Exp Brain Res (2010) 203:249–260
DOI 10.1007/s00221-010-2224-4

RESEARCH ARTICLE

Predicting the position of moving audiovisual stimuli

Steven L. Prime · Laurence R. Harris

Received: 10 August 2009 / Accepted: 9 March 2010 / Published online: 30 March 2010

Abstract Predicted motion (PM) tasks test the accuracy of predicting the future position of a moving target. Previous PM studies using audiovisual stimuli have suggested that observers rely primarily on visual motion cues. To clarify the role of auditory signals in predicting future positions of bimodal targets, we designed a novel PM task where spatial coincidence of audio and visual motion signals was varied in three conditions: auditory and visual motion stimuli were spatially correlated (*congruent condition*), the auditory motion stimulus was moving behind the visual motion stimulus (*sound-trailing condition*), or the auditory motion stimulus was moving ahead the visual motion stimulus (*sound-leading condition*). We manipulated target speed (5.5 or 11 cm/s), the time that the moving audiovisual stimulus was presented (500 or 750 ms viewing time), and the time the visual stimulus disappeared while the auditory stimulus continued to move by itself before prompting subjects to estimate the position of the visual stimulus would have traveled if it continued along with the auditory stimulus (750, 1,000, or 1,500 ms prediction time). We also included two unimodal control conditions: *visual-only* and *auditory-only*. Subjects ($n = 12$) typically overestimated the target position of *congruent* bimodal targets. In the *sound-trailing* and *sound-leading* conditions, pointing responses were biased in the direction of the auditory stimulus, showing that PM performance is not reliant solely upon visual motion cues. We conclude that putative cognitive extrapolation mechanisms assume spatial coherence of bimodal motion signals and may perform some averaging of these motion signals when they do not spatially coincide.

Keywords Prediction motion · Cross-modal processing · Multi-sensory integration · Motion perception

S. L. Prime · L. R. Harris (✉)
Department of Psychology, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada
e-mail: harris@yorku.ca

Present Address:
S. L. Prime
Department of Psychology, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

Introduction

Predicting some future position of a moving object is necessary when performing many everyday tasks such as catching a ball, crossing a busy street, or manoeuvring a car through traffic. One common experimental task used in the laboratory to study spatiotemporal predictions of moving objects is the predicted motion (PM) task (e.g., Alderson and Whiting 1974; Ellingstad 1967; Gottsdanker 1955; Huber and Krist 2004; Peterken et al. 1991; Slater-Hammel 1955). In a typical PM task, subjects judge when a moving stimulus would have reached a specified location after the stimulus disappears, assuming it continued at the same speed and in the same trajectory. To perform this task subjects need to extrapolate the stimulus's future position. They base their judgements on a variety of motion cues (DeLucia and Liddell 1998; DeLucia et al. 2003; Jagacinaki et al. 1983).

Most PM studies have only considered unimodal motion cues—either visual motion (Alderson and Whiting 1974; Ellingstad 1967; Gottsdanker 1955; Huber and Krist 2004; Peterken et al. 1991; Slater-Hammel 1955) or auditory motion (Rosenblum et al. 1987, 1993, 2000). Since tracking tasks in the real world often involve moving objects that can be tracked by both sight and sound (e.g., cars or mosquitoes), it is reasonable to suppose that observers

might use both auditory and visual information the better to predict an object's future position rather than relying only on one sensory modality. Though audiovisual motion perception has been extensively studied in a variety of other motion tasks (for recent review see: Soto-Faraco et al. 2004), few studies have investigated spatiotemporal predictions of audiovisual stimuli in a PM task.

To take advantage of all the information available to estimate the future position of moving stimuli, ideally the motion signals from all relevant modalities would be integrated. However, it appears that this is not done. The few bimodal PM studies that have been carried out show that spatiotemporal predictions of audiovisual stimuli are indeed more accurate than for unimodal auditory stimuli but are only equally as accurate as unimodal visual stimuli (Hofbauer et al. 2004; Schiff and Oldak 1990) i.e., the presence of auditory cues does not seem to enhance performance. These findings might suggest that subjects rely primarily on visual motion cues in a PM task, even when auditory information is available. However, despite this apparent lack of contribution by auditory motion cues embedded within a bimodal stimulus, bimodal PM performance does suffer when auditory cues are uncorrelated or incongruent with the visual motion signals (Gordon and Rosenblum 2005). This finding shows that auditory motion cues can indeed influence bimodal PM performance, if only to hinder it. However, these bimodal PM studies did not systematically vary the congruency of the bimodal target's auditory and visual motion signals. It remains unclear, therefore, how auditory motion signals might influence the spatiotemporal prediction of the position of a bimodal target.

In the present study, we introduce a new PM task explicitly designed to clarify the role of auditory signals in bimodal targets. Similar to previous PM tasks, subjects were briefly presented with a moving audiovisual target and required to predict the target's future position some time after the visual motion stimulus disappeared, assuming it continued moving at an unchanging speed and in an unchanging direction. Unlike the previously cited bimodal PM studies, in our experimental task the sound motion signals were allowed to continue uninterrupted after the visual target disappeared to simulate a typical real-world PM scenario where observers may no longer visually track a moving object because they looked away, blinked, or the object was obscured, but they can often still hear it moving. Also, we systematically varied the spatial correlation of the targets' bimodal motion signals. That is, auditory and visual motion signals were sometimes spatially displaced with respect to each other. Although other motion perception studies have shown that bimodal motion perception is most efficient when motion signals from separate modalities coincide both spatially and temporally (Lewald et al. 2001; Meyer et al. 2005; Soto-Faraco et al. 2002; Stein et al. 1988; Zapporoli and Reatto 1969), the effect of such coincidence has yet to be tested in a PM task.

Furthermore, previous PM studies usually required subjects to judge the target's time of arrival at a specified location by a button press (e.g., Gordon and Rosenblum 2005; Hofbauer et al. 2004; Huber and Krist 2004; Peterken et al. 1991; Rosenblum et al. 1993; Schiff and Oldak 1990; but c.f., Gottsdanker 1952, 1955; Wiener 1962). However, the time of a button press is variable (Bootsma 1989; McLeod et al. 1986) and may not be the best measure of PM performance (Tresilian 1995, 1999a). In contrast, our task required subjects to estimate the bimodal target's spatial position at the end of the trial by pointing with a laser pointer. In this way, we could differentiate between two possibilities when the audiovisual motion signals of the bimodal target do not spatially coincide. Subjects' spatial estimates may be biased towards one motion signal's (either auditory or visual) true position at the end of the trial. Such bias would most likely favour visual motion signals as implied by previous findings (Hofbauer et al. 2004; Schiff and Oldak 1990). Alternatively, subjects may estimate the bimodal target's spatial position as somewhere between the true positions of the two motion signals. This would show that motion predictions depend on some averaging of the two motion signals, suggesting some integration as shown in other motion perception studies (Ecker and Heller 2005; Manabe and Riquimaroux 2000; Sekuler et al. 1997; Watanabe and Shimojo 2001).

Methods

Subjects

A total of 12 subjects (7 males and 5 females; mean age 25.3 years) participated in this study. All subjects had normal or corrected-to-normal visual acuity and normal, uncorrected hearing according to self-report. All procedures were approved by the York University Human Participates Review Sub-Committee, and informed consent was obtained from each subject.

Apparatus

Stimulus presentation and data recording were rigorously controlled by specialized software, E-Prime (Psychology Software Tools, Pittsburgh, PA) and MatLab (MathWorks, Natick, MA), running on a personal computer. A laser pointer (GSI Lumonics, Billerica, LA) projected a visual stimulus onto a 2.4 × 2.1 m fabric screen spanning 67.4° horizontally by 64.5° vertically. Subjects sat 1 m directly in front of the screen, aligned with the screen's centre—

designated as "0 cm". Positions referring to the left of centre were labelled as 'negative' (e.g., −50 cm), and positions referring to the right of centre were labelled as 'positive' (e.g., +50 cm). Two speakers were positioned behind the screen, 2 m apart and centred with respect to the subject (−100 cm and +100 cm, i.e., ±45°). Subjects were unable to see the speakers and were naïve as to the true nature of the sound source that produced the moving sound. Even though subjects performed the task in a completely dark room, they also wore neutral 0.6od (transmittance of 25%) density filtered goggles to ensure all visual references that may have potentially aided pointing judgements were eliminated (e.g., edges of the screen) and that nothing else was visible. These goggles were wide enough so that the subjects' field of view included all our stimulus locations. Subjects' heads were stabilized using a chin and forehead-rest.

Stimuli

Lights

The visual target was a red laser spot subtending 0.5 cm (0.23°). The visual target's variables, all randomly determined, consisted of four start positions (−75, −25, +25, or +75 cm), two speeds (5.5 or 11 cm/s), two viewing-time intervals (500 or 750 ms), and three prediction-time intervals (i.e., the time after the stimulus disappeared: 750, 1,000, or 1,500 ms). The visual target's trajectory was horizontal at eye level across the screen, moving in the opposite direction from its starting displacement (i.e., if the visual target initially appeared on the left, the trajectory was towards the right, and vice versa). Subjects were presented with the visual target at the start of the trial, where it remained stationary until the subject pressed the computer mouse button. The visual target then moved for the duration of the viewing-time interval. After the viewing time, the laser target was turned off and subjects continued to 'track' the target during the prediction-time interval. The visual target's "true" position was calculated as the motion end-point that would have been reached if the target had continued to move during the prediction-time interval.

To estimate the target's position, subjects pointed using the red laser spot that appeared at the end of the trial and remained visible until the subject recorded their pointing response by pressing a mouse button. Subjects moved the spot via a computer mouse, which they moved around on a tabletop in front of them.

Sounds

All auditory stimuli were pure tones (1 kHz at 68 dB measured at the subject's head level). The perceived position of our auditory stimuli was simulated by changing the relative intensities of the sound played through two speakers at fixed locations 1 m on either side of the centre of the screen (c.f., Bauer 1961; Grantham 1986; Rosenblum et al. 1987). The sound was made to appear at a particular position for a particular subject by varying the relative intensities with which it was played through each speaker. As one speaker's intensity increased the other simultaneously decreased such that the overall intensity remained constant. This method was calibrated against visually defined locations for each subject as described later. Varying the sound intensity in both speakers produced a percept of a continuous sound moving across the screen in front of the observer. In other words, simulating auditory space between two speakers means there were only perceived locations of sounds, which can vary between subjects. We acknowledge that an actual moving sound source would have been ideal. However, simulated auditory motion was employed for practical reasons such as avoiding extraneous noise produced by a real moving sound source—for example, a speaker moving along a track.

To ensure that the simulated auditory positions were the same for each subject independent of possible inter-individual differences for example in wax build up in each ear, each experimental session began with a calibration sequence. Each calibration trial consisted of a one-second tone (1 kHz at 68 dB) simulated at different spatial locations, which corresponded to all the various start and target positions of the visual target, 52 positions in total (4 start positions and 48 possible end-point positions—see below). Each position was presented five times in random order. Subjects used the laser spot to judge the tone's spatial location pressing the mouse button when they thought they were pointing at the right place. The subjects' average pointing responses for each position were calculated on-line and used to assign the start and end-positions of the auditory motion stimuli in the subsequent experiment conditions.

Procedure

Bimodal conditions

Figure 1 illustrates the general experimental design of our main PM task. Bimodal trials began with the visual target presented stationary at one of the four starting positions (i.e., −75, −25, +25, or +75 cm), determined randomly. By pressing the computer mouse button, subjects triggered the simultaneous motion onset of the visual target and the auditory stimulus. Auditory motion was spatially aligned with the moving visual target using the sound calibration performed at the beginning of the experimental session. Both the visual target and auditory motion traveled in the same direction at the same speed (5.5 or 11 cm/s) during

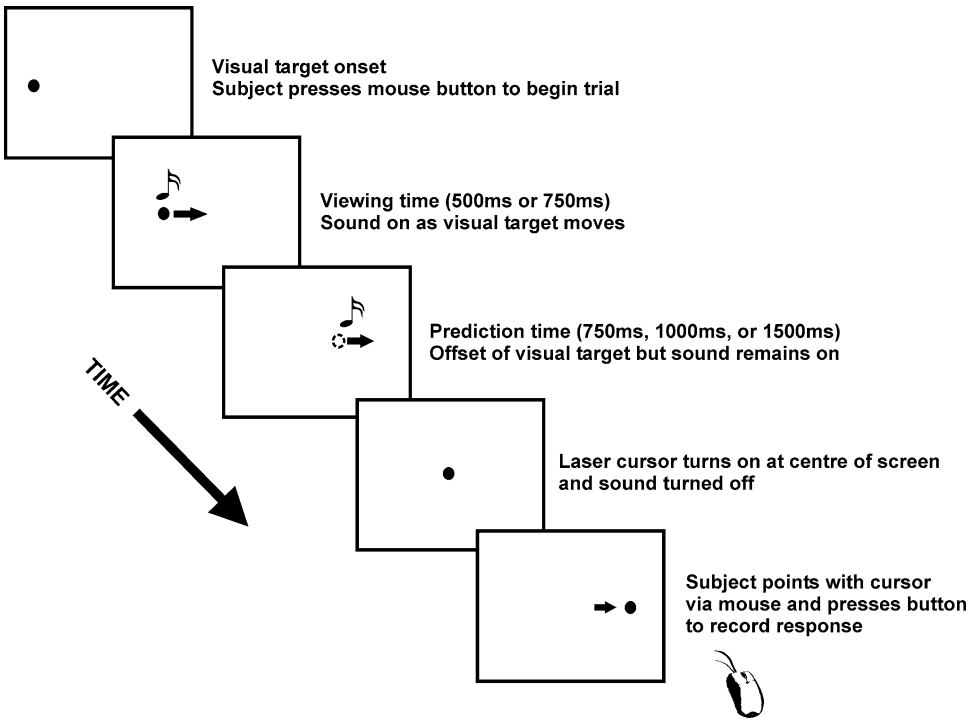

Fig. 1 General experimental paradigm. The visual target (shown as the *solid circle*) was presented at one of four possible start positions (25 or 75 cm either left or right of centre). The visual target remained stationary until triggered to move by the subject pressing the computer mouse button. Bimodal targets consisted of a visual target accompanied by a moving sound (depicted by the *musical note*). Alignment of the auditory motion stimuli relative to visual target varied among the three bimodal conditions—i.e., *congruent*, *sound-trailing*, *sound-leading*. Target speed and viewing time were randomly varied. The visual target was extinguished at the end of the viewing time but the moving sound continued through the prediction time. At the end of the trial, subjects were presented with a laser spot (*solid circle*) and used it to point, via the computer mouse, to the position they judged to be the final position of the visual target at the end of the trial (i.e., at the end of prediction time). The unimodal conditions were identical to the bimodal conditions except visual and auditory stimuli were presented independently

the viewing time (500 or 750 ms); both speed and viewing time were randomly determined. At the end of the viewing time, the visual target disappeared and the auditory motion continued uninterrupted by itself throughout the prediction-time interval. The rationale behind the auditory motion continuing after the visual target disappeared was to approximate a typical real-world PM scenario; that is, even though observers may no longer visually track a moving object because they looked away, blinked, or the object was obscured, observers can often still hear it moving. Subjects were instructed to visually track the bimodal stimulus throughout the viewing-time interval and attempt to continue 'tracking' the auditory motion during the prediction-time interval.

The auditory motion's offset at the end of the prediction interval was immediately followed by the laser spot appearing at the centre of the screen (designated as 0 cm). Subjects were instructed to predict the visual target's future position that it would have had at the end of the prediction time, i.e. the moment the auditory motion was turned off, assuming it had continued to move along with the sound during the prediction interval. Subjects responded by pointing the laser spot to the estimated location using the computer mouse and recorded their responses by pressing the computer mouse's button. Subjects were instructed to make their best guess if they were not sure. No feedback was provided during the experiment.

It is important to note that our task was not measuring time-to-contact (TTC). Typical TTC tasks involve subjects judging when an occluded moving object would either arrive at a designated location or collide with another occluded moving object (e.g., Bootsma and Oudejans 1993). In our task here, subjects had to extrapolate the future position of the visual target from the constant velocity and direction of the visual target during the viewing time and the auditory motion stimulus that was presented throughout the entire trial (viewing time + prediction time). Subjects had to estimate where the visual target was located at the end of the prediction time indicated either by the offset of the auditory motion signal (in the bimodal condition) or by the presentation of a tone (in the visual unimodal condition described below). Subjects did not know the end-positions of the audiovisual stimuli because the prediction time was randomized.

Bimodal trials were equally divided among *congruent* trials and two kinds of incongruent trials. *Congruent* trials consisted of spatially correlated bimodal motion signals where the visual target and the auditory motion stimulus traversed over the same spatial locations as determined by the calibration task (described in *Stimuli*) at the beginning of the experimental session. Incongruent trials consisted of displacing the auditory motion spatial start (and end) positions relative to those of the visual target. For *sound-leading* incongruent trials, the auditory motion stimulus was shifted to simulate a sound moving 10 cm ahead of the visual target. *Sound-trailing* incongruent trials consisted of shifting the auditory motion stimulus to simulate a sound moving 10 cm behind the visual target. Recall that shifted auditory motion positions were simulated by using intensity changes (c.f, Bauer 1961; Grantham 1986; Rosenblum et al. 1987). We selected 10 cm as the amount of displacement in the incongruent conditions so that the displacement would be large enough to make influences of one stimulus on the other clearly distinguishable while at the same time allowing the stimuli to be fused into a single percept. To ensure that subjects still perceived the audio-visual stimuli as being spatially correlated, we asked all subjects if they noticed any displacement between the sound and the light after the experiment. All of our subjects reported that they were unaware of the auditory motion being displaced with respect to the visual target. In every other respect, such as speed and duration, auditory motion was the same as *congruent* trials.

Unimodal conditions

Bimodal performance was compared to two unimodal conditions: *visual-only* and *auditory-only*. Both unimodal conditions were identical to the bimodal conditions except subjects tracked a unimodal motion stimulus. In *visual-only* trials, subjects tracked the visual target without accompanying sound motion. The speeds, viewing times, and prediction times of the visual target were the same as in the bimodal conditions. Subjects pointed to the visual target's estimated end-point at the end of the prediction-time interval as indicated by a short 50 ms auditory tone (1 kHz at 68 dB) presented from both speakers.

We also included an *auditory-only* condition as another control for the bimodal conditions. Comparing performance between the *auditory-only* condition and bimodal conditions allowed us to determine whether subjects were simply pointing to the sound in the bimodal conditions despite subjects being instructed to estimate the future location of the visual target that had disappeared. To that end, *auditory-only* trials were identical to the bimodal condition except without an accompanying visual target. Each trial began with a laser spot at the start location of the auditory motion stimulus. Upon pressing the mouse button, the laser spot was removed, and the auditory motion stimulus was triggered. The auditory motion stimulus was presented for durations, randomly selected, that were the equivalent of each of the total trial times of the bimodal trials (viewing time + prediction time). Subjects were required to point the laser spot to the location they judged as the last position of the auditory motion as indicated by the sound's offset. Note that *auditory-only* trials were not a PM task because subjects were presented with the entire trajectory of the auditory motion stimulus as in the bimodal conditions. Because this was a control condition, it was important to replicate the same auditory motion signals as used in the bimodal condition to conduct a proper comparison.

All conditions were presented in a block-design and were counter-balanced between subjects. Each block consisted of 240 trials that equally covered every combination of target parameters (speed, viewing time, and prediction time).

Results

Preliminary analyses of target start positions

Before proceeding to our main results, we show that pointing performance was not dependent on the target's start position. For each condition, we conducted a one-way ANOVA comparing the subjects' overall mean pointing errors between the four start positions across all target speeds, viewing times, and prediction times. In all three bimodal conditions and the *visual-only* condition, pointing errors were calculated as the relative difference between the subject's pointing position and where the visual target would have been at the time a response was requested (i.e. after viewing time + prediction time). In the *auditory-only* condition, pointing errors were calculated between pointing positions and the position of the auditory target at the end of the trial. These analyses all show that pointing errors were not determined by start position: *congruent* condition ($F_{(3,\ 44)} = 1.39, p = 0.26$), *sound-trailing* condition ($F_{(3,\ 44)} = 0.65,\ p = 0.59$), *sound-leading* condition ($F_{(3,\ 44)} = 0.42, p = 0.74$), *visual-only* condition ($F_{(3,\ 44)} = 1.26,\ p = 0.3$), nor the *auditory-only* condition ($F_{(3,\ 44)} = 0.28, p = 0.83$). Since target start position did not influence pointing performance, we collapsed start position in our subsequent data analyses and show normalized target positions in our following data figures.

Bimodal conditions

Figure 2 shows the main results of all three bimodal conditions (i.e., *congruent*, *sound-trailing*, and *sound-leading*).

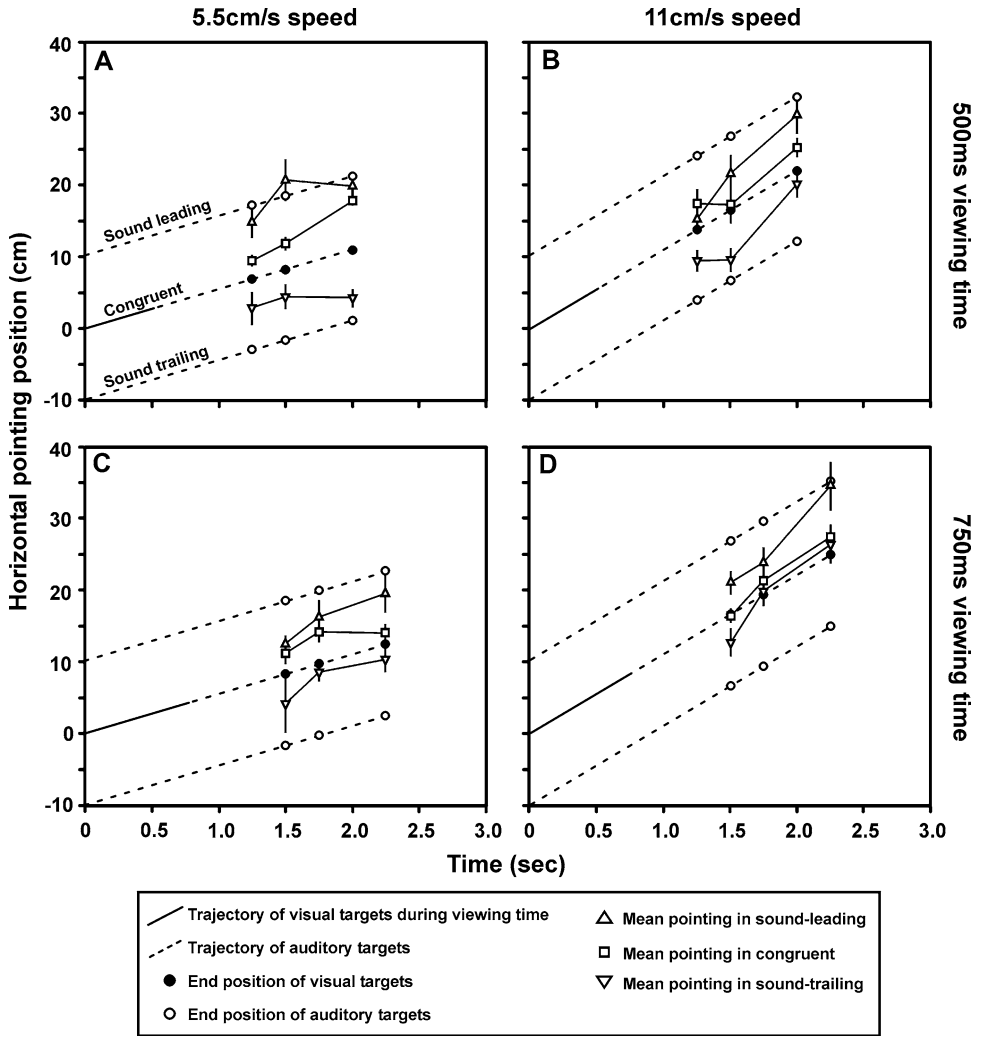

Fig. 2 Main results of bimodal conditions. *Panels* show subjects' mean horizontal pointing positions ($n = 12$) plotted against total trial time (viewing time + prediction time) for different target speeds and viewing times: 5.5 cm/s and 500 ms targets (A), 11 cm/s and 500 ms targets (B), and 5.5 cm/s and 750 ms targets (C), and 11 cm/s and 750 ms targets (D). *Solid lines* represent the visual target trajectories, and *dashed lines* represent the auditory target trajectories. Auditory target trajectories depended on bimodal condition. Actual target positions after different total trial times (viewing time + prediction time) are shown as *open circles* for auditory targets and *closed circles* for visual targets. Note that in the *congruent* condition, the end-positions of the visual and auditory targets overlapped. Corresponding to these target positions are the data curves that show the mean pointing positions for each target in the different bimodal conditions: *upright triangles* represent *sound-leading* condition, *squares* represent *congruent* condition, and the *inverted triangles* represent the *sound-trailing* condition. All target positions and mean pointing positions have been normalized to a start position of zero. *Error bars* represent one standard error

Overall mean pointing position is shown relative to where the visual target would have been at the time a response was requested (i.e. after viewing time + prediction time). Figure panels show the pointing responses of all three bimodal conditions at each speed (left panels for 5.5 cm/s speed and right panels for 11 cm/s) and viewing time (top panels for 500 ms and bottom panels for 750 ms). Solid lines and dashed lines depict the trajectories of the visual target and the auditory stimulus, respectively.

In general, subjects overestimated the visual target's end-point (closed circles) in the *congruent* condition (open squares) across different speeds, viewing times, and prediction times. Even though subjects were instructed to point to the estimated end-point of the visual target in the *sound-leading* condition (upright triangles), pointing responses were generally shifted towards the end-points of the auditory target (open circles). Similarly, in the *sound-trailing* condition (inverted triangles), subjects generally underestimated the visual target end-point position towards the auditory target's end-points. In other words, the auditory motion affected the predicted position of the visual target. Overall mean pointing errors among the three bimodal conditions were found to be significantly different ($F_{(2,22)} = 34.06$, $p < 0.01$). Planned comparisons revealed

that errors in the congruent condition were overall smaller than in the *sound-trailing* condition ($t_{(22)} = 6.49$, $p < 0.01$) and the *sound-leading* condition ($t_{(22)} = -3.79$, $p < 0.01$); also, pointing errors (relative to the visual target location) were overall smaller in the *sound-trailing* condition than in the *sound-leading* ($t_{(22)} = -8.58$, $p < 0.01$). Table 1(A) shows the magnitude of these over/under-estimations of pointing responses in the three bimodal conditions as mean pointing error (i.e., the difference between the pointing position and the visual target's end-point) for different speeds, viewing times, and prediction times. The magnitudes of these errors are considered below.

To determine whether pointing responses were affected by different target parameters, we conducted a three-way repeated measures ANOVA (speed × viewing time × prediction time) for each bimodal condition. In the *congruent* condition, we found a main effect for speed ($F_{(1,11)} = 6.54$; $p = 0.027$), subjects' mean pointing errors were smaller when the target was moving at 11 cm/s than when it was moving at 5.5 cm/s, and a main effect for viewing time ($F_{(1,11)} = 4.89$; $p = 0.04$), mean pointing errors were smallest at 750 than 500 ms. The main effect of prediction time was not significant ($F_{(2,22)} = 2.13$; $p = 0.16$). All interactions were non-significant except for the 3-way interaction ($F_{(2,22)} = 4.31$; $p = 0.03$).

In the *sound-trailing* condition, we found a main effect for viewing time ($F_{(1,11)} = 6.99$; $p = 0.02$), as in the congruent trials, subjects' pointing responses were closest to the visual target's end-point when visible for 750 ms. However, pointing responses in the sound-trailing condition were not affected by speed ($F_{(1,11)} = 0.98$; $p = 0.34$) or prediction time ($F_{(2,22)} = 0.78$; $p = 0.47$). All interactions of the *sound-trailing* condition were non-significant.

Lastly, a main effect was found for speed in the *sound-leading* condition ($F_{(1,11)} = 6.2$; $p = 0.03$), as in the congruent trials, pointing responses were closer to visual target's end-point at 11 cm/s. No main effects were found for speed ($F_{(1,11)} = 0.97$; $p = 0.35$) nor prediction time ($F_{(2,22)} = 3.02$; $p = 0.09$). The only significant interaction in the *sound-leading* condition was found between speed and viewing time ($F_{(1,11)} = 6.53$; $p = 0.03$).

Unimodal conditions

Figure 3 shows the results for both *visual-only* and *auditory-only* conditions. Subjects' overall mean pointing responses relative to unimodal targets' end-points are plotted against total trial time (viewing time + prediction time). Figure 3a and b show the *visual-only* results according to viewing time, 500 and 750 ms, respectively. The overall *auditory-only* results are shown in Fig. 3C.

Figure 3a and b show that subjects consistently overestimated the visual targets' end-points when the target was moving at 11 cm/s speed and underestimated their end-points at 5.5 cm/s speed. These pointing biases were confirmed statistically by analysing actual pointing positions between the two speeds ($F_{(1,22)} = 163.65$, $p < 0.01$). The magnitude of these over and under-estimations is shown as mean pointing errors in Table 1(B). Despite the pointing biases of over and under-estimations at different speeds, the actual mean pointing errors for targets moving 5.5 and 11 cm/s were not statistically different ($F_{(1,11)} = 1.92$, $p = 0.19$). Pointing errors were smallest (i.e., closer to the visual target's end-point) for the 750 ms viewing time ($F_{(1,11)} = 7.99$, $p = 0.016$). Nor did varying prediction time influence pointing responses ($F_{(2,22)} = 1.63$, $p = 0.43$).

Figure 3c and Table 1(C) summarize the mean pointing results and mean pointing errors of the *auditory-only* condition. Recall that, unlike the *visual-only* condition, there was no distinction between viewing time and prediction time since the auditory target was presented continuously throughout the trial. For *auditory-only* targets, no effect was found for speed ($F_{(1,11)} = 0.014$, $p = 0.91$) and, though there seems to be a general trend of errors decreasing as a function of trial time, the main effect of trial time was not significant ($F_{(4,44)} = 1.43$, $p = 0.24$).

Comparison of bimodal and unimodal conditions

Comparisons with respect to mean pointing errors were conducted between the bimodal data and the *visual-only* data. Significant differences were found for all three comparisons between the *visual-only* condition and the *congruent* condition ($F_{(1,11)} = 10.72$; $p < 0.01$), the *sound-trailing* ($F_{(1,11)} = 16.59$; $p < 0.01$), and the *sound-leading* ($F_{(1,11)} = 19.28$; $p < 0.01$). Overall, these results indicate that *visual-only* targets yielded smaller pointing errors than bimodal targets from all three bimodal conditions. Similar comparisons were made between the bimodal data and *auditory-only* data. A significant difference was found between the *auditory-only* and *congruent* conditions ($F_{(1,11)} = 7.36$; $p = 0.02$), indicating smaller pointing errors in the *congruent* condition relative to the *auditory-only* condition. Mean pointing errors in the *auditory-only* condition were not found statistically different between either the *sound-trailing* condition ($F_{(1,11)} = 0.27$; $p = 0.61$) or the *sound-leading* condition ($F_{(1,11)} = 1.05$; $p = 0.33$).

Mean variances of pointing responses

We calculated the average variance of pointing responses in all unimodal conditions and bimodal conditions. Variance was defined as the square of the standard deviations. The average variance in the *visual-only* condition was

Table 1 Mean pointing errors in all three bimodal conditions (A), visual-only condition (B), and auditory-only condition (C)

Target parameters			Mean pointing errors		
Speed	Viewing time	Prediction time	Congruent	Sound-leading	Sound-trailing
(A) Bimodal conditions					
5.5	500	750	2.7 (2.3)	8.0 (7.9)	−3.9 (7.8)
5.5	500	1,000	3.7 (3.1)	12.5 (9.7)	−3.7 (6.1)
5.5	500	1,500	6.9 (2)	8.9 (6.4)	−6.7 (4.4)
5.5	750	750	2.8 (4.9)	4.1 (4.3)	−4.2 (13.8)
5.5	750	1,000	4.4 (4.7)	6.5 (7.9)	−1.2 (4.4)
5.5	750	1,500	1.6 (3.7)	6.9 (8.7)	−2.0 (6.1)
11	500	750	3.8 (6.7)	1.5 (3.8)	−4.2 (5.3)
11	500	1,000	0.9 (9.1)	5.2 (9)	−6.9 (5.6)
11	500	1,500	3.3 (4.6)	7.8 (9.1)	−1.8 (6.1)
11	750	750	−0.2 (3.2)	4.4 (5.7)	−3.9 (6.6)
11	750	1,000	1.9 (3.2)	4.4 (7.3)	0.5 (7.1)
11	750	1,500	2.5 (4.1)	9.6 (11.6)	1.5 (9.1)

Speed	Viewing time	Prediction time	Mean pointing errors
(B) Visual-only condition			
5.5	500	750	−3.4 (2)
5.5	500	1,000	−3.5 (2.4)
5.5	500	1,500	−3.6 (3.2)
5.5	750	750	−2.1 (1.7)
5.5	750	1,000	−1.8 (2.2)
5.5	750	1,500	−0.9 (3.1)
11	500	750	2.1 (2)
11	500	1,000	2.9 (1)
11	500	1,500	2.6 (2.1)
11	750	750	2.0 (1.6)
11	750	1,000	2.3 (1.9)
11	750	1,500	2.8 (2.8)

Speed	Movement time	Mean pointing errors
(C) Auditory-only condition		
5.5	1,250	8.0 (3.6)
5.5	1,500	6.4 (6)
5.5	1,750	2.9 (7.7)
5.5	2,000	3.7 (8)
5.5	2,250	2.7 (4)
11	1,250	5.1 (2.8)
11	1,500	4.3 (4.3)
11	1,750	6.8 (6.8)
11	2,000	2.8 (2.8)
11	2,250	4.0 (5)

Errors (cm) are shown for bimodal targets and visual-only targets moving at different target speeds (cm/s), viewing times (ms), and prediction times (ms), and auditory-only targets moving at different speeds and total movement times (ms). Positive and negative pointing errors indicate over and under-estimations, respectively. Standard deviations are in parentheses

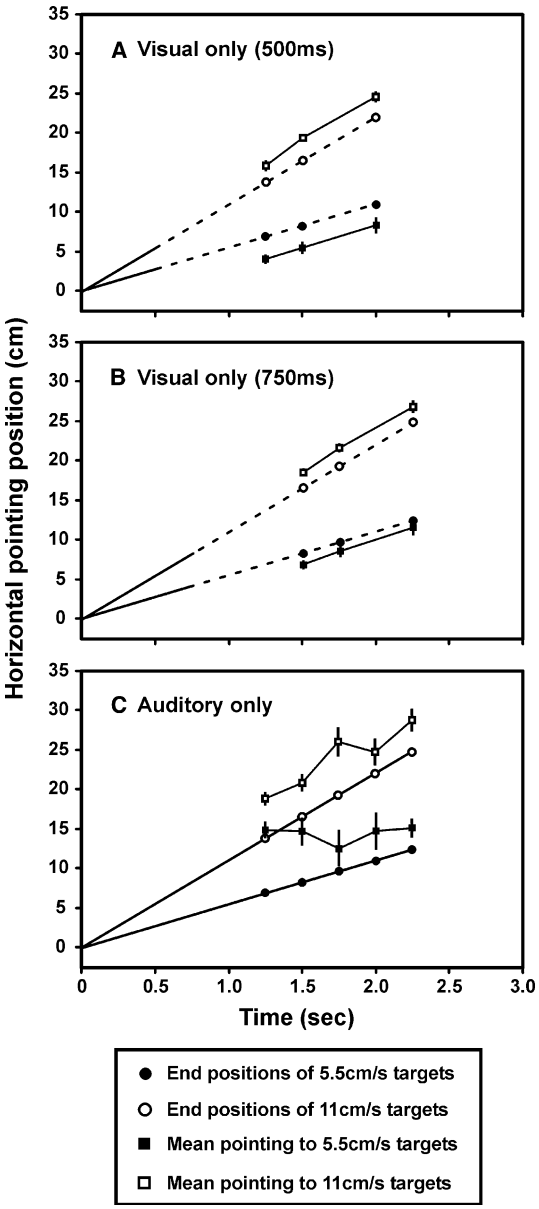

Fig. 3 Main results of unimodal conditions. Subjects' mean horizontal pointing positions ($n = 12$) are plotted against total trial time (viewing time + for visual and auditory targets. Panels **a** and **b** show the *visual-only* results according to viewing time, 500 and 750 ms, respectively. *Auditory-only* results are shown in panel **c**. Results are shown separately for both target speeds in each panel: the *shallow line* is the target trajectory at 5.5 cm/s and the steeper line at 11 cm/s. The *closed* and *open circles* on these *lines* represent the final target positions at each speed, 5.5 and 11 cm/s respectively, for different total trial times (viewing time + prediction time). In *panels* **a** and Fig. 2**b**, the visual target was visible during the viewing time (*solid line*), and visual target was turned off during the prediction time (*dashed line*). *Panel* **c** shows the trajectory of the auditory stimulus as it remains on throughout the trial. Corresponding data curves show the mean pointing positions for each target position at each speed, *closed squares* for 5.5 cm/s and *open squares* for 11 cm/s. All target positions (i.e., final position at the end of the trial) and mean pointing positions have been normalized to a start position of zero. *Error bars* represent one standard error

5.1 cm^2 (SE = ±0.8) and in the *auditory-only* condition was 30.9 cm^2 (SE = ±5.7). In contrast, the average variances in the *congruent* condition was 22.1 cm^2 (SE = ±6.4), in the *sound-trailing* was 53.3 cm^2 (SE = ±13.5), and in the *sound-leading* condition was 62.7 cm^2 (SE = ±9.8). The significance of these obtained variances will be discussed later when we consider how the audiovisual motion signals are integrated. An ANOVA of these variances yielded a significant effect ($F_{(4,55)} = 7.82$; $p < 0.01$). Tukey post hoc tests revealed significant differences for comparisons between the *visual-only* versus *sound-trailing* conditions, *visual-only* versus *sound-leading* conditions, and *congruent* versus *sound-leading* conditions ($p < 0.01$). All other comparisons were not significant.

Discussion

When visual and auditory stimuli were presented simultaneously to make bimodal targets, varying the spatial coincidence of the stimuli led to systematic differences in where the subject thought the target was after delays of between three quarters to one and a half seconds. Subjects slightly overestimated the position of all spatially congruent bimodal targets, even though the positions of visual-only targets were underestimated at the 5.5 cm/s speed. When the sound trailed or led the visual stimulus, the predicted location was shifted in the direction of the accompanying sound. Thus, we demonstrated for the first time that sound can have a direct effect on the predicted location of a visual target.

It is well established that sound localization can be biased towards a visual stimulus (the ventriloquist effect, Howard and Templeton 1966). Our experiments showed a novel version of the ventriloquism effect between an imagined visual target and an actual auditory one. Estimations of the visual target's future position in the incongruent conditions were biased in the direction of the auditory motion signal's end-point, indicating that auditory and visual motion signals were integrated when performing the bimodal task.

One might suggest that subjects were biased to ignore the visual target and rely only on the continuous auditory motion signal to solve the bimodal task. If subjects relied only on the auditory motion signal then we would predict pointing responses in the *sound-leading* condition, for example, would lie ahead of the actual end-positions of the auditory signal and resemble the data from the *auditory-only* control condition. However, this was not the case. Mean pointing responses in the *sound-leading* condition were found in between the end-points of the auditory and visual signals with bimodal targets of 5.5 cm/s, 500 ms

viewing time, and 1,000 ms prediction time being the only exception. These results indicate that the bimodal data cannot be explained by the argument that subjects followed only the sound, but instead suggest that both motion signals were taken into account to predict the future location of the visual target.

When tracking the location of a *visual-only* stimulus after it had disappeared, the predicted end-point depended on the target's speed, and how much of its trajectory had been seen. Curiously, how long it had to be remembered for (within the range of three quarters to one and a half seconds) did not seem to have an effect. Subjects underestimated target positions when targets were moving at 5.5 cm/s and overestimated target position when it was moving at 11 cm/s. The location of a moving *auditory-only* stimulus was always overestimated for the speeds and trial times used here.

Comparison with previous predicted motion studies

Most predicted motion (PM) studies have only considered unimodal targets—either visual targets (Alderson and Whiting 1974; Ellingstad 1967; Gottsdanker 1955; Huber and Krist 2004; Peterken et al. 1991; Slater-Hammel 1955) or auditory targets (Rosenblum et al. 1987, 1993, 2000). The few studies that have used bimodal stimuli in a PM task found no difference in the predicted position of targets made up of correlated audiovisual signals compared to predicting visual targets that were presented by themselves without an accompanying sound (Gordon and Rosenblum 2005; Hofbauer et al. 2004; Schiff and Oldak 1990). However, the study by Gordon and Rosenblum (2005) went further by showing that sound motion embedded in audiovisual stimuli is not completely ignored and can interfere with accurate PM performance although they did not show a facilitation of performance. What might account for the different results between the present study and these previous bimodal PM studies?

Our study is the first to look at the predicted spatial position of a bimodal target. Like most PM tasks, previous bimodal PM studies required subjects to make a temporal judgment of the target's time of arrival in a time-to-contact (TTC) paradigm. But timing judgements can be unreliable (McLeod et al. 1986) and may not be the best measure for testing PM performance (Tresilian 1999a). Spatially localizing an object's future position may involve a different set of computational processes than timing judgments and provide more precise estimates of PM performance (Bootsma 1989; Tresilian 1995). Some PM studies have been designed so that subjects have to make an explicit estimate of the target's future spatial position (e.g., DeLucia and Liddell 1998; Lyon and Waag 1995). In these studies, the moving object disappears for some variable time interval and then either reappears or a cue line is presented further along its motion path. Subjects make a two-alternative forced choice (2AFC) response indicating whether or not the object reappeared in the correct position or if it has passed the cue line. In our task, the visual target did not reappear, and no visual cue was presented: our subjects were not required to make relative spatial judgments of the visual target's position.

Even though subjects in a TTC task are only expected to make timing judgments about an occluded object's arrival or contact with another object, it has been suggested that TTC performance is not mediated solely by a clocking mechanism that uses temporal information to count down time of arrival, but that subjects might also use spatial information to track the object's virtual motion after it disappeared (DeLucia and Liddell 1998). Schiff and Oldak (1990) have argued that whether such a spatial extrapolation mechanism mediates TTC performance depends on one's theoretical perspective. However, even if both temporal and spatial information can be used in a TTC task, other researchers have argued that the way in which the information is used in response control is task-dependent (Bootsma 1989; Tresilian 1995, 1999b). That is, coupling the putative mechanisms for predicting future motion to the control of button-pressing responses to make temporal judgments likely involve different transformations than the control of pointing responses to make positional estimates. And thus, Tresilian (1995, 1999b) has suggested that caution should be taken when generalizing about the operations employed in one PM task to those in another.

Another source of difference from earlier studies might be task-related. In the studies by Schiff and Oldak (1990) and Gordon and Rosenblum (2005), subjects watched videos of an approaching car. After a brief presentation of the car's initial approach, the video and sound of the moving car were turned off. Subjects judged the time the car would pass the camera after a prediction-time interval. Although using videos of approaching vehicles might be considered a more realistic task, in both these studies the auditory signals were extinguished at the same time as the visual stimuli; thus, no motion information was provided during the prediction-time interval. This is also true for the only other bimodal PM study (Hofbauer et al. 2004), even though they designed their experimental task using simple stimuli similar to ours. However, in our task, the auditory motion signals continued throughout the prediction time (after the visual target was extinguished). We feel that this is closer to real-world scenarios where one can only briefly see a moving object but still hear it when the object is no longer visually available. That is, sight of a moving object may be interrupted by looking away or if the target is obstructed, but real-world auditory information is often continuously available and providing motion information.

For instance, crossing a street requires looking both ways for traffic. One may look to the left and see a car approaching from a distance. When looking to the right, the same car would no longer be in view but one may still hear it approaching, and thus, motion information is constantly provided about the car's changing position. Under these real-world conditions, our experiments indicate that the auditory motion can be used to help predict the target location.

Averaging of cross-modal motion signals in PM task

Performing a PM task requires some cognitive extrapolation mechanism that generates predictions of a future target position based on a variety of motion cues (DeLucia and Liddell 1998; DeLucia et al. 2003; Jagacinaki et al. 1983). Ideally, predicting the future position of audiovisual stimuli would take into account motion signals from all relevant sensory modalities. Previous PM studies on audiovisual stimuli found that coherent auditory motion signals did not help to improve accuracy of PM judgments (Hofbauer et al. 2004; Schiff and Oldak 1990), suggesting that the brain's extrapolation mechanism relies primarily on visual motion signals. But in our bimodal task here, estimations of the visual target's future position were shifted in the direction of the auditory motion signal when the bimodal signals did not spatially coincide. These findings indicate that rather than relying only on visual motion the brain's extrapolation computations do take motion cues from both sensory modalities into account.

How might the brain synthesize auditory and visual motion cues to extrapolate the future position of a moving audiovisual stimulus? The putative cognitive processes that govern extrapolation computations may operate with the same cross-modal synthesis rules that seem to apply to many multi-sensory systems, that is combining senses using a weighted average so as to minimize the variance in the combined signal (for examples see Dyde et al. 2006; Ernst and Banks 2002). However, our analysis of the variances did not show such a reduction; indeed, the bimodal variances were considerably larger than the unimodal variances. This lack of reduction in bimodal variances suggests that optimal integration of the motion signals is unlikely and the brain is performing some kind of averaging of independent motion signals to generate the best estimate of the future position of the audiovisual stimulus.

Cross-modal averaging of the independent motion signals is most evident in the *sound-trailing* and *sound-leading* conditions. However, spatially incongruent bimodal motion stimuli do not typical occur in the real world. It would be reasonable to suppose, then, that the brain deals with this spatial incongruency by assuming that the audiovisual motion signals do spatially coincide, thus, maintaining perceptual constancy. Our results show that the brain continues to apply perceptual constancy after the visual target disappears after being presented for only a portion of the time that the sound motion stimulus was presented (i.e., in our experiment, the visual target was presented from a third to half of time the sound motion was presented).

Conclusions

To summarize, we have shown that observers do use both auditory and visual motion cues to predict the future position of a moving target. Moreover, PM performance can be influenced by the spatial alignment of these bimodal motion cues. Typically, correlated audiovisual motion cues in the *congruent* condition yielded overestimated pointing responses of the target. Displacing the auditory motion relative to the visual target biased pointing responses in the direction of the auditory stimulus. We conclude that putative cognitive extrapolation mechanisms assume spatial coherence of bimodal motion signals and may perform some averaging of these independent motion signals when they do not spatially coincide.

References

Alderson GJK, Whiting HTA (1974) Prediction of linear motion. Hum Factors 16(5):495–502

Bauer BB (1961) Phasor analysis of some stereophonic phenomena. J Acoust Soc Am 33(11):1536–1539

Bootsma RJ (1989) Accuracy of perceptual processes subserving different perception-action systems. Q J Exp Psychol 41A:489–500

Bootsma RJ, Oudejans RRD (1993) Visual information about time to collision between two objects. J Exp Psychol Hum Percept Perform 19:1041–1052

DeLucia PR, Liddell GW (1998) Cognitive motion extrapolation and cognitive clocking in prediction motion tasks. J Exp Psychol Hum Percept Perform 24:901–914

DeLucia PR, Kaiser MK, Bush JM, Meyer LE, Sweet BT (2003) Information integration in judgements of time to contact. Q J Exp Psychol 56A:1165–1189

Dyde RT, Jenkin MR, Harris LR (2006) The subjective visual vertical and the perceptual upright. Exp Brain Res 173:612–622

Ecker AJ, Heller LM (2005) Auditory-visual interactions in the perception of a ball's path. Perception 34:59–75

Ellingstad VS (1967) Velocity estimation for briefly displayed targets. Percept Mot Skills 24:943–947

Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415:429–433

Gordon MS, Rosenblum LD (2005) Effects of intrastimulus modality change on audiovisual time-to-arrival judgments. Percept Psychophys 67(4):580–594

Gottsdanker RM (1952) The accuracy of prediction motion. J Exp Psychol 43:26–36

Gottsdanker RM (1955) A further study of prediction motion. Am J Psychol 68:432–437

Grantham DW (1986) Detection and discrimination of simulated motion of auditory targets in the horizontal plane. J Acoust Soc Am 79:1939–1949

Hofbauer M, Wuerger SM, Meyer GF, Roehrbein F, Schill K, Zetzsche C (2004) Catching audiovisual mice: predicting the arrival time of auditory-visual motion signals. Cogn Affect Behav Neurosci 4(2):241–250

Howard IP, Templeton WB (1966) Human spatial orientation. Wiley, New York

Huber S, Krist H (2004) When is the ball going to hit the ground? Duration estimates, eye movements, and mental imagery of object motion. J Exp Psychol Hum Percept Perform 30(3):431–444

Jagacinaki RJ, Johnson WW, Miller RA (1983) Quantifying the cognitive trajectories of extrapolated movements. J Exp Psychol Hum Percept Perform 9(1):43–57

Lewald J, Ehrenstein WH, Guski R (2001) Spatio-temporal constraints for auditory–visual integration. Behav Brain Res 121:69–79

Manabe K, Riquimaroux H (2000) Sound controls velocity perception of visual apparent motion. J Acoust Soc Jpn 21:171–174

McLeod P, McLaughlin C, Nimmo-Smith I (1986) Information encapsulation and automaticity: evidence from the visual control of finely-timed actions. In: Posner M, Malin O (eds) Attention & performance, vol XI. Lawrence Erlbaum, Hillsdale, pp 391–406

Meyer GF, Wuerger SM, Rohrbein F, Zetzsche C (2005) Low-level integration of auditory and visual motion signals requires spatial co-localisation. Exp Brain Res 166:538–547

Peterken C, Brown B, Bowman K (1991) Predicting the future position of a moving target. Perception 20:5–16

Rosenblum LD, Carello C, Pastore RE (1987) Relative effectiveness of three stimulus variables for locating a moving sound source. Perception 16:175–186

Rosenblum LD, Wuestefeld AP, Saldaña HM (1993) Auditory looming perception: influences on anticipatory judgments. Perception 22(12):1467–1482

Rosenblum LD, Gordon MS, Wuestefeld AP (2000) Effects of performance feedback and feedback withdrawal on auditory looming perception. Ecol Psychol 12(4):273–291

Schiff W, Oldak R (1990) Accuracy of judging time to arrival: effects of modality, trajectory, and gender. J Exp Psychol Hum Percept Perform 16(2):303–316

Sekuler R, Sekular AB, Lau R (1997) Sound alters visual motion perception. Nature 385:308

Slater-Hammel AT (1955) Estimation of movement as a function of the distance of movement perception and target distance. Percept Mot Skills 5:201–204

Soto-Faraco S, Lyons J, Gazzaniga M, Spence C, Kingstone A (2002) The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. Cogn Brain Res 14:139–146

Soto-Faraco S, Spence C, Lloyd D, Kingstone A (2004) Moving multisensory research along: motion perception across sensory modalities. Curr Direct Psychol Sci 13(1):29–32

Stein BE, Huneycutt WS, Meredith MA (1988) Neurons and behavior: the same rules of multisensory integration apply. Brain Res 448(2):355–358

Tresilian JR (1995) Perceptual and cognitive processes in time-to-contact estimation: analysis of prediction-motion and relative judgment tasks. Percept Psychophys 57(2):231–245

Tresilian JR (1999a) Analysis of recent empirical challenges to an account of interceptive timing. Percept Psychophys 61(3):515–528

Tresilian JR (1999b) Visually-timed action: time-out for 'tau'? Trends Cogn Sci 3(8):301–310

Watanabe K, Shimojo S (2001) When sound affects vision: effects of auditory grouping on visual motion perception. Psychol Sci 12(2):109–116

Wiener EL (1962) Motion prediction as a function of target speed and duration of presentation. J Appl Psychol 46:420–424

Zapporoli GC, Reatto LL (1969) The apparent movement between visual and acoustic stimulus and the problem of intermodal relations. Acta Pscyhol 29:256–267