Psychological Research (2013) 77:388–398
DOI 10.1007/s00426-012-0441-0

ORIGINAL ARTICLE

Synchronization with competing visual and auditory rhythms: bouncing ball meets metronome

Michael J. Hove · John R. Iversen · Allen Zhang · Bruno H. Repp

Received: 2 February 2012 / Accepted: 11 May 2012 / Published online: 26 May 2012
© Springer-Verlag 2012

Abstract Synchronization of finger taps with periodically flashing visual stimuli is known to be much more variable than synchronization with an auditory metronome. When one of these rhythms is the synchronization target and the other serves as a distracter at various temporal offsets, strong auditory dominance is observed. However, it has recently been shown that visuomotor synchronization improves substantially with moving stimuli such as a continuously bouncing ball. The present study pitted a bouncing ball against an auditory metronome in a target–distracter synchronization paradigm, with the participants being auditory experts (musicians) and visual experts (video gamers and ball players). Synchronization was still less variable with auditory than with visual target stimuli in both groups. For musicians, auditory stimuli tended to be more distracting than visual stimuli, whereas the opposite was the case for the visual experts. Overall, there was no main effect of distracter modality. Thus, a distracting spatiotemporal visual rhythm can be as effective as a distracting auditory rhythm in its capacity to perturb synchronous movement, but its effectiveness also depends on modality-specific expertise.

M. J. Hove
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

J. R. Iversen
The Neurosciences Institute, San Diego, CA, USA

A. Zhang
Yale University, New Haven, CT, USA

B. H. Repp (✉)
Haskins Laboratories, 300 George Street, New Haven, CT 06511-6624, USA
e-mail: repp@haskins.yale.edu

Introduction

The visual modality is generally superior to the auditory modality in terms of spatial perception. When stimuli conveying spatial information in the two modalities are brought into conflict, visual dominance is usually observed. This is illustrated by the well-known ventriloquist effect: A visual stimulus strongly biases the spatial localization of a sound, but not vice versa (Bermant & Welch, 1976; Bertelson & Aschersleben, 1998; Bertelson & Radeau, 1981; Vroomen, Bertelson, & de Gelder 2001; Warren, Welch, & McCarthy 1981). Only when the visual stimulus is severely degraded does the auditory stimulus become dominant in spatial localization judgments (Alais & Burr, 2004). In contrast, audition is generally superior to vision in terms of temporal processing. In situations where the two modalities convey conflicting temporal information, auditory dominance is usually observed, which has led to the term "temporal ventriloquism" (Burr, Banks, & Morrone 2009; Fendrich & Corballis, 2001; Morein-Zamir, Soto- Faraco, & Kingstone, 2003). When the auditory stimulus is highly ambiguous, however, a reversal of the dominance may occur (Wada, Kitagawa, & Noguchi, 2003).

To account for such findings, as far as they were known at the time, Welch and Warren (1980) proposed an often-cited "modality appropriateness hypothesis" according to which the visual modality is dominant for spatial processes while the auditory modality is dominant for temporal processes. In recent years, however, evidence has been accumulating in support of a Bayesian optimal integration hypothesis (Alais & Burr, 2004; Ernst & Banks, 2001; Körding & Wolpert, 2004; van Beers, Sittig, & Gon 1999), according to which not modality as such but the relative precision of the information provided by two stimuli dictates their relative weights and dominance. Although this hypothesis has been applied most often to multimodal integration of information, it can also be extended to multimodal conflict situations in which integration is to be avoided. Although the modality appropriateness hypothesis would be unrealistic if it completely disregarded the relative precision of information in the different modalities, it does predict visual spatial dominance and auditory temporal dominance in conflict situations where the visual and auditory stimuli are similar in their information precision. In contrast, the optimal integration hypothesis predicts no dominance in such situations. The present study examined these predictions in a new bimodal conflict situation.

One task that has been used to assess auditory dominance in temporal processing is sensorimotor synchronization. Repp and Penel (2002) asked participants to tap in synchrony with an isochronous sequence of flashes. The flashes occurred in synchrony with an isochronous sequence of tones that was to be ignored. Temporal perturbations were introduced simultaneously in both sequences by shifting coincident events in opposite directions. It was found that participants responded to the changes in the auditory sequence despite having been told to synchronize with the visual sequence. This could have been due to temporal ventriloquism in perception: Perhaps the shifted flash was perceived as shifted in the direction of the shifted tone. Subsequently, however, Kato and Konishi (2006) presented auditory and visual sequences in antiphase (separated by nearly 500 ms), which made direct perceptual interactions unlikely. Nevertheless, they again observed auditory dominance in the participants' responses to perturbations in the sequence that served as a distracter.

To further investigate auditory dominance in a synchronization task, Repp and Penel (2004) varied the phase relationship (temporal offset) between isochronous auditory and visual sequences (tones and flashes, respectively). Either sequence served as the target for synchronization, while the other sequence served as the distracter that was to be ignored. Using two auditory sequences distinguished by pitch, Repp (2003a, 2004) had previously found that taps deviated from target tones in the direction of distracter tones when the temporal offset was less than about 150 ms, and that the deviation was stronger toward leading than toward lagging distracter tones. Repp and Penel found huge distracter effects of this kind with visual targets and auditory distracters, but hardly any effects in the reverse condition.

These results were perhaps not overly surprising because it had long been known (and was confirmed by Repp & Penel, 2004) that a sequence of flashes is more difficult to synchronize with than a sequence of tones (Chen, Repp, & Patel 2002; Fraisse, 1948; Kolers & Brewster, 1985; Repp, 2003b). In particular, the variability of taps is greater in the visual task, and it is difficult to maintain synchrony if the event rate exceeds 2 Hz (Repp, 2003b). Thus, the visual stimuli in the study of Repp and Penel (2004) entered the competitive target–distracter paradigm with a severe handicap. Both the modality appropriateness hypothesis and the optimal integration hypothesis would predict strong auditory dominance in such a case.[1] However, the findings raised the question of whether there might be other kinds of rhythmic visual stimuli that afford more accurate synchronization and would be more competitive with auditory stimuli. It seemed that moving stimuli would be required so that the visual information becomes spatiotemporal rather than purely temporal. Research on interception of moving visual targets has shown that temporal accuracy is very high in such tasks (Bootsma & van Wieringen, 1990; McBeath, Shaffer, & Kaiser, 1996; Tresilian, 1994).

Some recent studies have employed moving visual rhythmic stimuli in unimodal synchronization (finger tapping) tasks. Hove and Keller (2010) compared synchronization with a flashing visual stimulus to synchronization with alternating images of a finger in raised and lowered positions, which generated apparent movement. Tapping was significantly less variable with the apparent movement stimuli than with the flashes. Hove, Spivey, and Krumhansl (2010) employed an auditory metronome, a flashing square, and images of a bar or a finger that genuinely moved up and down, albeit at a constant velocity. The main dependent variable was the percentage of trials showing successful synchronization, according to a specified criterion, at each of several relatively fast tempi (intervals ranging from 500 down to 240 ms). Synchronization with the moving stimuli was more successful than with the flashing stimulus, but still not nearly as good as with the auditory metronome.

In a recent study, Iversen, Patel, Nicodemus, and Emmorey (2012) went one step further by asking the participants to synchronize with a synthetic video of a bouncing ball (with velocity varying according to a rectified sinusoid), a flashing square, and an auditory metronome. Variability of asynchronies between taps and the target stimuli was lower for the bouncing ball than for the flashing square. In fact, the bouncing ball yielded a variability that was not significantly larger than that with the auditory metronome. This was the first time that a rhythmic visual stimulus had been shown to approach an auditory metronome in the precision of its temporal information. It might be noted that the tempo in that study (intervals of 600 ms) was slower than the tempi used by Hove et al. (2010). The

[1] For applications of the optimal integration hypothesis to cue integration in a multimodal synchronization task, see Wing, Doumas, and Welchman (2010), and Elliott, Wing, and Welchman (2010, 2011).

greater effectiveness of a bouncing ball than a flashing light was evidently due to its continuous motion that enabled observers to predict the point of impact more accurately. The vertical direction of the ball movement toward the impact may also be facilitative because of its directional congruence with the tapping movement (Hove et al., 2010).

These new results (already known to us in 2010) stimulated us to ask whether the bouncing ball would be fully competitive with an auditory metronome in a bimodal target–distracter synchronization task with varying temporal offsets (Repp & Penel, 2004). According to the optimal integration hypothesis, this should indeed be the case: If the two stimuli afford equally stable synchronization as unimodal targets (a finding we hoped to replicate), they should also be equally effective as cross-modal distracters. The modality appropriateness hypothesis, however, still predicts stronger distracter effects from auditory than from visual stimuli because temporal processing is involved.

We initially conducted the experiment with musically trained participants because they happened to be readily available, being regular participants in experiments in author BHR's laboratory. However, because musicians' daily experience with auditory rhythms in their music-making may bias them strongly in favor of the auditory modality, we subsequently tested a second group of participants who had little music training but special visual expertise: video gamers and ball players. Although these activities do not usually involve synchronization with a visual rhythm, they are likely to sensitize people to spatiotemporal information in the visual modality. Therefore, we expected these visual experts to be more sensitive to visual than to auditory rhythms, or at least to show a smaller auditory bias than musicians. By incorporating and comparing these two participant groups, our study addressed the question of whether expertise relying on one or the other modality has any effect on performance in the bimodal target–distracter paradigm, and perhaps on synchronization with the unimodal stimuli as well.

Methods

Participants

The musicians ($N = 12$) included nine graduate students from the Yale School of Music (five females, four males, ages 21–27) who were regular participants in rhythm and timing experiments at Haskins Laboratories and practiced their primary instruments (piano-2, violin, viola-2, trombone, harp, and guitar-2) 28 h per week on average and three music students from the Volkshochschule Leipzig (two females, one male, ages 22–24) who played viola, flute, and trombone at an advanced level and practiced 23 h per week on average. The visual experts ($N = 17$) included nine video gamers (two females, seven males, ages 18–30) who regularly played games that required manual dexterity and high temporal precision (e.g., racing and sports games, ego-shooters, actions-per-minute games such as Starcraft) and played 24 h per week on average, and eight ball players (two females, six males, ages 21–30) who were currently playing on teams in various ball sports (basketball, volleyball, and soccer) and played 9 h per week on average. These participants were recruited with flyers at the University of Leipzig and advertisements posted on a gamer e-mail list. They had less than 1 year of formal music training, were not currently playing an instrument, and had no prior experience in tapping tasks. As the average results for video gamers and ball players were quite similar, they were treated as a single group of visual experts. All the participants were paid for their participation.

Materials and equipment

The computer-generated video of a continuously bouncing ball was similar to that used by Iversen et al. (2012). A realistic image of a grey basketball, 2 cm in diameter, oscillated vertically against a black background. The ball moved 4 cm according to a rectified sinusoidal velocity function (i.e., with maximal velocity at impact), with a period of 600 ms. At its lowest position (the bounce) the ball touched a stationary white bar 3.2 cm wide and 2 mm high. The video was shown in a window 10 cm wide and 7.6 cm high at 60 frames per second. A program written in Max/MSP 4.6.3 controlled the experiment and recorded the times of the taps.

There were two conditions: auditory target with visual distracter (A-V) and visual target with auditory distracter (V-A). Each condition included 14 bimodal videos, created by adding a sequence of pure tones (262 Hz, 50 ms duration) with a period of 600 ms in various constant temporal relationships to the ball bounces, and one unimodal video (i.e., without distracter stimuli) in which tones sounded while the ball remained stationary in its highest position (in the A-V condition) or in which the bouncing ball was not accompanied by tones (in the V-A condition). The temporal offsets in the bimodal videos ranged from −250 ms (tone leading) to 300 ms (tone lagging) in steps of 50 ms (including 0), and including ±25 ms as well. During presentation, the temporal offsets were slightly jittered and/or shifted (by less than −10 ms on average) due to the fact that the video frames had to wait for the next refresh cycle of the LCD monitor (60 Hz in New Haven, 75 Hz in Leipzig). However, there were no visible irregularities in ball movement, and all offsets were similarly affected.

Each bimodal A-V video contained 42 tones and 34 bounces. While the first eight tones sounded, the ball remained stationary in its highest position. Conversely, each bimodal V-A video contained 42 bounces and 34 tones. The first eight bounces of the ball were accompanied by silence. The reason for these unimodal lead-ins was that with the MAX software the video started playing after a variable and unknown delay, which made it impossible to measure asynchronies between taps and target stimuli (tones or bounces) accurately. The lead-ins provided a unimodal baseline for asynchronies in each individual trial, which was subtracted from the asynchronies in the later bimodal part of the trial, thereby removing any constant delay (albeit at the cost of introducing additional variability).

In bimodal videos of both conditions, the white bar on which the ball bounced occasionally turned grey for a brief period (267 ms). This occurred 1–3 times randomly during each video but never in close succession.

Procedure

In both venues (New Haven and Leipzig), the participants sat in front of a computer monitor, listened over Sennheiser HD280 pro headphones, and tapped with the index or middle finger of their preferred hand on a Roland SPD-6 electronic percussion pad that they held on the lap. They started each trial by pressing the space bar; the video started playing some 3 s later. They were instructed to tap in synchrony with the target stimuli (tones or bounces, depending on the condition), starting with the third stimulus in each trial, and to ignore the distracter stimuli. To make sure that they kept watching the video in the A-V condition, the participants were required to report at the end of each trial how often the white bar on which the ball bounced had turned grey. This response was also required in the V-A condition and was made by clicking a numbered box on the computer screen.[2]

The A-V and V-A conditions constituted separate 1-h sessions whose order was counterbalanced. They were usually at least 1 week apart. Each session consisted of six blocks of 15 trials each (the different videos described above, presented in randomized order). At the end of each block, the participant saved the data in a file.

Analysis

We assessed distracter effects in two ways: in terms of (1) the mean and (2) the variability of the relative phase (or asynchrony) of the taps with the target stimuli as a function of the temporal offset between targets and distracters. The data for completely unimodal trials were analyzed separately.

We used circular statistical methods because they are less sensitive to outliers than standard mean asynchrony calculations. Each tap was mapped onto a unit circle in terms of its phase relative to the target stimuli. The first six taps in each trial coincided with unimodal target stimuli; we ignored the first two taps and computed the average relative phase of the remaining four, which yielded a unimodal baseline. We then calculated the average relative phase of all subsequent taps and subtracted the unimodal baseline from it. We converted these relative phase shifts (expressed in degrees or radians) to relative asynchronies in milliseconds because these values are easier to grasp and there is evidence from previous research that distracter effects are tied to absolute temporal separation, not relative phase (Repp, 2004).

We also computed the variability of relative phases for the subsequent taps in terms of their circular variance (CV).[3] Circular variance indexes the stability of tap-to-target coordination on a scale from 1 (unstable tapping with phases distributed uniformly around the unit circle) to 0 (perfectly stable tapping with a unimodal distribution of phases). The relative asynchronies and CVs were averaged across the six repetitions of each trial type (i.e., across blocks) for each participant. If the tapping in a trial was highly variable (defined here as CV > 0.15; 3 % of all trials), that trial was not included in the average relative asynchrony calculation for that trial type, but it was included in the average CV calculation.

Results

Musicians

Relative asynchronies

Changes in the mean relative asynchrony as a function of target–distracter offset constituted our primary measure of distracter effects. Since distracters attract taps (Repp, 2003a, b, 2004; Repp & Penel, 2004), the taps were

[2] Participants' accuracy in those reports was considered satisfactory (92.3 % correct on average for musicians, 95.5 % correct for visual experts). One musician forgot to report the numbers but affirmed that she had watched the videos at all times; her data were retained.

[3] An approximation to the linear standard deviation (SD) can be obtained by first calculating the circular standard deviation CSD = sqrt(2 × CV) (Fisher, 1993, p. 33) and then, since we had previously calculated the linear SD for the nine Yale musicians, determining the exact relationship between CSD and SD by linear regression of the mean values of these participants. The equation turned out to be SD = 98.36 × CSD – 0.45, $R^2 = 0.999$. According to that formula, CVs of 0.02, 0.03, and 0.04 correspond to SDs of 19.2, 23.6, and 27.4 ms, respectively.

expected to occur earlier when the distracter preceded the target, resulting in more negative asynchronies, and to occur later when the distracter followed the target, resulting in more positive asynchronies relative to the trials in which the target and distracter were exactly simultaneous. These effects were expected to be maximal at a certain temporal offset (in the vicinity of ±100 ms, according to the previous studies just cited) and to decrease as the offset increased further. The question of interest was whether the magnitude of these distracter effects would be different in the A-V and V-A conditions. The relative asynchrony in the unimodal trials was of little interest; it was expected to be close to zero unless the asynchrony changed between the early (baseline) and later portions of a trial.

Figure 1a shows the musicians' mean relative asynchrony as a function of distracter lead/lag in the A-V condition. The dashed horizontal line is drawn through the mean relative asynchrony in the simultaneous bimodal condition (zero lag). Relative to that reference, visual distracters had a decidedly asymmetric effect, with leading distracters attracting the taps more strongly than lagging distracters did.[4] The maximal distracter effects occurred at somewhat longer temporal separations than predicted.

Figure 1b shows the musicians' relative asynchrony data for the V-A condition, which exhibit a more complex pattern. Unexpectedly, the distracter function has two peaks and two valleys, suggesting that auditory distracters had an effect not only when they deviated from an in-phase relationship (lead/lag of 0) but also when they deviated from an anti-phase relationship (lead/lag of 300 ms). In each case, lagging auditory distracters seemed to have a stronger effect than leading distracters.

To quantify the relative strength of the visual and auditory distracter effects, we calculated the range of values of the mean relative asynchrony (i.e., the maximum minus the minimum) in the distracter functions for each individual participant and compared the mean ranges. They were not significantly different, $t(11) = 0.63$, $p = 0.544$. Hence, according to this measure there was no difference in the magnitude of auditory and visual distracter effects for the musicians.

The mean relative asynchronies in the unimodal auditory and visual trials were not significantly different from zero, as expected. Thus the asynchrony did not seem to change between the initial and later parts of a trial. The mean relative asynchrony in the simultaneous A-V condition (Fig. 1a) was likewise near zero. However, the mean relative asynchrony in the simultaneous V-A condition (Fig. 1b) was significantly negative (−15.2 ms), $t(11) = 3.62$, $p < 0.005$. This suggests that the simultaneous auditory distracter advanced the taps, perhaps because the real mean unimodal asynchronies, which we could not measure directly here, were more negative for the auditory than for the visual stimuli. (Iversen et al., 2012, however, had found the opposite.)

Circular variance

The bimodal CVs were expected to be smallest in the vicinity of zero lead/lag, to increase with lead/lag up to the point of maximal distraction, and then to decrease (i.e., to vary as an M-shaped function of lead/lag; cf. Repp, 2004). The magnitude of the change in CV (i.e., the range of CV values across all leads/lags) provided a second measure of distracter effectiveness. The CV in the unimodal trials was of interest: Did we replicate the finding of Iversen et al. (2012) of equally tight synchronization with auditory and visual stimuli?

Figure 1c shows the musicians' mean CV as a function of visual distracter lead/lag in the A-V condition. The function has roughly the expected shape, with a minimum near zero and an increase up to ±150 ms, though there was not much of a decrease at longer leads/lags.

Figure 1d shows the analogous data for the V-A condition. Here the M-shape is more pronounced, though here too, variability at the extremes (lead/lag of 300 ms) remained higher than in the simultaneous condition. Interestingly, the CV function does not give any indication of the complex shape of the asynchrony function in Fig. 2b. The error bars indicate large individual differences in the CV for visual target stimuli.

The mean unimodal CV was significantly higher for visual than for auditory stimuli, 0.031 versus 0.020, $t(11) = 3.18$, $p = 0.009$. Thus, we did not replicate exactly the non-significant result of Iversen et al. (2012), although their data did show a tendency favoring the auditory modality. The significant difference may reflect the auditory expertise of our musician participants. The mean bimodal CVs, too, were higher in the V-A than in the A-V condition, $t(11) = 2.76$, $p = 0.019$. The mean CV in the

[4] It could be argued that the mean relative asynchrony for the 300 ms distracter lead/lag (this data point being duplicated at −300 and +300 ms in the figure) is a better reference because distracter effects should be minimal at the separation of half a cycle. With that reference, the distracter effects appear more nearly symmetric, but then it would seem that visual distracters exerted an effect at the zero lag, making taps occur later than they otherwise would. A possible reason for this could be that the point of subjective simultaneity of tones and ball bounces actually corresponded to a slight lead of the ball bounce, so that the bounces were perceived as lagging the tones when they were physically simultaneous (cf. Arrighi, Alais, & Burr, 2005, 2006; Petrini et al., 2009). A related possibility is that the real mean asynchrony (which we could not assess because of the video delays) was less negative (or more positive) for unimodal bounces than for unimodal tones, so the asynchrony shifted in the positive direction when the two stimuli occurred simultaneously. However, in the Iversen et al. (2012) study, the asynchrony for the bouncing ball was considerably *more* negative than for tones: −75 ms versus −8 ms.

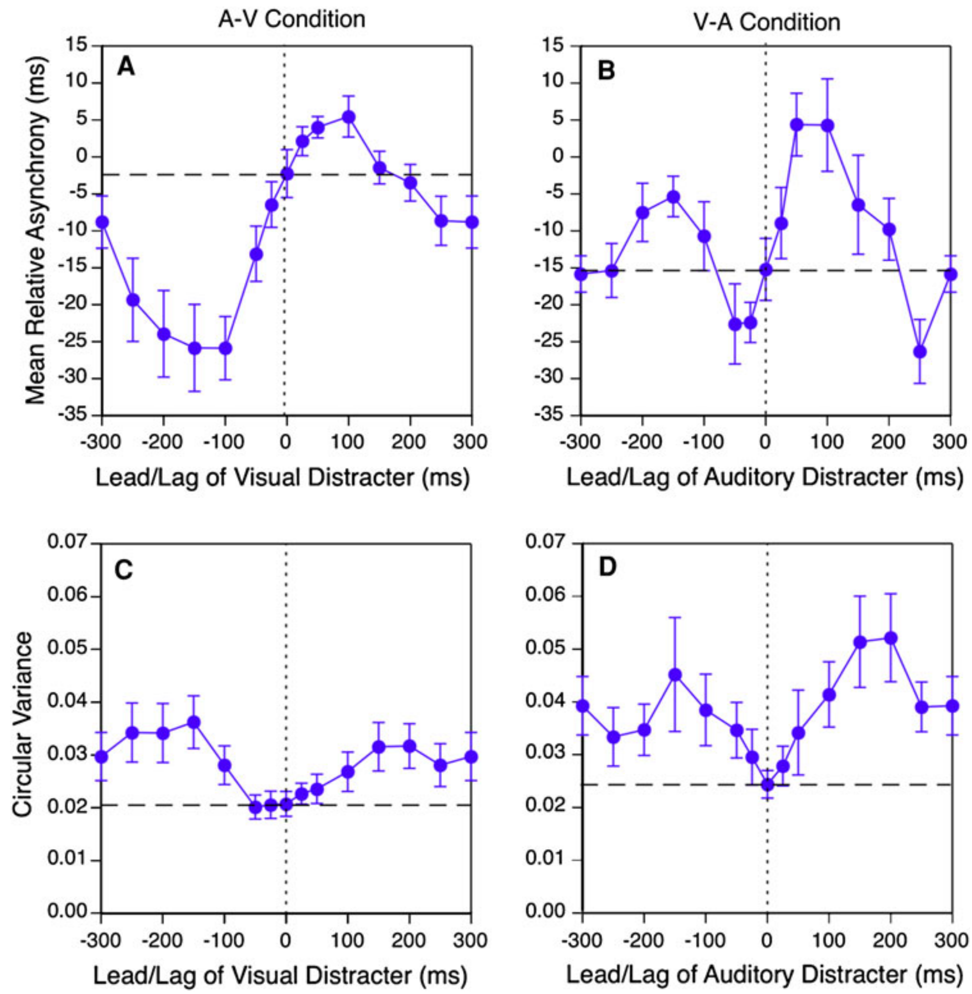

Fig. 1 Results of musicians: Mean relative asynchrony (**a**, **b**) and circular variance (**c**, **d**) as a function of distracter lead/lag in the A-V and V-A conditions. The *dashed horizontal line* is drawn through the simultaneous bimodal condition. The *error bars* represent standard errors

simultaneous A-V condition (0.021) was similar to that in unimodal auditory trials, but the mean CV in the simultaneous V-A condition (0.024) was significantly lower than the mean unimodal visual CV, $t(11) = 2.53$, $p = 0.028$, suggesting that the simultaneous auditory distracters stabilized synchronization with visual targets. Finally, the range of bimodal CV variation across temporal offsets (maximum minus minimum CV) was larger in the V-A condition than in the A-V condition, $t(11) = 2.38$, $p = 0.037$. Thus, according to this measure, musicians were affected more by auditory than by visual distracters.

To see whether the difference between the unimodal visual and auditory CVs for individual participants predicted the relative strength of their visual and auditory distracter effects in bimodal trials, correlations were computed between the unimodal CV differences (visual minus auditory) and the differences between our two measures of the distracter effect: the bimodal relative asynchrony range and the CV range (V-A minus A-V). The first correlation was near zero, $r(10) = 0.05$, but the second correlation was significant, $r(10) = 0.67$, $p < 0.05$. This means that a musician whose synchronization was more variable with unimodal visual stimuli than with unimodal auditory stimuli also tended to show a greater range of variability for bimodal visual targets than for bimodal auditory targets, suggesting greater interference from auditory than from visual distracters.

Visual experts

Relative asynchronies

Figure 2a shows the visual experts' data for the A-V condition. The distracter function had the expected shape and, like that of the musicians, suggested greater effects of leading than of lagging visual distracters. Note, however, the different scale of the ordinate: These distracter effects were much larger than those shown by the musicians! (A statistical comparison between the two participant groups will be presented later.)

In the V-A condition (Figure 2b), the visual experts' distracter function had a simple shape, similar to the A-V condition and unlike that of the musicians. There was no indication here of any effect of anti-phase auditory distracters. The auditory distracter effects seem smaller

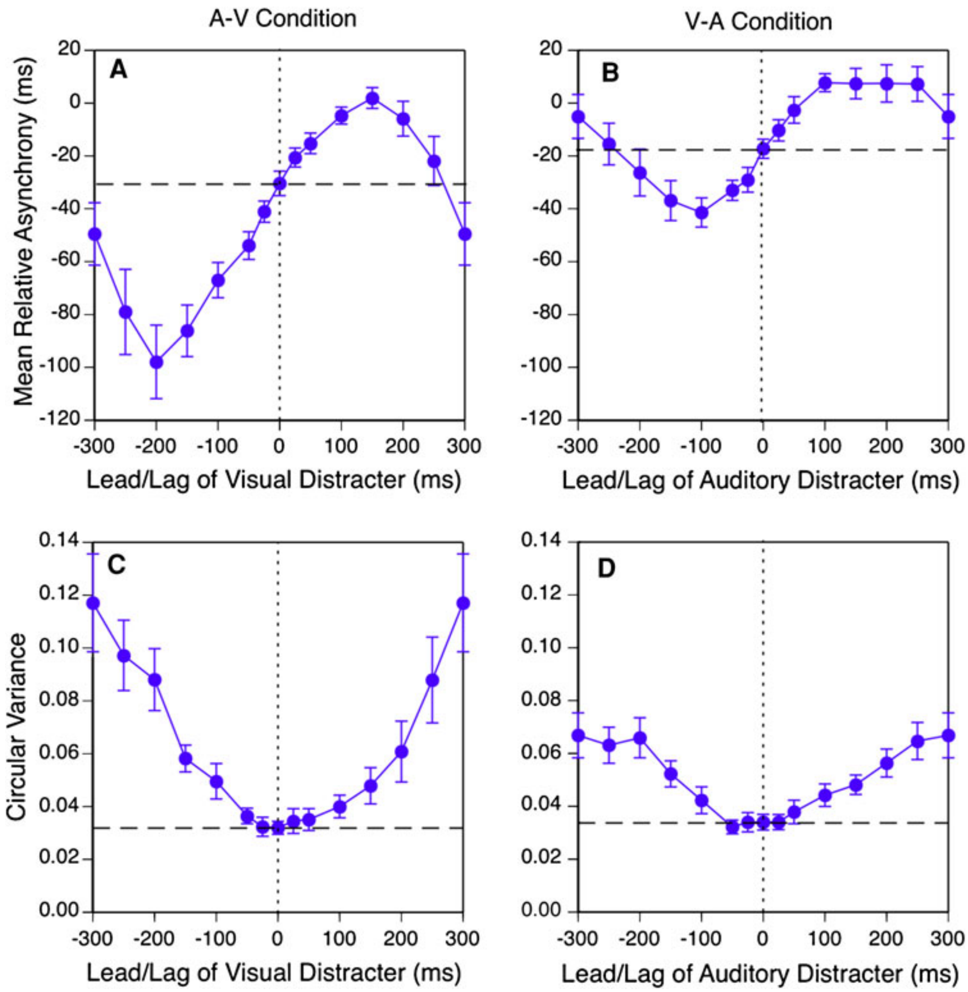

Fig. 2 Results of visual experts: Mean relative asynchrony (**a**, **b**) and circular variance (**c**, **d**) as a function of distracter lead/lag in the A-V and V-A conditions. The *dashed horizontal line* is drawn through the simultaneous bimodal condition. Note that the scales of the ordinates are different from those in Fig. 1. The *error bars* represent standard errors

than the visual ones, but a comparison of the mean ranges of relative asynchronies revealed no significant difference, $t(16) = 1.53$, $p = 0.146$. This was due to large individual differences, especially in the A-V condition.

The mean unimodal relative asynchronies were −18.5 and −13.6 ms for auditory and visual targets, respectively, and the mean relative asynchronies in the simultaneous A-V and V-A conditions were also significantly negative (−30.4 and −17.3 ms, respectively). This indicates an increasing tendency to tap ahead of target stimuli, following the initial six taps.

Figure 2c, d shows the visual experts' mean CVs. Unlike the musicians' CV functions, these functions were not M-shaped but more nearly U-shaped, especially in the A-V condition. Evidently, distracters in anti-phase destabilized synchronization performance of these non-musicians. The mean CV was larger in the A-V than in the V-A condition, though not significantly so, $t(16) = 1.80$, $p = 0.090$, again because of large individual differences. However, the mean range of the CV was significantly larger in the A-V than in the V-A condition, $t(16) = 2.69$, $p = 0.016$, indicating greater interference from visual than from auditory distracters.

The visual experts' mean CV was slightly larger for unimodal visual than for unimodal auditory stimuli, 0.042 versus 0.034, but the difference was not significant, $t(16) = 1.68$, $p = 0.113$. As with the musicians, the mean CV was smaller in the simultaneous V-A condition than in the unimodal visual condition, though again the difference was not significant here, $t(16) = 1.60$, $p = 0.130$. The mean CV in the simultaneous A-V condition closely matched the unimodal auditory CV.

Of the correlations between the unimodal CV difference and the two measures of the relative strengths of bimodal distracters, the first was marginally significant, $r(15) = 0.48$, $p = 0.05$, and the second was non-significant but not much lower, $r(15) = 0.39$, $p > 0.10$. Thus there was potentially a relationship between unimodal synchronization variability and bimodal modality dominance.

Comparison of musicians and visual experts

We compared the two measures of distracter effects: the range of mean relative asynchronies and the range of CVs across all bimodal temporal offsets. For the range of mean relative asynchronies, a mixed-model ANOVA with distracter modality (auditory, visual) as the within-participant variable and group (musicians, visual experts) as the between-participant variable showed the distracter effects to be significantly smaller for the musicians than for the visual experts, $F(1,27) = 33.54$, $p < 0.001$. There was no main effect of distracter modality, $F(1, 27) = 1.06$, $p = 0.313$, and the interaction was also non-significant, $F(1, 27) = 2.05$, $p = 0.164$. For the range of bimodal CVs, a similar ANOVA again showed the distracter effects to be clearly smaller for musicians than visual experts, $F(1, 27) = 22.60$, $p < 0.001$. There was no main effect of distracter modality, but the interaction was significant, $F(1, 27) = 9.07$, $p = 0.006$, confirming our earlier observation that according to this second measure visual experts showed larger visual than auditory distracter effects, whereas musicians showed the opposite.

A third ANOVA of the same type was conducted on the CV for unimodal trials. The musicians tapped with significantly less variability than the visual experts, $F(1, 27) = 9.38$, $p = 0.005$, and synchronization was less variable with auditory than with visual stimuli, $F(1, 27) = 8.71$, $p = 0.006$. There was no significant interaction $F(1, 27) = 0.197$, $p = 0.66$, hence no effect of modality-specific expertise.

Finally, the mean bimodal CV was analyzed. It was significantly lower for musicians than for visual experts, $F(1, 27) = 17.35$, $p < 0.001$, but there was no main effect of modality, $F(1, 27) = 0.00$, $p > 0.98$. Instead, the interaction was significant, $F(1, 27) = 7.29$, $p = 0.012$, because musicians showed higher variability with visual target stimuli, whereas visual experts showed higher variability with auditory target stimuli. However, this was not a cross-over interaction: While musicians obviously had much lower variability than visual experts with auditory stimuli, they also tended to have lower variability with visual stimuli, $F(1, 27) = 3.97$, $p = 0.057$.

Discussion

In this study we used a bimodal target–distracter synchronization paradigm to pit a periodically moving visual stimulus, a bouncing ball, against an auditory metronome. Recent findings by Iversen et al. (2012) had led us to expect that the variability of synchronization with the unimodal visual and auditory stimuli would be similar. If so, the optimal multimodal integration (or competition) hypothesis predicts that the two stimuli would be equally effective distracters in the bimodal task, whereas the modality appropriateness hypothesis still predicts auditory dominance. We discuss first the overall results for both participant groups combined. Then we consider effects of expertise.

To begin with, we did not replicate the finding of stimulus equivalence: Variability of unimodal synchronization was significantly higher with visual than with auditory stimuli, although the difference was small and due primarily to the musicians. Our results do not really conflict with those of Iversen et al. (2012) because these authors, too, had found a tendency in the same direction, only it did not reach significance. However, given that we found a significant difference, the optimal integration hypothesis now predicts slight auditory dominance in the bimodal situation.

We did not find such dominance. If anything, there was a tendency for the visual stimuli to be more distracting, but the modality difference in distracter effects was not significant overall, neither in terms of mean relative asynchrony nor in terms of CV. This result is problematic for both hypotheses, but especially for the modality appropriateness hypothesis: A visual stimulus that was slightly inferior to the auditory stimulus as a synchronization target nevertheless proved to be an equal competitor. It could be argued, however, that the modality appropriateness hypothesis concerns only purely temporal or purely spatial stimuli. While the auditory metronome was purely temporal and discrete, the bouncing ball provided continuous spatiotemporal information that may have engaged capabilities for which the visual modality is specialized. Therefore, the results perhaps do not speak directly to the modality appropriateness hypothesis. According to the optimal integration hypothesis, the auditory stimulus should have been somewhat more distracting than the visual one. While that was not the case overall, some weak support for the hypothesis came from the observation that, within each participant group, the unimodal variability difference between auditory and visual stimuli tended to predict the relative strength of these stimuli as distracters in the bimodal task. Possibly, the fact that there was a secondary visual task (counting how often the white bar turned grey) but no secondary auditory task could have worked in favor of the visual stimuli as distracters. While this procedure was necessary to force the participants to keep their gaze on the bouncing ball and (in the absence of continuous gaze monitoring) to provide objective proof that they did so, it may have diverted attentional resources from the auditory stimuli.

Regardless of these considerations, what is clear from the present results is that a bouncing ball can be as effective as an auditory metronome in entraining participants' movements, in agreement with Iversen et al. (2012). Moving stimuli used in the previous study by Hove et al. (2010) may have been less effective because they had less realistic (linear) velocity profiles and were presented at faster tempi. However, those stimuli were not tested in bimodal competition with an auditory metronome, and it is possible that they would be stronger competitors than might be predicted from unimodal synchronization performance. Nevertheless, our results suggest that a periodically bouncing ball can convey significant rhythmic information and can strongly engage human movement. Possibly, such a visual stimulus could even convey more complex (non-isochronous) rhythms and a metrical beat, contrary to what Patel et al. (2005) concluded about visual flashes. This remains to be investigated. Note that more complex rhythmic information conveyed by a bouncing ball would again have spatiotemporal, not purely temporal correlates, due to the tight relationship between movement amplitude (trajectory length) and cycle duration when velocity is appropriately constrained.

We also investigated the role of expertise by testing groups of auditory and visual experts. We should acknowledge immediately that the musicians were not simply auditory experts but auditory sensorimotor synchronization experts, mainly from playing music with a metronome or in ensembles, and they even had some experience synchronizing with visual stimuli such as a conductor's baton. (Their participation in previous laboratory experiments requiring synchronization with auditory stimuli probably added little to their expertise.) Their predominantly auditory expertise was reflected in their significantly lower tapping variability with the auditory metronome, compared to the visual experts. However, they were also less variable than the visual experts in tapping with the bouncing ball, especially in the unimodal condition. The visual experts' visual expertise seemed to be of little help in synchronizing with the bouncing ball, probably because their expertise was not with periodic visual stimuli but rather with non-rhythmic interception of moving objects. In contrast, the musicians' expertise in sensorimotor synchronization seemed to extend to the visual rhythm conveyed by the bouncing ball. This is noteworthy because a previous study (Repp, 2003b) provided little indication that synchronizing with visual flashes benefited from musical experience. (However, see Krause et al., 2010, for a contrary result.) Unlike the purely temporal information conveyed by these stimuli, the spatiotemporal rhythm of the bouncing ball may be more akin to that of an auditory metronome in that it engages similar internal processes. A recent fMRI study indicates that timing networks, including the putamen, are similarly active during synchronization with auditory and spatiotemporal visual stimuli, but not with flashing visual stimuli (Hove, Fairhurst, Kotz, & Keller, in preparation).

Effects of expertise were also evident in the cross-modal distracter effects. The musicians were generally less affected by distracters than were the visual experts, which may be due to the musicians' general rhythmic skill. More specifically, however, some of the non-musician visual experts were much more distracted by visual than by auditory stimuli, whereas the musicians tended to show the opposite. These effects were statistically reliable only in terms of the increases in tapping variability caused by the distracters. The other measure of distracter effects, namely the attraction of the taps to leading and lagging distracter stimuli, did not show a significant group difference, due to large inter-individual variability. Nevertheless, it can be concluded that the pre-experimental auditory or visual experience of our participants was reflected in their relative ability to ignore auditory or visual distracter stimuli.

A few secondary findings warrant discussion as well. Only musicians' taps were attracted to auditory stimuli that were close to being in anti-phase with visual target stimuli. Related results have been reported previously by Repp (2004) and also by Kato and Konishi (2006), although the participants in the latter study were not identified as musicians. Our results suggest that, when synchronizing with visual targets that alternated with auditory distracters, the musicians automatically took the auditory distracters as an additional temporal reference. Curiously, however, this tendency was not reflected in the musicians' pattern of variability. Visual experts showed no hint of being attracted toward anti-phase distracters, but their variability was increased greatly when auditory distracters were in anti-phase. Visual distracters in anti-phase with auditory targets also increased variability, particularly in the visual experts. In neither group was there a tendency for anti-phase visual distracters to attract taps, even though the bouncing balls (unlike the metronome) reversed direction midway between two bounces. Evidently that extra visual event did not play any special role.

Both groups of participants showed reduced variability when auditory distracters coincided with visual targets, relative to unimodal visual targets. The reduced variability was similar to the variability with unimodal auditory targets, suggesting that participants effectively synchronized with the auditory distracters (cf. Repp & Penel, 2002). When visual distracters coincided with auditory targets, however, the variability remained similar to that with unimodal auditory targets. These specific findings suggest auditory dominance in the bimodal simultaneous condition, in both participant groups.

The pattern of changes in asynchrony with distracter leads/lags gave rise to a suspicion that the subjective point of synchrony between tones and bounces occurred when the bounces led the tones by a few tens of milliseconds and/or that the unimodal asynchronies were more negative for tones than for bounces, which would be consistent with a faster processing time or earlier perceived time of occurrence of tones compared to bounces. While technical limitations in this study prevented us from measuring actual asynchronies, there is good evidence in the literature on cross-modal temporal order and synchrony judgment that auditory stimuli indeed need to lead the lowest point in the trajectory of a periodically moving visual stimulus to be perceived as synchronous (Arrighi et al., 2005, 2006; Petrini et al., 2009).

In conclusion, this study presents new evidence that a visual spatiotemporal rhythm can entrain human movement nearly as well as an auditory metronome and can serve as an effective cross-modal distracter during synchronization with a metronome. The potential of more complex spatiotemporal stimuli to convey rhythmic and metrical information remains to be explored.

Acknowledgments This research was supported by National Science Foundation Grant BCS-0924206 to BHR and by the Max Planck Society. The authors are grateful to Yi-Huang Su for helpful comments on the manuscript

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257–262.

Arrighi, R., Alais, D., & Burr, D. (2005). Neural latencies do not explain the auditory and audio-visual flash-lag effect. *Vision Research, 45*, 2917–2925.

Arrighi, R., Alais, D., & Burr, D. (2006). Perceptual synchrony of audiovisual streams for natural and artificial motion sequences. *Journal of Vision, 6*, 260–268.

Bermant, R. I., & Welch, R. B. (1976). Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. *Perceptual and Motor Skills, 43*, 487–493.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin and Review, 5*, 482–489.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception and Psychophysics, 29*, 578–584.

Bootsma, R. J., & van Wieringen, P. C. W. (1990). Timing an attacking forehand drive in table tennis. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 21–29.

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in perception of interval duration. *Experimental Brain Research, 198*, 49–57.

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.

Elliott, M. T., Wing, A. M., & Welchman, A. E. (2010). Multisensory cues improve sensorimotor synchronisation. *European Journal of Neuroscience, 31*, 1–8.

Elliott, M. T., Wing, A. M., & Welchman, A. E. (2011). The effect of ageing on multisensory integration for the control of movement timing. *Experimental Brain Research, 213*, 291–298.

Ernst, M., & Banks, M. (2001). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429–433.

Fendrich, R., & Corballis, P. (2001). The temporal cross-capture of audition and vision. *Attention, Perception & Psychophysics, 1*, 719–725.

Fisher, N. I. (1993). *Statistical analysis of circular data*. Cambridge, UK: Cambridge University Press.

Fraisse, P. (1948). Rythmes auditifs et rythmes visuels (Visual and auditory rhythms). *L'Anneé Psychologique, 49*, 21–41.

Hove, M. J., Fairhurst, M. T., Kotz, S. A., & Keller, P. E. (in preparation). Synchronizing with auditory and visual rhythms: a reassessment of modality differences with fMRI.

Hove, M. J., & Keller, P. E. (2010). Spatiotemporal relations and movement trajectories in visuomotor synchronization. *Music Perception, 28*, 15–26.

Hove, M. J., Spivey, M. J., & Krumhansl, C. L. (2010). Compatibility of motion facilitates visuomotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 1525–1534.

Iversen, J. R., Patel, A. D., Nicodemus, B., & Emmorey, K. (2012). *Synchronization to auditory and visual rhythms in hearing and deaf individuals.* Manuscript submitted for publication.

Kato, M., & Konishi, Y. (2006). Auditory dominance in the error correction process: a synchronized tapping study. *Brain Research, 1084*, 115–122.

Kolers, P. A., & Brewster, J. M. (1985). Rhythms and responses. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 150–167.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Acta Psychologica, 133*, 28–37.

Krause, V., Pollok, B., & Schnitzler, A. (2010). Perception in action: The impact of sensory information on sensorimotor synchronization in musicians and non-musicians. *Acta Psychologica, 133*, 28–37.

McBeath, M. K., Shaffer, D. M., & Kaiser, M. K. (1996). On catching fly balls. *Science, 273*, 256–259.

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Cognitive Brain Research, 17*, 154–163.

Patel, A. D., Iversen, J. R., Chen, Y., & Repp, B. H. (2005). The influence of metricality and modality on synchronization with a beat. *Experimental Brain Research, 163*, 226–238.

Petrini, K., Dahl, S., Roccesso, D., Waadeland, C. H., Avanzini, F., Puce, A., et al. (2009). Multisensory integration of drumming actions: Musical expertise affects perceived audiovisual asynchrony. *Experimental Brain Research, 198*, 339–352.

Repp, B. H. (2003a). 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. (2003b). 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, 35*, 355–370.

Repp, B. H. (2004). On the nature of phase attraction in sensorimotor synchronization with interleaved auditory sequences. *Human Movement Science, 23*, 389–413.

Repp, B. H., & Penel, A. (2002). Auditory domination in temporal processing: New evidence from synchronization with simultataneous visual and auditory sequences. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 1085–1099.

Repp, B. H., & Penel, A. (2004). Rhythmic movement is attracted more strongly to auditory than visual rhythms. *Psychological Research, 68*, 252–270.

Tresilian, J. R. (1994). Perceptual and motor processes in interceptive timing. *Human Movement Science, 13*, 335–373.

van Beers, R., Sittig, A., & Gon, J. (1999). Integration of proprioceptive and visual position-information: an experimentally supported model. *Journal of Neurophysiology, 81*, 1355–1364.

Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception and Psychophysics, 73*, 651–659.

Wada, Y., Kitagawa, N., & Noguchi, K. (2003). Audio-visual integration in temporal perception. *International Journal of Psychophysiology, 50*, 117–124.

Warren, D., Welch, R., & McCarthy, T. (1981). The role of visual-auditory compellingness in the ventriloquist effect: Implications for transitivity among the spatial senses. *Perception and Psychophysics, 30*, 557–564.

Welch, R., & Warren, D. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*, 638–667.

Wing, A., Doumas, M., & Welchman, A. E. (2010). Combining multisensory temporal information for movement synchronisation. *Experimental Brain Research, 200*, 277–282.