

Bruno H. Repp · Amandine Penel

Rhythmic movement is attracted more strongly to auditory than to visual rhythms

Received: 4 February 2003 / Accepted: 24 June 2003 / Published online: 3 September 2003
© Springer-Verlag 2003

Abstract People often move in synchrony with auditory rhythms (e.g., music), whereas synchronization of movement with purely visual rhythms is rare. In two experiments, this apparent attraction of movement to auditory rhythms was investigated by requiring participants to tap their index finger in synchrony with an isochronous auditory (tone) or visual (flashing light) target sequence while a distractor sequence was presented in the other modality at one of various phase relationships. The obtained asynchronies and their variability showed that auditory distractors strongly attracted participants' taps, whereas visual distractors had much weaker effects, if any. This asymmetry held regardless of the spatial congruence or relative salience of the stimuli in the two modalities. When different irregular timing patterns were imposed on target and distractor sequences, participants' taps tended to track the timing pattern of auditory distractor sequences when they were approximately in phase with visual target sequences, but not the reverse. These results confirm that rhythmic movement is more strongly attracted to auditory than to visual rhythms. To the extent that this is an innate proclivity, it may have been an important factor in the evolution of music.

or deliberate foot tapping, body swaying, or dancing. Spontaneous sensorimotor synchronization with music, songs, or speech has been observed intermittently in infants and toddlers (Condon & Sander, 1974; Moog, 1976, cited in Hargreaves, 1986; Trevarthen, 1999–2000), and young children's ability to synchronize deliberate actions with a musical beat increases rapidly with age (Drake, 1997; Drake, Jones, & Baruch, 2000; Fraise, Pichot, & Clairouin, 1949; Rainbow & Owen, 1979). By contrast, people rarely move in synchrony with rhythms defined by purely visual stimuli, certainly not spontaneously. For example, people are not likely to tap their foot when they observe musicians playing or people dancing on a muted television set. Rhythmic visual sequences are less often encountered in daily life than auditory rhythmic sequences such as music, and people rarely seek them out for pleasure. Sound rather than light is the preferred medium for rhythmic stimulation. One reason why music has evolved in the auditory rather than the visual modality may be that auditory rhythms inevitably engage our body more than visual rhythms do (Fraise, 1948).

There is considerable evidence that human sensitivity to purely temporal information is greater in the auditory than in the visual modality. Psychophysical experiments have demonstrated that temporal discrimination is poorer in vision than in audition (e.g., Goodfellow, 1934; Grondin, 1993; Grondin, Meilleur-Wells, Ouellette, & Macar, 1998; Grondin, Ouellet, & Roussel, 2001; Grondin & Rousseau, 1991; Rousseau, Poirier, & Lemyre, 1983). Also, discrimination and reproduction of temporal patterns are superior in the auditory modality (Gault & Goodfellow, 1938; Glenberg & Jona, 1991; Glenberg, Mann, Altman, Forman, & Procise, 1989). When auditory and visual stimuli are in conflict with respect to their number, timing, duration, or rate, observers' judgments are typically more strongly influenced by the auditory than by the visual temporal information (Goldstone, Boardman, & Lhamon, 1959; Fendrich & Corballis, 2001; Morein-Zamir, Soto-Faraco, & Kingstone, 2003; Recanzone, 2003; Shams, Kamitani, & Shimojo, 2000; Walker & Scott, 1981).

Introduction

Background

When people listen to music, they readily move in synchrony with its beat. This may be evident in spontaneous

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

A. Penel
Cold Spring Harbor Laboratory,
Cold Spring Harbor, NY 11724, USA

This asymmetry extends to the temporal control of action. When participants are required to tap in synchrony with isochronous auditory or visual sequences (tones vs. flashes) of equal tempo, the variability of the asynchronies between taps and sequence events is much greater with the visual sequences (Fraisse, 1948; Kolers & Brewster, 1985; Repp & Penel, 2002). However, when participants are instructed to synchronize their taps with flashes that are accompanied by tones that are to be ignored, the variability of the asynchronies is as low as in tapping to the tones alone (Chen, Repp, & Patel, 2002; Repp & Penel, 2002). This indicates that the motor activity is controlled by the auditory input, even when attention is focused on the visual modality. Repp and Penel (2002) also found that the automatic phase correction process triggered by phase perturbations in simultaneous auditory and visual sequences responds more strongly to auditory than to visual perturbations.

The present study

The previous research on sensorimotor synchronization suggests that rhythmic motor behavior is indeed attracted more strongly to auditory than to visual sequences. However, the experiments sampled only a limited range of conditions. In order to achieve greater generality, we conducted two new experiments that extended the previous research in a number of ways. Specifically, as elaborated below, we used a slower sequence tempo, required participants to attend to one or the other modality, and varied the phase relationship between auditory and visual sequences. In Experiment 1, we also manipulated the temporal structure of the sequences, and in Experiment 2 we varied the relative salience of the two modalities. Between experiments, we manipulated the spatial congruence of visual and auditory stimuli, as well as the presence or absence of auditory feedback from the taps. Experiment 1 constituted a detailed investigation with attention to individual differences, whereas Experiment 2 was a more compact study that addressed specific methodological issues. The main question throughout was whether auditory distractors affect synchronization with visual target sequences more than visual distractors affect synchronization with auditory target sequences.

Sequence tempo

Repp and Penel (2002) and Chen et al. (2002) used isochronous sequences with an event inter-onset interval (IOI) of 488 ms.¹ In both studies, some participants had difficulty synchronizing with visual sequences and had to be excluded. It was discovered subsequently (Repp, in press) that the mean “synchronization threshold”

¹ The IOI was described nominally as 500 ms in Repp and Penel (2002), with a footnote pointing out that it was in fact 2.4% shorter.

(defined as the IOI at which people are able to synchronize successfully in 50% of trials, each comprising 48 taps) is at the surprisingly high IOI value of about 450 ms for sequences of isochronous light flashes. At shorter IOIs, attempts to synchronize usually result in linear phase drift, indicating lack of sensorimotor coordination. By contrast, the mean synchronization threshold for auditory sequences is at an IOI of about 120 ms (see also Bartlett & Bartlett, 1959). Thus, the visual sequences of the previous studies were dangerously close to the synchronization threshold and hence were likely to give rise to unstable behavior, even in those participants who had few unsuccessful trials. Seen in this light, the observed auditory bias seems less surprising. The present experiments used visual and auditory sequences with a constant IOI of 625 ms, a value that is well above the visual synchronization threshold of most participants. Nevertheless, it was expected that variability would be much higher in tapping to visual than to auditory sequences, and that a strong bias toward the auditory modality would persist in a bimodal conflict situation.

Focus of attention

The participants in the studies of Repp and Penel (2002) and Chen et al. (2002) were told to focus their attention always on the visual modality when visual and auditory sequences were presented simultaneously. Although dominance of the auditory modality in such a condition is a strong finding, it is conceivable that visual sequences would have an equally distracting effect if attention were directed to the auditory modality. Therefore, the present experiments included both attention conditions. In Experiment 1, we trusted our participants to follow instructions when told to keep looking at visual distractors. In Experiment 2, however, we obtained tangible evidence of visual monitoring by omitting single flashes in some trials and requiring participants to report these omissions. Even so, we expected effects of visual distractors to be much smaller than those of auditory distractors.

Relative phase

The previous studies used visual and auditory sequences that were exactly in phase when presented simultaneously.² The present experiments used a more elaborate target/distractor paradigm (Repp, 2003) in which sequences of equal tempo were presented at a number of

² After Experiment 1 had been completed, two studies in which relative phase was varied came to our attention. Peryer, Sloboda, and Nte (2002) used isochronous visual target sequences in combination with faster or slower auditory distractor sequences. The auditory sequences significantly increased the variability of synchronization performance. Aschersleben and Bertelson (Gisa Aschersleben, personal communication, June 2002) have conducted an experiment apparently similar to ours. However, we have not yet seen a written report.

different cross-modal phase relationships. It was expected that, regardless of relative phase, auditory distractors would have larger effects than visual distractors on both the mean asynchronies between taps and target stimuli and on their variability. However, these effects were also expected to vary systematically as a function of relative phase. One reason is that the attraction exerted by a distractor sequence is likely to decrease if the temporal distance between target and distractor events exceeds about 100 ms, the temporal window proposed for auditory-visual integration (Lewald, Ehrenstein, & Guski, 2001; Lewald & Guski, 2003; Shams et al., 2000; Slutsky & Recanzone, 2001).³ Mean asynchronies were expected to vary positively with the relative phase of the distractor sequence in the vicinity of zero, being more negative in the presence of leading distractors (negative relative phases) and less negative or more positive in the presence of lagging distractors (positive relative phases). The magnitude of this systematic variation, a measure of the degree of attraction to the distractor sequence, was expected to be greater for auditory than for visual distractors. At larger relative phases (those corresponding to absolute temporal separations larger than 100 ms), however, this attraction was expected to decrease, and the mean asynchrony was expected to eventually revert to a value similar to that obtained at zero relative phase. Thus, we expected mean asynchrony to vary as a roughly sinusoidal function of relative phase (or, rather, temporal separation). This hypothesis was examined closely only in Experiment 1, as Experiment 2 used a restricted range of relative phases.

As regards the variability of the asynchronies between taps and target events, the strongest auditory distractor effect was expected to occur at zero relative phase, where Repp and Penel (2002) and Chen et al. (2002) had found complete dominance of the auditory modality. This effect was expected to wear off as the relative phase deviated from zero in either direction. Because the variability of the asynchronies was expected to be larger for unimodal visual than for unimodal auditory sequences, auditory distractors would have the effect of reducing variability near zero relative phase (a U-shaped function), whereas visual distractors would either have no effect or would increase variability near zero relative phase (an inverted U-shaped function).

Temporal modulation

Experiment 1 also included a condition in which the sequences were temporally irregular, such that the

³ That is, we were prepared to interpret effects of relative phase as effects of absolute temporal distance. For reasons of convenience, however, we used normalized relative phase (ranging from -0.5 to 0.5) rather than temporal separation in milliseconds as the independent variable. Whether relative phase as such plays a role in governing the magnitude of distractor effects is a question we did not address directly in the present study because this would have required varying the IOI duration, resulting in an excessively large design.

deviations from isochrony went in opposite directions in the target and distractor sequences. By generalizing Repp and Penel's (2002) technique of introducing conflicting local phase perturbations in simultaneous auditory and visual sequences, this condition provided a third measure of auditory bias or dominance. It is known that attempts to synchronize with a sequence containing unpredictable temporal variation results in tracking behavior. The inter-tap intervals (ITIs) mimic the IOIs at a lag of one interval, so that there is a positive lag-1 cross-correlation (Michon, 1967; Repp, 2002b). Because the distractor sequence had the inverse temporal pattern of the target sequence, an influence of the distractor would be manifest in a reduction or even reversal of the lag-1 cross-correlation between the ITIs and the IOIs of the target sequence. This influence was expected to be more pronounced in the case of auditory distractors than in the case of visual distractors. It was also expected to be largest at zero relative phase and to decrease with increases in the absolute temporal separation between target and distractor sequences (i.e., a U-shaped function of relative phase).

Stimulus salience

When the relative intensities of visual and auditory stimuli are not controlled, as was the case in our Experiment 1 and in most previous studies, it is possible that the tones are more salient perceptually than the flashes and therefore attract participants' attention. In Experiment 2, we first adjusted the loudness of the tones until their subjective salience seemed to match that of the flashes, and then presented the tones at that level and also at a lower level. If relative stimulus salience plays a role, auditory dominance should be reduced with softer tones.

Spatial congruence

Recent studies by Spence and colleagues (Spence, Shore, & Klein, 2001; Spence, Baddeley, Zampini, James, & Shore, 2003) have drawn attention to the fact that many studies involving auditory and visual stimuli (except those concerned specifically with spatial localization) have presented these stimuli in different locations, with the auditory stimuli typically being presented over headphones, as in our Experiment 1. Some studies have found no effect of spatial disparity on temporal judgments (Lewald & Guski, 2003; Recanzone, 2003; Shams et al., 2000), while others have found such effects (Harrington & Peck, 1998; Spence et al., 2001, 2003). Although spatial separation should make it easier to attend selectively to either modality and thus could not account for the effectiveness of auditory distractors, it could be cited as a reason for the ineffectiveness of visual distractors. In Experiment 2, we opted for spatial congruence of auditory and visual stimuli. This was expected to result in stronger cross-modal integration and

hence in stronger cross-modal distractor effects. However, we did not expect it to change the relative dominance of the auditory modality.

Auditory feedback

It is common for taps to produce a noise upon impact with a hard surface. This was also the case in our Experiment 1. The resulting intra-modal tap-tone asynchronies may well provide more accurate error feedback than the cross-modal tap-flash asynchronies, and this may cause a bias in favor of the auditory modality. To remove this bias, Experiment 2 used a quiet tapping device, so that all asynchronies were cross-modal (kinesthetic-auditory and kinesthetic-visual respectively).⁴ We still expected to find auditory dominance.

Experiment 1

Method

Design

The experiment consisted of four sessions that resulted from the combination of two attention conditions and two temporal structure conditions. In Sessions 1 and 2, the sequences were isochronous, whereas in Sessions 3 and 4 they were irregular (jittered). In Sessions 1 and 3, the targets were visual and the distractors were auditory, whereas in Sessions 2 and 4 the targets were auditory and the distractors were visual. The fixed order of the attention conditions was motivated by the thought that visual distractors would be more likely to have an effect if they had first served as targets. The conditions with irregular sequences constituted a replication and extension of the conditions with regular sequences, and therefore followed them. In each session, there were five blocks of 24 trials.

Participants

Five paid volunteers (4 women) aged 18–25 and the first author (B.R., age 57) participated.⁵ All had considerable experience in synchronization tasks, as well as substantial musical training (at least 6 years), which enabled them to tap with low variability.

Materials and equipment

All sequences were produced from pre-assembled MIDI (musical instrument digital interface) instructions under the control of a Macintosh Quadra 660AV computer, using a program written in MAX, a graphic programming language for MIDI applications.

Auditory sequences consisted of identical high-pitched (C8, 4168 Hz) synthetic piano tones (“pings”) of equal intensity, with-

out any nominal duration (i.e., “note offset”) specified in the MIDI instructions. The tones had sharp attacks that included key impact noise and decayed within about 100 ms. They were produced on a Roland RD-250 s digital piano via a MIDI interface and were heard over Sennheiser HD540 II headphones.

Visual sequences consisted of flashes emitted by the “out” indicator of an Opcode II MIDI Translator box, a circular green light-emitting diode 3 mm in diameter which responded to MIDI messages. Because the flashes were difficult to see in bright light, the experiment was conducted in the dark. The lights of the windowless room were switched off, and the computer screen was dimmed and covered with a piece of cardboard, leaving a gap at the bottom that provided dim illumination for the computer keyboard. The box with the flashing light was placed on the rear edge of the keyboard. Under these conditions, the flashes were clearly visible and separated from each other. MIDI instructions consisting exclusively of “note on” messages were used to activate the light.

Target sequences consisted of 32 events, and distractor sequences of 27 events. Each trial began with five target events. The distractor sequence, if any, started in the vicinity of the sixth target event at a particular temporal separation, which was maintained until the end of the trial. In Sessions 1 and 2, all sequences were isochronous with a nominal IOI of 640 ms.⁶ The following nominal temporal displacements between target and distractor sequences were used, with the corresponding normalized relative phases (ranging from $-.5$ to $.5$) given in parentheses: 0 ms (0), ± 10 ms ($\pm .015625$), ± 20 ms ($\pm .03125$), ± 40 ms ($\pm .0625$), ± 80 ms ($\pm .125$), ± 120 ms ($\pm .1875$), ± 160 ms ($\pm .25$), ± 200 ms ($\pm .3125$), ± 240 ms ($\pm .375$), ± 280 ms ($\pm .4375$), and 320 ms ($.5$, equivalent to $-.5$). In one block of 24 trials, each of these 20 relative phases occurred once; the remaining 4 trials were control (baseline) trials without any distractor sequence. The 24 trials occurred in different random orders in each of five blocks.

The only difference between Sessions 3 and 4 and the preceding two sessions was that the sequences were temporally modulated (jittered), starting with the seventh target event and the associated (second) distractor event. Five fixed patterns of deviations, one for each block of trials, were generated by generating a random series of numbers ranging from -25 ms to $+25$ ms and normalizing it to have a mean of zero. Within each block, all target sequences had the same pattern of deviations (obtained by adding the series of random numbers to the onset times), whereas all distractor sequences had the inverse pattern (obtained by subtracting the same series of numbers from the onset times). The average IOI of each jittered pattern was thus exactly 640 ms, whereas the standard deviation of the IOIs ranged from 16 to 23 ms.⁷ The phase relationship between a target sequence and a distractor sequence was defined by the initial distractor event relative to the sixth target event, both of which were unperturbed. Although phase relationships were not constant thereafter because of the temporal modulations, their average values within trials were the same as in the isochronous sequences of Sessions 1 and 2.

Procedure

Sessions were typically 1 week apart and were self-paced, lasting about 1 h. Each trial block was started by the participant clicking the mouse; thereafter, successive trials were started (with a 2-s

⁴ We are assuming that peripherally visible movements of the finger contribute little to error correction in synchronization. If they do play a role, this would work in favor of the visual modality.

⁵ The data of two additional participants were excluded, in one case (not analyzed) because of recording artifacts and in the other case because of atypical and highly variable results that were difficult to make sense of. For example, that participant tapped after rather than before visual stimuli (i.e., with positive asynchronies), which is quite unusual.

⁶ It is known from earlier waveform measurements that the output is 2.4% faster than specified in MAX (cf. Footnote 1). Thus, the actual IOI was 625 ms. (Note that relative phases are independent of this scaling factor.) All millisecond values reported in the Method and Results sections are nominal, as specified or recorded by the MAX program. Actual values were 2.4% shorter. Apart from this scaling factor, input and output timing are believed to be accurate within ± 1 ms.

⁷ According to the first author’s informal impressions, the deviations from isochrony were detectable but not particularly salient in auditory sequences. They were undetectable in visual sequences (cf. Repp Penel, 2002).

delay) by pressing the space bar of the computer keyboard. Data were saved during breaks between blocks.

Participants sat in front of the Macintosh computer in the dark room, listened to the auditory sequences over the headphones at a comfortable intensity, watched the flashing light, and tapped with their right index finger (all participants were right-handed) on a Roland SPD-6 electronic percussion pad which they held on their lap. The impact of the finger on the rubber pad produced an audible thud whose loudness varied in proportion to the tapping force. (No digital sound output from the percussion pad was heard.) Participants were instructed to start tapping with the third tone of each target sequence and to continue tapping in synchrony with the target events until the end of each trial while ignoring the distractor events. It was emphasized that closing one's eyes or looking away from visual distractors would defeat the purpose of the experiment, and this was well understood by the participants.

Results

All results are based on the taps to target events 9–32 (taps 7–30); that is, taps to the initial lead-in targets, to the first two targets accompanied by distractors, and to the last target were excluded. Because the number of participants was small and there were some interesting individual differences, the results are presented for individual participants.

Unimodal control trials

In each session, there were 20 control trials in which the target sequence occurred alone. The results from these trials give an indication of differences between unimodal visual and auditory conditions in terms of mean asynchronies and mean within-trial standard deviations of asynchronies (referred to simply as variability in the following). Statements about the significance of within-participant differences are based on the criterion of non-

overlapping single-standard-error confidence intervals (about $p < .05$).

The mean asynchronies are shown in Table 1A. Most of them were negative, which reflects the well-known anticipation tendency in sensorimotor synchronization (see, e.g., Aschersleben, 2002), although in some cases there was no anticipation tendency. For 5 of the 6 participants (all but A.M.), the anticipation tendency was significantly greater with visual than with auditory isochronous sequences (Sessions 1 and 2), which agrees with results obtained previously for the same IOI (Repp, in press). However, only 3 participants (H.R., S.V., V.T.) continued to show this significant difference with jittered sequences (Sessions 3 and 4), and A.M. now showed a significant difference in the opposite direction. Unexpectedly, all participants showed significant differences between Sessions 1 and 3 (visual targets), but for 3 the difference went in one direction, and for 3 it went in the other direction. The average asynchronies for auditory targets were more consistent across sessions; nevertheless, 2 participants (S.V., V.T.) showed significantly more negative, and 1 (A.M.) showed significantly less negative, asynchronies in Session 4 than in Session 2. These results suggest that mean asynchrony is not a very reliable measure of modality differences.

Much more consistent modality differences were obtained in terms of variability, as shown in Table 1B. As expected, variability was much greater in synchronization with visual than with auditory targets, and this difference was highly reliable for all participants, for both isochronous and modulated sequences. (Note also in Table 1A that all participants showed larger between-trial standard errors for visual than for auditory sequences.) Furthermore, and not surprisingly, all participants showed numerically larger variability when the sequences were temporally modulated than when they

Table 1 Results for unimodal control trials in Experiment 1. Standard errors are in parentheses. *IOI* inter-onset interval, *ITI* inter-tap interval

Participant	Session 1 Visual target	Session 2 Auditory target	Session 3 Visual target	Session 4 Auditory target
A. Mean asynchronies				
A.M.	-15 (7.1)	-21 (1.9)	1 (4.1)	-13 (1.4)
B.R.	-53 (3.2)	-38 (2.4)	-39 (3.7)	-43 (2.9)
H.R.	-23 (6.5)	3 (1.6)	-51 (9.1)	0 (2.2)
J.S.	-69 (5.8)	-47 (3.8)	-50 (5.5)	-53 (5.3)
S.V.	-35 (9.8)	3 (1.9)	-60 (6.4)	-24 (4.1)
V.T.	-76 (5.6)	3 (1.9)	-132 (8.3)	-5 (2.5)
B. Mean within-trial standard deviations of asynchronies				
A.M.	29.5 (1.7)	17.4 (0.7)	30.8 (2.3)	20.3 (0.7)
B.R.	24.3 (1.2)	14.8 (0.8)	24.8 (1.6)	17.5 (0.8)
H.R.	34.9 (2.9)	20.8 (0.9)	37.9 (3.0)	24.1 (1.1)
J.S.	30.9 (1.6)	20.3 (1.3)	32.1 (2.3)	25.6 (1.1)
S.V.	25.2 (1.6)	14.8 (0.7)	27.5 (1.9)	20.0 (1.5)
V.T.	31.2 (2.9)	20.4 (1.9)	35.9 (2.6)	21.3 (0.8)
C. Lag-1 cross-correlations between IOIs and ITIs				
A.M.	.17 (.05)	.43 (.05)		
B.R.	.19 (.05)	.36 (.05)		
H.R.	.10 (.05)	.37 (.04)		
J.S.	.11 (.04)	.43 (.05)		
S.V.	.11 (.06)	.39 (.05)		
V.T.	.05 (.05)	.41 (.05)		

were isochronous. However, this difference did not reach significance for any participant in the case of visual targets (Session 1 vs. Session 3), whereas it was significant for 5 participants (all but V.T.) in the case of auditory targets (Session 2 vs. Session 4). This indicates that participants were less affected by the temporal structure of jittered visual than of jittered auditory target sequences.

This observation is confirmed by the lag-1 cross-correlations for control sequences in Sessions 3 and 4, which are shown in Table 1C. All participants showed positive correlations, but the correlations for visual sequences (Session 3) were very small and not significantly different from zero in 2 cases (S.V., V.T.). By contrast, the correlations for auditory sequences (Session 4) were significantly larger and clearly different from zero in every case. This provides evidence that modulated

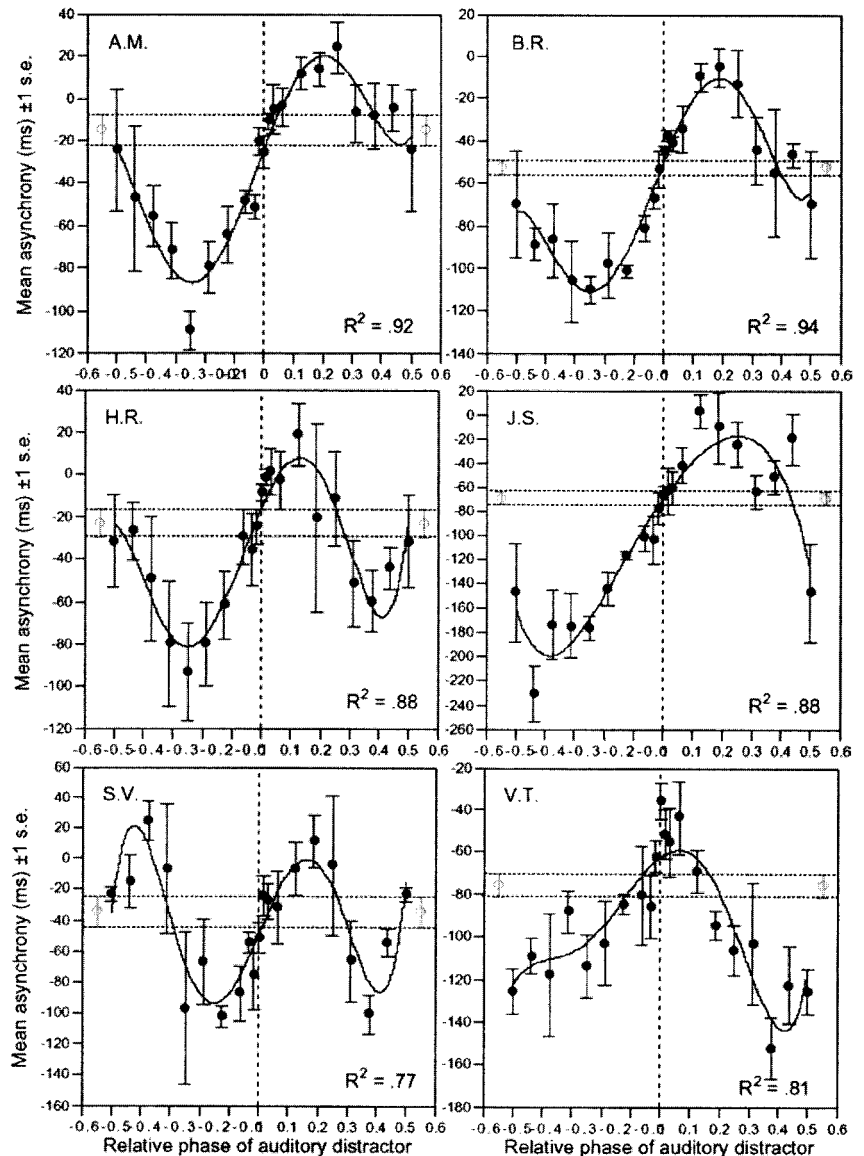
auditory sequences were tracked more closely than were visual sequences. Even the auditory correlations were relatively low, however, because the modulations were small relative to the inherent variability of the taps.

In summary, the results for control trials show much higher variability and poorer tracking of temporal modulations, as well as a tendency towards more negative asynchronies, with visual than with auditory sequences.

Mean asynchrony as a function of relative phase in bimodal trials

Figure 1 shows the effects of auditory distractors on individual participants' asynchronies between taps and visual targets in isochronous sequences (Session 1).

Fig. 1 Mean asynchrony as a function of the normalized relative phase between auditory distractor and visual target, with standard error bars, for individual participants (Session 1). *Open circles* represent control trials



Mean asynchrony, with standard error bars based on five trials, is plotted as a function of the phase of the distractor sequence relative to the target sequence in the normalized range from -0.5 to 0.5 . A negative relative phase indicates that the distractors preceded the targets, just as a negative mean asynchrony indicates that the taps preceded the targets. The data for antiphase distractors are plotted twice, at relative phases of -0.5 and 0.5 , respectively. The open circle with standard errors, which is also plotted twice (arbitrarily, at the meaningless relative phase of ± 0.55), represents the mean asynchrony in visual control (baseline) trials. The dotted horizontal lines indicate the single-standard-error confidence interval of the baseline value. For a distractor effect to be considered significant (at about $p < .05$), its standard error bar must not invade the baseline confidence interval. The solid curve is a fifth-order polynomial fitted to the data, to indicate their general trend. Although a third-order polynomial would have been adequate to trace the predicted quasi-sinusoidal function, a fifth-order polynomial was chosen to capture the more complex pattern of results shown by some participants. The variance accounted for by the curve fit (R^2) is also shown.

It is evident that all participants showed significant effects of auditory distractors at most relative phases. The positive slope of the fitted function in the vicinity of zero indicates that the taps were attracted to both leading and lagging distractor tones. That is, asynchronies became more negative when the distractors preceded the targets, and less negative (or more positive) when the distractors lagged behind the targets. There was an asymmetry, however, in that the attraction exerted by leading distractors was usually stronger than that exerted by lagging distractors, as had also been observed by Repp (2003) within the auditory modality for sequences differing in pitch. These distractor effects typically increased up to relative phases of about $\pm .2$ (± 128 ms temporal separation) and then either ceased to increase or decreased. The way in which the distractor effect depended on relative phase beyond $\pm .2$ differed considerably among participants. Only 2 (A.M., B.R.) showed the expected sinusoidal shape of the overall function, in which the distractor effect approached zero when targets and distractors were in antiphase. Participant H.R. showed similar results at negative relative phases, but showed a tendency to be repelled by distractors (i.e., asynchronies deviated in the direction opposite to the distractor's relative phase) at positive relative phases between $.2$ and $.5$. Participant S.V. showed such repulsion at both positive and negative relative phases between $\pm .2$ and $\pm .5$. What these repulsion effects suggest is that, for these participants, antiphase distractors served as a reference for tapping in phase with the target sequence. When the distractors deviated from antiphase, the taps shifted in the same direction as that deviation, which means they were attracted to the temporal position that was in antiphase with the distractors. Participants J.S. and V.T., in con-

trast to the others, had significantly more negative asynchronies when the distractors were in antiphase than when there were no distractors at all (baseline). The reason for this is not clear, but it may be related to the fact that these two participants had rather large negative baseline asynchronies.

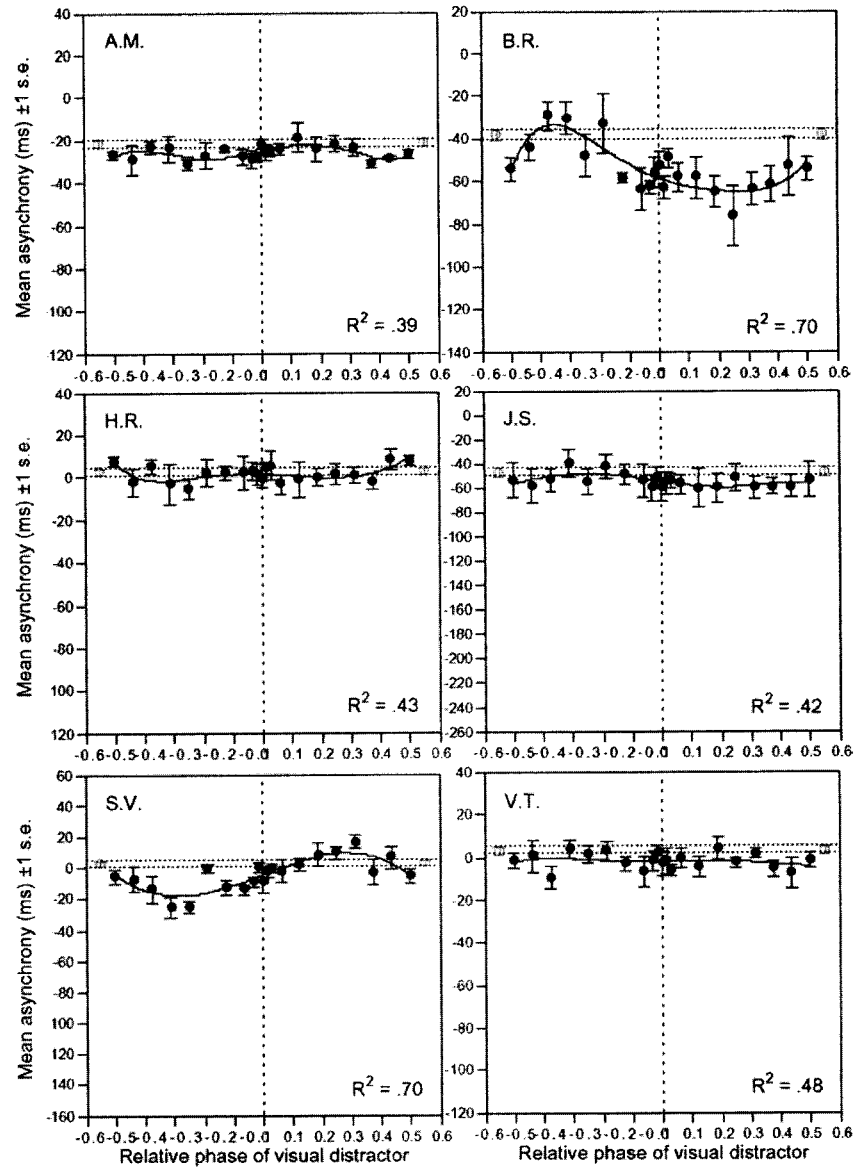
Five participants (all but A.M.) had shown significantly less negative asynchronies for auditory than for visual control sequences (Table 1A). Therefore, a positive shift of the asynchronies (relative to baseline) at relative phases at and near zero might have been expected, indicating that participants tended to synchronize partially or wholly with the auditory distractors. However, for only 1 participant (V.T.) was the zero intercept of the fitted function above the baseline confidence region. Thus, even though the relative changes in mean asynchrony as a function of relative phase clearly reflected distractor effects, there was no consistent absolute shift toward the auditory baseline. Repp and Penel (2002) also had found inconsistent results for asynchronies at zero relative phase.

The results of Session 3 (temporally modulated sequences) were substantially similar to those of Session 1 and are therefore not shown in a separate figure. The only noteworthy differences were that, for participants S.V. and V.T., the asynchronies in visual control trials were more negative than in Session 1 while the data for bimodal trials were similar, so that the distractor effects were now stronger in the positive than in the negative direction. The similarity of the data in Sessions 1 and 3 is captured by their correlation, which ranged from $.62$ ($d.f. = 19, p < .01$) to $.87$ for individual participants.

Figure 2 shows the effects of visual distractors on the asynchronies between taps and auditory targets (for isochronous sequences, Session 2). To highlight the contrast with the Session 1 results, the data of each participant are plotted on the same y-axis scale as in Fig. 1. It is clear that the effects of visual distractors were much smaller than those of auditory ones, as predicted. Four participants (A.M., H.R., J.S., V.T.) showed no systematic distractor effects, only an overall tendency toward slightly more negative asynchronies in the presence of visual distractors. One participant (B.R.) showed a larger negative shift of the asynchronies at all positive and at small negative relative phases, which paradoxically suggests repulsion by lagging visual distractors. Only S.V. showed a sinusoidal pattern indicating attraction of taps to visual distractors, again with an asymmetry favoring leading distractors. However, even for him the attraction was much weaker than that to auditory distractors when the targets were visual, as shown in Fig. 1.

The effects of visual distractors, if any, in Session 4 (modulated sequences) were similar to those in Session 2 and are therefore not shown. The only noteworthy difference was that participant V.T. now showed an increased negativity of asynchronies at small (both negative and positive) relative phases, which may be related to her large negative asynchronies for visual

Fig. 2 Mean asynchrony as a function of the normalized relative phase between visual distractor and auditory target, with standard error bars, for individual participants (Session 2). *Open circles* represent control trials



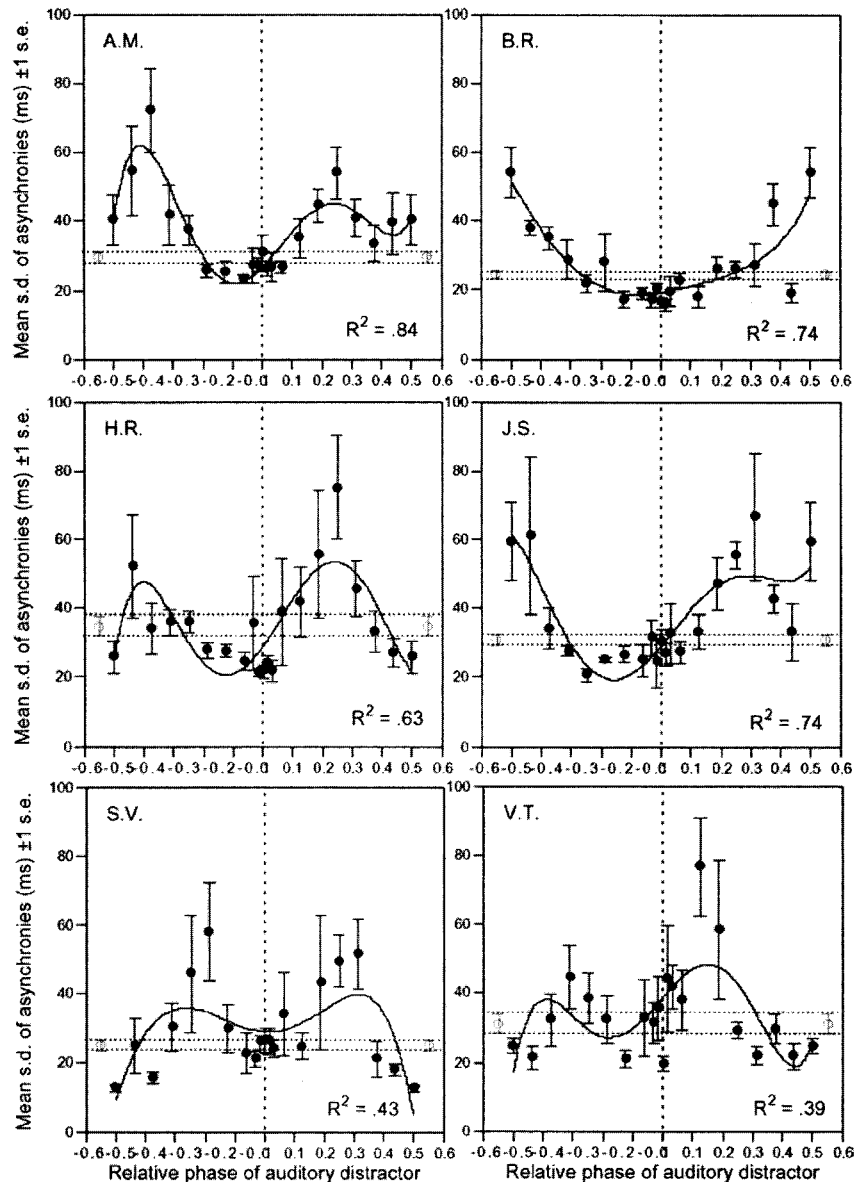
control stimuli. The between-session correlations of individual data ranged from $-.03$ to $.56$ (S.V., d.f. = 19, $p < .01$). The correlations for B.R. and J.S. also reached significance ($p < .05$).

Variability as a function of relative phase in bimodal trials

Figure 3 presents the variability results for visual targets and auditory distractors (Session 1). Each panel shows a participant's mean within-trial standard deviation as a function of relative phase, with standard error bars based on five trials. The variability of the visual control trials is also shown, and a fifth-order polynomial function has been fitted to the data as an aid to the eye.

Because earlier studies had found complete auditory dominance for variability when visual and auditory sequences coincided (Chen et al., 2002; Repp & Penel, 2002), it was expected that variability would be reduced substantially near zero relative phase and then would return to baseline at larger relative phases, resulting in a U-shaped function. This may have been a naive prediction because in fact no participant showed this pattern. Although some (B.R., H.R.) showed a reduction in variability in the vicinity of zero relative phase, all showed an *increase* beyond the already large baseline variability at larger relative phases. This indicates that, at these larger temporal separations, the auditory distractor sequence destabilized synchronization with the visual target sequence. Presumably, this was due to a reduced strength of attraction, so that participants

Fig. 3 Mean within-trial standard deviation of asynchronies as a function of the normalized relative phase between auditory distractor and visual target, with standard error bars, for individual participants (Session 1). *Open circles* represent control trials



began to vacillate between the target and distractor sequences.

There were considerable individual differences in the way the standard deviation varied as a function of relative phase, and these differences seemed to derive from different results for antiphase distractors. Three participants (H.R., V.T., and especially S.V.) showed decreased variability with antiphase distractors, whereas the other 3 showed increased variability, very strikingly so in the cases of B.R. and J.S. It appears that the former participants did not try to ignore temporally distant auditory distractors but rather used them as an aid by tapping in antiphase with them, whereas the latter participants tried to ignore them, but with little success. H.R. and S.V. had also shown a tendency for antiphase distractors to function as repellers (or antiphase attractors) with regard

to asynchronies (Fig. 1). Their variability functions (and that of A.M. as well) are reminiscent of the “seagull effect” in variability of bimanual coordination at different relative phases, which is characterized by variability minima at zero and ± 0.5 relative phase (Kelso, 1995; Semjen & Ivry, 2001; Tuller & Kelso, 1989; Yamashita, Kawato, & Suzuki, 1980).

The variability results for modulated sequences (Session 3) were substantially similar to those of Session 1 and are therefore not shown in a separate figure; only the absolute magnitude of the distractor effects was somewhat smaller. Although this could have been due to practice, it may also indicate that modulated distractors were somewhat easier to ignore. Individual between-session correlations of standard deviations ranged from .44 (d.f. = 19, $p < .05$) to .75.

The variability results for isochronous auditory targets and visual distractors (Session 2) are shown in Fig. 4. As with the asynchrony results, the variation as a function of relative phase was much smaller than in Session 1, but instead of highlighting this fact by plotting the data on the same scale as in Fig. 3 (as was done in Fig. 2 vs. Fig. 1), the ordinate scale has been magnified, so that the data can be inspected more closely. The difference in the magnitude of the effects can be gauged from the fact that the distractor effects look about the same size or smaller than in Fig. 3, even though the y-axis has been expanded by a factor of 5. Only 3 participants (B.R., H.R., S.V.) showed a seagull-like pattern here, and only B.R. showed substantially increased variability with antiphase distractors. Participants B.R. and S.V., it will be recalled, also showed systematic

effects of visual distractors on asynchronies (Fig. 2). The remaining 3 participants did not seem to be affected systematically by visual distractors.

The variability results of Session 4 showed similarities to those of Session 2 for some participants (mainly B.R. and S.V.) but not for others who did not show systematic visual distractor effects. For most participants there was an increase in the scatter of data points and in their standard errors, indicating less reliable results overall. This was probably a consequence of the temporal modulation of the target and distractor sequences. The effects of visual distractors were, if anything, smaller than in Session 2. The individual between-session correlations ranged from $-.37$ to $.51$, with only the highest value reaching significance (B.R., $d.f. = 19, p < .05$).

Fig. 4 Mean within-trial standard deviation of asynchronies as a function of the normalized relative phase between visual distractor and auditory target, with standard error bars, for individual participants (Session 2). *Open circles* represent control trials

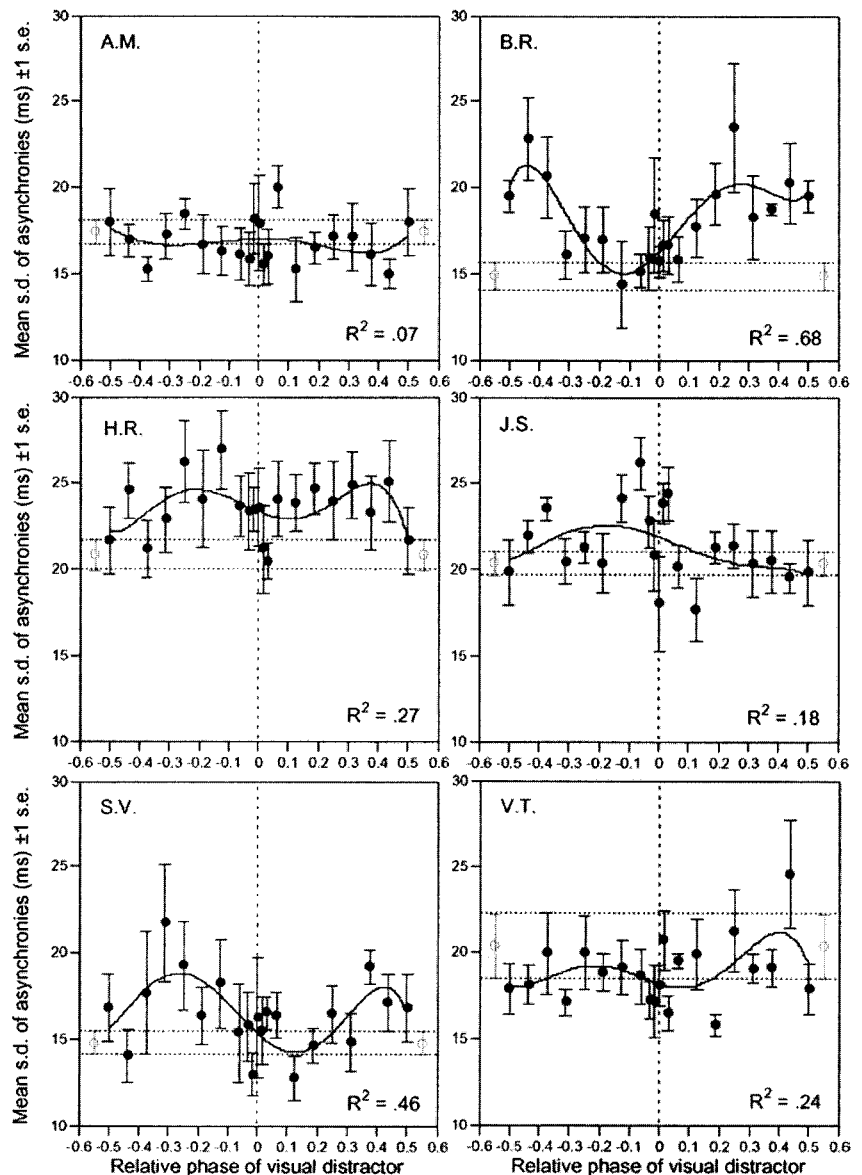
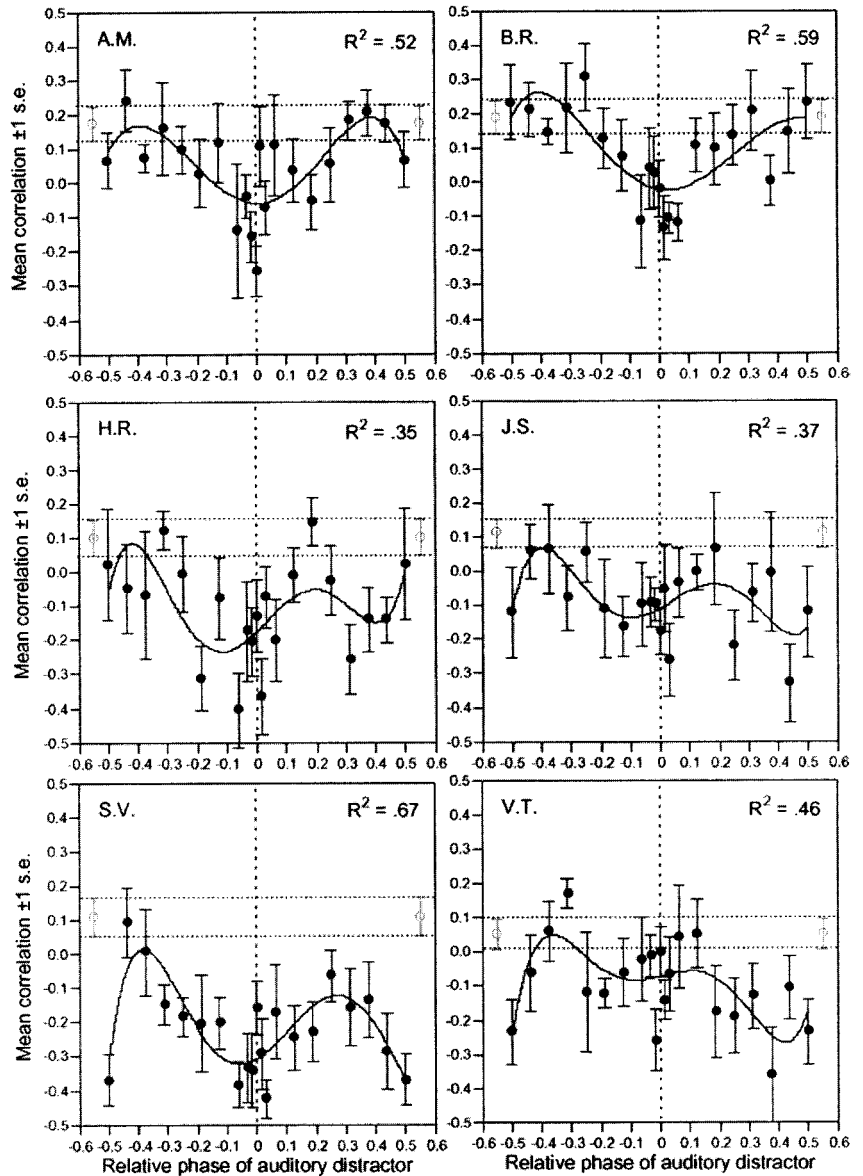


Fig. 5 Mean lag-1 cross-correlation between inter-tap intervals (ITIs) and visual target inter-onset intervals (IOIs) as a function of the normalized relative phase between auditory distractor and visual target, with standard error bars, for individual participants (Session 3). *Open circles* represent control trials



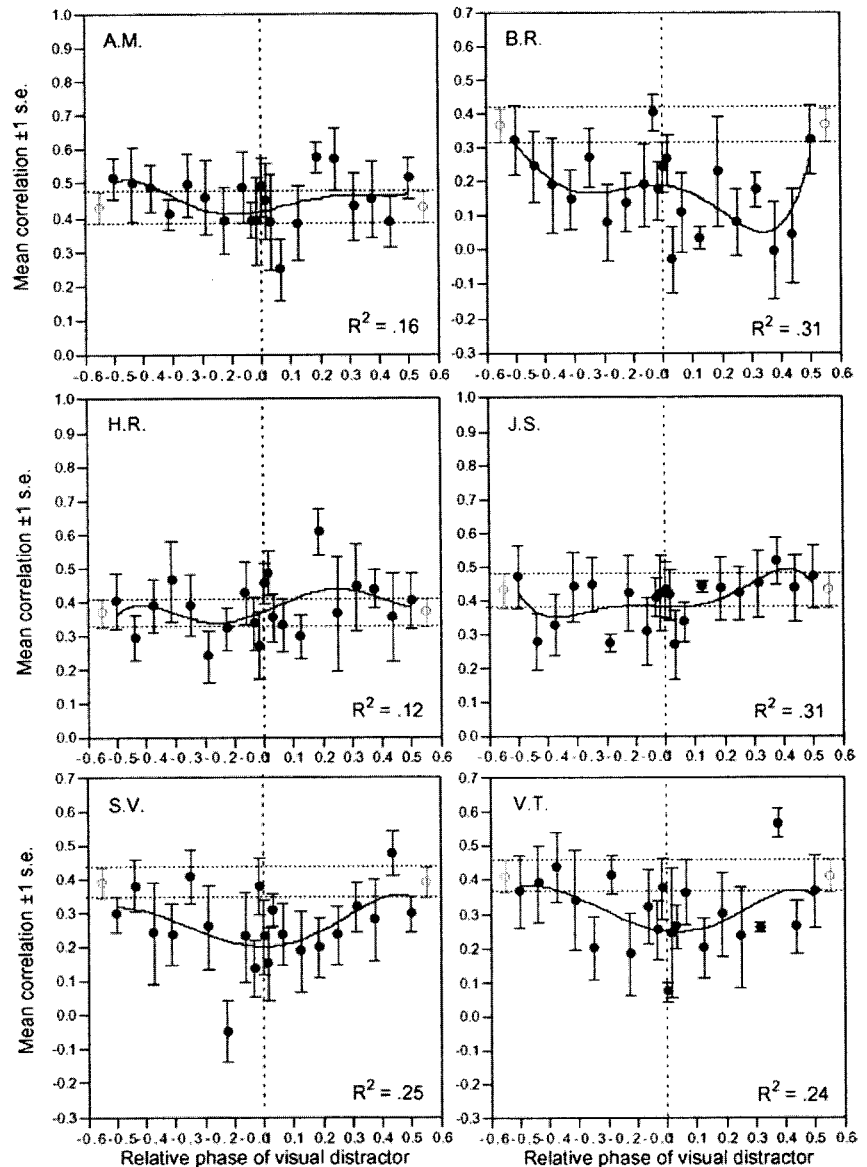
Lag-1 cross-correlation as a function of relative phase

Figure 5 shows the lag-1 cross-correlation values for jittered visual targets (Session 3) as a function of the average relative phase of jittered auditory distractors, in the format familiar from previous figures. It is evident that all participants had significantly reduced (in fact, negative) correlations at small relative phases, relative to the baseline provided by the visual control sequences. At larger relative phases, the correlations tended to be closer to the baseline. Close to antiphase, however, there were again large individual differences. Three participants (J.S., S.V., V.T.) showed significantly reduced, negative correlations there, whereas the other three had correlations close to baseline. As a result, the former participants evinced a seagull-like

function, whereas the latter had a more simple U-shaped function or something in between. Two participants in the former group (S.V., V.T.), but not the third (J.S.), had also shown reduced variability at antiphase (Fig. 3) and a tendency for antiphase distractors to repel taps (Fig. 1).

Finally, Fig. 6 shows the lag-1 cross-correlations for jittered auditory targets with jittered visual distractors (Session 4). Three participants (A.M., H.R., J.S.) did not show any systematic effect of visual distractors. For the other 3, the visual distractor lowered the correlations according to a roughly U-shaped function of relative phase. These same participants also showed an effect of visual distractors on asynchronies in Session 4 (not shown in a figure).

Fig. 6 Mean lag-1 cross-correlation between ITIs and auditory target IOIs as a function of the normalized relative phase between visual distractor and auditory target, with standard error bars, for individual participants (Session 4). *Open circles* represent control trials



Experiment 2

Experiment 1 provided striking new evidence for auditory dominance in a synchronization task. However, several questions could be raised about the generality of the findings, and some of these questions were addressed in Experiment 2 with a reduced number and range of phase relationships between targets and distractors.

The most important question concerned the relative salience of the stimuli in the two modalities, particularly the role of the relative intensity of the tones. Could it be that the auditory dominance observed in Experiment 1 and earlier studies was due merely to the specific intensity of the tones, which made them more salient than the flashing light? To address this nagging issue, we presented the tones at two intensity levels, at

one of which they seemed about as salient as the lights, and at the other, less salient. For auditory dominance to have some generality, it should be found even when the tones are subjectively less salient than the lights.

A second concern was that auditory feedback from the taps may have favored the auditory modality, as mentioned in the Introduction. Therefore, we eliminated this feedback in Experiment 2. A third concern, at least for some skeptical readers, might be that some participants in Experiment 1 had not looked at the visual distractors. Although we consider this unlikely, we did obtain a measure of participants' attention to the visual sequences in Experiment 2. Furthermore, we decided to present tones and lights from the same position in space, even though we did not expect this to have a dramatic effect on the results. Finally, we

had noticed that conducting the experiment in a dark room led to artifacts caused by eye movements (viz., illusory motion of the light source in the absence of a spatial reference frame), which may have interfered with synchronization. In Experiment 2, therefore, we merely dimmed the light in the room, which reduced these motion artifacts while retaining good visibility of the flashes.

Method

Participants

The 9 participants included 6 paid volunteers (5 women, aged 18–30), one research assistant (S.H., female, age 44), and both authors (B.R., age 58, and A.P., age 29). One additional participant's data were not analyzed because of many missing taps. Two of the volunteers and B.R. had participated in Experiment 1. Most participants had substantial musical training (10 or more years of instruction), but 2 (1 being S.H.) had had only a few years of instruction. S.H. also had little tapping experience, whereas all others were (or had been some time ago, in A.P.'s case) regular participants in synchronization experiments.

Design

The experiment consisted of a single session in which the two attention conditions (auditory or visual targets) were presented in succession, with their order counterbalanced across participants. Each attention condition comprised five blocks of 12 trials. The visual target condition additionally included a practice block of 12 trials at the beginning.

Materials and equipment

The sequences were the same as in Experiment 1, but only five temporal displacements between target and distractor sequences were used (normalized relative phases are given in parentheses): 0 ms (0), ± 64 ms ($\pm .1$), and ± 128 ms ($\pm .2$). In a block of 12 trials, each of these five relative phases occurred once in combination with each of two tone intensity levels (see below), and the remaining two trials were control trials without any distractor sequence. The 12 trials occurred in different random orders in each of the five blocks of a condition. In 5 of the 12 visual target sequences and in 4 of the 10 visual distractor sequences in a block, a single flash was missing. Its position varied but was neither close to the beginning nor close to the end of the sequence.

In contrast to Experiment 1, the tones from the digital piano were not delivered over headphones but were fed through an Optimus SA-155 amplifier to a Realistic Minimus-7 loudspeaker, which was placed to the left of the computer, about 60 cm from the participants. The amplifier volume control was kept fixed throughout the experiment. The MIDI Translator box containing the LED was placed in front of the loudspeaker. The room lights were turned off, but dim illumination was provided by a 60-W hooded desk lamp placed on a table about 2 m to the right and facing away from the experimental equipment. During data collection, the computer monitor was rotated 90°, so it faced away from the loudspeaker and did not shine into participants' eyes. The flashes were clearly visible under these conditions, though not as bright subjectively as in Experiment 1.

Before starting the experiment, and with the equipment set up as just described, the two authors and the research assistant independently adjusted the output volume control of the digital piano

until a sequence of isochronous tones seemed to them as salient as a sequence of simultaneous flashes. They arrived at very nearly the same setting, and that setting was then maintained for all participants, including themselves. The adjusted intensity level of the tones will be referred to as "loud" in the following, although it was actually fairly soft and lower than the level of the tones heard over headphones in Experiment 1. A second, even lower intensity level, referred to as "soft" in the following, was obtained by specifying the MIDI velocity in the sequence instructions as 30 units lower. These tones were quite soft but still well above the detection threshold. According to earlier acoustic measurements of the digital piano output, albeit of tones with lower pitch (Repp, 1997: Fig. 1), the difference between loud and soft tones amounted to about 10 dB.

Procedure

After reading a block of trials into the computer memory, participants rotated the computer monitor to the right, as described above, and turned their head toward the loudspeaker with the LED in front of it. Depression of a key on the computer keyboard started the first sequence in a block. Participants were instructed to start tapping with the third event in the target sequence and keep tapping in synchrony with the target events while ignoring the distractor sequence, which started later. They tapped by rhythmically depressing and releasing a white key on a Fatar Studio 37 MIDI controller (a quiet three-octave piano keyboard), which they held on their lap. Participants were instructed to keep their finger in contact with the chosen key and tap gently, so that no slapping or impact sounds were produced. They were alerted to the possible presence of gaps in the visual sequences and were asked to tap through any gaps in visual target sequences. (Auditory sequences never contained gaps.) After each trial, they pressed the "0" key on the computer keyboard if they had detected a gap in the visual sequence, and the space bar otherwise. The next sequence started 2 s after this key depression. A chime signaled the end of a block. Between blocks, participants returned the computer monitor to its normal orientation to save the data and read in the next block.

Results

Gap detection

On average, participants missed .8 gaps (range: 0 to 3) out of 45 and gave 2.6 false-alarm responses (range: 0 to 11) to the 65 trials that did not contain a gap. Thus, gap detection was almost perfect, which provides evidence that the participants monitored the visual sequences, whether they were targets or distractors. The few mistakes were probably due to eye blinks.

Phase drift

Phase drift was defined as asynchronies that exceeded half the IOI and kept increasing in absolute magnitude. One participant (S.H.), the only one without previous experience in synchronization experiments, had some difficulty staying in phase with visual target sequences in control trials and in bimodal trials with relative phases of -0.2 and $+0.2$, showing phase drift in 11 out of 30 trials. These trials were excluded. A single phase drift trial was excluded from one other participant's data.

Table 2 Results for unimodal control trials in Experiment 2. Standard errors are in parentheses. *Combined* average of results for loud and soft auditory targets

Participant	Target			
	Visual	Auditory (loud)	Auditory (soft)	Combined
A. Mean asynchronies				
A.P.	-127 (7.2)	-133 (7.9)	-150 (7.5)	-141 (5.8)
B.R.	-29 (4.8)	-45 (3.2)	-37 (3.3)	-41 (2.5)
H.R.	-57 (7.6)	-14 (4.5)	-11 (2.6)	-12 (2.5)
N.K.	30 (15.8)	-42 (2.9)	-44 (5.9)	-43 (3.1)
S.L.	-53 (5.6)	-47 (1.6)	-61 (2.6)	-54 (2.7)
S.K.	-37 (9.0)	-52 (4.2)	-60 (7.4)	-56 (4.3)
S.H.	-120 (22.5)	-96 (4.1)	-119 (7.6)	-108 (5.6)
V.T.	-102 (6.3)	-76 (3.2)	-86 (8.6)	-81 (4.6)
V.N.	-70 (7.4)	-64 (8.5)	-65 (6.0)	-65 (4.9)
B. Mean within-trial standard deviations of asynchronies				
A.P.	29.1 (2.2)	18.8 (1.4)	19.6 (1.4)	19.2 (1.0)
B.R.	17.2 (1.1)	14.4 (1.3)	11.7 (0.6)	13.1 (0.8)
H.R.	35.7 (4.6)	15.7 (1.0)	19.5 (2.1)	17.6 (1.2)
N.K.	35.8 (3.3)	13.6 (1.5)	15.0 (1.7)	14.3 (1.1)
S.L.	29.1 (2.9)	17.7 (0.2)	16.2 (0.3)	17.0 (0.3)
S.K.	33.6 (2.6)	26.6 (2.7)	20.0 (1.5)	23.3 (1.8)
S.H.	69.3 (10.2)	18.9 (2.7)	22.2 (1.9)	20.5 (1.6)
V.T.	28.6 (2.0)	17.2 (0.6)	22.7 (3.8)	20.0 (2.0)
V.N.	52.4 (5.0)	25.2 (3.6)	26.9 (3.4)	26.1 (2.4)

Unimodal control trials

In each attention condition (visual or auditory targets), there were 10 control trials in which the target sequence occurred alone. For auditory targets, 5 control trials were loud and the other 5 were soft. The mean asynchronies for visual and auditory control trials are shown in Table 2A. In general, they were more negative than in Experiment 1, which we attribute to the absence of auditory feedback from the taps and to the different tapping device used. (Taps were electronically registered before bottom contact of the response key, which increased the negative asynchronies by 10–20 ms.) This anticipation tendency was larger in synchronization with soft than with loud tones for 7 participants, although it was significant for only 3 of them. Two participants showed a difference in the opposite direction, one of which was significant. We averaged the mean asynchronies for soft and loud tones (last column in Table 2A) to compare them with the asynchronies for visual stimuli (first column). Five participants had more negative asynchronies for visual than with auditory sequences (significant in two cases), whereas significant effects in the opposite direction were shown by the remaining 4 participants. Again, it seems that mean asynchrony is not a reliable measure of modality differences.

As in Experiment 1, much more consistent differences between modalities were obtained in terms of variability, as shown in Table 2B. Again, standard deviations were much larger in synchronization with visual than with auditory targets, and the difference was highly significant for all participants. (Table 2A also shows larger between-trial standard errors for visual than for auditory sequences, as in Experiment 1.) There was no consistent difference between loud and soft auditory

sequences: 6 participants showed higher variability with soft than with loud auditory sequences (significant in 2 cases), whereas 3 showed the opposite difference (significant in all 3 cases).

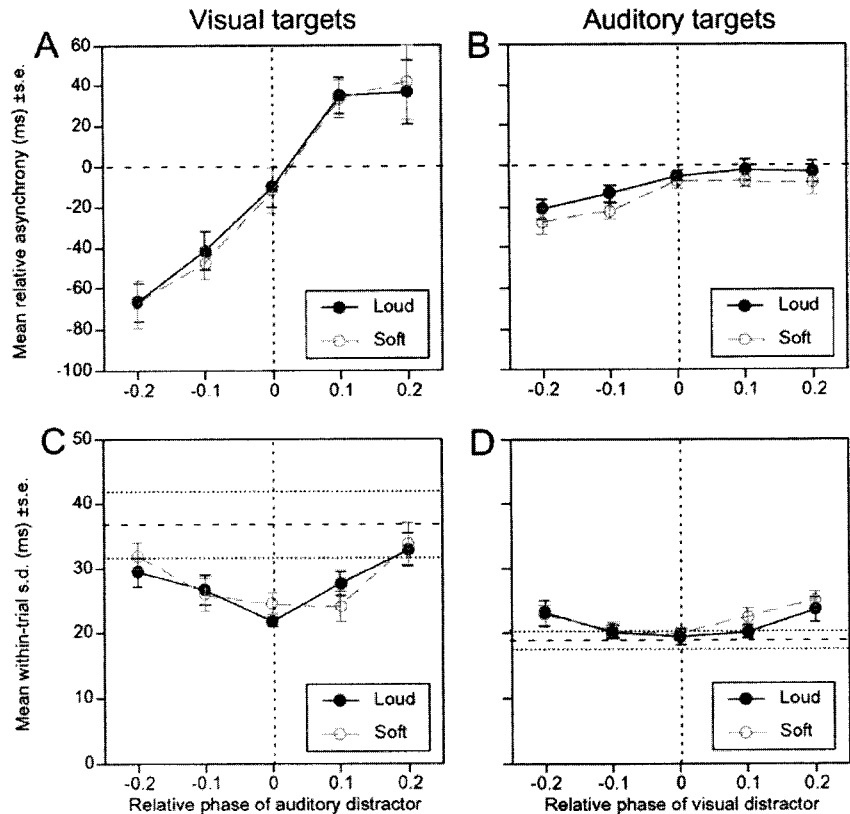
In summary, the results for control trials confirmed those of Experiment 1 in that they showed much higher variability in synchronization with visual than with auditory sequences, even under the changed conditions of the present experiment.

Mean asynchrony as a function of relative phase in bimodal trials

For the sake of conciseness, we present the main results averaged across participants, mentioning individual differences only when necessary. To exclude the large individual differences in absolute asynchronies from the between-participant standard errors, we expressed the asynchronies of each participant as deviations from his or her mean asynchrony for the appropriate control trials (i.e., as relative asynchronies).

Figure 7A shows the effects of auditory distractors on the mean relative asynchronies between taps and visual targets for loud and soft distractors separately. Similarly, Fig. 7B shows the effects of visual distractors on the asynchronies between taps and auditory targets for loud and soft targets separately. The horizontal line represents the zero baseline. It can be seen that the results of Experiment 1 were largely replicated, with auditory distractors having a much stronger effect than visual distractors. Again, leading distractors had a stronger effect than lagging distractors. Experiment 2 adds the new result that the intensity of the tones made little difference, regardless of whether the tones served as targets or as distractors. These observations were confirmed in a

Fig. 7A, B Mean relative asynchronies and **C, D** mean within-trial standard deviations of asynchronies as a function of normalized relative phase for visual and auditory targets in Experiment 2



repeated-measures ANOVA, which only showed a significant effect of relative phase, $F(4, 32) = 26.5$, $p < .0001$, and an interaction between relative phase and target/distractor type, $F(4, 32) = 12.4$, $p < .0001$. A separate ANOVA on the auditory target data showed that the effect of relative phase was significant, $F(4, 32) = 9.35$, $p < .0001$. Here there was also a main effect of tone intensity, $F(1, 8) = 5.89$, $p < .05$, due to slightly more negative relative asynchronies for softer tones.

To determine whether the presence of gaps in about 40% of the visual sequences had any effect on the results, a separate ANOVA compared trials with gaps and trials without gaps. There were no significant effects involving this variable.

Inspection of individual data revealed some departures from the average pattern. One of the authors (A.P.) was the only participant to show a relatively small effect of auditory distractors but a strong effect of leading visual distractors. One other participant (S.H.) showed a strong effect of leading visual distractors, but only when synchronizing with soft tones; paradoxically, she also showed a larger effect of soft than of loud auditory distractors. Most other participants were hardly affected by visual distractors, but all showed large auditory distractor effects. Two participants showed no effect of lagging auditory distractors at a relative phase of .2, one of them only when the distractor tones were loud.

Variability as a function of relative phase in bimodal trials

The standard deviations are shown in Fig. 7C and D for loud and soft tones separately. The horizontal lines represent the variability of the control trials plus/minus one standard error. Auditory distractors at zero relative phase reduced the variability for visual targets almost to the value for auditory control trials, which indicates strong auditory dominance. Variability increased at larger relative phases, resulting in a U-shaped function. By contrast, visual distractors did not increase variability when they coincided with auditory targets, and variability increased only slightly at the larger relative phases, resulting in a very shallow U-shaped function. Again, the intensity of the tones seemed to make little difference. An ANOVA revealed significant effects of target/distractor type, $F(1, 8) = 50.8$, $p < .0001$, reflecting greater variability with visual than with auditory targets, and of relative phase, $F(4, 32) = 26.5$, $p < .0001$. The interaction was not significant, however, and there were no significant effects involving tone intensity. A separate ANOVA on the auditory target data showed that the small effect of relative phase was significant, $F(4, 32) = 6.18$, $p < .0001$. Another separate ANOVA contrasting trials with and without gaps in the visual sequences did not reveal any significant effects of that variable.

Inspection of individual data revealed some departures from the average pattern, but all participants showed a greater effect of auditory than of visual distractors (including A.P. and S.H. when synchronizing with soft tones).

General discussion

This study examined the relative attraction of rhythmic action to auditory and visual stimulus sequences in a sensorimotor synchronization task. It attempted to replicate some of the findings of earlier studies (Chen et al., 2002; Repp & Penel, 2002) using a slower sequence tempo, while also significantly extending the experimental design to include different attentional conditions, different phase relationships between target and distractor sequences, modulated as well as isochronous sequences, different levels of relative salience, and different spatial relationships of the two modalities. The results demonstrate convincingly that auditory distractors affect synchronization with visual target sequences more than visual distractors affect synchronization with auditory target sequences.

Sequence tempo

The use of a slower sequence tempo than in previous studies had the desired consequence that no participant (except the least experienced one in Experiment 2) had any difficulty synchronizing with visual sequences, so that almost no data had to be excluded. Nevertheless, variability was much higher for unimodal visual than for unimodal auditory sequences; this was true for all participants in both experiments. This finding replicates the results of Kolers and Brewster (1985), Repp and Penel (2002), Chen et al. (2002), and Repp (in press), among others. Clearly, this modality difference is not specific to any sequence tempo, nor is it dependent on particular levels of relative stimulus salience (Experiment 2). Chen et al. (2002) showed by means of spectral analysis that the modality difference in variability reflects a difference in low-frequency (long-term) drift. Local (short-term) variation, such as might be due to ongoing phase correction (Pressing, 1998), does not seem to exhibit a modality difference, although Repp and Penel (2002) found weaker phase correction with visual than with auditory sequences when the phase correction was unintended.

Attentional focus

In our earlier studies, we had been satisfied with demonstrating auditory dominance when participants were told to focus on the visual modality in the presence of auditory distractors. By including the reverse attentional condition in the present study, we gave the visual

modality a chance to exert a distractor effect. With one exception (participant A.P. in Experiment 2), these visual distractor effects were much smaller than the auditory ones, and several participants were totally unaffected by visual distractors. Overall, visual distractor effects seemed slightly larger in Experiment 2 than in Experiment 1. This may have been due to the requirement of monitoring the visual sequences for missing flashes. Although we have no reason to believe that any participants in Experiment 1 closed their eyes or looked away from the visual stimuli, active monitoring in Experiment 2 may have increased the attention allocated to the visual modality. However, this increased attention was not sufficient to overcome the dominance of the auditory modality (except in one case, as noted).

Relative phase

The primary independent variable in this study was the relative phase between target and distractor sequences, which varied between 0 and $\pm .5$ in Experiment 1 and between 0 and $\pm .2$ in Experiment 2. In presenting the results, we used relative phase mainly for convenience, not because we believe it is the actual controlling variable in this paradigm. On the contrary, recent experiments conducted within the auditory modality (Repp, in preparation) suggest that it is absolute temporal separation, not relative phase, that governs distractor effects. Although we did not test this hypothesis directly in the present study, our findings are consistent with it. Auditory attractor effects began to wear off around relative phases of $\pm .2$ (128 ms), which is in the same range as the temporal window proposed for auditory-visual integration (Lewald et al., 2001; Lewald & Guski, 2003; Shams et al., 2000; Slutsky & Recanzone, 2001).

Relative phase did play a role, however, in sensorimotor coordination, as was evident in the variable results obtained with auditory antiphase distractors. For some participants, the distractor effect disappeared in that condition and variability increased, suggesting that they were able to ignore the distractors, but at a considerable cost. Others chose instead to tap in antiphase with antiphase distractors, which resulted in low variability and reversed distractor effects in the vicinity of antiphase. These individual differences with regard to the effect of antiphase distractors contrast with the universal attraction to auditory distractors at small relative phases, which is consistent with the generally greater stability (hence, smaller inter-individual variability) of in-phase than antiphase relationships in sensorimotor and bimanual coordination (see, e.g., Haken, Kelso, & Bunz, 1985; Kelso, DelColle, & Schöner, 1990; Yamanishi et al., 1980).

Experiment 2 replicated what was the most striking manifestation of auditory dominance in our earlier studies (Chen et al., 2002; Repp & Penel, 2002), namely the reduction of variability, relative to unimodal visual sequences, when tapping with visual targets that are

synchronous with auditory distractors. This reduction was less evident in Experiment 1, for reasons that are not quite clear. An increase in variability occurred at relative phases beyond $\pm .2$, which we attribute to vacillation between the target and distractor sequences as they become easier to separate perceptually.

Temporal modulation

Experiment 1 introduced irregular temporal modulations in target and distractor sequences. Although these modulations remained fixed during a block of trials, they were small and impossible to memorize and predict. The lag-1 cross-correlations for unimodal sequences showed that participants' inter-tap intervals tracked the sequence IOIs, as expected (Michon, 1967; Repp, 2002b). This is believed to be a consequence of automatic phase correction. The correlations were much smaller for visual than for auditory unimodal sequences, presumably because of greater variability of the taps and/or because of less effective phase correction with visual sequences. The correlations for bimodal sequences added to the evidence for auditory dominance. When visual targets and auditory distractors occurred at small relative phases, there was a strong tendency to track the auditory rather than the visual temporal pattern. In the reverse modality assignment, only some participants showed a slight reduction in their tracking of the auditory pattern.

Stimulus salience

The most important manipulation in Experiment 2 was the adjustment and variation of the intensity of the tones. In many previous cross-modal studies, including our own, the relative salience of the stimuli in the two modalities was not controlled. In fact, Spence et al. (2001, p. 803) express skepticism about matching the intensities of stimuli in different modalities. We were encouraged by the close agreement of three participants' adjustments of tone intensity to match the salience of the lights. However, the validity and generality of this adjustment was far less crucial than our use of two different intensities for the tones, a manipulation that obviously affected their relative salience. The fact that tone intensity had no effect at all on the magnitude of the distractor effects and on auditory dominance in particular suggests that the dominance is indeed a function of modality and not of relative stimulus salience (as long as the stimuli are clearly above threshold). Recently, Recanzone (2003) reported that the effect of auditory sequence rate on the perception of visual sequence rate ("auditory driving") is likewise independent of tone intensity.

Spatial congruence and auditory feedback

Strong auditory dominance was obtained both with spatially disparate (Experiment 1) and spatially con-

gruent (Experiment 2) bimodal sequences. Tighter cross-modal integration of spatially congruent stimuli may have contributed to the significant visual distractor effects obtained in Experiment 2, and to the greater reduction in variability when visual targets coincided with auditory distractors. However, auditory distractor effects on asynchronies were about as large in Experiment 2 as in Experiment 1. The absence of auditory feedback from the taps seemed to have no impact on the results other than causing more negative asynchronies overall.

Could there be visual dominance for temporal information?

An important issue awaiting investigation is whether the relative attraction of movement to visual rhythms can be increased by introducing spatial variation and movement, especially biological motion. A flashing light is not a common visual experience, whereas moving objects and organisms are ubiquitous. Moreover, for spatial information, dominance of vision over audition has been demonstrated, as in the ventriloquist effect (e.g., Bertelson & Aschersleben, 1998; Slutsky & Recanzone, 2001). Would auditory dominance vanish when tone sequences are combined with rhythmic visual movements such as a video of a conductor beating time or of a finger tapping? Brass, Bekkering, and Prinz (2001) have shown that observation of finger movements facilitates congruent actions in terms of reaction time. Would such visible movements stimulate and attract rhythmic action the way music does? Probably not, but the experiments need to be done. We suspect that Fraise (1948) was right in noting a fundamental difference between auditory and visual rhythms, in that only the former have the capability of inducing rhythmic action.

A reasonable working hypothesis is that, in a conflict situation, the modality that affords greater behavioral stability will dominate. So far, there is no evidence that any form of repetitive visual stimulation can give rise to synchronized movement of equal or lower variability than a simple auditory sequence can. If such a form of visual stimulation exists, it may be competitive or even gain the upper hand in an auditory-visual conflict situation.

The locus of cross-modal interaction

Information from different modalities converges at many different sites in the brain (Driver & Spence, 2000; Stein, 1998). The different inputs may be perceived independently and combined at a cognitive level, or the input in one modality may affect the perception of that in the other modality (and vice versa). As Fendrich and Corballis (2001) and Morein-Zamir et al. (2003), among others, have shown, auditory temporal information does indeed affect the visual perception of timing. Thus, the

effect of auditory distractors on synchronization with visual targets could be mediated by a perceptual interaction. On the other hand, several earlier findings have suggested that control of synchronized action is independent of perception: phase correction is independent of the detection threshold for phase perturbations (Repp, 2000, 2001, 2002a; Repp & Penel, 2002); auditory dominance in auditory-visual conflict situations is independent of large individual differences in cross-modal perceptual interactions (Repp & Penel, 2002); and influences of extended temporal context are evident only in perception, not in synchronization (Repp, 2002c). These findings suggest an alternative interpretation of distractor effects, namely that they reflect involuntary dual synchronization with competing sequences whose timing is registered veridically for action control, prior to any perceptual interaction. However, our hypothesis that absolute temporal separation, rather than relative phase, is the true controlling variable of distractor effects seems more compatible with a perceptual interaction account, because what temporal separation limits is perceptual interaction. This issue requires further investigation, which is in progress.

Conclusions

Three dependent variables—asynchronies, their standard deviations, and lag-1 cross-correlations with jittered sequences—consistently show that auditory distractors affect synchronization with visual targets much more than visual distractors affect synchronization with auditory targets. These findings extend earlier results and support the claim that rhythmic movement is attracted more strongly to auditory than to visual rhythmic stimuli. To what extent this attraction may be a *consequence* of extensive exposure to music and other rhythmic auditory stimuli is not known. This would be difficult to investigate because virtually everybody has been exposed much more to auditory than to visual rhythmic stimulation, beginning even before birth.⁸ It seems likely that we are dealing here with a very basic human propensity, which is one of the reasons why music evolved in the auditory modality. The purpose of music is to move people, both literally and metaphorically, and our ancestors may have discovered long ago that this is achieved more easily in audition than in vision.

Acknowledgements This research was supported by NIH grant MH-51230 awarded to the first author, and by a Cold Spring Harbor Laboratory post-doctoral fellowship awarded to the second author. Thanks are due to Helen Sayward and Susan Holleran for help with data analysis, and to Alex Aksentijevic, Peter Keller, and Ani Patel for helpful comments on an earlier version of the manuscript.

⁸ We are not aware of any observations that deaf individuals synchronize more readily with visual stimuli than hearing individuals, although this would be worth investigating.

References

- Aschersleben, G. (2002). Temporal control of movements in sensorimotor synchronization. *Brain and Cognition*, *48*, 66–79.
- Bartlett, N. R., & Bartlett, S. C. (1959). Synchronization of a motor response with an anticipated sensory event. *Psychological Review*, *66*, 203–218.
- Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, *5*, 482–489.
- Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica*, *106*, 3–22.
- Chen, Y., Repp, B. H., & Patel, A. D. (2002). Spectral decomposition of variability in synchronization and continuation tapping: Comparisons between auditory and visual pacing and feedback conditions. *Human Movement Science*, *21*, 515–532.
- Condon, W. S., & Sander, L. W. (1974). Neonate movement is synchronized with adult speech: Interactional participation and language acquisition. *Science*, *183*, 99–101.
- Drake, C. (1997). Motor and perceptually preferred synchronization by children and adults: Binary and ternary ratios. *Polish Quarterly of Developmental Psychology*, *3*, 43–61.
- Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, *77*, 251–288.
- Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, *10*, R731–R735.
- Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, *63*, 719–725.
- Fraisse, P. (1948). Rythmes auditifs et rythmes visuels. [Visual and auditory rhythms.] *L'Année Psychologique*, *49*, 21–41.
- Fraisse, P., Pichon, P., & Clairouin, G. (1949). Les aptitudes rythmiques: Etude comparée des oligophrènes et des enfants normaux. [Rhythmical aptitudes: A comparative study of mentally retarded and normal children.] *Journal de Psychologie Normale et Pathologique*, *42*, 309–330.
- Gault, R. H., & Goodfellow, L. D. (1938). An empirical comparison of audition, vision, and touch in the discrimination of temporal patterns and ability to reproduce them. *Journal of General Psychology*, *18*, 41–47.
- Glenberg, A., & Jona, M. (1991). Temporal coding in rhythm tasks revealed by modality effects. *Memory & Cognition*, *19*, 514–522.
- Glenberg, A., Mann, S., Altman, L., Forman, T., & Prociase, S. (1989). Modality effects in the coding and reproduction of rhythms. *Memory & Cognition*, *17*, 373–383.
- Goldstone, S., Boardman, W. K., & Lhamon, W. T. (1959). Intersensory comparisons of temporal judgments. *Journal of Experimental Psychology*, *57*, 243–248.
- Goodfellow, L. D. (1934). An empirical comparison of audition, vision, and touch in the discrimination of short intervals of time. *American Journal of Psychology*, *46*, 243–258.
- Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception & Psychophysics*, *54*, 383–394.
- Grondin, S., & Rousseau, R. (1991). Judging the relative duration of multimodal short empty time intervals. *Perception & Psychophysics*, *49*, 245–256.
- Grondin, S., Meilleur-Wells, G., Ouellette, C., & Macar, F. (1998). Sensory effects on judgments of short time-intervals. *Psychological Research*, *61*, 261–268.
- Grondin, S., Ouellet, B., & Roussel, M.-E. (2001). About optimal timing and stability of Weber fraction for duration discrimination. *Acoustical Science and Technology*, *22*, 370–372.
- Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A theoretical model of phase transitions in human hand movements. *Biological Cybernetics*, *51*, 347–356.
- Hargreaves, D. J. (1986). *The developmental psychology of music*. Cambridge, U.K.: Cambridge University Press.

- Harrington, L. K., & Peck, C. K. (1998). Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Experimental Brain Research*, *122*, 247–252.
- Kelso, J. A. S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge, MA: MIT Press.
- Kelso, J. A. S., DelColle, J. D., & Schöner, G. (1990). Action-perception as a pattern formation process. In M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 138–169). Hillsdale, NJ: Erlbaum.
- Kolers, P. A., & Brewster, J. M. (1985). Rhythms and responses. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 150–167.
- Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, *16*, 468–478.
- Lewald, J., Ehrenstein, W. H., & Guski, R. (2001). Spatio-temporal constraints for auditory-visual integration. *Behavioral Brain Research*, *121*, 69–79.
- Michon, J. A. (1967). *Timing in temporal tracking*. Assen, NL: Van Gorcum.
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, *17*, 154–163.
- Peryer, G., Sloboda, & Nte, S. (2002). How is the synchronisation of tapping to a visual isochronous pulse affected by an interfering auditory pulse? In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney, 2002* (pp. 783–786). Adelaide, Australia: Causal Productions (CD-ROM).
- Pressing, J. (1998). Error correction processes in temporal pattern production. *Journal of Mathematical Psychology*, *42*, 63–101.
- Rainbow, E. L., & Owen, D. (1979). A progress report on a three year investigation of the rhythmic ability of pre-school aged children. *Bulletin of the Council for Research in Music Education*, *59*, 84–86.
- Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology*, *89*, 1078–1093.
- Repp, B. H. (1997). Acoustics, perception, and production of legato articulation on a computer-controlled grand piano. *Journal of the Acoustical Society of America*, *102*, 1878–1890.
- 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. (2002a). 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. (2002b). The embodiment of musical structure: Effects of musical context on sensorimotor synchronization with complex timing patterns. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action: Attention and Performance XIX* (pp. 245–265). Oxford, U.K.: Oxford University Press.
- Repp, B. H. (2002c). Perception of timing is more context sensitive than sensorimotor synchronization. *Perception & Psychophysics*, *64*, 703–716.
- 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. (in press). Rate limits in sensorimotor synchronization with auditory and visual sequences: The synchronization threshold and the benefits and costs of interval subdivision. *Journal of Motor Behavior*.
- Repp, B. H., & Penel, A. (2002). Auditory dominance in temporal processing: New evidence from synchronization with simultaneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 1085–1099.
- Rousseau, R., Poirier, J., & Lemyre, L. (1983). Duration discrimination of empty time intervals marked by intermodal pulses. *Perception & Psychophysics*, *34*, 541–548.
- Semjen, A., & Ivry, R. B. (2001). The coupled oscillator model of between-hand coordination in alternate-hand tapping: A reappraisal. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 251–265.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*, 788.
- Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *NeuroReport*, *12*, 7–12.
- Spence, C., Shore, D. I., & Klein, R. M. (2001). Multisensory prior entry. *Journal of Experimental Psychology: General*, *130*, 799–832.
- Spence, C., Baddeley, R., Zampini, M., James, R., & Shore, D. I. (2003). Multisensory temporal order judgments: When two locations are better than one. *Perception & Psychophysics*, *65*, 318–328.
- Stein, B. E. (1998). Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Experimental Brain Research*, *123*, 124–135.
- Trevarthen, C. (1999–2000). Musicality and the intrinsic motive pulse: evidence from human psychobiology and infant communication. *Musicae Scientiae* (special issue), 155–215.
- Tuller, B., & Kelso, J. A. S. (1989). Environmentally-specified patterns of movement coordination in normal and split-brain subjects. *Experimental Brain Research*, *75*, 306–316.
- Walker, J. T., & Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1327–1339.
- Yamanishi, Y., Kawato, M., & Suzuki, R. (1980). Two coupled oscillators as a model for the coordinated finger tapping by both hands. *Biological Cybernetics*, *37*, 219–225.