Citation: Freeman, E. D. & Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. Current Biology, 18(16), pp. 1262-1266. doi: 10.1016/j.cub.2008.07.066

Hearing moves seeing:
Direction of visual apparent-motion driven solely by timing of a static sound

Elliot Freeman (1,2)

Jon Driver (2)

1. Department of Psychology, Brunel University.
2. UCL Institute of Cognitive Neuroscience, University College London

Correspondence to:
Dr E. Freeman,
Brunel University,
Uxbridge, UB8 3PH, UK.
Tel: +44 (0)18952 66106
e-mail : elliot.freeman@brunel.ac.uk

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 inter-flash intervals are shortest [4]). In our new multisensory condition, inter-flash 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 (Fig. 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 $vSOA_{LR}$ = 333ms and $vSOA_{RL}$ = 666ms, but leftwards dominates for the converse timing (Figs 1b & c). In our main multisensory experiment, the vSOA between alternating flashes was constant and symmetric at 500ms (Fig. 1d), but each flash could be paired with an auditory beep (from a central loudspeaker), one beep lagging a particular flash onset by 83ms and the next beep leading the next flash onset by the same asynchrony (see Figs. 1e/f). 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 (Figs 1e/f), by the same amount as for the asymmetric visual SOAs in the purely visual version (cf. Figs 1b/c, and see on-line animated supplementary demonstration). Given the phenomenon of 'temporal ventriloquism' [2,3], we hypothesized 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). In addition to measuring "on-line" 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).

On-line judgements (Experiment 1) are summarized as average proportion of 'rightwards' responses across all 8 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 datapoints in Fig 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 datapoints in Fig 2a) on judgements of visual motion direction. This multisensory effect arose even though absolutely no biasing auditory motion was presented [unlike 7,8,9,10,11], 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 [2x2 interaction: $F(1,7) = 0.68$, n.s.]; see Fig 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 83ms 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 were tended to be slow, in accord with the task of making unspeeded judgements for cyclical stimuli (mean latency in the Visual Timing conditions was 3.17sec, SE 0.55; for the Auditory Timing condition the corresponding mean was 2.67sec, SE 0.33, which did not differ significantly [$t(7) = 1.51$, ns]. While unspeeded 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 Supplementary Materials, and brief description 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 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 reconsidered 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 Supplementary Materials and Fig S1 for sample psychometric functions from illustrative individuals).

The group results from this aftereffect procedure are summarized in main 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 following 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 Fig 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 effects to manipulations of visual timing. Moreover, the aftereffect cannot be readily explained by response bias.

Supplementary Materials provide further details of three additional control experiments, summarised 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 which *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).

While 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. 13,14, see 15], but here we showed that auditory timing alone can determine *direction* for 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. 7,8,9], although others reported more positive outcomes [10,11,16]. 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 Supplementary Materials). 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' [17], 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 non-directional phenomenal ratings hitherto.

Other recent studies show that auditory onsets can cause anomalous visual phenomena [1,18,19], 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 ±83ms 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 [16] 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, while 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 on-line 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.

Methods

Subjects

8 healthy young adults (5 male) participated for financial compensation. None had previous experience of similar tasks, and all were naïve to the purposes of this study.

Apparatus and stimuli

The visual display was a 21'' CRT (Sony GDM-F520), using video mode 1024 x 768 at 75hz, viewed in a dark chamber from 57cm. Auditory signals were emitted by a single PC speaker positioned 30cm below the display, out of view of the observer. On-line eye-tracking (100hz) was provided by an infra-red 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 centre 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 480Hz sine-wave carrier, sampled at 22kHz 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 Fig. 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 secs, followed by an aftereffect measure lasting 5 secs, but all subsequent exposure and aftereffect pairs in a block lasted for 5 seconds 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, as 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 500ms empty fixation display, by a silent 5-second after-effect phase, and then a blank interval which remained until a two-alternative forced choice response was made using the same response keys. For these after-effect phases, observers were instructed to only respond after the visual sequence had terminated.

At the start of each trial, both bars were presented simultaneously in silence for 500ms. 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 1000ms. Each 1000ms event-cycle comprised a Left and Right bar each appearing in alternation for 200ms with empty gaps of 300 ms (Figs. 1a, 1d). In each Visual Timing exposure sequence, the SOA between Left and Right bars ($vSOA_{LR}$) was set to either 500 + 166ms (=666 ms) leading to leftward apparent motion (see Fig 1b); or to 500 - 166ms (=333 ms) for rightward apparent motion (see Fig. 1c). Note that the sum of $vSOA_{RL}$ and $vSOA_{LR}$ was always

1000ms 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 500ms (Fig. 1d) but 60 ms auditory beeps from a fixed central source were each paired with left or right flashes, with an 83ms lag or lead in onset, relative to the paired flash. When one beep lagged the onset of the left bar by 83ms, 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 $aSOA_{RL}$ (interval between sounds paired to left and right flashes) was equal to 333 ms and $aSOA_{LR}$ was 666 ms; the opposite motion arose when $aSOA_{LR}$ was 333ms, and $aSOA_{RL}$ was 666 ms (Fig. 1e/f). During aftereffect sequences (Experiment 2), no sounds were played but $vSOA_{RL}$ was varied pseudorandomly in five steps across a range of 500±83ms.

Prior to the experiment, all observers were initially shown silent sequences with a gross visual SOA asymmetry (e.g. $vSOA_{LR}$ or $vSOA_{RL}$ = 267ms) 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 nor 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 using the ILAB toolbox for Matlab [21], filtering for blinks and periods of signal loss. Horizontal and Vertical eye positions were then each analysed 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.

Behavioural data were obtained for observer's responses in each of the exposure (Experiment 1) and aftereffect phases (Experiment 2), 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±83ms), and initially considered as a psychometric function of the visual SOA in the test stimulus (see supplementary 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. 500ms), with positive values indicating shifts in the predicted aftereffect direction given the prior exposure phase (see Fig 2b). The effect of exposure modality on these nulling-point shifts was then analysed by ANOVA.

Acknowledgements

This research was funded by the BBSRC (JD & EF), plus the MRC and the Wellcome Trust (JD). JD 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.

References

1. Gebhard, J.W., Mowbray, G.H. (1959). On discriminating the rate of visual flicker and auditory flutter. Am. J. Psychol. *72:521-9.,* 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., 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., 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., Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. Curr. Biol. *14,* 257-262.
8. Bertelson, P., Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Percept. Psychophys. *29,* 578-584.
9. Meyer, G.F., 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., 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., 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. Kitagawa, N., Ichihara, S. (2002). Hearing visual motion in depth. Nature *416,* 172-174.

17. Welch, R.B., Warren, D.H. (1980). Immediate perceptual response to intersensory discrepancy. Psychol Bull. *88,* 638-667.

18. Shams, L., Kamitani, Y. and Shimojo, S. (2000). Illusions. What you see is what you hear. Nature *408,* 788.

19. Sekuler, R., Sekuler, A.B. and Lau, R. (1997). Sound alters visual motion perception. Nature *385,* 308.

20. Ghazanfar, A.A., 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.

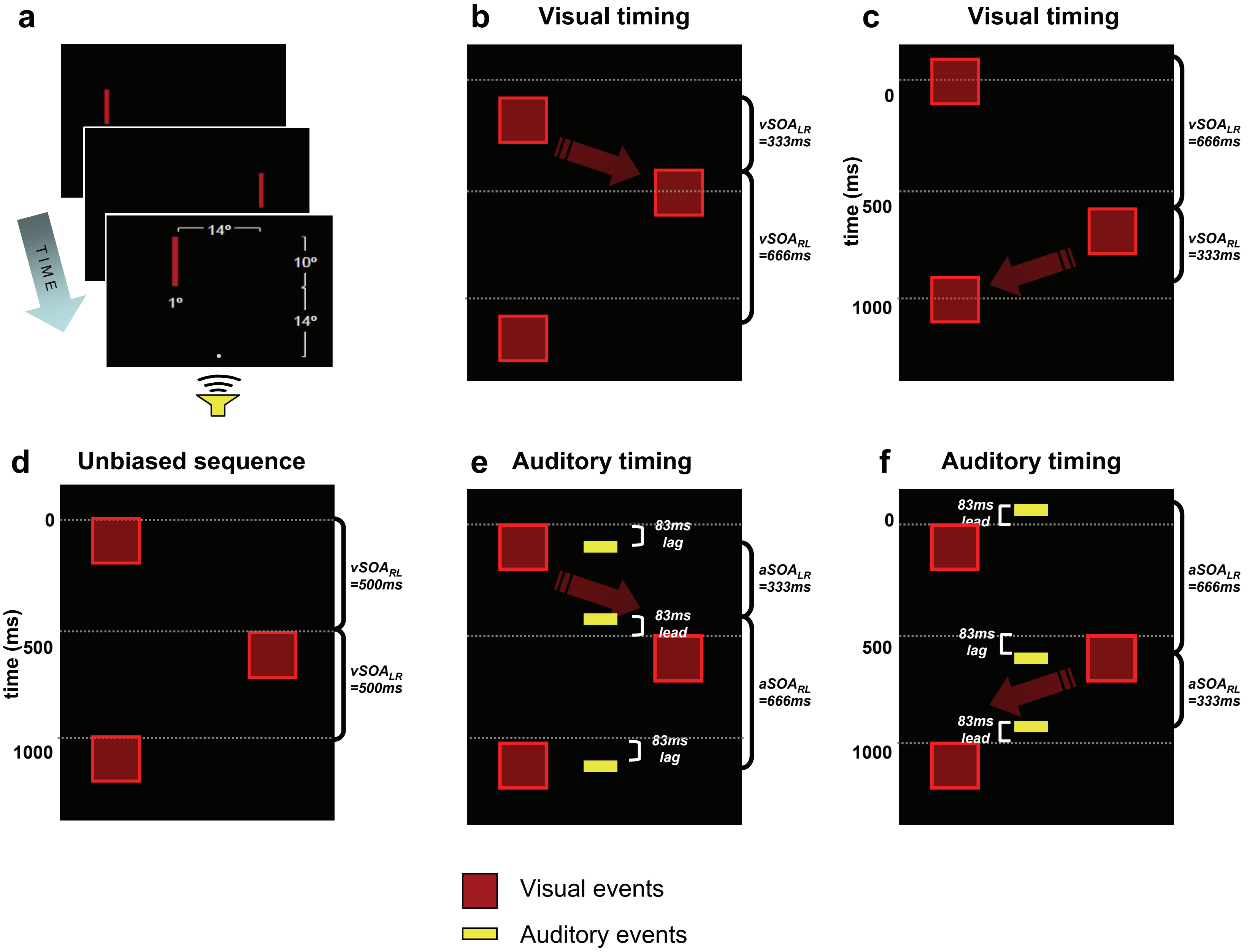
Figure captions

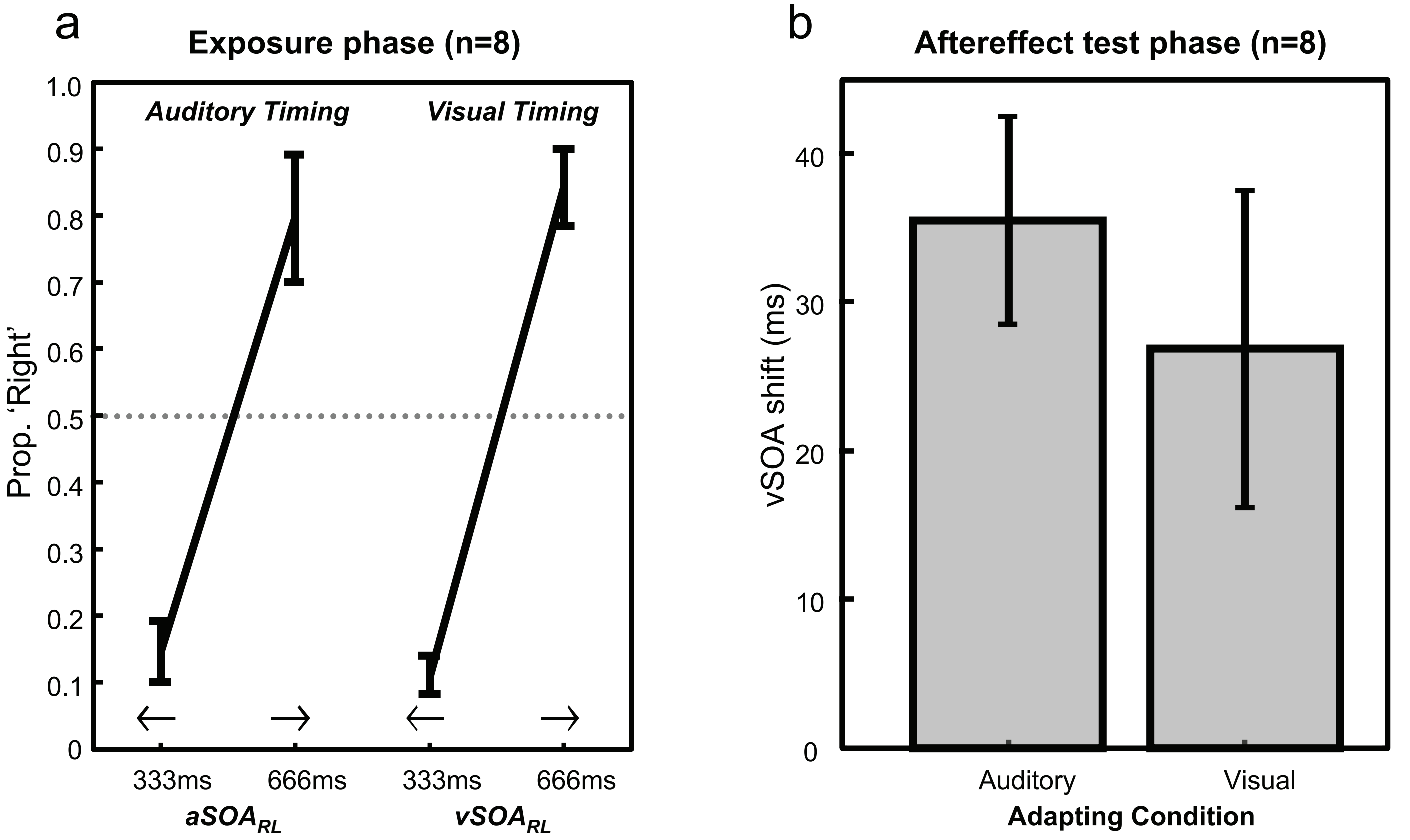
Figure 1

(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-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) 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 right bar by 83ms, with the next sound leading onset of the left bar by the same interval. Red arrow cartoons the hypothesised 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 upper arrow) instead. See supplementary movie for an audio-visual demonstration of cases d-f.

Figure 2

(a) Exposure phase: Proportion of 'rightwards' responses, as a function of auditory SOA_{RL} (left pair of datapoints) 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) following 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.


a
TIME
14º
10º
1º
14º
b
Visual timing
$vSOA_{LR}$ =333ms
$vSOA_{RL}$ =666ms
c
Visual timing
0
500
1000
time (ms)
$vSOA_{LR}$ =666ms
$vSOA_{RL}$ =333ms
d
Unbiased sequence
0
500
1000
time (ms)
$vSOA_{RL}$ =500ms
$vSOA_{LR}$ =500ms
e
Auditory timing
83ms lag
83ms lead
83ms lag
$aSOA_{LR}$ =333ms
$aSOA_{RL}$ =666ms
f
Auditory timing
0
500
1000
83ms lead
83ms lag
83ms lead
$aSOA_{LR}$ =666ms
$aSOA_{RL}$ =333ms
Visual events
Auditory events

a
Exposure phase (n=8)
Auditory Timing
Visual Timing
Prop. 'Right'
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
←
→
←
→
333ms
666ms
333ms
666ms
aSOA_RL
vSOA_RL
b
Aftereffect test phase (n=8)
vSOA shift (ms)
40
30
20
10
0
Auditory
Visual
Adapting Condition

Supplementary Materials

Experiment 2: Aftereffect measure

Figure S1 shows illustrative psychometric functions for three observers in the silent test phase (Experiment 2 in main paper) of the adaptation paradigm, plotting the proportion of 'rightwards' responses as a function of the visual SOA between right and left flashes during silence. Different columns in this figure correspond to the three different observers illustrated. Separate rows show data acquired following prior adaptation under exposure to conditions manipulating either (a) the visual SOA, in silence; or (b) just the auditory SOA with a constant symmetrical visual SOA during exposure. Filled-red or open-green symbols/curves refer respectively to conditions where rightwards or leftwards motion was predicted to be induced by the timing manipulation during the preceding exposed adaptation sequence. Maximum-Likelihood best-fitting logistic functions are superimposed, with upward arrows along the x-axis indicating the determined 'null point' vSOA at which 'rightwards' vs 'leftwards' reports are equiprobable (i.e. the point along the x-azis at which the function passes through 0.5 on the y-axis, as marked by grey horizontal line). The magnitude of the aftereffect was quantified following this curve-fitting analysis, by measuring the absolute deviation of the observed null points from the veridical null point (as marked by the vertical grey line at vSOA 500ms, where ambiguous motion would be predicted given no aftereffect). Note that while, as might be expected, the observations are somewhat 'noisy' in individuals unlike the overall group effects (see main text), nevertheless for every individual a shift in the inferred null points is apparent between the two opposite directions of prior exposure, in the predicted direction, regardless of whether this preceding exposure was purely visual (upper graphs) or audio-visual with just auditory timing varied (lower graphs)

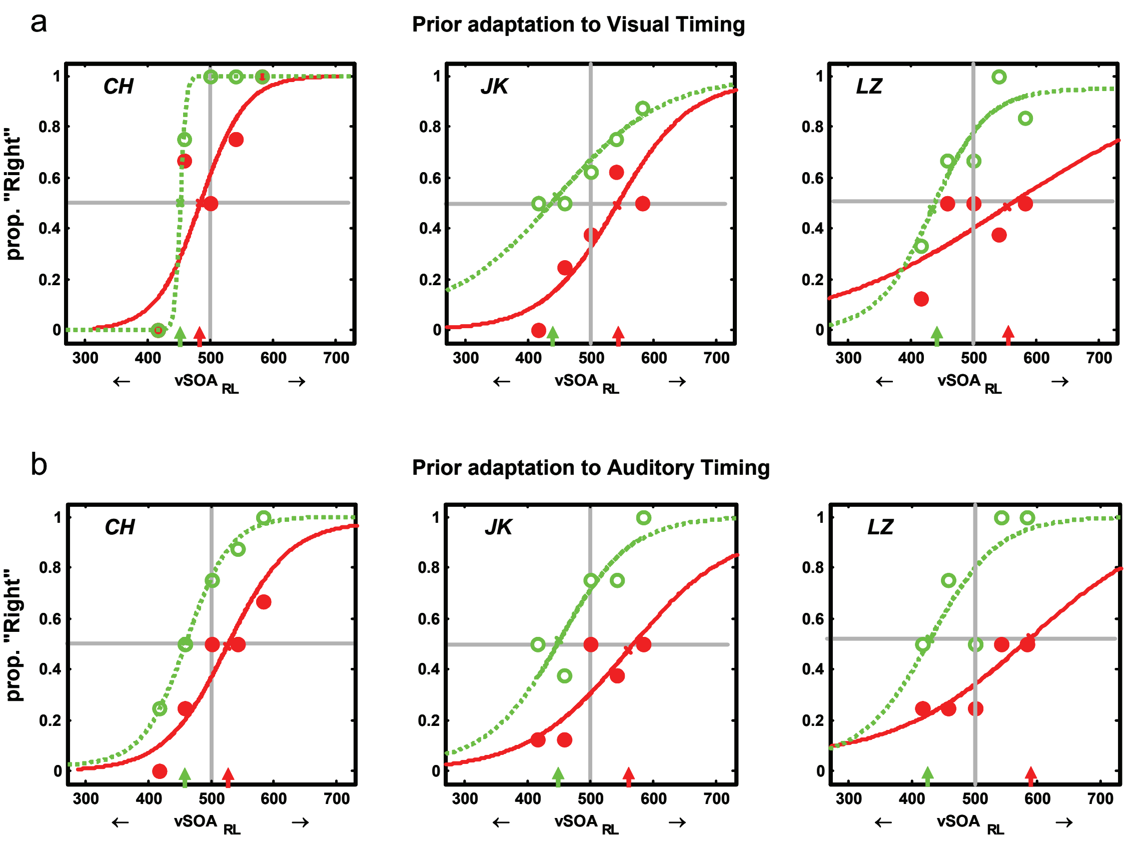


Figure S1 Illustrative psychometric functions for after-effects in Experiment 2 from three individual observers, for visual-timing (upper row) and auditory-timing (lower row) conditions in terms of <u>prior</u> exposure (green colouring for Leftwards direction of prior exposure, red for the Rightwards prior motion), but now always measured visually during subsequent silence.

Experiment 3: Auditory effect on temporal discrimination of visual intervals

The results from the main Experiments 1 and 2 appear consistent with our suggestion that temporal ventriloquism can lead to sound-induced illusory shifts of visual timing, which in turn can then induce visual motion percepts. However it might be countered that the results from Experiments 1 and 2 (main paper) did not directly demonstrate a change in perceived visual timing itself, as per previous demonstrations of audio-visual temporal ventriloquism [1,2]. Accordingly, here we used a temporal-discrimination paradigm to test more directly whether our beeps can indeed influence perceived timing of our flashes.

Methods

Five new naïve subjects participated. On each of 320 trials, observers had to indicate which of two successive flash-pairs had a longer interval between the two flashes within a pair (or which had a shorter interval, for 2 of the 5 subjects). In the critical condition, the first audio-visual 'reference' interval contained a two-flash Reference sequence (L-R or R-L direction varying between blocks) with fixed visual SOA of 500ms and auditory offsets of +/- 83.33ms, as in main Experiment 1; while the second 'test' interval comprised a silent two-flash Test sequence in the same direction, with pseudo-randomly varying visual SOA (500±111ms, across 5 levels). The point of subjective equality (at which test-reference discrimination is at 50% chance) was estimated for each observer from logistic fits of psychometric functions. For completeness, we also had subjects compare synchronous audio-visual Reference sequences to silent Test intervals (see accompanying supplementary movies illustrating each condition).

Results and Discussion

We found that the fixed duration of a two-flash asynchronous audio-visual reference sequence (*'Flash-Beep : Beep-Flash'*) was consistently judged as shorter than a silent test sequence with the same vSOA. Indeed, on average the test sequence had to be 66ms (SE 30ms) shorter than the reference to achieve subjective equivalence [t(4)= 2.43, p=0.035 one-tailed]. In a control condition with synchronous audio-visual presentations the test was judged on average to be 17ms (SE 45ms) longer than the reference but this was not a significant deviation from veridical timing [t(4)= 0.42, n.s.]. These results are therefore consistent with a change in perceived duration for the interval between flashes of ~83ms between asynchronous versus synchronous audio-visual conditions [t(4)= 3.33, *p*=0.03].

Experiment 4: Manipulation of pitch to assess potential role of auditory grouping

One potential rival to our temporal-ventriloquism suggestion, for the present findings of an auditory-timing effect upon visual motion, is that those beep-pairs with the shorter aSOA might provide some form of auditory ‘perceptual grouping’ cue, potentially promoting the interpretation of one flash pair as a single moving object that coheres across space and time.

As one approach to assessing this, Experiment 4 manipulated auditory perceptual grouping via pitch. Evenly-alternating flashes were now paired with synchronous beeps, but these pitches changed pitch after each pair (from 480Hz to 880 Hz, or vice-versa). A strong version of the potential auditory-perceptual-grouping account should presumably predict apparent motion to be reported predominantly between flash-pairs corresponding to pitch-pairs with the same pitch. For comparison, we also replicated the asynchronous auditory-timing conditions from Experiment 1 (see accompanying supplementary movies).

Methods

6 observers (4 male) participated. All were highly experienced at psychological experiments, and all but one were naïve about the purpose of the present study and had not participated in any previous similar studies.

In Experiments 4 (and also Experiment 5, see below) each event-cycle was now shorter, lasting 750ms. vSOAs were thus 375ms, and any audio-visual lead or lag was scaled accordingly, when present (i.e. now 62.5ms). We used this new timing because, in intervening pilot observations (including on our ourselves) we found that visual apparent motion induced by auditory manipulations (in timing) was subjectively even more compelling for these slightly faster cycles. We successfully replicated our on-line measure of any impact from auditory timing on visual motion here (analogous to main Experiment 1), in addition to testing any impact of the new pitch modulation.

In the critical new pitch-grouping condition, flashes were accompanied by synchronous beeps that were either 480Hz and 880Hz in pitch, with this pitch alternating every two successive sounds. In the auditory-timing replication condition, 480Hz beeps both lagged and led flashes by 62.5ms (in a 'flash-beep : beep-flash' sequence, as per Experiment 1). In conditions intended to induce rightwards motion, $aSOA_{LR}$ was 250ms and $aSOA_{RL}$ was 500ms, and opposite timings were used to induce leftwards motion. On each trial, a fixation point was followed by a flash-sequence lasting 2000ms, after which a 'Left'-'Right' unspeeded forced-choice was elicited using the standard keyboard cursor keys. Each observer performed 30 trials for each condition.

Results

We found no significant effect of pitch modulation [t(5)= 1.96, n.s.]. In comparison, the analogous comparison to the asynchronous auditory-timing conditions from Experiment 1 found our new motion-induced-by-auditory-timing effect strongly once again (t(5)= 3.32, *p*=0.02) with this differing significantly from the null outcome for pitch alone [two-way interaction of pitch/timing x directional effect: F(1,5) = 14.30, *p*=0.012]; see Figure S2a.

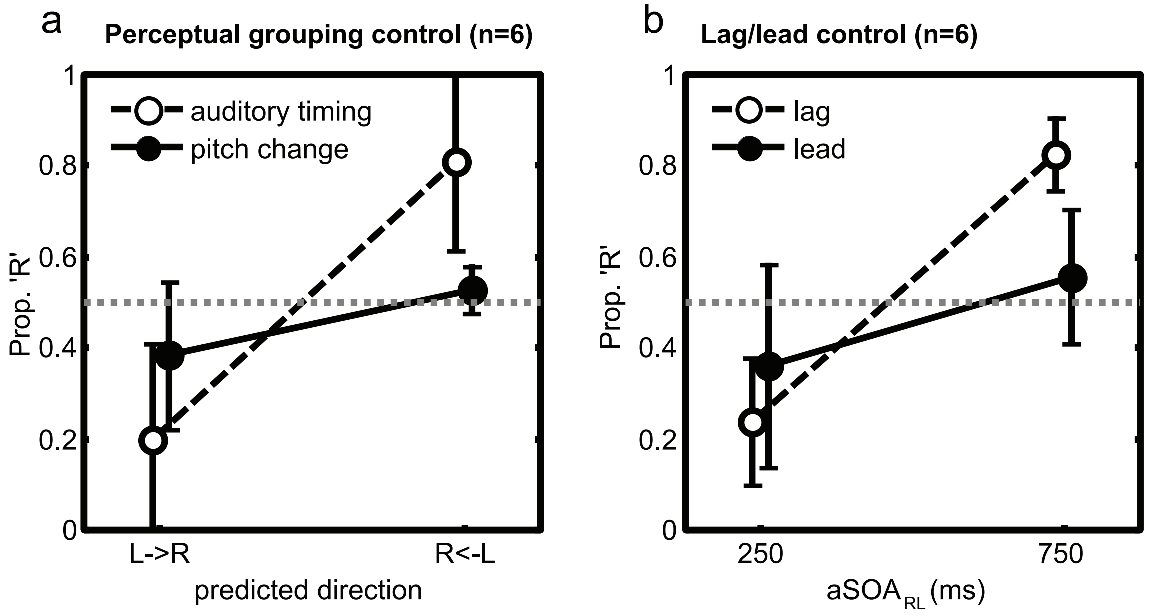


Figure S2 (a) Results from Experiment 4, a control study to test whether auditory perceptual-grouping induced by pitch can produce analogous effects on visual motion to auditory timing. Proportion of 'rightwards' responses is shown as a function of the direction predicted given either our standard auditory-timing manipulation (analogous manipulation as in Experiment 1, new replication of motion-induced-by-auditory-timing effect for Experiment 4 shown as open symbols), versus the new pitch manipulation (filled symbols), which did not produce the same effect. (b) Results from Experiment 5, a control study testing whether lagging beeps have more effect than leading on visual motion. Proportion of 'rightwards' responses is shown as a function of auditory SOA, for auditory sequences shifted later relative to the visual events (Lag condition, open symbols) or earlier instead (Lead condition, filled symbols), respectively by +60 or -60 degrees of phase, relative to our standard auditory-timing conditions in main Experiments 1 (and in Experiment 4, see open symbols in (a) here). Note that the Lag condition replicates our effect of motion-induced-by-auditory timing, but the effect of the Lead condition is much weaker.

Experiment 5: 'lead-lag' manipulation of auditory phase-offset relative to vision

A further possible 'high-level' account of our main findings, which is a potential alternative to our own suggestion in terms of temporal ventriloquism, is that any auditory events that intervene between one pair of flashes might merely draw attention to the interval between those flashes. Thus for a '*Right-Flash, Left-Flash Beep, Beep Right-Flash*' cycle, this might conceivably result in rightwards motion via possible attentional tracking [3-5] that might even produce subsequent aftereffects [5].

Experiment 5 sought to pit this hypothesis against our own temporal ventriloquism proposal. We compared two new conditions (see Fig S3) in which the phase of the auditory sequence was now shifted in time, either backwards or forwards equally, relative to the visual sequence (+/- 60 degree phase-shift relative to the arrangement in Experiment 1, c.f. Fig 1e/f in the main paper). Both of the new situations (Fig S3) created audio-visual sequences where flash and beep onsets now alternated (i.e. *'Flash-Beep, Flash-Beep*), unlike the original case (which had been '*Flash-Beep, Beep-Flash*' in Experiments 1 and 2, and likewise for the auditory-timing conditions in Experiment 4).In both of the new conditions, one beep always perfectly bisected the interval between a flash-pair (henceforth we call this the 'quarter-cycle' beep); the other beep either slightly *lagged* its respective flash in one condition, or slightly *led* in the other condition (see Fig S3a and S3b, respectively).

Note that the only difference between these conditions is the phase of the auditory sequence relative to the visual (as conveyed here by showing ongoing cycles of events). In the experiment itself, all of these audio-visual relations could reflect the opposite pairing with right-left or left-right visual flashes instead (i.e. only half of the possibilities are shown in Fig S3 for simplicity, with the quarter-cycle beeps that bisect intervals between flash onsets always being shown for left-right flash-onset intervals here, although on a random half of the trials these would instead bisect the right-left flash-onset intervals).

If intervening beeps serve merely to attract attention and thereby 'mark' the interval between a given flash-pair, then the same perceptual effect should presumably be predicted for both Lag and Lead conditions in the new situation, as beeps now always occured between successive flashes in both cases. By contrast, our own account in terms of temporal ventriloquism may lead one to expect stronger auditory-driven apparent motion for the Lag condition than the Lead (as indicated by the red arrow in Fig S3a), following prior evidence that audio-visual integration (including temporal ventriloquism) is typically stronger for sounds which *lag* rather than lead a visual event [2,6]. The reader may assess the experience of these different conditions in the accompanying supplementary demonstration movie.

Methods

Six experienced observers participated, four of whom had also taken part in Experiment 4. The two new male observers were naïve. Two conditions were compared, 'Lag' versus 'Lead', in which the phase of the standard auditory sequence (as per Experiment 1, see Figure 1e/f in main paper) was now shifted 62.5ms later or earlier respectively, relative to the visual event-cycle (+/- 60 degree shift in phase relative to Expt. 1 and also Expt. 4). Duration of auditory signals was 25ms. There were 60 trials for the lead condition and likewise for the lag condition, randomly intermingled, with each trials comprising a 2000ms audio-visual sequence.

Results and Discussion

Consistent with our own expectation, results showed strong sound-induced apparent visual motion in the Lag condition (i.e. t(5)= 6.28, $p<0.01$, see Figure S2b, open symbols), indicating that beeps could still be bound with their preceding flashes. In contrast, the directional influence of sounds was not significant in the Lead condition that had an equal but opposite phase-shift [t(5)= 1.24, n.s.], with this null effect for Lead (see filled symbols in Fig S2b) differing significantly from the positive effect for Lag (two-way interaction of lead/lag x directional effect, F(1,5) = 8.72, $p<0.05$]. These phase-shift effects appear to concur with our proposed account in term of temporal ventriloquism, while following less naturally from a simple attentional account in terms of any beep between visual onsets ‘marking’ that interval.

The actual direction of motion observed in the Lag condition (see red arrow in Figure S3a) may result from both the designated lagging beep *and* the quarter-cycle beep (which also lags its respective flash in Fig S3a) each attracting their preceding flashes later in time, so that those flashes appear to occur at (or closer to) the time of those beep onsets. Hence the perceived interval between those events may then be shorter, leading to the directional motion percept. In the Lead condition (Figure S3b), the same direction of motion might have been expected if the leading beep can play a strictly equivalent role to the lagging beep in the other condition. But the observed absence of such a consistent direction effect appears consistent with the established asymmetry of temporal ventriloquism, whereby lagging sounds tend to attract visual onsets more strongly than leading sounds, [2,6].

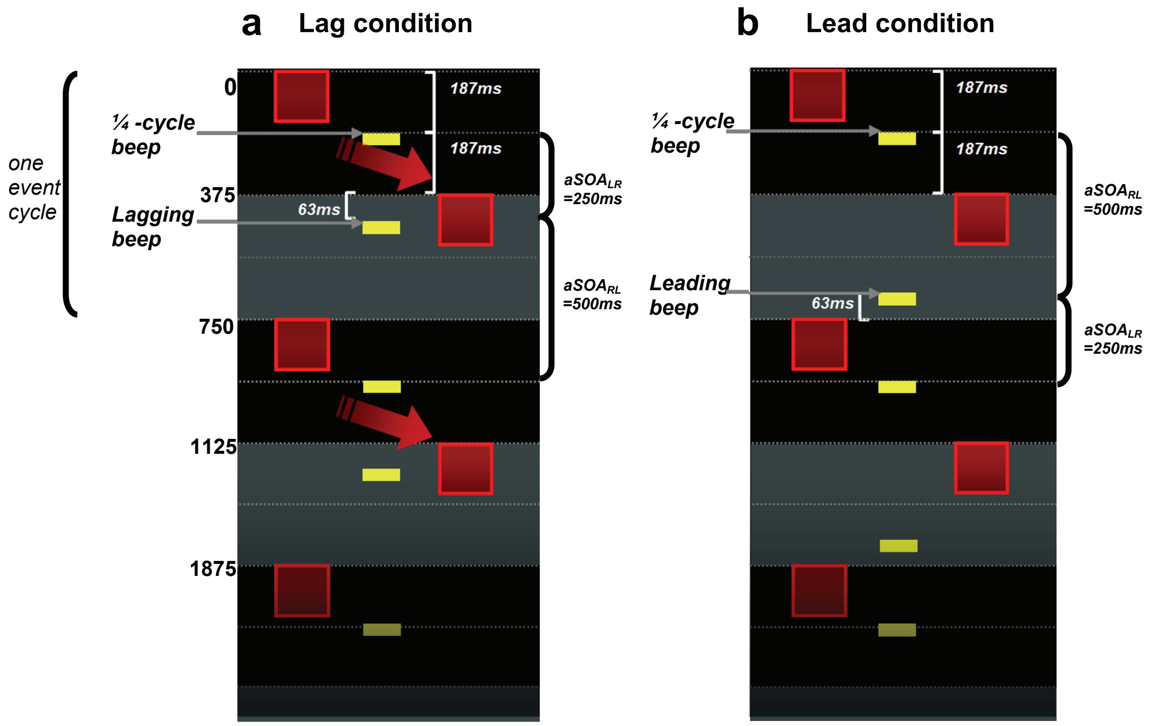


Figure S3: Space-Time diagrams for the new (a) Lag condition and (b) Lead condition, with visual events again indicated by red squares, and auditory events by yellow bars. The timing of auditory onsets was now phase shifted relative to visual events (by +/- 60 degrees of phase, relative to our standard audio-visual sequences in main Experiment 1 (cf. Figures 1 e/f in main paper) and in the auditory-timing conditions of Experiment 4. However, note that the auditory sequences themselves were unchanged in their temporal pattern (i.e. once again a relatively short-interval between successive beeps alternated with a longer-interval, as conveyed here by showing multiple ongoing cycles). Note that both the Lag and the Lead conditions had alternate beeps 'bisect' the interval between successive flash onsets, as shown here for the first beep from top in both cases. Note also that each beep now fell in a different inter-flash interval (these intervals are emphasized here by using alternating gray and black backgrounds, whereas the background was constant in the experiment). Moreover, every inter-flash interval now contained a single beep, quite unlike the situation in Experiment 1 (cf. main Fig 1e/f) or in the auditory-timing conditions of Experiment 4. Taken together, these aspects would lead one to expect equivalent motion effects for Lead and Lag conditions, if all that matters is whether or not a beep falls within an inter-flash interval. By contrast, our finding that consistent directional motion arose only in the Lag condition, as per the red arrow in (a), seems consistent with a temporal-ventriloquism account, since lagging sounds typically produce more ventriloquism than leading [2,6].

References

1. Gebhard, J.W., Mowbray, G.H. (1959). On discriminating the rate of visual flicker and auditory flutter. Am. J. Psychol. *72:521-9.***,** 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. Cavanagh, P. (1992). Attention-based motion perception. Science *257***,** 1563-1565.

4. Lu, Z.L., Sperling, G. (1995). Attention-generated apparent motion. Nature *377***,** 237-239.

5. Culham, J.C., Verstraten, F.A., Ashida, H. and Cavanagh, P. (2000). Independent aftereffects of attention and motion. Neuron *28***,** 607-615.

6. Dixon, N.F., Spitz, L. (1980). The detection of auditory visual desynchrony. Perception *9***,** 719-721.

Supplementary Movies

expt1.wmv

This movie shows examples of rightwards and then leftwards apparent motion induced purely by audio-visual asynchrony (equivalent to the Auditory Timing condition in Experiment 1, c.f. Figs 1e/f in main paper). Finally the sound is switched off, revealing ambiguous motion (c.f. Fig 1d). Note that the same regular visual timing is displayed throughout (i.e. an equal inter-flash interval for both left-to-right and right-to-left flashes).

expt3_asynch.wmv & expt3_synch.wmv

Typical trials in the duration discrimination paradigm of control Experiment 3 (see Supplementary Materials) for asynchronous audio-visual events, and synchronous (control condition). First interval is the reference, second is the test. Visual SOAs themselves are identical for both intervals in these examples.

expt4_asynch.wmv & expt4_synch.wmv

Typical trials in control Experiment 4 (see Supplementary Materials): the critical alternating-beep pitch condition with synchronous audio-visual events, and the corresponding control condition with asynchronous audio-visual events and no pitch modulation. We found no significant impact of the pitch manipulation.

expt5_Lag.wmv & expt5_Lead.wmv

Demonstrations for the Lag and Lead conditions of control Experiment 5 (see Supplementary Materials). Consistent rightwards motion is more predominant for the Lag condition, with more bistable or ambiguous motion for the Lead condition, thus leading to a consistent directional effect only for the Lag condition.