

Direction of Visual Apparent Motion Driven Solely by Timing of a Static Sound

Elliot Freeman^{1,2,*} and Jon Driver²

¹Department of Psychology
Brunel University
Uxbridge, UB8 3PH
United Kingdom

²UCL Institute of Cognitive Neuroscience
University College London
London, WC1N 3AR
United Kingdom

Summary

In temporal ventriloquism, auditory events can illusorily attract perceived timing of a visual onset [1–3]. We investigated whether timing of a static sound can also influence spatio-temporal processing of visual apparent motion, induced here by visual bars alternating between opposite hemifields. Perceived direction typically depends on the relative interval in timing between visual left-right and right-left flashes (e.g., rightwards motion dominating when left-to-right interflash intervals are shortest [4]). In our new multisensory condition, interflash intervals were equal, but auditory beeps could slightly lag the right flash, yet slightly lead the left flash, or vice versa. This auditory timing strongly influenced perceived visual motion direction, despite providing no spatial auditory motion signal whatsoever. Moreover, prolonged adaptation to such auditorily driven apparent motion produced a robust visual motion aftereffect in the opposite direction, when measured in subsequent silence. Control experiments argued against accounts in terms of possible auditory grouping, or possible attention capture. We suggest that the motion arises because the sounds change perceived visual timing, as we separately confirmed. Our results provide a new demonstration of multisensory influences on sensory-specific perception [5], with timing of a static sound influencing spatio-temporal processing of visual motion direction.

Results and Discussion

In classical demonstrations of visual apparent motion [6], motion can be perceived between two alternately flashing bars (Figure 1A). The dominant direction of perceived motion can depend critically on visual timing, specifically the visual stimulus onset asynchrony between left-then-right flashes ($vSOA_{LR}$), relative to right-then-left flashes [4]. Thus, rightwards motion will typically dominate when $vSOA_{LR} = 333$ ms and $vSOA_{RL} = 666$ ms, but leftwards dominates for the converse timing (Figures 1B and 1C). In our main multisensory experiment, the $vSOA$ between alternating flashes was constant and symmetric at 500 ms (Figure 1D), but each flash could be paired with an auditory beep (from a central loudspeaker), one beep lagging a particular flash onset by 83 ms and the next beep leading the next flash onset by the same asynchrony

(see Figures 1E and 1F). The beeps thus had their own asymmetric auditory SOAs (e.g., $aSOA_{LR} = 333$ ms and $aSOA_{RL} = 666$ ms, or vice versa), which differed between conditions (Figures 1E and 1F) by the same amount as for the asymmetric visual SOAs in the purely visual version (cf. Figures 1B and 1C, and see Movie S1 available online). Given the phenomenon of “temporal ventriloquism” [2, 3], we hypothesized that this might influence the perceived timing of visual onsets to create an illusory inequality between *perceived* $vSOA_{RL}$ and $vSOA_{LR}$ (despite the veridical visual equality, in our multisensory conditions). Our new question was whether such putative illusory shifts in visual onsets, determined solely by auditory timing, could determine the *direction* of perceived visual motion, analogous to when manipulating visual timing instead (readers may assess the auditory-timing effect from the demonstration Movie S1). In addition to measuring “online” percepts of visual motion during exposure with beeps (Experiment 1), we also tested for any subsequent aftereffects produced by such exposure, on later judgements for purely visual displays in silence (Experiment 2).

Online judgements (Experiment 1) are summarized as average proportion of “rightwards” responses across all eight observers in Figure 2A, for displays in which either the actual visual timing or instead just the auditory timing were manipulated. As classically found [4], observers reported rightwards motion more than leftwards when $vSOA_{LR}$ was less than $vSOA_{RL}$ (i.e., with asymmetric visual timing, rightmost two data points in Figure 2A). The new finding was that when using symmetric visual timing instead, and manipulating solely the auditory timing, this too had an analogous impact (leftmost two data points in Figure 2A) on judgements of visual motion direction. This multisensory effect arose even though absolutely no biasing auditory motion was presented (unlike [7–11]), only a change in auditory *timing* relative to the regular visual events. The main effect of our timing manipulation was highly significant [$F(1,7) = 60.01$, $p > 0.0001$] and the magnitude of this effect did not depend on the modality manipulated [2×2 interaction: $F(1,7) = 0.68$, n.s.]; see Figure 2A. Thus, manipulating just the auditory timing (i.e., to determine which flash was slightly lagged by a sound and which flash was slightly led by a sound, by 83 ms in both cases), had an analogous impact on perception of visual motion direction as a corresponding change in visual timing itself.

Response latencies for the judgements in Experiment 1 tended to be slow, in accord with the task of making unsped-up judgements for cyclical stimuli (mean latency in the visual timing conditions was 3.17 s, SE 0.55; for the auditory timing condition, the corresponding mean was 2.67 s, SE 0.33, which did not differ significantly [$t(7) = 1.51$, ns]). Although unsped-up tasks potentially allow time for top-down influences, we conducted several further control experiments that argue against specific top-down accounts, based on putative perceptual grouping or attentional capture (see Supplemental Data, and brief description later below).

First, however, we further assessed the influence of auditory timing on processing of visual motion, via a quantitative measure of the strength of any subsequent visual aftereffects (Experiment 2). The stimulus sequences from which we had

*Correspondence: elliott.freeman@brunel.ac.uk

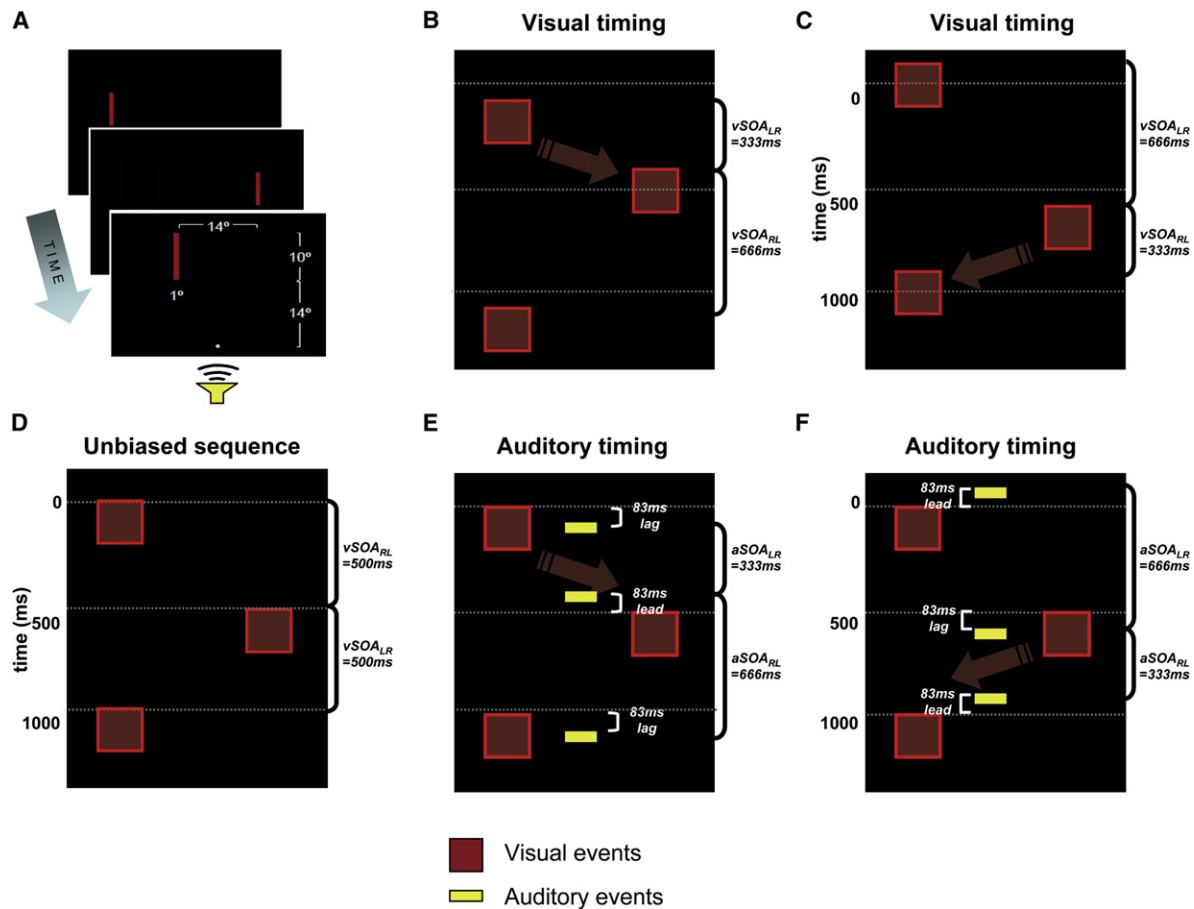


Figure 1. Audio-Visual Timing Diagrams

(A) Visual stimuli were vertical red bars presented alternately in upper left (L) or upper right (R) quadrant on a black background. Auditory stimuli were emitted by a single central speaker below the display.
 (B and C) Schematic representation of the visual timing conditions. Space is represented horizontally and time vertically. In (B), the veridical visual stimulus onset asynchrony (vSOA) between L and R bars is shorter (333 ms) than for the return R-to-L direction at 666 ms (i.e., $vSOA_{RL} > vSOA_{LR}$), inducing a percept of rightwards apparent motion (cartooned by the red arrow). In (C) the complementary case of $vSOA_{RL} < vSOA_{LR}$ inducing leftwards motion percepts is shown.
 (D) Unbiased visual sequence with equal intervals between L-to-R and R-to-L flashes (all vSOAs set at 500 ms).
 (E) Auditory timing exposure sequence, with same symmetric visual SOAs as in (D), but now with two auditory signals (represented here by central yellow bars); an initial sound lagging the onset of the left bar by 83 ms, with the next sound leading onset of the right bar by the same interval. Red arrow cartoons the hypothesized rightwards direction of motion induced auditorily by illusory shifts in visual timing.
 (F) The opposite timing relationship between sounds and left or right flashes, predicted to induce leftwards apparent motion (see red arrow) instead.
 See [Movie S1](#) for an audio-visual demonstration of cases (D)–(F).

obtained the above results (i.e., with symmetric visual timing plus sounds that manipulated auditory timing; or during vision-only with asymmetries in visual timing), can now be re-considered as adapting “exposure” phases. We had by design interleaved these with short silent test sequences, used to measure any visual aftereffects induced by the preceding, more prolonged exposure phase. We measured any visual aftereffects in silence with a psychophysical nulling procedure, in which visual SOA was varied pseudorandomly over 5 equally spaced values, according to the method of constant stimuli. The goal was to obtain a measure of the strength of any directional aftereffects, by finding the value of $vSOA_{LR}$ at which any illusory aftereffect may now be perfectly cancelled (so that the observer then reports leftwards and rightwards motion with equal frequency; see [Supplemental Data](#) and [Figure S1](#) for sample psychometric functions from illustrative individuals).

The group results from this aftereffect procedure are summarized in [Figure 2B](#), which plots the mean shift of nulling

point vSOA, with positive values denoting a shift in the direction predicted for an aftereffect given the prior exposure direction (i.e., with an opposite aftereffect predicted [4]). After a prolonged exposure phase that had a particular asymmetry in visual SOAs, the mean nulling point when measuring any aftereffect was indeed shifted in the opposite direction to the predominant exposure motion. The new result was that an analogous visual aftereffect was produced after exposure to symmetric visual timing, now with just *auditory* timing (i.e., slight lags and leads of the beeps by 83 ms, relative to the paired flashes) having crossmodally induced the directional motion percept during exposure. Statistical tests confirmed a significant main effect of adapting direction [$F(1,7) = 18.30$, $p < .004$], but no interaction with the modality manipulated during exposure [$F(1,7) = 1.73$, n.s.]; see [Figure 2B](#). This provides further evidence that manipulating just the auditory timing of a static sound (without any auditory motion whatsoever) can impact on visual motion processing, with apparently similar

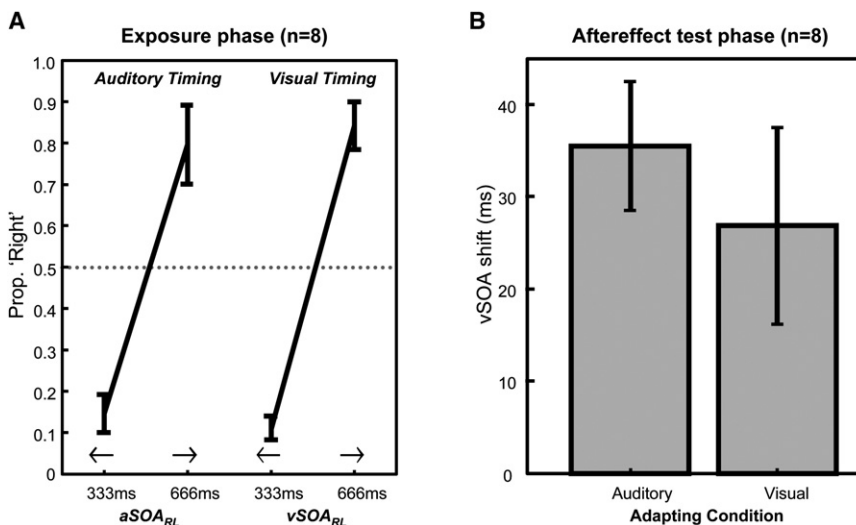


Figure 2. Results for Experiments 1 and 2

(A) Exposure phase: Proportion of “rightwards” responses, as a function of auditory SOA_{RL} (left pair of data points) or visual SOA_{RL} (right pair of data points), indicating the analogous effects of purely auditory timing or purely visual timing on perceived direction of visual motion (perceived direction indicated schematically by arrows at bottom) in Experiment 1.

(B) Aftereffect results from silent test phases (Experiment 2) after each exposure phase. Bars represent the shift in the mean nulling point (with positive values in the predicted aftereffect direction), as a function of the modality whose timing was manipulated in the preceding exposure sequence.

All error bars represent one unit of standard error of the means.

effects to manipulations of visual timing. Moreover, the aftereffect cannot be readily explained by response bias.

Supplemental Data provide further details of three additional control experiments, summarized briefly here. Experiment 3 used a duration-discrimination task to confirm that our manipulation of beep timing did indeed influence the perceived timing of flashes, as we expected. Experiments 4 and 5 provide some evidence against specific alternative “high-level” accounts of the phenomena, in terms of possible influences from perceptual grouping, or potential attention-capture, that might otherwise potentially have made some intervals between flashes more salient than others, without impacting on perceived visual timing. The final experiment also demonstrates that beeps that lag flashes influence visual motion more effectively than beeps that lead in time, in accord with an observed asymmetry in past studies of temporal ventriloquism [2, 12] (which unlike here did not consider the implications for motion).

Although several prior studies reported other possible audio-visual interactions in motion processing, none made our current point that changes solely in auditory timing (for static sounds, without any auditory motion) can influence the perceived direction of visual motion, probably via an influence on apparent visual timing. Some studies showed that static sounds and their timing can bias subjective ratings of apparent motion strength (e.g., [13–15]), but here we showed that auditory timing alone can determine *direction* of perceived visual motion. Our new paradigm also allowed us to compare the impact from auditory timing versus visual timing on perceived motion direction *and* on aftereffects for visual motion, finding these impacts to be analogous.

Several prior studies had investigated whether audition can influence visual motion direction but reported little or no effect (e.g., [7–9]), although others reported more positive outcomes [10, 11]. However, all those studies had manipulated the spatial relation between sound and vision. Here we did not present any auditory motion, manipulating only the timing of beeps from a single source, so that they could slightly lag or lead visual events, to produce “temporal ventriloquism” [1–3], particularly when lagging (see Experiment 5 in Supplemental Data). The observed influences of audition on perceived timing for vision (and thereby on visual motion direction as shown here) may accord with hypothesis of “modality appropriateness” [16], given that audition typically has superior temporal resolution to vision. Our results provide decisive new evidence for

the recent proposal [13] that auditorily driven temporal ventriloquism may drive visual motion, an idea that was solely based on nondirectional phenomenal ratings hitherto.

Other recent studies show that auditory onsets can cause anomalous visual phenomena [1, 17, 18], such as one flash appearing to be two if presented with two sounds in quick succession. Our own study bears only an abstract similarity to that (in showing an auditory influence on visual perception), differing in many specifics of the methods used and conclusions reached (e.g., here we varied only the relatively subtle ± 83 ms audio-visual timing relation, not presence versus absence of a second beep; moreover we measured the impact of auditory timing on perceived direction of visual motion, and we were able to show an auditorily determined visual aftereffect). Only one previous study [19] to our knowledge has tested whether audition can determine some visual-motion aftereffects, but found none. Unlike here, that study concerned motion in depth rather than laterally and presented auditory motion during exposure (i.e., an apparently looming sound), whereas here we manipulated only auditory timing rather than auditory motion. Thus, whereas our study accords generally with a burgeoning multisensory literature [5, 20] in showing that the senses are more intimately linked than traditionally thought, our specific demonstrations and conclusions here are novel, in showing that timing of a static auditory stimulus can determine direction of perceived visual motion and induce directional visual aftereffects. We found that changing just the timing of static auditory events can substantially affect processing of visual motion, to affect online visual perception of motion direction and also to induce visual aftereffects. We propose that these motion effects may reflect the well-established influence of auditory timing on perceived visual timing [1–3]. Our new effects illustrate that auditory timing information can penetrate substantially into visual processing of motion.

Experimental Procedures

Subjects

Eight healthy young adults (five male) participated for financial compensation. None had previous experience of similar tasks, and all were naive to the purposes of this study.

Apparatus and Stimuli

The visual display was a 21” CRT (Sony GDM-F520), with video mode 1024 × 768 at 75 Hz, viewed in a dark chamber from 57 cm. Auditory signals

were emitted by a single PC speaker positioned 30 cm below the display, out of view of the observer. Online eye tracking (100 Hz) was provided by an infrared video camera set into the chin-rest (Cambridge Research Systems, Reading), trained on the left eye. Experimental control was provided by a PC running Matlab 7.1 with the Cogent Graphics toolbox (<http://www.vislab.ucl.ac.uk/Cogent2000>) and the CRS Video Eyetracker Toolbox. A white fixation point (diameter 0.4 deg, luminance 120 cdm^{-2}) was displayed 13 degrees below the center of the display before and during each trial. Visual bar stimuli were red (CIE chromaticity coordinates $x = .611$, $y = .35$) with luminance of 8 cdm^{-2} on a black background (0.2 cdm^{-2}). The auditory signal comprised a rectangular-windowed 480 Hz sine-wave carrier, sampled at 22 kHz with 8-bit quantization.

Procedure

In Experiments 1 and 2 (run as different parts of the same study), visual timing or audio timing conditions (see Figure 1 for illustrations) were blocked in a pseudo-random order that was counterbalanced over subjects. Direction of “exposure” motion (whether determined by visual or auditory SOAs) was also counterbalanced across block orders. Each observer performed 4–8 blocks for each of the four conditions (i.e., leftwards versus rightwards motion induced by visual or auditory timing). Each block consisted of 20 pairs of exposure-then-subsequent-aftereffect phases. The duration of the first exposure phase was 30 s, followed by an aftereffect measure lasting 5 s, but all subsequent exposure and aftereffect pairs in a block lasted for 5 s each. In order to engage continued attention during exposure, observers were instructed to continually hold down one of two response keys (left-arrow or right-arrow on a standard PC keyboard), corresponding to the direction of motion they were currently experiencing (or no key if ambiguous). For simplicity, only the initial keypress was used for analysis, because this gave similar results to more complex analyses. Very rarely (1% of trials) was no key depressed. Exposure phases (with continuous responding throughout) were followed, after a 500 ms empty fixation display, by a silent 5 s aftereffect phase, and then a blank interval that remained until a two-alternative forced choice response was made with the same response keys. For these aftereffect phases, observers were instructed to respond only after the visual sequence had terminated.

At the start of each trial, both bars were presented simultaneously in silence for 500 ms. A sequence of visual flashes and auditory beeps then commenced with the offset of one randomly selected bar. Commencing with an offset helped to avoid a phenomenon observed in piloting, with sequences that originally began with a single visual onset, whereby motion tended to be reported initially in a direction away from that initial onset regardless of visual SOA.

In Experiments 1 and 2, each event-cycle lasted 1000 ms. Each 1000 ms event-cycle comprised a left and right bar, each appearing in alternation for 200 ms with empty gaps of 300 ms (Figures 1A and 1D). In each visual timing exposure sequence, the SOA between left and right bars ($vSOA_{LR}$) was set to either $500 + 166$ ms (= 666 ms) leading to leftward apparent motion (see Figure 1C); or to $500 - 166$ ms (= 333 ms) for rightward apparent motion (see Figure 1B). Note that the sum of $vSOA_{RL}$ and $vSOA_{LR}$ was always 1000 ms in Experiments 1 and 2, hence a short $vSOA_{RL}$ entailed a long $vSOA_{LR}$, and vice versa. For the auditory timing exposure sequences, $vSOAs$ were fixed symmetrically at 500 ms (Figure 1D), but 60 ms auditory beeps from a fixed central source were each paired with left or right flashes, with an 83 ms lag or lead in onset, relative to the paired flash. When one beep lagged the onset of the left bar by 83 ms, while the next beep preceded the right bar by the same interval, and so on, we found that this could induce rightwards visual motion. In this case, $aSOA_{LR}$ (interval between sounds paired to left and right flashes) was equal to 333 ms and $aSOA_{RL}$ was 666 ms; the opposite motion arose when $aSOA_{RL}$ was 333 ms and $aSOA_{LR}$ was 666 ms (Figures 1E and 1F). During aftereffect sequences (Experiment 2), no sounds were played but $vSOA_{RL}$ was varied pseudorandomly in five steps across a range of 500 ± 83 ms.

Prior to the experiment, all observers were initially shown silent sequences with a gross visual SOA asymmetry (e.g., $vSOA_{LR}$ or $vSOA_{RL} = 267$ ms) and told that these displays could sometimes appear to have either leftwards or rightwards motion. Without further prompting, all observers could readily discriminate between SOAs consistent with either leftward or rightward motion and likewise for more subtle visual SOA asymmetries.

No feedback was given during the main experiment, apart from occasional reminders to maintain central fixation during stimulus displays. Observers were not told that this was an experiment concerning adaptation or aftereffects and were given no information leading them to expect that any one sequence should affect their perception of a subsequent one in any particular direction.

Data Analysis

Eye data were processed with the ILAB toolbox for Matlab [21], filtering for blinks and periods of signal loss. Horizontal and vertical eye positions were then each analyzed in an ANOVA with repeated-measures factors of apparent-motion-direction and inducing-modality during exposure (i.e., visual or auditory SOA asymmetries). There were no significant differences in mean eye position between conditions.

Behavioral data were obtained for observer’s responses in each of the exposure (Experiment 1) and aftereffect (Experiment 2) phases, split by exposure condition (auditory or visual timing manipulated) and the SOA during exposure (i.e., predicted to induce either rightward or leftward motion). For the exposure phases, responses were summarized according to the percentage of “rightwards” responses (see Figure 2A). For aftereffect phases (which tested a range of visual SOAs in silence), the percentage of reporting “right” apparent motion versus “left” was calculated for each visual SOA (varied over 5 values across a range of 500 ± 83 ms) and initially considered as a psychometric function of the visual SOA in the test stimulus (see Figure S1 for examples). The nulling point at which a given $vSOA$ results in equal probability of “left” versus “right” responses was then estimated via maximum-likelihood logistic fit of the psychometric function. Shifts in nulling-point $vSOA_{LR}$ were summarized relative to the unadapted nulling point $vSOA$ (i.e., 500 ms), with positive values indicating shifts in the predicted aftereffect direction given the prior exposure phase (see Figure 2B). The effect of exposure modality on these nulling-point shifts was then analyzed by ANOVA.

Supplemental Data

Supplemental Data include Supplemental Experimental Procedures, three figures, and seven movies and are available at <http://www.current-biology.com/cgi/content/full/18/16/1262/DC1/>.

Acknowledgments

This research was funded by the BBSRC (J.D. and E.F.), plus the MRC and the Wellcome Trust (J.D.). J.D. also holds a Royal Society Leverhulme Trust Senior Research Fellowship. We are grateful to two anonymous reviewers for insightful comments and to Christopher Bolton for assistance in data collection. Thanks also to Toemme Noesselt for useful initial discussions.

Received: January 11, 2008

Revised: July 14, 2008

Accepted: July 15, 2008

Published online: August 21, 2008

References

1. Gebhard, J.W., and Mowbray, G.H. (1959). On discriminating the rate of visual flicker and auditory flutter. *Am. J. Psychol.* 72, 521–529.
2. Morein-Zamir, S., Soto-Faraco, S., and Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Brain Res. Cogn. Brain Res.* 17, 154–163.
3. Vroomen, J., and de Gelder, B. (2004). Temporal ventriloquism: Sound modulates the flash-lag effect. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 513–518.
4. von Grunau, M.W. (1986). A motion aftereffect for long-range stroboscopic apparent motion. *Percept. Psychophys.* 40, 31–38.
5. Driver, J., and Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Curr. Biol.* 10, R731–R735.
6. Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Z Psychol* 61, 161–265.
7. Alais, D., and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262.
8. Bertelson, P., and Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Percept. Psychophys.* 29, 578–584.
9. Meyer, G.F., and Wuerger, S.M. (2001). Cross-modal integration of auditory and visual motion signals. *Neuroreport* 12, 2557–2560.
10. Brooks, A., van der Zwan, R., Billard, A., Petreska, B., Clarke, S., and Blanke, O. (2007). Auditory motion affects visual biological motion processing. *Neuropsychologia* 45, 523–530.
11. Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., and Kingstone, A. (2002). The ventriloquist in motion: Illusory capture of dynamic

- information across sensory modalities. *Brain Res. Cogn. Brain Res.* *14*, 139–146.
12. Dixon, N.F., and Spitz, L. (1980). The detection of auditory visual desynchrony. *Perception* *9*, 719–721.
 13. Getzmann, S. (2007). The effect of brief auditory stimuli on visual apparent motion. *Perception* *36*, 1089–1103.
 14. Staal, H.E., and Donderi, D.C. (1983). The effect of sound on visual apparent movement. *Am. J. Psychol.* *96*, 95–105.
 15. Soto-Faraco, S., Kingstone, A., and Spence, C. (2003). Multisensory contributions to the perception of motion. *Neuropsychologia* *41*, 1847–1862.
 16. Welch, R.B., and Warren, D.H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* *88*, 638–667.
 17. Shams, L., Kamitani, Y., and Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature* *408*, 788.
 18. Sekuler, R., Sekuler, A.B., and Lau, R. (1997). Sound alters visual motion perception. *Nature* *385*, 308.
 19. Kitagawa, N., and Ichihara, S. (2002). Hearing visual motion in depth. *Nature* *416*, 172–174.
 20. Ghazanfar, A.A., and Schroeder, C.E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* *10*, 278–285.
 21. Gitelman, D.R. (2002). ILAB: A program for postexperimental eye movement analysis. *Behav. Res. Methods Instrum. Comput.* *34*, 605–612.