...) sequences likewise contained 91 A tones. B tones started after the 9th A tone, the IOI then being either 150 ms (fast) or 200 ms (slow). In sequences with a narrow (2 st) pitch separation, the pitches were E7 (2637 Hz) and F#7 (2960 Hz), with either the higher or the lower tones serving as A tones. In sequences with a wide (10 st) pitch separation, the pitches were C7 and A#7, again with either the lower or the higher tones serving as A tones. All tones were of equal nominal intensity (MIDI velocity) and had no specified duration but decayed rapidly.

Each sequence contained eight shifted A tones, one at each of the following shift magnitudes: -8%, -6%, -4%, -2%, 2%, 4%, 6%, and 8% of the IOI. Their position in the sequence was semi-random, with at least five unperturbed A tones intervening between two consecutive shifted tones and the first shift occurring on the 12th A tone at the earliest. The sequences were arranged into eight blocks of five sequences for each tempo condition, with different orders within each block. Each block contained four composite sequences, one of each type (narrow or wide pitch separation, low or high A tones), and one baseline sequence (low or high). Results for low and high baseline sequences were combined in all analyses because each type was presented only four times, and no effect of pitch was expected in those sequences.

Programs written in MAX 4.6.3 controlled the experiment. The tones were produced on a Roland RD-250s digital piano that was connected to the computer via a MOTU Fastlane MIDI interface. The sounds were heard over Sennheiser HD540 II headphones. Participants tapped on a Roland SPD-6 electronic percussion pad.

Design and procedure

Participants came for four sessions on different days. Sessions 1 and 3 required tapping; Sessions 2 and 4 were the corresponding perception tasks. Half of the participants did the fast tempo condition (Sessions 1–2) before the slow one (Sessions 3–4); the other participants were assigned the reverse order. Three tapping sessions had to be repeated later because a substantial number of taps had not been registered, due to insufficient tapping force.

In the tapping sessions, participants started each sequence by pressing the space bar, started tapping with the third tone, and continued tapping in synchrony with the A tones while ignoring any B tones and any deviations from temporal regularity in the sequence. In the perception sessions, they listened to the identical sequences and pressed the down-arrow key as quickly as possible whenever they heard any deviation from temporal regularity. Response latencies were measured from the onset of each shifted A tone.

Results

Perception

Responses with latencies between 200 and 1,200 ms were considered hits; others were false alarms. The mean number of false alarms per sequence was about 0.25, except for a higher rate in slow baseline sequences (0.79) that was mainly due to a single participant. Given that only about 1 in 10 A tones was a detection target, the probability of a false alarm being classified as a hit was quite low. Some responses classified as false alarms, however, were probably hits with long latencies.

Hit percentages naturally increased with the magnitude of the shift. They ranged from an average 9.7% for shifts of $\pm 2\%$ to 93.2% for shifts of $\pm 8\%$. The mean hit percentages in the different conditions are shown in Figures 2A and 2B. As predicted, correct detection responses increased beyond baseline levels at the narrow pitch separation (2 st), but they dropped back to baseline at the wide pitch separation (10 st), at least at the faster tempo. Collapsing over the pitch variable (LH and HL in the figure), a 2×3 repeated measures analysis of variance (ANOVA) with the variables of tempo (fast,
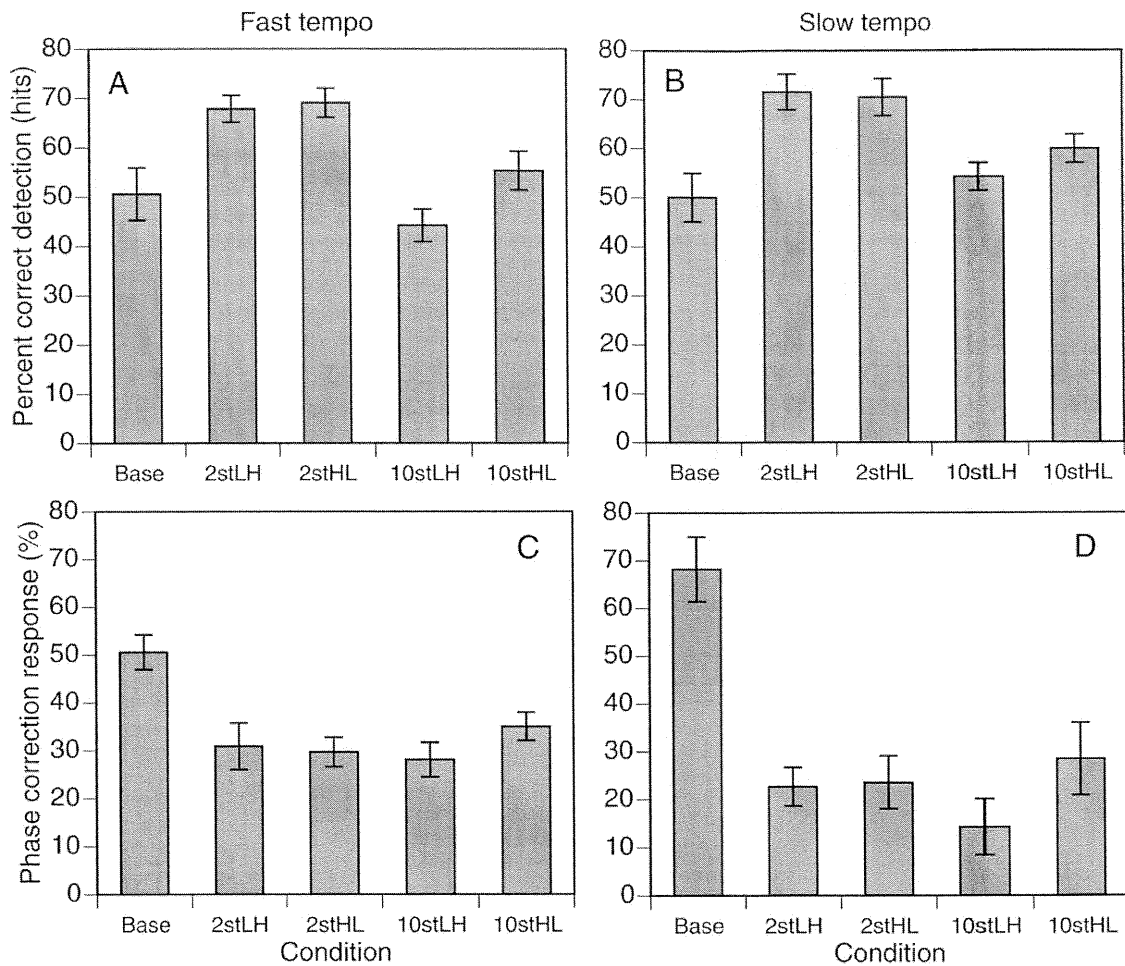
Figure 2. *Results for the perception task (Panels A, B) and the tapping task (Panels C, D) of Experiment 1. Base = baseline. Condition names indicate pitch separation (2 st or 10 st) and pitch of A and B tones (LH = low–high; HL = high–low).*

slow) and condition (baseline, 2 st, 10 st) yielded a highly reliable main effect of condition, $F(2, 18) = 33.17$, $p < .001$, as well as a significant interaction, $F(2, 18) = 4.42$, $p = .031$ (Greenhouse–Geisser corrected). To unpack the interaction, two separate 2×2 ANOVAs were conducted. Comparing the baseline with the 2-st condition, there was only a main effect of condition, $F(1, 9) = 40.09$, $p < .001$, due to better performance in the 2-st condition, and no interaction. Comparing the baseline with the

10-st condition, there was no significant main effect of condition, which indicates a return of performance to baseline levels, $F(1, 9) = 1.34$, $p = .277$. However, there was a significant interaction, $F(1, 9) = 7.32$, $p = .024$, due to better performance in the 10-st condition at the slow than at the fast tempo.

In another ANOVA ($2 \times 2 \times 2$), the baseline condition was excluded, and pitch (LH, HL) was included as a third variable. Besides the obvious main effect of condition, $F(1,$

9) = 67.30, $p < .001$, there was a main effect of pitch, $F(1, 9) = 13.28$, $p = .005$, as well as a Condition × Pitch interaction, $F(1, 9) = 14.22$, $p = .004$: In the 10-st condition detection performance was better when A tones were high (HL) than when they were low (LH), and this difference was larger at the fast than at the slow tempo. The main effect of tempo was marginally significant, $F(1, 9) = 5.33$, $p = .046$. The Tempo × Condition interaction was not significant, $F(1, 9) = 2.73$, $p = .133$.

Tapping

Figures 2C and 2D show the mean PCR (expressed as a percentage of event onset shift magnitude) in the various conditions. As predicted, the PCR was reduced in composite sequences relative to the baseline condition, and this difference was more pronounced at the slow tempo; this replicates findings of Repp (2008a). In a 2 × 3 ANOVA, the main effect of condition was highly reliable, $F(2, 18) = 34.26$, $p < .001$, as was the Tempo × Condition interaction, $F(2, 18) = 17.73$, $p = .001$. However, a 2 × 2 × 2 ANOVA on the 2-st and 10-st conditions, with the baseline omitted, revealed no main effect of condition, $F(1, 9) = 0.16$, $p = .906$. Thus, there was no indication of any return of the PCR toward baseline levels at the wide pitch separation, at either tempo. Despite a suggestion of an effect of pitch (LH vs. HL) in 10-st sequences, no effect in a 2 × 2 × 2 ANOVA reached significance.

EXPERIMENT 2

Experiment 1 revealed that, while perception returns to near-baseline levels at a 10-st pitch separation, especially at the fast tempo, the PCR remains strongly reduced at a 10-st pitch separation, regardless of tempo. Is it possible that the PCR is truly immune to stream segregation? Experiment 2 addressed this question by using an extreme pitch separation, 48 st. A perception task was included as well.

Method

Participants

Nine of the 10 participants of Experiment 1 returned.

Materials and equipment

These were the same as those in Experiment 1, except for the following differences. In baseline sequences, the pitch of the tones was either C3 (131 Hz) or C7 (2,093 Hz). Composite sequences were made up of these same tones (a 48-st pitch separation), with either the low or the high tones serving as target (A) tones. Because of the slow decay of low piano tones, a nominal duration of 40 ms was specified for all tones in the MIDI instructions. (The actual duration was slightly longer because of damped decay.) Sequences were arranged into eight blocks of six randomly ordered trials each (one baseline sequence and two composite sequences at each of two tempi). Unlike Experiment 1, the tempi were mixed here rather than blocked.

Design and procedure

Participants came for two sessions, with the perception task following the tapping task. The procedure was the same as that in Experiment 1.

Results

Perception

Hits were defined as previously. The mean false-alarm rate per sequence was 0.32. The mean hit percentages are shown in Figures 3A and 3B. It is clear that there was no facilitation of detection at the 48-st pitch separation; on the contrary, there was some interference from B tones, especially at the fast tempo. Accordingly, a 2 × 2 ANOVA comparing the baseline with the 48-st condition (collapsing over pitch) yielded a significant main effect of condition, $F(1, 8) = 9.84$, $p = .014$, and a Tempo × Condition interaction, $F(1, 8) = 12.65$, $p = .007$. There was also more interference when the B tones were low (HL) than when they were high (LH). (Low tones may have been perceived as louder because they
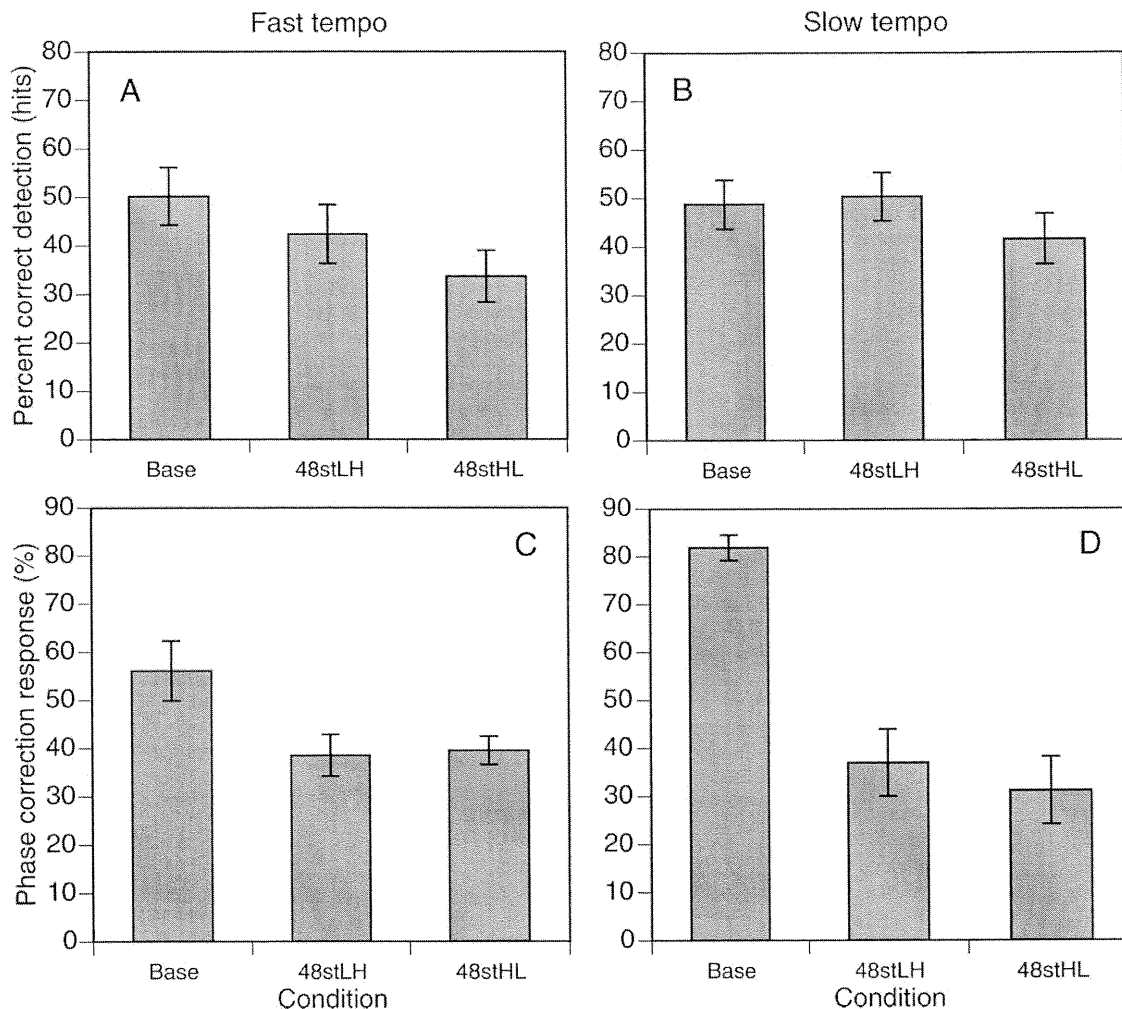
Figure 3. *Results for the perception task (Panels A, B) and the tapping task (Panels C, D) of Experiment 2. Base = baseline. Condition names indicate pitch separation (48 st) and pitch of A and B tones (LH = low–high; HL = high–low).*

had a much slower decay than the high tones and thus delivered more energy during their short duration.) A 2×2 ANOVA on the 48-st condition alone revealed significant main effects of tempo, $F(1, 8) = 11.64$, $p = .009$, and of pitch, $F(1, 8) = 13.20$, $p = .007$.

Tapping

The PCR results are shown in Figures 3C and 3D. Remarkably, they were quite similar to the results of Experiment 1. The PCR was reduced in composite

sequences, and much more so at the slow than at the fast tempo. A 2×2 ANOVA comparing the baseline with the 48-st condition yielded a significant main effect of condition, $F(1, 8) = 30.67$, $p = .001$, and a significant Tempo $\times$ Condition interaction, $F(1, 8) = 11.39$, $p = .010$. A 2×2 ANOVA on the 48-st condition alone did not yield any significant effects.

In a final $2 \times 2 \times 2$ ANOVA, the baseline and 2-st conditions of Experiment 1 were compared with the baseline and 48-st conditions of

Experiment 2, with the variables being experiment, tempo, and condition (baseline vs. composite). As expected, there was a significant main effect of condition, $F(1, 8) = 48.26$, $p < .001$, as well as a Tempo $\times$ Condition interaction, $F(1, 8) = 17.43$, $p = .003$. In addition, there were main effects of tempo, $F(1, 8) = 9.61$, $p = .015$, and experiment, $F(1, 8) = 7.25$, $p = .027$. Importantly, however, the Experiment $\times$ Condition interaction was quite negligible, $F(1, 8) = 0.06$, $p = .821$. This demonstrates that the reduction of the PCR was just as large at the 48-st separation as at the 2-st separation. This was true at both tempi, as the triple interaction was likewise far from significance, $F(1, 8) = 0.81$, $p = .395$.

DISCUSSION

The present results demonstrate that perception of temporal deviations depends strongly on the pitch difference between the A and B tones, whereas the PCR in synchronized tapping does not. How should this be explained?

One theoretical possibility is that perceptual stream segregation did not occur in the tapping task. However, this is highly implausible because the task demands actually favoured segregation in tapping (the instructions were to tap in synchrony with the A tones and ignore the B tones), whereas they encouraged integration in perception (because integration facilitates detection of temporal irregularities). Moreover, even though direct perceptual judgements of stream segregation were not obtained in this study, extrapolation from the data of van Noorden (1975) and Sussman et al. (2007) suggests that the 48-st pitch separation, if not the 10-st separation, was well beyond the temporal coherence boundary.

It could be argued that, instead of focusing their attention on the A tones and tapping in phase with them, participants focused on the B tones and tapped in (generalized) antiphase with them (i.e., in the gap between successive pairs of B tones). Such a strategy might account for a reduced influence of shifted A tones on the PCR, but it too is implausible because it is contrary to instructions,

and because tapping in antiphase with pairs of B tones is quite difficult at the tempi used here (Repp, 2005a), whereas tapping in phase with A tones is relatively simple. However, it is undoubtedly true that participants were tapping de facto in antiphase with the B tones at the same time that they were tapping in phase with the A tones. This observation holds the key to what seems to be the correct explanation: A and B tones both served as temporal references in synchronization and jointly affected the PCR, despite their large pitch difference. In other words, pitch differences seem to be irrelevant to coordination of action with external rhythms; the relevant perceptual information is contained solely in the timing of event onsets. In effect, in the synchronization task two perceptually segregated streams become *functionally integrated* through their joint effects on action timing.

Thus, an integrated, purely temporal representation of rhythm for action control may exist in the brain simultaneously with a multidimensional, potentially segregated representation for conscious perception. The intriguing electrophysiological results by Winkler et al. (2005), mentioned in the Introduction, also suggest this possibility, which is furthermore consistent with emerging evidence for separate auditory processing streams serving conscious perception and action control (e.g., Arnott, Binns, Grady, & Alain, 2004; Thaut, Kenyon, Schauer, & McIntosh, 1999).

One important qualification of the present results is that they were obtained with highly trained musicians. It remains to be seen whether similar results can be obtained from nonmusicians, who may find the tapping task challenging. Interestingly, a single participant (the author, highly practised in synchronization tasks) did show a return of the PCR to baseline at the 48-st separation. This raises questions regarding the roles of age, musical training, and task experience, which require further investigation.

In conclusion, the present results suggest that auditory stream segregation is irrelevant to musicians' rhythmic sensorimotor coordination. Their coordination seems to rely simultaneously on all available sources of auditory rhythmic stimulation,

whether perceptually segregated or not. Such functional integration may be useful in ensemble performance with diverse instruments. Whether functional integration extends to situations in which distractor events are nonrhythmic or exhibit radical timbre differences remains to be explored.

REFERENCES

Aglioti, S., DeSouza, J. F., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology, 5*, 679–685.

Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *NeuroImage, 22*, 401–408.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA: MIT Press.

Carey, D. P. (2001). Do action systems resist visual illusions? *Trends in Cognitive Sciences, 5*, 109–113.

Ganel, T., Tanzer, M., & Goodale, M. A. (2008). A double dissociation between action and perception in the context of visual illusions: Opposite effects of real and illusory size. *Psychological Science, 19*, 221–225.

Hafke, H. Z. (2008). Nonconscious control of fundamental voice frequency. *Journal of the Acoustical Society of America, 123*, 273–278.

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.

Loui, P., Guenther, F. H., Mathys, C., & Schlaug, G. (2008). Action-perception mismatch in tone-deafness. *Current Biology, 18*, R331–R332.

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

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

Repp, B. H. (2001). Phase correction, phase resetting, and phase shifts after subliminal timing perturbations in sensorimotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 600–621.

Repp, B. H. (2002). Automaticity and voluntary control of phase correction following event onset shifts in sensorimotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 410–430.

Repp, B. H. (2003). Phase attraction in sensorimotor synchronization with auditory sequences: Effects of single and periodic distractors on synchronization accuracy. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 290–309.

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

Repp, B. H. (2006). Does an auditory perceptual illusion affect on-line auditory action control? The case of (de)accentuation and synchronization. *Experimental Brain Research, 168*, 493–504.

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

Repp, B. H. (2008b). Perfect phase correction in sensorimotor synchronization with slow auditory sequences. *Journal of Motor Behavior, 40*, 363–367.

Smeets, J. B., & Brenner, E. (2006). 10 years of illusions. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1501–1504.

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.

Thaut, M. H., Kenyon, G. P., Schauer, M. L., & McIntosh, G. C. (1999). The connection between rhythmicity and brain function. *IEEE Engineering in Medicine and Biology, 18*, 101–108.

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.

Vorberg, D., & Schulze, H.-H. (2002). Linear phase-correction in synchronization: Predictions, parameter estimation, and simulations. *Journal of Mathematical Psychology, 46*, 56–87.

Winkler, I., Takegata, R., & Sussman, E. (2005). Event-related brain potentials reveal multiple stages in the perceptual organization of sound. *Cognitive Brain Research, 25*, 291–299.