Temporal auditory capture does not affect the time course of saccadic mislocalization of visual stimuli

Paola Binda
Department of Psychology, Università Vita-Salute San Raffaele, Milano, Italy, & Italian Institute of Technology, IIT Network, Research Unit of Molecular Neuroscience, Genova, Italy

M. Concetta Morrone
Department of Physiological Sciences, Università di Pisa, Pisa, Italy, & Scientific Institute Stella Maris, Calambrone, Pisa, Italy

David C. Burr
Department of Psychology, Università Degli Studi di Firenze, Firenze, Italy, CNR Neuroscience Institute, Pisa, Italy, & Department of Psychology, University of Western Australia, Stirling Hw., Nedlands, Perth, Western Australia, Australia

Irrelevant sounds can "capture" visual stimuli to change their apparent timing, a phenomenon sometimes termed "temporal ventriloquism". Here we ask whether this auditory capture can alter the time course of spatial mislocalization of visual stimuli during saccades. We first show that during saccades, sounds affect the apparent timing of visual flashes, even more strongly than during fixation. However, this capture does not affect the dynamics of perisaccadic visual distortions. Sounds presented 50 ms before or after a visual bar (that change perceived timing of the bars by more than 40 ms) had no measurable effect on the time courses of spatial mislocalization of the bars, in four subjects. Control studies showed that with barely visible, low-contrast stimuli, leading, but not trailing, sounds can have a small effect on mislocalization, most likely attributable to attentional effects rather than auditory capture. These findings support previous studies showing that integration of multisensory information occurs at a relatively late stage of sensory processing, after visual representations have undergone the distortions induced by saccades.

Keywords: spatial vision, temporal vision, eye movements, multisensory, saccadic mislocalization, space–time

Citation: Binda, P., Morrone, M. C., & Burr, D. C. (2010). Temporal auditory capture does not affect the time course of saccadic mislocalization of visual stimuli. *Journal of Vision, 10*(2):7, 1–13, http://journalofvision.org/10/2/7/, doi:10.1167/10.2.7.

Introduction

When the spatial locations of visual and auditory stimuli are in conflict, vision usually dominates, a well-known phenomenon termed the "ventriloquist effect" (Mateeff, Hohnsbein, & Noack, 1985; Pick, Warren, & Hay, 1969; Radeau, 1994; Stekelenburg & Vroomen, 2009; Warren, Welch, & McCarthy, 1981). Many explanations have been advanced for the ventriloquist effect, the most successful being that it results from optimal cue combination: if information across senses is weighted according to the statistical reliability of the various sensory signals, vision will dominate perceived location because it specifies location more precisely than audition (Alais & Burr, 2004). Indeed, when visual stimuli are degraded, either by blurring (Alais & Burr, 2004) or by presenting them during saccades (Binda, Bruno, Burr, & Morrone, 2007), auditory information becomes more important for spatial location (see also Burr, Binda, & Gori, in press). Similar arguments have been made successfully for combination of various forms of multimodal information (e.g., Clark & Yuille, 1990; Ernst & Banks, 2002; Ghahramani, 1995).

While vision dominates over hearing in spatial vision, hearing dominates vision in the perception of time, a phenomenon termed auditory driving (Fendrich & Corballis, 2001; Gebhard & Mowbray, 1959; Recanzone, 2003; Shipley, 1964; Welch, DuttonHurt, & Warren, 1986) or "temporal ventriloquism" (Aschersleben & Bertelson, 2003; Bertelson & Aschersleben, 2003; Burr, Banks, & Morrone, 2009; Hartcher-O'Brien & Alais, 2007; Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Sounds can also alter the perception of a sequence of visual events, inducing the illusory perception of extra visual stimuli (Shams, Kamitani, & Shimojo, 2000, 2002). Sounds not only alter the perceived timing of visual flashes but, in some instances, can improve visual discrimination, for example by increasing their perceived temporal separation (Morein-Zamir et al., 2003; Parise &

doi: 10.1167/10.2.7 Received August 17, 2009; published February 11, 2010 ISSN 1534-7362

Spence, 2009). Auditory driving can also improve orientation discrimination of visual bars (Berger, Martelli, & Pelli, 2003), by increasing the number of representations of the stimulus, which in turn is known to improve discrimination performance (Verghese & Stone, 1995). That hearing dominates over vision in determining perceived event time is qualitatively consistent with the "Bayesian" account of multisensory integration, since auditory temporal cues are much more precise than the visual cues (Burr et al., 2009; Morein-Zamir et al., 2003; Recanzone, 2003). However, unlike spatial integration (Alais & Burr, 2004; Ernst & Banks, 2002), the quantitative predictions of the Bayesian model were found to be less than perfect (Burr et al., 2009).

We have previously studied the effect of saccadic eye movements on the integration of spatial auditory and visual signals (Binda et al., 2007). Saccades—frequent, rapid ballistic eye movements—have many consequences for perception, one being that stimuli flashed briefly around the time of saccades are grossly mislocalized, systematically displaced toward the saccadic landing point (Honda, 1989; Matin & Pearce, 1965; Morrone, Ross, & Burr, 1997; Ross, Morrone, & Burr, 1997; Ross, Morrone, Goldberg, & Burr, 2001). Importantly, the magnitude of the errors depends on the exact time of stimulus presentation relative to the saccade onset, peaking at saccadic onset and following clearly defined dynamics. Binda et al. (2007) studied the spatial ventriloquist effect during saccades, showing that saccades reduced considerably the weighting of the visual stimuli flashed together with sounds, with the results well predicted by reliability-based optimal integration. This suggests that saccades act on the visual signal, changing its reliability, before cross-modal visual-auditory integration.

In the present study, we investigated the effect of saccades on the perceived timing of audiovisual stimuli and tested whether an asynchronous sound source, known to advance or retard the perceived timing of a flash, could also affect the time course of saccadic mislocalization. If integration of visual and auditory cues occurs at an early stage (as suggested by Berger et al., 2003; Shams et al., 2000; Shams, Kamitani, Thompson, & Shimojo, 2001), before saccades perturb visual representations, then the dynamics of perisaccadic distortion should be altered by a sound that "temporally captures" the visual stimulus, shifted toward the sound. On the other hand, if visual representations are already distorted when they are integrated with signals from other modalities, no change in the dynamics of perisaccadic mislocalization should be observed.

A recent study (Maij, Brenner, & Smeets, 2009) has also addressed this issue, with inconclusive results. They reported that sounds can influence the time course of saccadic mislocalization, but only if they precede the visual stimulus, and then only by a fraction of the amount predicted. One possibility is that the small effects of Maij et al. were not due to auditory capture, but other mechanisms such as uncertainty and attention, known to affect timing. Indeed the results of the present study show that under conditions where auditory capture is strong, sounds do not affect the dynamics of spatial mislocalization, suggesting that the mislocalization precedes visual-acoustic integration. Control experiments with low-contrast stimuli resolve the apparent discrepancy with Maij et al.

These results have been presented at IMRF (Sydney, June 2007) and VSS (Naples, Florida, pre-conference symposium "Action for perception: Functional significance of eye movements for vision", May 9, 2008).

Methods

Apparatus and subjects

Experiments were performed in a quiet, dimly lit room. Subjects sat with their head stabilized by a chin rest 30 cm from the screen of a CRT color monitor (Barco Calibrator), which subtended 70 deg by 50 deg of visual angle. The monitor was driven at a resolution of 464 × 243 pixels and refresh rate of 250 Hz by a visual stimulus generator (Cambridge Research Systems VSG2/5) housed in Personal Computer and controlled by Matlab (Mathworks, Natick, MA). Visual stimuli were presented against a red background (Commission International de l'Eclairage (CIE) coordinates: $x = 0.624$; $y = 0.344$; luminance: 18 cd/m^2). Auditory stimuli were generated by the computer sound board and gated to a speaker placed above the monitor via a digitally controlled switch.

Four subjects participated in the experiment (one author and three naive to the goals of the experiment), all with normal or corrected-to-normal vision, and normal hearing. Experimental procedures were approved by the local ethics committees and are in line with the declaration of Helsinki.

Eye movements

Horizontal eye movements were recorded by an infrared limbus eye tracker (ASL 310), calibrated before each session. The infrared sensor was mounted below the left eye on plastic goggles through which subjects viewed the display binocularly. The PC sampled eye position at 1000 Hz and stored the trace in digital form. In offline analysis, saccadic onset was determined by an automated fitting procedure and checked by eye. The experimenter also checked the quality of saccades and, when necessary, discarded the trial (less than 10% of trials, in the presence of a corrective saccade or with unsteady fixation).

Data analysis

Analyses and data fitting were performed with custom software. Psychometric functions were fit with cumulative

Gaussian distributions, using the Maximum Likelihood method (Watson, 1979). Linear fit were performed with the standard Matlab function (Mathworks, Natick, MA), weighting data points by their squared standard errors. Standard errors of the (Gaussian or linear) fit parameters were estimated by bootstrap (Efron & Tibshirani, 1993; 1000 repetitions). Comparisons between values were performed with a Bootstrap sign tests with 5000 repetitions.

Experiment 1: Audiovisual temporal bisection task

This experiment used a similar technique to that of Burr et al. (2009) to measure auditory capture during saccades. At the beginning of each trial, subjects fixated a 1 deg black dot presented 10 deg left of screen center. After a variable delay of 1000 ms on average (SD: 200 ms), another 1 deg diameter black dot (the saccadic target) was presented 10 deg right of center and subjects saccaded to it. At about the time of saccadic onset, a bimodal audiovisual stimulus was delivered. The test stimulus comprised a green vertical bar (2 deg × 50 deg, CIE coordinates: x = 0.286; y = 0.585; luminance: 55 cd/m^2) flashed at the center of the screen (see Figure 1, upper panel) for one monitor frame and by a 16-ms white noise burst of about 60 dB at the subject's distance. The flash and the sound were asynchronous, separated by Δ ms, where Δ = ±50 ms or +10 ms. The different asynchronies were randomly intermixed within sessions. Note that, differently from Burr et al. (2009), we define Δ as the full interval between the flash and the sound and we take the time of flash presentation as measure of the stimulus presentation time.

The stimulus sequence comprised three successive audiovisual stimuli, a test with audiovisual conflict (described above), flanked by two temporal "markers" similar to the test except the audio and visual components were synchronous, separated in time by 800 ms. The test was presented between the markers, jittered by a Gaussian distribution with mean centered on the middle of the flankers and standard deviation of 80 ms (Figure 1). Subjects judged in two-alternative forced choice which of

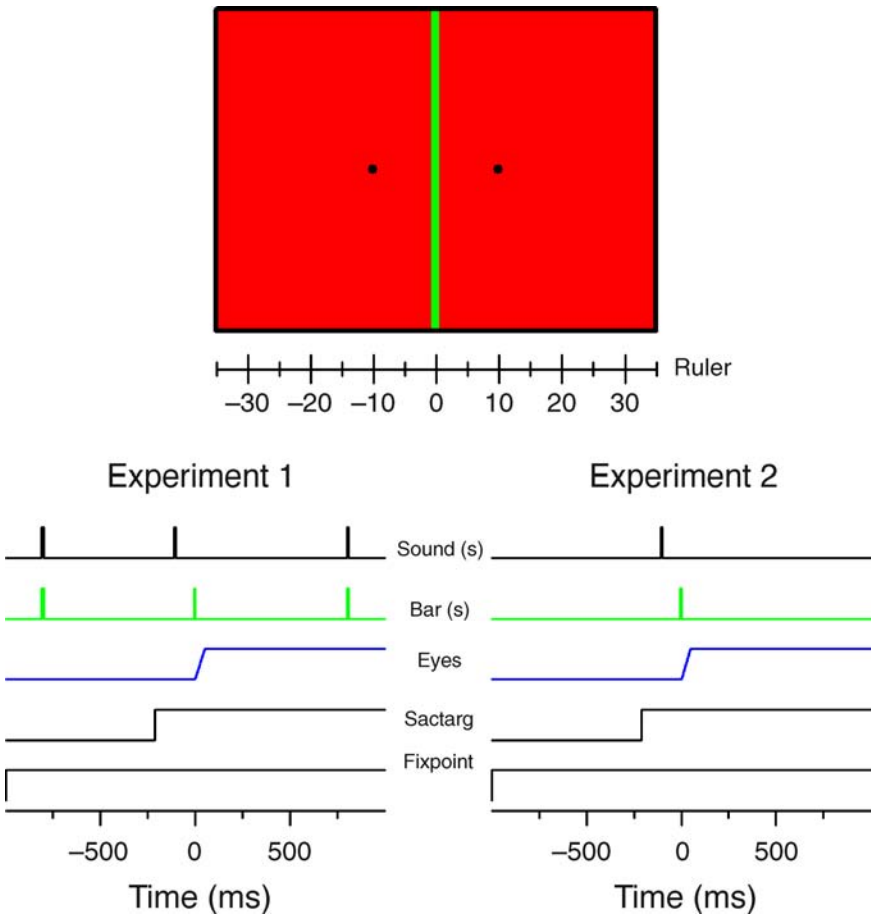

Figure 1. Spatial arrangement of the stimuli (upper panel): a vertical green bar was flashed at the center of the screen against a red background, sometimes accompanied by a noise burst played from a speaker placed above the monitor screen. At the beginning of each sitting, the ruler was displayed (while the calibration of the eye tracker was performed) for subjects to memorize. The lower panels show the time course of presentations, identical in the two experiments except there were no temporal markers in Experiment 2.

the two markers appeared to be temporally closer to the test. Only trials where the flash was presented within ±25 ms from the saccadic onset were analyzed. For each subject and condition, we measured psychometric functions (like those presented in Figures 2A and 2B), plotting the proportion of "closer to the first marker" responses against the flash presentation time relative to the middle of the markers. The median of the curves estimates the Point of Subjective Bisection (PSB), the time of the flash when the stimulus was perceived as bisecting the two markers. Each subject completed a minimum of 9 sessions, each of 50 trials, and 3 sessions of 50 trials for the control condition (steady fixation).

To estimate the relative weight of visual and auditory cues on perceived time of the stimulus, we regressed PSB values against conflict (Δ). From the estimated slope of the regression of PSB against Δ, we derive the weights of the visual (w_V) and auditory cues (w_A; see also Burr et al., 2009). Because we define the time of stimulus presentation as the flash presentation time, visual capture will result in PSBs being independent of Δ, and auditory capture in PSBs equal to Δ. Formally, the time of the auditory stimulus is Δ and the time of the flash is 0 (relative to the time taken to define the bimodal stimulus presentation, i.e., the flash time): assuming optimal combination of signals, the predicted timing of the audiovisual stimulus, $\hat{T}_{AV}(\Delta)$, is given by the weighted sum of the two signals, plus possible biases (b) such as temporal order effects:

$$\hat{T}_{AV}(\Delta) = 0w_V + \Delta w_A + b = \Delta w_A + b \quad (1)$$

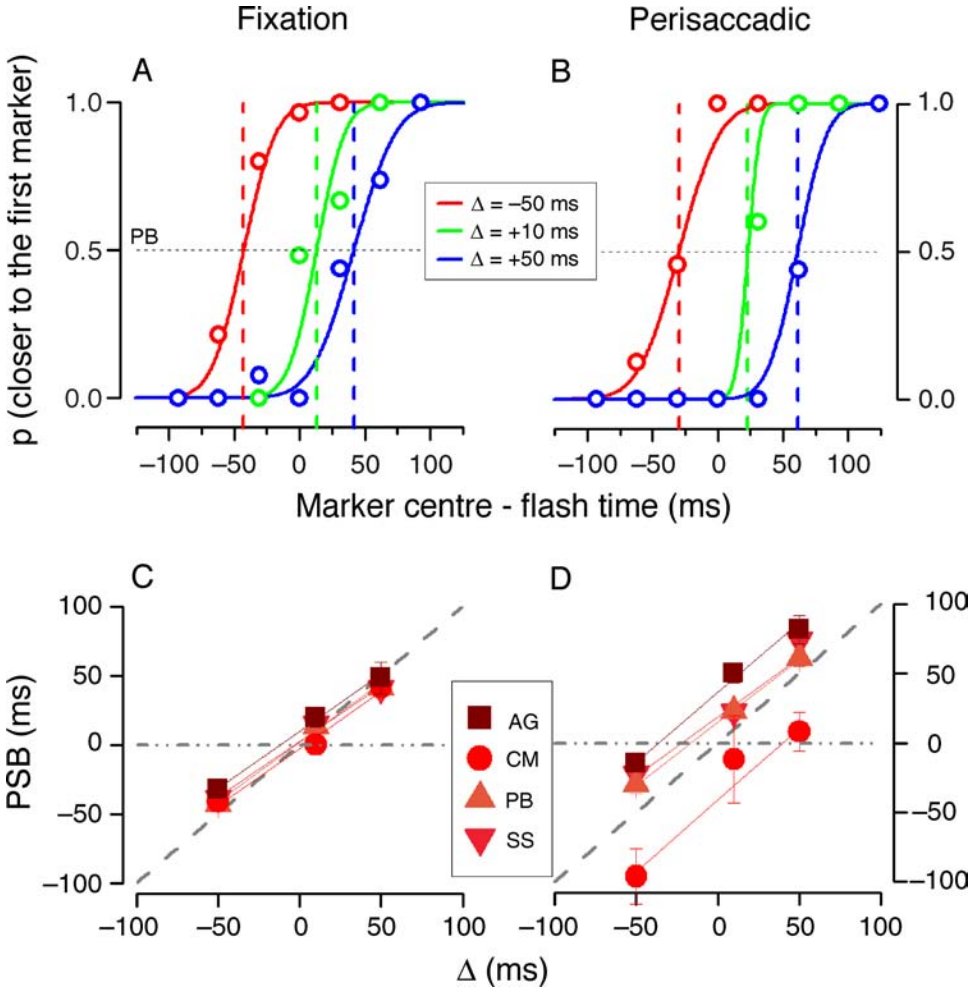

Figure 2. Results from the bimodal temporal bisection task (Experiment 1). Psychometric curves (A) for the three tested flash–sound asynchronies in fixation and (B) for bars flashed between 25 ms before and after saccade onset. The proportion of trials in which the stimulus was perceived as closer to the first temporal marker is plotted against the stimulus presentation time (the delay of marker center relative to flash time) and the distribution fit with a cumulative Gaussian function (different colors refer to different flash–sound separations). (C, D) The median of the curves estimates PSB values (point of subjective bisection), plotted as a function of Δ in the lower panels. The gray dashed line shows total auditory dominance (unitary slope), and the dash-dotted line (flat) shows visual dominance. Data points from the four subjects are reported with different colored symbols, with errors showing ±1 *SEM* calculated by bootstrap; the thin color-coded lines report the linear fit of each subject.

Differentiating with respect to Δ:

$$\hat{T}'_{AV}(\Delta) = w_A \quad (2)$$

where $\hat{T}'_{AV}(\Delta)$ is the slope of the regression of PSB against Δ. The visual weight w_V is $1 - w_A$. The estimates of visual and auditory weights obtained in this way were used to predict the shifts in mislocalization time course measured in the next experiment.

Experiment 2: Spatial localization of a bimodal audiovisual stimulus

This experiment examined the effect of a sound on the time course of saccadic mislocalization. The procedure was similar to that of Experiment 1, except that no temporal markers were presented, and subjects were required to report the apparent position of the bar relative to a memorized ruler displayed at the beginning of each session (Figure 1). Four different stimuli were randomly intermingled, a bar with no accompanying sound, or a bar-sound with the same audiovisual separations of Experiment 1 (±50 ms and +10 ms). Trials were binned into contiguous 10-ms intervals relative to saccadic onset and perceived stimulus location computed as the average reported location in each time bin (minimum of 5, average of 15 trials per bin). We also estimated a continuous localization curve by computing the average perceived stimulus location within a square temporal window 25 ms wide, sliding along the timings of stimulus presentation by steps of 1 ms.

Experiment 3: Manipulation of stimulus contrast

As mentioned in the Introduction section, Maij et al. (2009) have performed a similar study to this, reporting different results. To address this discrepancy, we repeated our experiment under conditions more similar to theirs, with weaker visual stimuli. For this we dropped the frame rate of the monitor to 100 Hz (to allow higher spatial resolution: 800 × 600), set the background to the maximum luminance obtainable (CIE coordinates: x = 0.338; y = 0.364; luminance: 90 cd/m^2), and illuminated the room to about 500 lux. Subjects executed 8 deg saccades, following a fixation dot (diameter: 0.5 deg; luminance: 0.7 cd/m^2) that appeared at the center of the screen, jumped to 8 deg right (or, in separate sessions, left) of center and disappeared after ~200 ms, just before the flash stimulus was presented. In this case, the flash was a small dot (diameter: 0.5 deg) presented for one frame and defined only by luminance contrast. In separate sessions, the flash could have high contrast (90%), or low, near-threshold contrast, yielding 50–75% correct response (contrast 15% for PB; 75% for MCM). The flash was usually displayed 3 deg beyond the saccadic target (producing a mislocalization against the direction of the saccade) and in one case 3 deg before the saccadic target. A noise burst (60 dB at the subject's distance; 20-ms duration) was played simultaneously with or 50 ms before/after the flash (intermingled across trials). Subjects reported perceived flash location by adjusting the position of the mouse pointer and left-clicking when satisfied (or right-clicking if they failed to detect the stimulus).

Two subjects (two authors) participated in this experiment.

Results

Experiment 1: Auditory capture during saccades

The first experiment measured audiovisual temporal capture in fixation and during saccades. Subjects reported which flanking marker appeared closer in time to an audiovisual test stimulus (a flash and a sound presented asynchronously). Presentation time of the test relative to the markers was varied to produce psychometric functions like the examples shown in Figures 2A and 2B. The curves plot the proportion of "test closer to the first marker" responses against the presentation time of the flash (the visual component of the test stimulus). The effect of audiovisual conflict is clear from inspection of the curves: if judgments were uniquely dependent on the visual stimulus, the three curves would have been aligned to each other. Instead, negative conflict (leading sound, red curves) shifted the curve leftward to more negative time values, while positive conflicts (lagging sound, blue curves) shifted the curves rightward by an amount nearly equal to the flash–sound separation, implying that perceived time is determined primarily by sound (agreeing with Burr et al., 2009). The auditory dominance is apparent in both fixation (A) and saccade (B) conditions.

To quantify the magnitude of the effect, we calculated the point of subjective bisection (PSB) from the median of the best-fitting cumulative Gaussian and plot this as a function of the conflict Δ (Figures 2C and 2D). The PSBs were regressed against Δ (weighing each point by its squared standard error); regressions are shown by the thin color-coded lines. The slope of the regression lines estimates the weight of auditory temporal information (Equations 1 and 2 in Methods section) and of visual information (auditory and visual weights sum to 1). For all four subjects, in both fixation and saccade conditions, slopes are close to unity, implying strong auditory dominance of perceived time (visual dominance would have resulted in a flat line of zero slope; equal visual and auditory weights would have produced a slope of 0.5).

Figure 3 plots the estimated visual weights in the saccadic condition against those in fixation, with the black star representing the average visual weight (0.17 in fixation and 0.05 during saccades). For all four subjects, the data lie below the equality line, implying lower visual weight (and therefore higher auditory weight) for saccades than in fixation. The difference in weights in the saccadic and fixation conditions was statistically significant (p = 0.049[1]). This result is consistent with the idea that saccades affect visual stimuli before the integration with auditory stimuli: if saccades acted after integration, they should not affect the integration weights.

For three out of four subjects, the intercept of the best-fitting regression is more positive during saccades than during fixation (Figures 2C and 2D). The mean difference was 18.2 ms, which is statistically significant ($p < 0.001$, bootstrap sign test on pooled data). This implies that for all flash–sound separations, the audiovisual stimulus is perceived as delayed when it is presented at the time of a saccade. This is consistent with other studies from our laboratory (Binda, Cicchini, Burr, & Morrone, 2009)

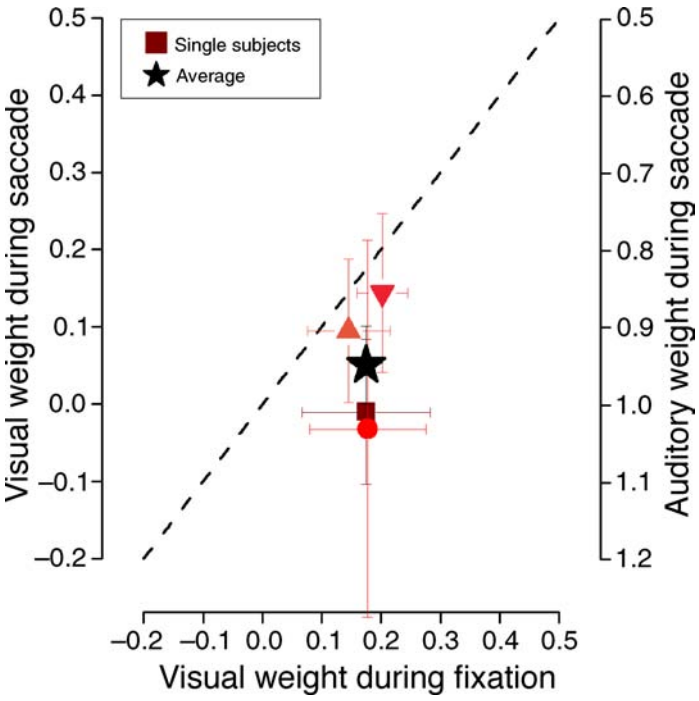

Figure 3. Visual and auditory weights during saccades are plotted against those in fixation. Weights were derived from the slope of the best-fitting regressions of Figures 2C and 2D (Equations 1 and 2 in Methods section). Errors show ±1 *SEM*, calculated by bootstrap. Visual weights (left ordinate) and auditory weights (right ordinate) sum to 1. Both during fixation and during saccades, visual weights are far lower than auditory weights, implying strong auditory capture. All data points lay below the equality line, suggesting that auditory capture is even stronger during saccades than during fixation. The star reports the average across subjects (the average visual weight during saccades is 0.05; the auditory weight is 0.95).

showing that saccades cause a delay in the perceived time of the flash. The delay reported in that study (for purely visual stimuli) was 50–100 ms, while here it is only ~20 ms. The reduced delay is to be expected, as audition is dominant in temporal judgments (especially during saccades), thereby reducing (but not completely eliminating) the saccade-induced delay (similarly to the effect on perceived location: Binda et al., 2007).

Experiment 2: Effect of auditory capture on time course of saccadic mislocalization

Experiment 1 clearly shows that a sound presented near the time of a flash captures it in time, both during fixation and—even more so—during saccades. Given an estimated visual weight of 0.05, a sound presented 50 ms before or after a flash should displace the flash forward or backward in time by about 47.5 ms. Here we ask whether this bias of perceived flash time affects the time course of saccadic mislocalization.

This experiment was like the previous, except there were no temporal markers and subjects were required to report the apparent position of the stimulus, rather than apparent timing. Four conditions were randomly intermingled within sessions: flashes presented with no accompanying sound, and flashes presented with a sound with the same offsets of Experiment 1 (±50 ms and +10 ms). Figure 4 shows the results, plotting perceived position against flash presentation time (relative to the saccade onset), with different colors representing different flash–sound separations. It is evident that curves for all audiovisual offsets follow very similar time courses, with no tendency whatsoever of a shift toward the auditory stimulus. The upper insert shows the shift predicted by auditory capture. The curves are splines of the average of all data in the vision-only condition, shifted by the amount predicted by auditory capture (the audiovisual offset times the auditory weight). The predictions clearly do not match the measured results.

For each subject and audiovisual delay, we quantified the amount the sound shifted the curve by sliding it along the time axis to find the point where it coincided best with the no-sound condition. We smoothed first the mislocalization data for the no-sound condition to a continuous curve of 1-ms resolution by computing the average perceived stimulus location within a square temporal window 25 ms wide, moved along the stimulus timings by 1-ms steps. For each subject and condition, we compared the mislocalization data (raw data) to the smoothed no-sound curve, calculating the mean-squared residuals for the data set. We repeated this procedure for positive and negative time shifts (20-ms range, 1-ms resolution) and considered the best shift that yielded the lowest mean-squared residual (error bars calculated by bootstrapping the whole procedure). Figure 5 plots these values separately for each subject as a function of audiovisual asynchrony and compares them with the

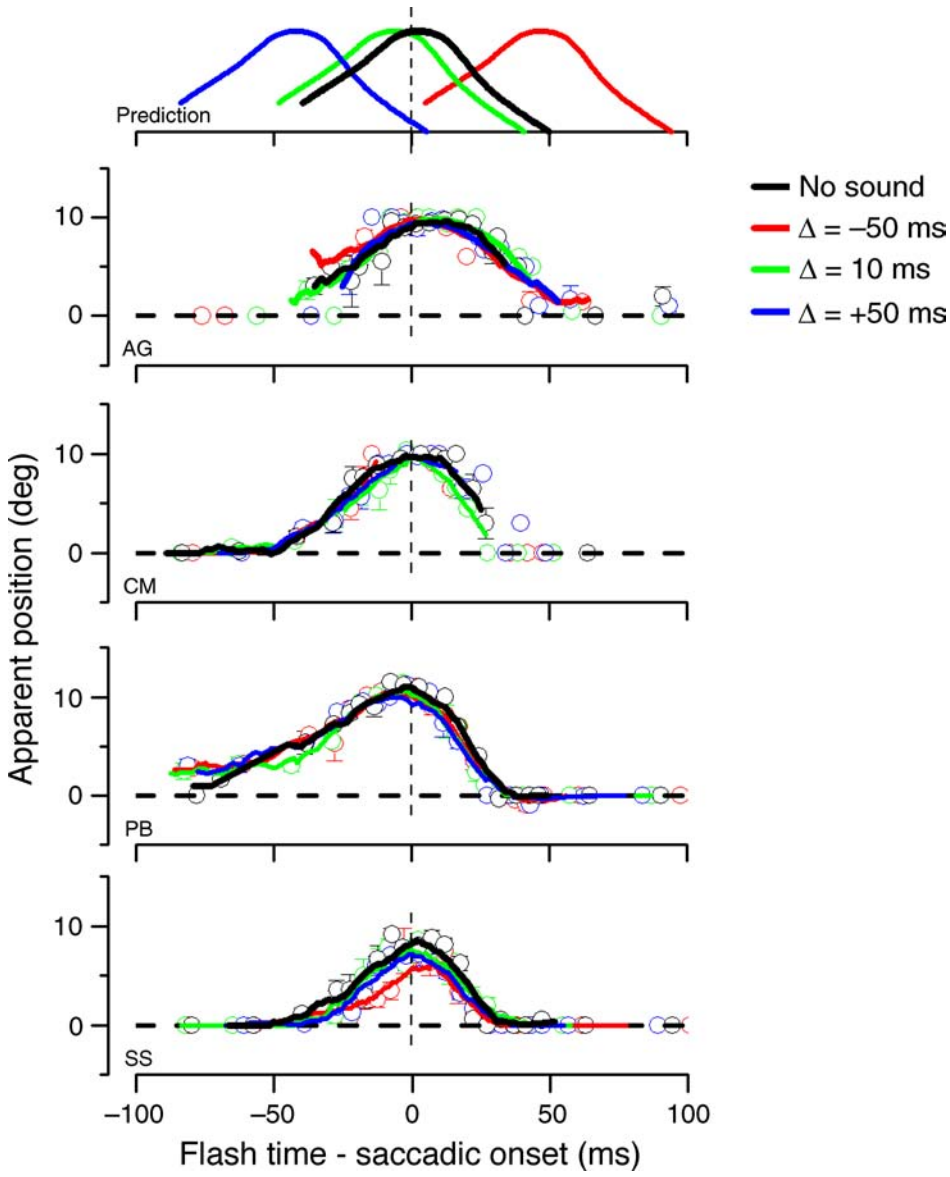

Figure 4. Dynamics of perisaccadic flash mislocalization, with and without an asynchronous auditory presentation accompanying the flash. The upper panel illustrates the predictions from Experiment 1 if sound were to capture vision before saccadic mislocalization. A sample curve (black) is a b-spline interpolation of the average mislocalization curve for the no-sound condition across subjects; the same curve is shifted by a variable delay (computed on the basis of the strength of the auditory capture estimated in Experiment 1) and predicts the mislocalization of a flash preceded or followed by a sound (red, green, and blue curves, see Methods section). The lower panels report data from the four tested subjects. Data points (hollow symbols) give the average reported flash location in 10-ms time bins (with minimum of 5, average of 15 trials per bin); continuous curves report the running average of perceived location (computed by taking the average perceived stimulus location within a square temporal window 25 ms wide, sliding along the timings of stimulus presentation by steps of 1 ms).

predictions (color-coded lines without data). Clearly, the data show no tendency to follow the predictions nor to deviate from zero. The observed dependency on audio-visual asynchrony, given by the regression of displacement against asynchrony, was -0.004 ± 0.016 ms on average and in no subject was it statistically different from zero (bootstrap sign test).

We conclude that while the perceived timing of a perisaccadic flash can be substantially biased by an asynchronous auditory presentation, the dynamics of saccadic mislocalization remains essentially the same as when no sound is delivered.

Experiment 3: Manipulation of stimulus contrast

Although Experiment 2 showed clearly that sounds do not affect the dynamics of flash mislocalization, a recent

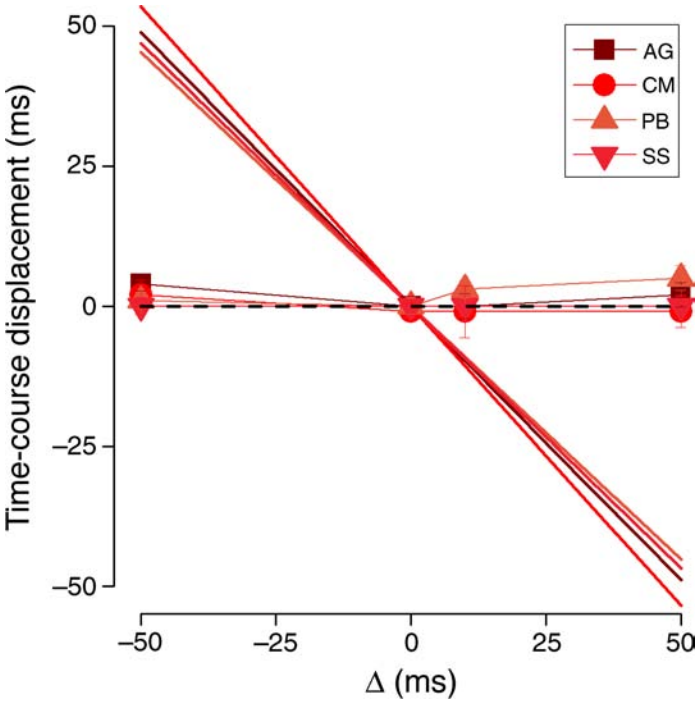

Figure 5. The effect of asynchronous sounds on time course of saccadic mislocalization. Displacement of the time course (calculated by minimizing squared residual errors) is plotted against the audiovisual asynchrony, separately for each subject. Error bars show ±1 *SEM* calculated by bootstrap of the fitting (1000 iterations). The color-coded lines without symbols show the predictions in shift, based on the auditory weights calculated from Figure 2. The data clearly show no tendency to follow the predictions. The slopes of the best-fitting regression lines are −0.02, −0.03, 0.03, 0.00, none is statistically different from 0 (*p*-values computed with bootstrap sign test: 0.12, 0.07, 0.22, 0.3).

study (Maij et al., 2009) reported that a leading sound can displace slightly the mislocalization time course. To address this discrepancy, we performed a further experiment to attempt to replicate Maij et al.'s result, by reducing the contrast of our stimuli to make them less visible.

As in Maij et al., subjects localized a 0.5 deg dark spot presented against a light-gray background, at the same vertical position as the saccadic target and 3 deg apart from it. We also allowed a "no-flash-detected" response. The flash was accompanied by a noise burst (Δ = ±50 or 0 ms, intermixed between trials). For one subject, we tested (in separate sessions) the mislocalization with both rightward and leftward saccades and with two stimuli positions (before or beyond the saccadic target). Being small and defined by luminance contrast only, the flash provided a much weaker visual signal than the 2 × 50 deg colored bar of our previous experiment. However, its position and timing were fairly predictable in each session, unlike Maij et al.'s study where the flash appeared at a variable location after a sequence of saccades in random directions.

Two stimulus contrast levels were tested: low (near detection threshold) and high (allowing nearly perfect detection). The results for the two conditions are reported in Figures 6 and 7, respectively. The leftmost panels show the mislocalization time courses for the three tested Δ values. The rightmost panels report the percent of trials where subjects were able to detect the stimulus (i.e., those trials from which the localization curves on the leftmost panels were computed); detection rates were computed by pooling trials across the time course. Stimuli presented 3 deg beyond the saccadic target (Figures 6, 7C, and 7E)

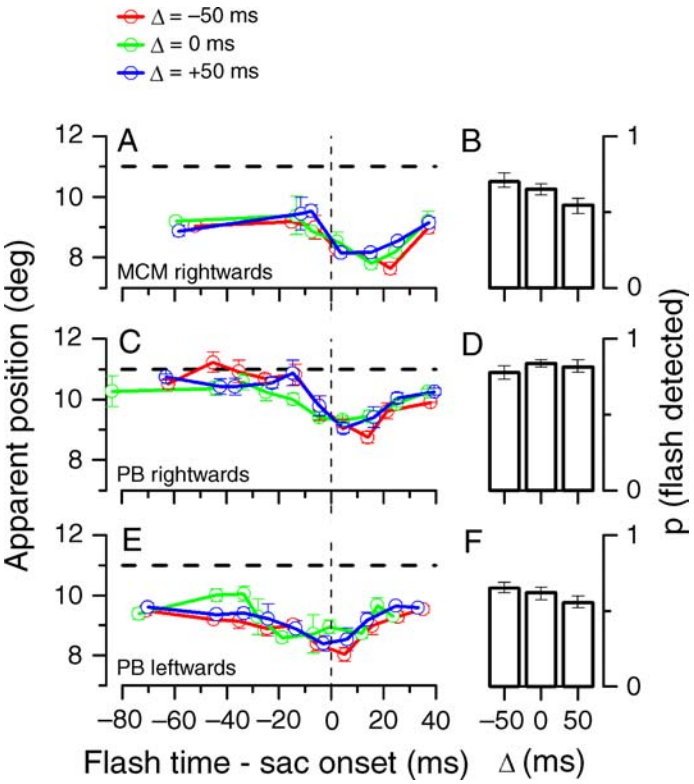

Figure 6. The effect of sound on mislocalization of low-contrast stimuli. (Left) Mislocalization curves for (A) MCM rightward saccades, (C) PB rightward saccades, and (E) PB leftward saccades. In all cases, the stimulus flash appeared 3 deg beyond the saccadic target (flash position marked by the dashed horizontal line, saccadic target at 8 deg). Data points give the average reported flash location in contiguous time bins (with minimum of 5, average of 15 trials per bin), joined by straight lines. Red symbols and lines refer to sounds leading by 50 ms, blue to sounds lagging by 50 ms, and green to simultaneous flash and sounds. (Right) The proportion of trials in which the subject detected the presence of the stimulus, pooling trials across the time course for the three tested Δ values. Otherwise, subjects did not attempt to report its location. Error bars are ±1 *SEM* computed by bootstrap.

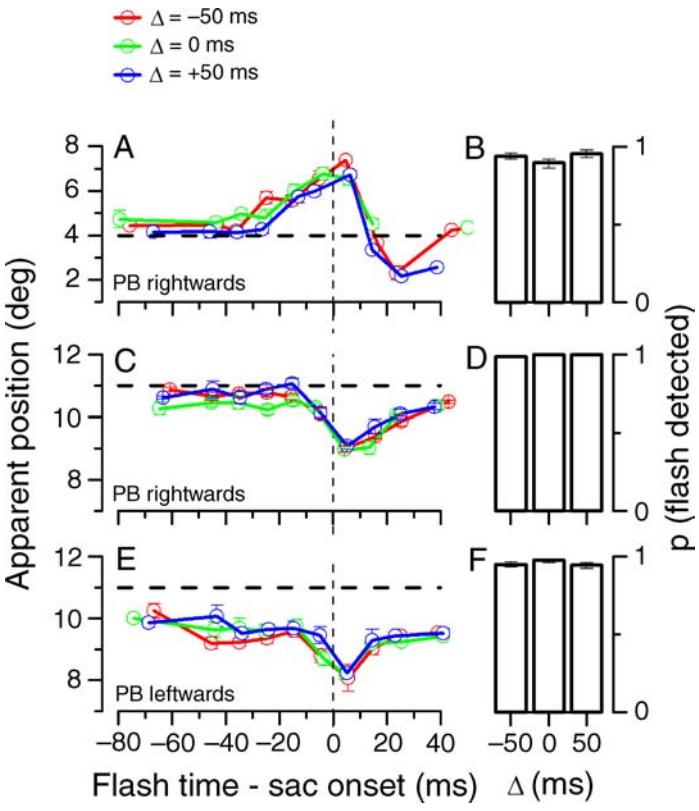

Figure 7. As for Figure 6 (without subject MCM), except the flashed stimuli were displayed at 90% contrast, hence almost always detected.

were mislocalized against the direction of the saccade, as expected, while stimuli displayed 3 deg before the saccadic target were mislocalized in the direction of the saccade (Figure 7A), as in Experiment 2.

For low-contrast stimuli (Figure 6), sounds did have a small effect on the mislocalization time course. In agreement with Maij et al.'s (2009) report, we found that a leading sound ($\Delta = -50$ ms) produces a small displacement of the mislocalization time course, mostly for the part of the time course where stimuli were flashed during the saccade (positive flash-saccadic onset delays). Importantly, there was a tendency of leading sounds to improve flash detection, especially where flash contrast was near detection threshold (Figures 6B and 6F). However, lagging sounds ($\Delta = +50$ ms) did not displace the curves (agreeing with Maij et al., 2009), nor did they affect detection rate.

For higher contrast stimuli (detected at ~100% rate), mislocalization time courses were all aligned with each other, irrespectively of sounds, for leftward and rightward saccades, and for stimuli presented before and beyond the saccadic target (Figure 7), agreeing with our main results (Figure 4). This shows that the small and asymmetrical effect occurs only for stimuli at near-threshold contrast. As Experiment 1 shows that the higher contrast stimuli are captured by sounds, it would seem that the small effect for low-contrast stimuli is due to other phenomena, discussed below.

Discussion

We investigated the effect of asynchronous auditory stimuli on the time course of spatial mislocalization of brief visual stimuli. We replicated previous findings (Aschersleben & Bertelson, 2003; Bertelson & Aschersleben, 2003; Burr et al., 2009; Hartcher-O'Brien & Alais, 2007; Morein-Zamir et al., 2003) showing that auditory stimuli exert a strong temporal attraction over flashed visual stimuli and further showed that auditory dominance increases at the time of saccades. However, despite auditory capture being strong during saccades, the displacement of flash perceived time did not result in any measurable displacement of the time course of saccadic mislocalization.

It is to be expected that the dominance of audition over vision should increase during saccades, assuming, as evidence suggests (Binda et al., 2009; Morrone, Ross, & Burr, 2005), that temporal visual localization becomes less precise during saccades. We have recently described a similar phenomenon in the spatial domain, where vision usually dominates in determining the perceived location of an audiovisual stimulus (Alais & Burr, 2004); as saccades reduce the spatial resolution of vision, the strength of visual capture is reduced during saccades (Binda et al., 2007).

Experiment 1 also showed a slight delay in the perceived time of a perisaccadic bimodal stimulus, relative to fixation. This finding is consistent with recent evidence that saccades cause a strong delay in visual stimuli (Binda et al., 2009). Interestingly, the delay reported in normal conditions was in the order of 50–100 ms, whereas here it is only about 20 ms. Again, this is consistent with the concept of optimal integration between senses: as the perceived time of the composite stimulus was determined only partly by vision, the presence of the auditory stimulus reduced considerably the saccade-induced delay.

Maij et al. (2009) recently reported an experiment studying the effects of sounds on saccadic mislocalization, with different results from those reported here. While subjects made free saccades, they were presented with a flash, along with a leading or trailing sound. Sounds presented after flashes had no effect on the mislocalization time course, but leading sounds caused a small shift of about 15 ms in the mislocalization curve. The authors developed a multistage model to account for this asymmetrical effect based on estimates of auditory weights.

Several details in their procedure were different from those of our main experiment. Firstly, their technique for estimating auditory weights differed from ours. Rather

than following the standard practice (e.g., Ernst & Banks, 2002) of using bimodal markers (so both auditory and visual information are equally available), Maij et al. used visual flashes as markers, biasing the task toward vision. This probably lead to an overestimation of visual weight, which they report to be around 0.5 (equal to the auditory weight), very different from the near-zero weights of previous studies (Burr et al., 2009; Hartcher-O'Brien & Alais, 2007) that predict the phenomena of temporal ventriloquism (Aschersleben & Bertelson, 2003; Bertelson & Aschersleben, 2003; Burr et al., 2009; Hartcher-O'Brien & Alais, 2007; Morein-Zamir et al., 2003), "auditory driving" (Fendrich & Corballis, 2001; Gebhard & Mowbray, 1959; Recanzone, 2003; Shipley, 1964; Welch et al., 1986), and illusory multiple flashes (Shams et al., 2000, 2002). However, even with this questionable prediction of equal auditory and visual temporal dominance, Maij et al.'s model still predicts a displacement of time courses larger than that they observed.

Another major difference in the studies was the choice of stimuli. The stimuli for our primary studies (Figure 4) were large, high-contrast bars, modulated in both luminance and color, presented at fairly predictable positions. Those of Maij et al., however, were small dot stimuli (1/500th the area of ours) presented at unpredictable times and positions, modulated only in luminance (which is suppressed during saccades, e.g., Diamond, Ross, & Morrone, 2000), therefore likely to evoke weak neural responses. We therefore repeated our study with barely visible stimuli (Experiment 3) and showed that, under these conditions, leading (but not trailing) sounds displaced the mislocalization time course by some 10–20 ms, similar to what was observed by Maij et al. (2009). Interestingly, the leading (but not trailing) sounds also enhanced detection of the visual stimuli. However, when the contrast of the stimuli was increased above threshold, neither leading nor trailing sounds had any effect on mislocalization time courses (even though clear auditory capture could be demonstrated under these conditions). Thus we concur that sounds can have a small effect on the time course of flash mislocalization, but only under very specific conditions: either very low-contrast stimuli (like our Experiment 3) or small and spatiotemporally unpredictable (as in Maij et al.'s study). In addition, even in these specific conditions, neither the magnitude nor the symmetry of the effect is predicted quantitatively by auditory capture.

The asymmetry of the effect, together with the fact that it occurs only for weak stimuli, suggests an alternative explanation for Maij et al.'s result: sounds preceding the flash may have acted as a "cue", enhancing the visual signal (Driver & Spence, 1998; McDonald, Teder-Salejarvi, & Hillyard, 2000; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008) and thereby causing faster processing. As observed by Titchener (1908), attended stimuli are processed more quickly than non-attended stimuli, leading to the so-called prior-entry effect. Electrophysiological studies support this observation, showing that attending to stimuli accelerates the neural response by about 15 ms (Di Russo, Martinez, & Hillyard, 2003; Di Russo & Spinelli, 1999). This decrease in response latency for attended stimuli could explain the small and asymmetric displacement of time courses reported by Maij et al. Sounds will cue flashes, drawing attention to them, only if they coincide with or precede them. The cueing would be far less effective for high-contrast stimuli, and indeed no effect is observed under these conditions. That the leading (but not trailing) sounds increased detection performance is further evidence that attention-like processes are involved.

Taken together, the experiments reported in the present study suggest that perisaccadic mislocalization occurs at an earlier level of processing than integration of visual and auditory cues (spatial or temporal). For the visual weight to change with saccades, the saccades need to exert their influence before audiovisual combination. In addition, if auditory capture occurred before perisaccadic mislocalization, it should affect the time course of mislocalization, dragging the visual stimuli closer or farther in time from saccadic onset. The results reported here agree with and complement our previous study (Binda et al., 2007) that showed that the localization of perisaccadic bimodal stimuli presented at the time of saccades can be quantitatively predicted by assuming an optimal integration of auditory and visual cues, with the accuracy and precision of visual signals changing dynamically. For this prediction to work, it is necessary that the visual signals are distorted by saccades before the site of integration with other sensory cues. They must be biased and imprecise (as suggested by visual measurements) when combined with auditory information, or the bimodal localization could not be predicted from the visual and auditory time courses. This would also explain why their dynamics remains unaltered despite the perceived time of the flash is captured by an auditory presentation.

Visual signals are initially encoded in a retinotopic frame of reference, which shifts each time the eyes move. For vision to remain stable across saccades, retinotopic representations need to be converted into gaze-invariant (allocentric) maps that take into account the position of eye gaze. We propose that, in the case of rapid gaze shifts, eye position information fails in accuracy and precision, resulting in systematic localization errors and in the decrease of localization precision (Binda et al., 2007). Audition, on the other hand, encodes stimuli in craniotopic coordinates, so the spatial cues from the two modalities need to be converted into a common format before integration. Neurophysiological studies have revealed that a variety of frames of reference are used to encode auditory signals (in A1 and the inferior colliculus: Groh, Trause, Underhill, Clark, & Inati, 2001;

Werner-Reiss, Kelly, Trause, Underhill, & Groh, 2003) and audiovisual signals, spanning the full range between eye-centered and head-centered (in the intraparietal sulcus, Mullette-Gillman, Cohen, & Groh, 2005; Schlack, Sterbing-D'Angelo, Hartung, Hoffmann, & Bremmer, 2005). However, for the purpose of localizing stimuli at the time of saccades, a convenient format is craniotopic, stable across eye movements. Inaccurate and imprecise eye position signals will ultimately lead to a distorted representation of those visual signals that constitute the input to the process of multisensory integration. Deneve, Latham, and Pouget (2001) demonstrated that a class of neural networks can both integrate optimally multisensory signals and convert each signal into a new reference frame. In principle, such a network can simulate our findings in both the unimodal and the bimodal conditions, assuming that the output of the network is required to be craniotopic in all cases, and that eye position input is inaccurate and imprecise. A detailed model of how this could occur is presented in Binda et al. (2007).

We have argued, on many occasions (Binda et al., 2009; Burr & Morrone, 2006; Burr, Tozzi, & Morrone, 2007; Morrone, Ross, & Burr, in press), that the neural processing of visual space and time is closely linked. Time does not seem to be determined by a generic clock, but specific to each sense, and to each spatial position (Johnston, Arnold, & Nishida, 2006). Furthermore, the spatial selectivity is in allocentric, not retinotopic coordinates (Burr et al., 2007). The results reported here reinforce further this position, showing that the effects of saccades on space and time precede cross-sensory integration, so the integration of both dimensions can occur in an allocentric space that takes into account the eye movement.

Acknowledgments

This research was supported by the Italian Ministry of Universities and Research and by EC Projects "MEMORY" (FP6 NEST) and "STANIB" (FP7 ERC).

Commercial relationships: none.
Corresponding author: Paola Binda.
Email: p.binda1@in.cnr.it.
Address: Istituto di Neuroscienze del CNR, Via Moruzzi 1, 56124 Pisa, Italy.

Footnote

[1]For significance testing, the data from the four subjects were aligned by subtracting the intercept of each regression line, then pooled; the slope of the linear regression for saccade and fixation were compared with a bootstrap sign test.

References

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

Aschersleben, G., & Bertelson, P. (2003). Temporal ventriloquism: Cross-modal interaction on the time dimension: 2. Evidence from sensorimotor synchronization. *International Journal of Psychophysiology, 50,* 157–163. [PubMed]

Berger, T. D., Martelli, M., & Pelli, D. G. (2003). Flicker flutter: Is an illusory event as good as the real thing? *Journal of Vision, 3*(6):1, 406–412, http://journalofvision.org/3/6/1/, doi:10.1167/3.6.1. [PubMed] [Article]

Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Cross-modal interaction on the time dimension: 1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology, 50,* 147–155. [PubMed]

Binda, P., Bruno, A., Burr, D. C., & Morrone, M. C. (2007). Fusion of visual and auditory stimuli during saccades: A Bayesian explanation for perisaccadic distortions. *Journal of Neuroscience, 27,* 8525–8532. [PubMed] [Article]

Binda, P., Cicchini, G. M., Burr, D. C., & Morrone, M. C. (2009). Spatiotemporal distortions of visual perception at the time of saccades. *Journal of Neuroscience, 29,* 13147–13157. [PubMed]

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

Burr, D., & Morrone, C. (2006). Time perception: Space–time in the brain. *Current Biology, 16,* R171–R173. [PubMed]

Burr, D., Tozzi, A., & Morrone, M. C. (2007). Neural mechanisms for timing visual events are spatially selective in real-world coordinates. *Nature Neuroscience, 10,* 423–425. [PubMed]

Burr, D. C., Binda, P., & Gori, M. (in press). Combining vision with audition and touch, in adults and in children. In J. Trommershäuser, M. Landy, & K. Körding (Eds.), *Sensory cue integration*. Oxford, UK: Oxford University Press.

Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems*. Boston, MA: Kluwer.

Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience, 4,* 826–831. [PubMed]

Diamond, M. R., Ross, J., & Morrone, M. C. (2000). Extraretinal control of saccadic suppression. *Journal of Neuroscience, 20,* 3449–3455. [PubMed] [Article]

Di Russo, F., Martinez, A., & Hillyard, S. A. (2003). Source analysis of event-related cortical activity during visuo-spatial attention. *Cerebral Cortex, 13,* 486–499. [PubMed]

Di Russo, F., & Spinelli, D. (1999). Electrophysiological evidence for an early attentional mechanism in visual processing in humans. *Vision Research, 39,* 2975–2985. [PubMed]

Driver, J., & Spence, C. (1998). Cross-modal attention. *Current Opinion Neurobiology, 8,* 245–253. [PubMed]

Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. New York: Chapman & Hall.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415,* 429–433. [PubMed]

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics, 63,* 719–725. [PubMed] [Article]

Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *American Journal of Experimental Psychology, 72,* 521–528. [PubMed]

Ghahramani, Z. (1995). *Computation and psychophysics of sensorimotor integration*. Cambridge, MA: Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology.

Groh, J. M., Trause, A. S., Underhill, A. M., Clark, K. R., & Inati, S. (2001). Eye position influences auditory responses in primate inferior colliculus. *Neuron, 29,* 509–518. [PubMed]

Hartcher-O'Brien, J., & Alais, D. (2007). *Temporal ventriloquism: Perceptual shifts forwards and backwards in time predicted by the maximum likelihood model*. 8th Annual Meeting of the International Multisensory Research Forum, University of Sydney, Australia.

Honda, H. (1989). Perceptual localization of visual stimuli flashed during saccades. *Perception & Psychophysics, 45,* 162–174. [PubMed]

Johnston, A., Arnold, D. H., & Nishida, S. (2006). Spatially localized distortions of event time. *Current Biology, 16,* 472–479. [PubMed]

Maij, F., Brenner, E., & Smeets, J. B. (2009). Temporal information can influence spatial localization. *Journal of Neurophysiology, 102,* 490–495. [PubMed]

Mateeff, S., Hohnsbein, J., & Noack, T. (1985). Dynamic visual capture: Apparent auditory motion induced by a moving visual target. *Perception, 14,* 721–727. [PubMed]

Matin, L., & Pearce, D. G. (1965). Visual perception of direction for stimuli flashed during voluntary saccadic eye movements. *Science, 148,* 1485–1488. [PubMed]

McDonald, J. J., Teder-Salejarvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407,* 906–908. [PubMed]

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

Morrone, M. C., Ross, J., & Burr, D. (2005). Saccadic eye movements cause compression of time as well as space. *Nature Neuroscience, 8,* 950–954. [PubMed]

Morrone, M. C., Ross, J., & Burr, D. C. (1997). Apparent position of visual targets during real and simulated saccadic eye movements. *Journal of Neuroscience, 17,* 7941–7953. [PubMed] [Article]

Morrone, M. C., Ross, J., & Burr, D. C. (in press). Keeping vision stable: Rapid updating of spatiotopic receptive fields may cause relativistic-like effects. In R. Nijhawan (Ed.), *Problems of space and time in perception and action*. Cambridge, UK: CUP.

Mullette-Gillman, O. A., Cohen, Y. E., & Groh, J. M. (2005). Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. *Journal of Neurophysiology, 94,* 2331–2352. [PubMed] [Article]

Parise, C. V., & Spence, C. (2009). 'When birds of a feather flock together': Synesthetic correspondences modulate audiovisual integration in non-synesthetes. *PLoS One, 4,* e5664. [PubMed] [Article]

Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgements of spatial direction. *Perceptions & Psychophysics, 6,* 203–205.

Radeau, M. (1994). Ventriloquism against audio-visual speech: Or, where Japanese-speaking barn owls might help. *Current Psychology Cognitive, 13,* 124–140. [PubMed]

Recanzone, G. H. (2003). Auditory influences on visual temporal rate perception. *Journal of Neurophysiology, 89,* 1078–1093. [PubMed] [Article]

Ross, J., Morrone, M. C., & Burr, D. C. (1997). Compression of visual space before saccades. *Nature, 386,* 598–601. [PubMed]

Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends Neuroscience, 24,* 113–121.

Schlack, A., Sterbing-D'Angelo, S. J., Hartung, K., Hoffmann, K. P., & Bremmer, F. (2005). Multisensory space representations in the macaque ventral intraparietal area. *Journal of Neuroscience, 25,* 4616–4625. [PubMed] [Article]

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature, 408,* 788. [PubMed]

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain Research and Cognitive Brain Research, 14,* 147–152. [PubMed]

Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport, 12,* 3849–3852. [PubMed]

Shipley, T. (1964). Auditory flutter-driving of visual flicker. *Science, 145,* 1328–1330. [PubMed]

Stekelenburg, J. J., & Vroomen, J. (2009). Neural correlates of audiovisual motion capture. *Experimental Brain Research, 198,* 383–390. [PubMed] [Article]

Titchener, E. B. (1908). *Lectures on the elementary psychology of feeling and attention*. New York: MacMillan.

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology Human Perception and Performance, 34,* 1053–1065. [PubMed]

Verghese, P., & Stone, L. S. (1995). Combining speed information across space. *Vision Research, 35,* 2811–2823. [PubMed]

Warren, D. H., Welch, R. B., & McCarthy, T. J. (1981). The role of visual-auditory "compellingness" in the ventriloquism effect: Implications for transitivity among the spatial senses. *Perception & Psychophysics, 30,* 557–564. [PubMed]

Watson, A. B. (1979). Probability summation over time. *Vision Research, 19,* 515–522. [PubMed]

Welch, R. B., DuttonHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics, 39,* 294–300. [PubMed]

Werner-Reiss, U., Kelly, K. A., Trause, A. S., Underhill, A. M., & Groh, J. M. (2003). Eye position affects activity in primary auditory cortex of primates. *Current Biology, 13,* 554–562. [PubMed]