

Fusion of Visual and Auditory Stimuli during Saccades: A Bayesian Explanation for Perisaccadic Distortions

Paola Binda,^{1,2} Aurelio Bruno,^{1,3} David C. Burr,^{4,5} and Maria C. Morrone^{1,6}

¹Department of Psychology, Università Vita-Salute San Raffaele, 20132 Milan, Italy, ²Istituto Italiano di Tecnologia, 16163 Genova, Italy, ³Department of Psychology, University College London, London WC1E, United Kingdom, ⁴Department of Psychology, Università Degli Studi di Firenze, 89 Florence, Italy, ⁵Department of Psychology, University of Western Australia, Nedlands Perth, Western Australia, Australia, and ⁶Istituto Nazionale di Neuroscienze, 20132 Milan, Italy

Brief stimuli presented near the onset of saccades are grossly mislocalized in space. In this study, we investigated whether the Bayesian hypothesis of optimal sensory fusion could account for the mislocalization. We required subjects to localize visual, auditory, and audiovisual stimuli at the time of saccades (compared with an earlier presented target). During fixation, vision dominates and spatially “captures” the auditory stimulus (the ventriloquist effect). But for perisaccadic presentations, auditory localization becomes more important, so the mislocalized visual stimulus is seen closer to its veridical position. The precision of the bimodal localization (as measured by localization thresholds or just-noticeable difference) was better than either the visual or acoustic stimulus presented in isolation. Both the perceived position of the bimodal stimuli and the improved precision were well predicted by assuming statistically optimal Bayesian-like combination of visual and auditory signals. Furthermore, the time course of localization was well predicted by the Bayesian approach. We present a detailed model that simulates the time-course data, assuming that perceived position is given by the sum of retinal position and a sluggish noisy eye-position signal, obtained by integrating optimally the output of two populations of neural activity: one centered at the current point of gaze, the other centered at the future point of gaze.

Key words: vision; saccades; cross-modal fusion; Bayesian analysis; audition; psychophysics

Introduction

When asked to report the position of a visual stimulus flashed just before a saccade, subjects mislocated it, primarily in the direction of the saccade (Matin and Pearce, 1965; Mateeff, 1978; Honda, 1989). Typically, localization errors start up to 100 ms before saccadic onset and continue well after the saccade finishes, being maximal at saccadic onset. The dynamics of the effect led Matin (1972) to suggest that perceived position is given by the sum of retinal position, together with a sluggish extraretinal position signal.

In recent times, Bayes' (1764) theorem has been successfully applied to perceptual research, accounting for how many separate sensory cues can be integrated in a statistically optimal manner to yield the most probable form of the perception (Clark and Yuille, 1990), even if this leads to illusions (Weiss et al., 2002). This approach has been applied successfully to many fields of perception, particularly multimodal perception (Ernst and Banks, 2002; Ernst and Bulthoff, 2004). An interesting example is the so-called “ventriloquist effect,” in which a sound appears to emanate from a puppet rather than the more distant ventriloquist; in other circumstances, however, the sound can “capture”

the visual stimulus and cause it to be seen in an erroneous position (Alais and Burr, 2004). The Bayesian framework is well suited to explain these cross-modal effects, in which the perceived position of an event is given by combining auditory and visual information, after weighting each source according to its reliability (intertrial consistency). Normally, vision is more reliable than audition, and dominates; but if the visual stimulus is blurred, audition can become more reliable and predominant. Importantly, when both forms of information are present, precision for localizing the stimuli improves, although they are seen in erroneous positions (Alais and Burr, 2004).

The purpose of this study was to test whether the transient spatial distortions at the time of saccades may be consistent with statistically optimal “Bayesian” integration [as has been applied to other aspects of saccadic misperceptions (Niemeier et al., 2003)]. We take advantage of the fact that saccades have little effect on auditory space perception (Harris and Lieberman, 1996; Klingenhoefer and Bremmer, 2004) to investigate audiovisual localization in normal and saccadic conditions. We find that the auditory signal becomes more important for localization at the time of saccades, suggesting that the visual signal has become transiently noisy, and therefore receives less weight. Importantly, localization precision is better for the multimodal presentation than for the unimodal presentation, a strong prediction of the Bayesian theory of information fusion. We also show that, by assuming the optimal integration of sensorimotor information, the illusory mislocalization of visual stimuli during saccades can be well modeled.

Received Feb. 17, 2007; revised June 22, 2007; accepted June 22, 2007.

This work was supported by the Italian Ministry of University and Research, the Australian National Health and Medical Research Council, and the European Commission Sixth Framework Program (NEST, MEMORY). We are particularly grateful to Dr. Marco Cicchini for advice and help with the modeling.

Correspondence should be addressed to Dr. David Burr, Department of Psychology, Università Degli Studi di Firenze, 89 Florence, Italy. E-mail: dave@in.cnr.it.

DOI:10.1523/JNEUROSCI.0737-07.2007

Copyright © 2007 Society for Neuroscience 0270-6474/07/278525-08\$15.00/0

Materials and Methods

Subjects and apparatus. Four subjects (two authors and two observers naive to the goals of the experiment) participated in the entire experiment; in one of the experimental conditions (main experiment, visual stimulation), three additional naive subjects were tested. All participants had normal or corrected-to-normal vision and normal hearing. They were seated facing a vast quasi-hemispheric screen (Vision Station VS102; Elumens, Cary, NC) subtending $\sim 140 \times 140^\circ$ at the viewing distance of 60 cm, illuminated by a liquid crystal display projector (Epson/Elumens SPIClops API; refresh rate, 67 Hz), with head movement minimized by a chin rest. Ten small speakers (Philips MMS 221/00) were mounted behind the screen, at 8° intervals (the minimum interval allowed by their overall size).

Stimuli. Visual stimuli were briefly displayed (15 ms) as 2° blue blobs [Commission Internationale de l'Éclairage (CIE) coordinates: $x = 0.142$, $x = 0.065$; luminance, 1.21 cd/m^2] presented against a red background (CIE coordinates: $x = 0.59$, $y = 0.35$; mean luminance, 2.6 cd/m^2). Auditory stimuli were short (15 ms) bursts of white noise of 40 dB intensity, played through one of the 10 speakers (selected by computer-controlled digital switch). Both visual and auditory stimuli were presented 8° above gaze level, at various eccentricities. For bimodal presentations, the stimuli were presented together, within a tolerance of 7.5 ms (verified by direct measurement). Visual stimuli were generated by a Framstore (VSG 2/4F; Cambridge Research System, Cambridge, UK) housed in a personal computer (PC) controlled by Matlab programs. Auditory stimuli were generated by the computer soundboard and gated to the appropriate speaker via a switch controlled by digital output from the VSG board.

Task and procedure. As illustrated in Figure 1, trials began with subjects fixating a 1° dark fixation spot (F_0), presented at $x = -16^\circ$, $y = 0^\circ$ [where (0, 0) refers to straight ahead, with positive implying rightward or upward]. After 400 ms, the fixation spot was extinguished, and a saccadic target (F_1 , also 1° dark spot) was displayed at $(+16^\circ, 0^\circ)$, to which subjects saccaded immediately. Besides the fixation spot and the saccadic target, two additional stimuli were presented sequentially: a probe stimulus, displayed above F_0 ($-16^\circ, 8^\circ$) well before the eye movement, and a test stimulus, presented at variable eccentricities, some time after the appearance of the saccadic target (arriving before, during, or after the eye movement). Subjects reported the perceived location of the test stimulus relative to the probe (two-alternative forced choice left/right judgments; responses were voiced and recorded by the experimenter). The eccentricity of the test was varied by steps of 8° (the minimum distance between adjacent positions of auditory stimuli), using the adaptive QUEST method [which adjusted the eccentricity for each trial on the basis of the subject's previous responses; for more details, see Watson and Pelli (1983)]. The probe and the test stimuli could be visual, acoustic, or bimodal congruent audiovisual. In the latter case, each was composed of two unimodal presentations aligned in space: the probe stimulus comprised a blob and a sound, both presented above F_0 ($-16^\circ, 8^\circ$); the test stimulus comprised a blob and a sound, both presented in 1 of 10 locations of the speakers. For each subject and stimulus type, a minimum of two sessions of 100 trials each were run. Visual, auditory, and bimodal stimuli were tested in separate sessions, interleaved in a pseudo-randomized order; visual localization was also tested in the three additional subjects. Data were fitted with cumulative Gaussian functions (such as those in Fig. 2). The median gives estimates of the point of subjective equality (PSE; the position where the test was perceived aligned with the probe), and the difference between the PSE and probe

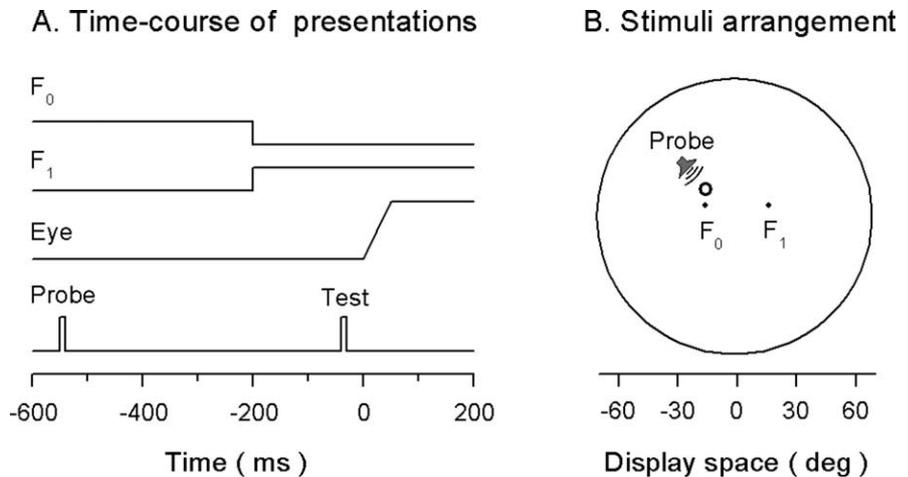


Figure 1. *A, B*, Time course (*A*) and spatial arrangement (*B*) of presentations. At the beginning of each trial, the fixation spot (F_0) appeared 16° to the left of the center of the screen. After 400 ms, it was replaced by the saccadic target (F_1 , presented 16° to the right of the center of the screen), to which subjects saccaded as quickly as possible. Two brief (15 ms) stimuli were presented sequentially, one 600 ms before the other: the first was the probe, displayed at a fixed position well before the saccadic onset; the second was the test, presented at a variable position near saccadic onset. Subjects reported whether the test seemed to be to the left or right of the probe. The two stimuli were both either unimodal (visual or auditory) or bimodal (visual and auditory simultaneously). The bimodal probe was always congruent (visual and auditory components presented at the same location, just above the fixation spot, as shown in *B*); the bimodal test (data not shown) could be congruent or conflicting (with the two components at different eccentricities).

position defines the localization bias; the SD gives the localization threshold, or just-noticeable difference.

We also tested a control condition, aiming to study unimodal and bimodal localization in a situation that mimicked that observed in the main experiment, but in the absence of eye movements. Procedure, task, data collection/analysis, and participants were all the same as in the main experiment, except that subjects maintained fixation at F_0 ($-16^\circ, 0^\circ$) and that the eccentricity of visual test stimuli was varied by steps of 1° (to obtain precise measurements of visual localization thresholds; hardware constraints prevented us from reducing the step size also for auditory stimuli). The stimulus arrangement was chosen on the basis of our expectations on the effects of eye movements on unimodal localization. We expected saccades to introduce systematic errors in visual but not auditory localization, having two consequences: (1) that the visual PSE would be systematically different from the auditory PSE and, more importantly, (2) that visual and auditory cues to the location of perisaccadic bimodal test stimuli would be systematically conflicting (although the stimuli were physically congruent). To mimic the difference in PSE values, we presented the visual probe stimulus in a different location than the auditory probe: the visual probe was moved to $(-24^\circ, 8^\circ)$, whereas the auditory probe remained at $(-16^\circ, 8^\circ)$. To mimic the conflict between unimodal cues to the location of the bimodal test stimulus, we introduced a physical displacement between its visual and auditory components (the average displacement was -8° , meaning that the auditory stimulus was, on average, 8° to the left of the visual stimulus), while maintaining the bimodal probe a physically congruent stimulus. This was presented at the same location of the visual probe stimulus ($-24^\circ, 8^\circ$), so that the bimodal PSE would be close to the visual PSE in case of visual dominance and close to the auditory PSE in case of auditory dominance.

Eye-movement measurement. Eye movements were recorded by means of an infrared limbus eye tracker (HVS SP150; horizontal resolution, 0.01° ; precision, 0.1°): the infrared sensor was mounted below the left eye on transparent wraparound plastic goggles, through which subjects viewed the display binocularly. There were practically no visual references because the display was very large (borders $\sim 70^\circ$ eccentricity, very hard to see), and the sensor was mounted below the eye without obscuring the view. The PC sampled eye position at 1000 Hz and stored the trace in digital form. The program identified the onset of the eye movement by convolving the trace with a difference-of-Gaussian function and searching for maximum (this method marks the cusp of the eye trace, so that its

estimate of the saccadic onset is not systematically delayed, as are estimates based on velocity derivation; in contrast, it requires the eye trace to be stored before being analyzed, so estimates of saccadic onset can only be provided at the end of the trial). In a later analysis, the experimenter checked the quality of the saccades and, when necessary (in ~2% of trials, in which the saccade was preceded by a blink, or fixation was noisy, or in the presence of 50 Hz noise), adjusted the estimate of the saccadic onset by visually estimating the point of interception between presaccade and during-saccade segments of the eye trace or eliminated the trial. Very few saccades were, furthermore, excluded from the analysis, for the presence of a corrective saccade.

Modeling audiovisual fusion. We assume that auditory and visual spatial cues are conditionally independent and are combined in a statistically optimal manner (Ernst and Banks, 2002; Alais and Burr, 2004). The combined audiovisual estimate of spatial position \hat{S}_{AV} is given by the following:

$$\hat{S}_{AV} = w_A \hat{S}_A + w_V \hat{S}_V, \quad (1)$$

where \hat{S}_A and \hat{S}_V are auditory and visual estimates of spatial location. w_A and w_V are the weights by which the unimodal estimates are scaled, inversely related to the variances (squared thresholds) σ_A^2 and σ_V^2 for audition and vision:

$$w_A = k/\sigma_A^2, w_V = k/\sigma_V^2, \text{ where } k = 1/(1/\sigma_A^2 + 1/\sigma_V^2), \quad (2)$$

ensuring that the weights sum to unity.

The predicted bimodal thresholds $\hat{\sigma}_{AV}^2$ are given by the following:

$$\hat{\sigma}_{AV} = \sqrt{\sigma_V^2 \sigma_A^2 / (\sigma_V^2 + \sigma_A^2)}. \quad (3)$$

$\hat{\sigma}_{AV}$ will always be less than $\min(\sigma_V, \sigma_A)$, and the reduction in threshold will be greatest ($\sqrt{2}$) when $\sigma_V \cong \sigma_A$.

Results

Spatial localization

Subjects were asked to report whether a test stimulus appeared to the left or the right of a previously presented probe. In different sessions, the two stimuli were presented either unimodally (visually or acoustically) or bimodally (visually and acoustically), either during fixation or at the time of a saccade. In the latter case, the probe stimulus was always presented well before (at least 500 ms) saccadic onset, whereas the test could fall before, during, or after the eye movement. In all conditions, subjects reported perceiving bimodal stimuli as single unified events, both when they were physically congruent (a blob and a sound presented at the same location) and when they were conflicting (locations of blob and sound differed by 8°, on average).

Example results from subject P.B. together with the best-fitting psychometric functions are shown in Figure 2. During fixation (Fig. 2A), the localization of visual and auditory stimuli was unbiased, meaning that the PSE corresponded to the physical position of the probe stimuli. The width (SD) of the curves gives an estimate of threshold, ~1° for visual stimuli and between 5° and 11° for auditory stimuli. Bimodal localization was studied by comparing the apparent position of a conflicting test stimulus with a bimodal congruent probe. The precision of the localization for bimodal stimuli was always similar to that for visual stimuli (as predicted by Eq. 3; see Materials and Methods). As predicted by Equation 1, conflicting stimuli were always localized toward the location of the visual component, regardless of the position of the auditory component, as Alais and Burr (2004) showed for small, unblurred stimuli.

Figure 2B shows example curves for stimuli presented within the interval of -25 to 0 ms (where 0 refers to saccadic onset). Perisaccadically, visual localization was grossly biased in the direction of the saccade (in subject P.B., the PSE was displaced

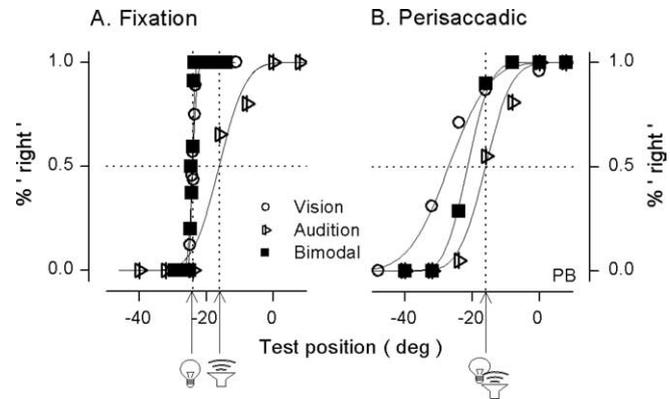


Figure 2. Localization of visual, auditory, and bimodal stimuli under various conditions for subject P.B. Data for the three types of stimuli are reported as psychometric curves (see legend), fitted with cumulative Gaussian functions (solid lines), the median and SD of which estimate, respectively, the PSE and the just-noticeable difference (or threshold). Below the abscissae, the lamp and the speaker indicate the positions of the visual and the auditory probe stimuli: the difference between the PSE and the position of the probe defines the localization bias. **A**, During fixation, unimodal localization was veridical (the visual and the auditory PSE coincide with the positions of the visual and the auditory probes), visual localization was more precise than auditory localization, and bimodal conflicting stimuli were localized where their visual component was actually displayed. **B**, For stimuli presented perisaccadically (in the interval of 25–0 ms before saccadic onset), the auditory localization performance remained unchanged, whereas visual localization became less precise and grossly biased (the test was mislocalized in the direction of the eye movement). Thus, vision and audition provided conflicting cues as to the location of bimodal perisaccadic stimuli (although physically congruent), like the bimodal conflicting stimuli during fixation. However, under these conditions, the perceived position of bimodal stimuli was intermediate between the perceived positions of their visual and auditory components. In this subject, the precision of visual localization during saccades got similar to that of auditory localization, and bimodal perisaccadic stimuli tended to be localized more precisely than either their unimodal component presented alone (the bimodal psychometric curve was steeper than the visual and the auditory curves).

~15° from probe position), and the localization was less precise (the curve is far broader) than during fixation. However, auditory localization remained approximately veridical (PSE near the physical location of the probe) and as precise as during fixation. Bimodal stimuli, although physically congruent, were mislocalized in the direction of the eye movement, like visual stimuli. However, the mislocalization was less than for unimodal visual stimuli. Importantly, the precision of the localization (given by the slope of the curves) was better than either the visual or the auditory unimodal case, as predicted by Equation 3 in Materials and Methods.

Figure 3 shows all individual results for perisaccadic localization, plotting localization threshold against the corresponding values of bias. Figure 3, *A* and *B*, reports data from perisaccadic visual and auditory localization, respectively (dashed lines indicate average values observed during fixation). In all subjects, auditory localization was nearly unbiased, with precision similar to that during fixation. In three of the four subjects who completed all experimental conditions, as well as in three naive subjects who were additionally tested in this condition (see figure legend), visual localization was grossly biased and ~10 times less precise than during fixation. In one subject (G.A.), precision was even worse, but his bias (in the opposite direction) was not significant.

Figure 3C reports the perisaccadic bimodal localization performance (filled symbols), with data from each subject in a distinct panel (see text insets); unimodal localization is also represented for comparison (open symbols). As predicted by Equation 3 in Materials and Methods, bimodal thresholds are either lower or similar to unimodal thresholds. The relative distance between

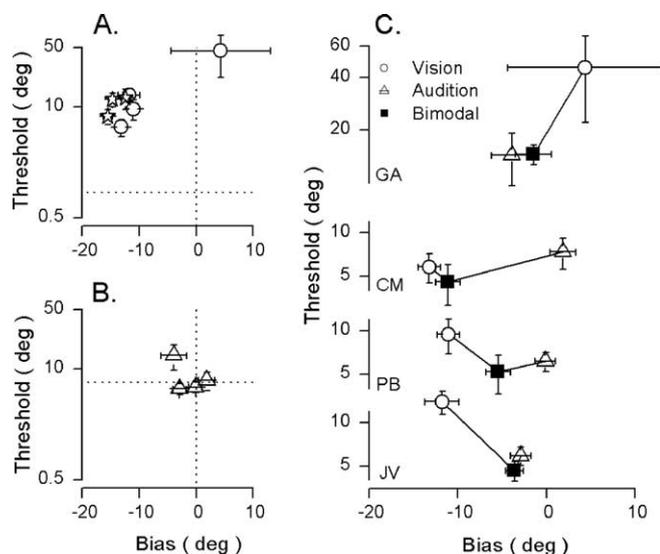


Figure 3. Localization of perisaccadic stimuli in all subjects: individual values of threshold are plotted against the corresponding values of bias, with bars representing SEM estimated with the bootstrap procedure (500 repetitions). **A**, Visual localization. In the four subjects participating in the entire experiment (circles), as well as in three additionally tested subjects (stars), the localization of perisaccadic visual stimuli was far less precise than during fixation (average thresholds and biases observed during fixation are reported by the dashed lines). Moreover, in all but one subject (G.A.), localization was grossly biased in the direction of the saccade (in G.A. the bias was small and in the opposite direction). **B**, Auditory localization. In all subjects, the localization of perisaccadic auditory stimuli was similar to that during fixation (as reported by the dashed lines). The average value of perisaccadic auditory bias is not significantly different from the average value observed during fixation ($p = 0.35$), and the average perisaccadic threshold does not differ significantly from the average threshold observed in fixation ($p = 0.44$) (probabilities were estimated by a bootstrap t test, 5000 repetitions). **C**, Bimodal localization (filled symbols); visual and auditory localization performances are also reported for comparison (hollow symbols). Data from the four subjects are reported in separate panels (see initials); notice that the scales for the ordinates are different than in **A** and **B**. They also differ across subjects, but proportions are preserved. As predicted by Equations 1–3 in Materials and Methods, bimodal stimuli were always localized with similar or higher precision than visual or auditory stimuli presented alone, and the value of bimodal localization bias is closer to that of visual or auditory bias, depending on the relative precision of the two modalities.

bimodal and visual (or auditory) values of bias (i.e., the distance on the abscissa between the full symbol and the hollow ones) gives a direct estimate of the relative dominance of visual (or auditory) cues in determining each subject's bimodal localization performance. For example, in subject J.V., auditory localization is far more precise than visual localization: as predicted by Equations 1 and 2 in Materials and Methods, bimodal localization is dominated by auditory cues.

Figure 4 plots the observed bimodal biases (**A**) and thresholds (**B**) against the corresponding predicted values. In all subjects, including the atypical subject G.A., the perceived position of bimodal stimuli and the precision of bimodal localization are well predicted by assuming Bayesian integration of visual and auditory spatial cues. The average biases and thresholds for unimodal (visual and auditory) and bimodal (observed and predicted) localization are shown in Figure 5: to minimize intersubject variability, thresholds were all normalized to the observed bimodal value. For both bias and thresholds, measured bimodal values are very close to the predicted values and significantly different from the unimodal values; as predicted by Equation 3 in Materials and Methods, the average bimodal localization threshold is significantly lower than either visual or auditory thresholds.

All results shown so far were for perisaccadic stimulus presentation, -25 to 0 ms before saccadic onset. But a stronger test

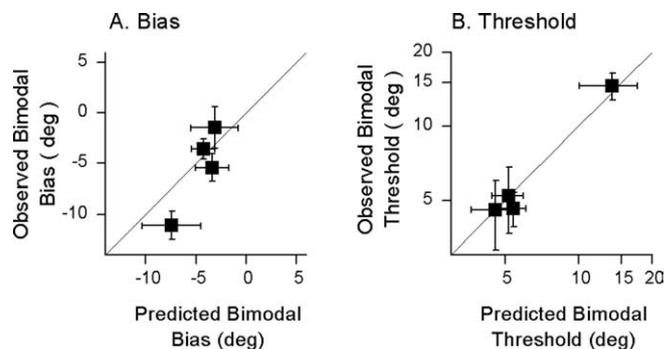


Figure 4. **A**, **B**, Data from all subjects in the bimodal condition: observed values of bias (**A**) and threshold (**B**) are plotted against the corresponding predicted values. All the points lay around the bisection of the axes, implying that, for all subjects, the observed bimodal localization performance was very close to that predicted by assuming a Bayesian integration of visual and auditory spatial cues (see Eqs. 1–3 in Materials and Methods).

would be for the Bayesian approach to predict the time course of the mislocalization effect. Figure 6 plots the time course of the average bias and threshold for three subjects (excluding G.A., for whom there was very little data across the time course). Visual mislocalization followed the characteristic dynamics (Morrone et al., 1997; Ross et al., 1997), rising to a maximum near the time of saccadic onset. Discrimination thresholds follow a similar time course, reaching a maximum just before saccadic onset. Stimuli presented early or late had localization biases and thresholds similar to fixation (Fig. 6, dashed lines). Auditory localization threshold and bias were virtually unaffected by the saccades, showing no systematic dependency on time relative to saccadic onset. Bimodal localization (Fig. 6, middle panels) followed a similar time course to the vision results, with less variation in bias and thresholds. Importantly, the variation over time followed closely the Bayesian predictions from the unimodal visual and auditory curves (Fig. 6, middle panels, thick gray curves) (see Eqs. 1–3 in Materials and Methods).

Modeling

The previous section shows that the time courses of both bias and precision of bimodal localization are well predicted by a simple Bayesian model of optimally weighted combination of the unimodal signals. Because auditory PSEs and thresholds were virtually unaffected by saccades, the variation in the bimodal performance was basically caused by the perisaccadic mislocalization and decrease in precision of vision. Here, we model the time course of visual localization, showing how the action of simple physiologically plausible mechanisms can account for the data.

The experimental task required subjects to localize stimuli relative to an external reference (the probe) while making horizontal saccades and therefore to estimate stimuli positions in head-centered coordinates (a coordinate system insensitive to gaze shifts). In these conditions, the head-centered location of a stimulus over time $[X(t)]$ can be derived from its retinotopic location estimate $[x(t)]$, through a simple translation of the horizontal dimension of the coordinates by a function $E(t)$, representing the change of eye position in time (Eq. 4).

$$X(t) = x(t) + E(t). \quad (4)$$

Notice that correct (i.e., unbiased) head-centered location estimates would require a correct estimate of the function $E(t)$, with the latter presumably provided by an extraretinal eye-position signal (Robinson, 1975; Dassonville et al., 1992; Morrone et al.,

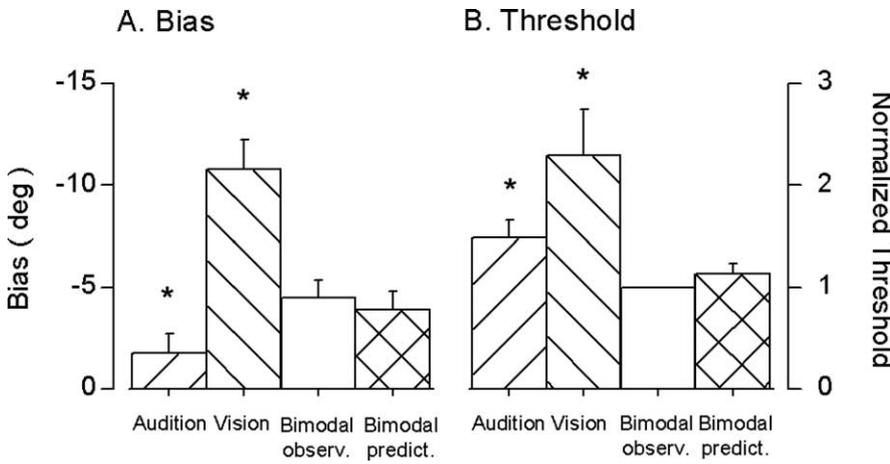


Figure 5. Average localization performance observed (observ.) for perisaccadic visual, auditory, and bimodal stimuli and predicted (predict.) for bimodal stimuli (Eqs. 1–3 in Materials and Methods). **A**, The average values of bias. **B**, The average thresholds (before being averaged, thresholds from each subject were normalized to the bimodal observed threshold). Error bars indicate SEM. Observed values for the bimodal condition are very similar to the predicted ones and significantly different from those observed in the unimodal conditions; as predicted by Equation 3 in Materials and Methods, the bimodal localization threshold is significantly lower than both visual and auditory thresholds (* $p < 0.05$, bootstrap t test with 5000 repetitions).

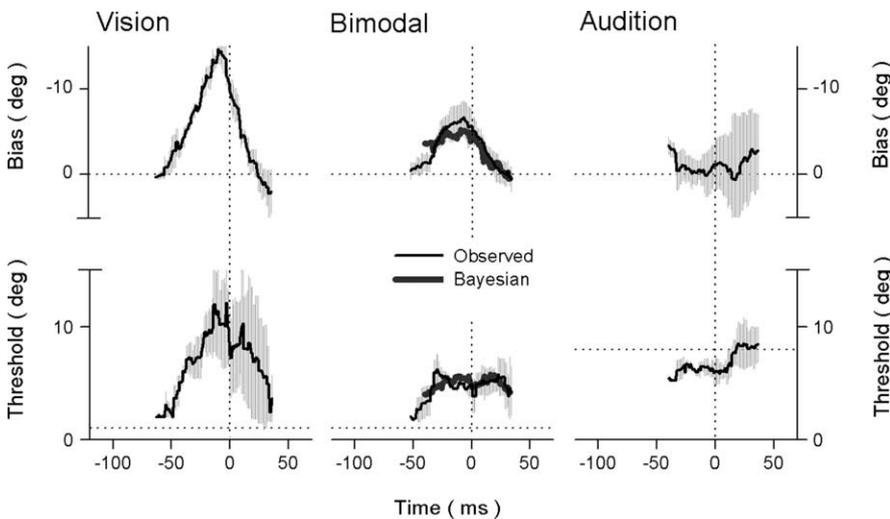


Figure 6. Time course of localization biases and thresholds. For each subject and stimulus type, trials were binned by the delay between test stimulus presentation and saccadic onset; bias and threshold values were estimated for consecutive subsets of trials, selected by a running window of 100-trial width. Average data from three subjects (G.A. excluded) are reported by the thin black lines, with gray shadings indicating SEM. Visual localization (left) became progressively more biased and less precise as the test presentation approached saccadic onset; the bimodal localization performance (middle) had a qualitatively similar time course, but the variations of bias and threshold (well predicted by assuming a Bayesian integration of visual and auditory cues; thick gray lines) were constrained to a smaller range. Auditory localization remained fairly constant and similar as during fixation (right).

1997; Deneve et al., 2001; Niemeier et al., 2003; Pola, 2004). On the contrary, the dynamics of the perisaccadic mislocalization effect (Fig. 6) (Matin and Pearce, 1965; Honda, 1989; Morrone et al., 1997) suggest that the $E(t)$ signal does not simply follow the actual eye trace but anticipates the saccade by at least 50 ms. It has been suggested that such anticipation may reflect the presence, before the actual saccadic onset, of a signal representing gaze at the intended postsaccadic location (an intention-to-move signal) (Dassonville et al., 1992; Bahcall and Kowler, 1999). We propose that an anticipatory signal $E(t)$ may result from the combination of two competing gaze-related signals, one indicating gaze in the presaccadic direction and the other indicating gaze at the future or intended postsaccadic position.

More specifically, we assume $E(t)$ to be the maximum likeli-

hood estimate of two gaze-related activities, each arising from a population of neurons signaling gaze at a particular direction. During normal fixation, neurons signaling gaze in the current direction (the presaccadic location, indicated as F_0) will be highly active, and all others are practically dormant. Around the time of saccades, the activation of neurons at the current fixation decreases, and those representing the new postsaccadic direction (the saccadic target F_1) begin to respond.

We model gaze-related population activities as Gaussian distributions centered at F_0 and F_1 , the SDs σ_0 and σ_1 of which vary over time (Eqs. 5a, 5b):

$$f_0(x,t) = \frac{1}{\sqrt{2\pi\sigma_0(t)}} e^{-\frac{(x-F_0)^2}{2\sigma_0^2(t)}} \quad (5a)$$

$$f_1(x,t) = \frac{1}{\sqrt{2\pi\sigma_1(t)}} e^{-\frac{(x-F_1)^2}{2\sigma_1^2(t)}} \quad (5b)$$

Notice that the maximum height of the distributions (i.e., maximum population activity) is inversely proportional to their SDs (Eqs. 6a, 6b) and therefore varies over time too:

$$\max(f_0(x,t)) \propto 1/\sigma_0(t) \quad (6a)$$

$$\max(f_1(x,t)) \propto 1/\sigma_1(t) \quad (6b)$$

Because eye position is assumed to be computed from the maximum likelihood estimate of $f_0(x,t)$ and $f_1(x,t)$, at each time the estimated position of the eyes [$E(t)$] is given by the weighted sum of the positions $x = F_0$ and $x = F_1$, with the weights inversely related to the variances σ_0^2 and σ_1^2 of the two distributions (Eqs. 7, 8), and the SD of the $E(t)$ signal [$\sigma(E(t))$] will be proportional to the product of σ_0 and σ_1 (Eq. 9):

$$E(t) = F_0 w_0(t) + F_1 w_1(t) \quad (7)$$

$$w_0(t) = k/(\sigma_0^2(t)), \quad w_1(t) = k/(\sigma_1^2(t))$$

$$\text{where } k = 1/(1/\sigma_0^2(t) + 1/\sigma_1^2(t)) \quad (8)$$

$$\sigma(E(t)) = \sigma_0(t)\sigma_1(t)/\sqrt{\sigma_0^2(t) + \sigma_1^2(t)}. \quad (9)$$

Thus, $E(t)$ and $\sigma(E(t))$ are completely determined once the maxima of the two population activities (or the variances σ_0^2 and σ_1^2) are known; notice, however, that $E(t)$ depends only on the relative strength of the two activities, whereas $\sigma(E(t))$ depends also on their absolute strengths; for example, if both populations are strongly activated, $\sigma(E(t))$ will be low, if both are weakly activated, $\sigma(E(t))$ will be high, but in any case, $E(t)$ will signal midway between F_0 and F_1 .

Because we have no physiological data on how the strengths of

these putative gaze-related activities may vary over time, we adopt another approach, working backward from our data to predict the gaze functions. As stated above, we assume that the eye-position signal is computed from the maximum likelihood estimate of gaze-related activities and calculate how these activities should vary over time to produce our empirical data.

The results of this simulation are illustrated in Figure 7. Figure 7A–D shows the transition of the eye-position signal from initial fixation (F_0) to the saccadic target (F_1): A shows the time courses of the maximum activities at F_0 (light gray line) and at F_1 (dark gray line), and B–D show the spatial distributions of the responses at three times. Well before the saccade (Fig. 7B), there is a clear, sharply tuned signal at F_0 (light gray curve), causing the maximum likelihood estimate of gaze position (thin black curve) to be centered there. This signal gradually decreases while the signal at F_1 begins to build up. Well after saccadic offset (Fig. 7D), the strong signal for eye position at F_1 (dark gray curve) dominates the maximum likelihood estimate, but between the two extremes (Fig. 7C), both signals contribute to the maximum likelihood estimate, causing it to be between the two fixation points. Importantly, during this intermediate period when both responses are low, the signal has high variance and hence low reliability.

Figure 7E shows the time course of the eye-position signal $E(t)$ and its associated SD $[\sigma(E(t))]$. $E(t)$ moves progressively from F_0 to F_1 , signaling midway for an extended period. During this period, the SD of the signal $\sigma(E(t))$ is particularly high. Figure 7F shows the effects of the saccade on retinotopic position estimates $x(t)$: the saccade displaces the retinal image, and hence the position signal arising from it, but we assume that it does not affect the reliability of estimated retinal positions of the stimuli. Figure 7G shows the sum of the two signals. As can be seen, it closely follows the data, both those for the changes in localization bias and for the changes in precision.

On one hand, the fact that the model fits the data well is unsurprising, given that the parameters of the model were estimated from psychophysical results. On the other hand, however, it shows that a system of this sort can predict these results. With two basic assumptions, that perceived position is given by the sum of an eye-position signal and retinal information and that the eye-position signal is computed as the maximum likelihood estimate of two Gaussian-shaped population activities, centered at F_0 and F_1 , the time course of activity at F_0 and F_1 are completely constrained by the measured values of localization bias and threshold, with no other

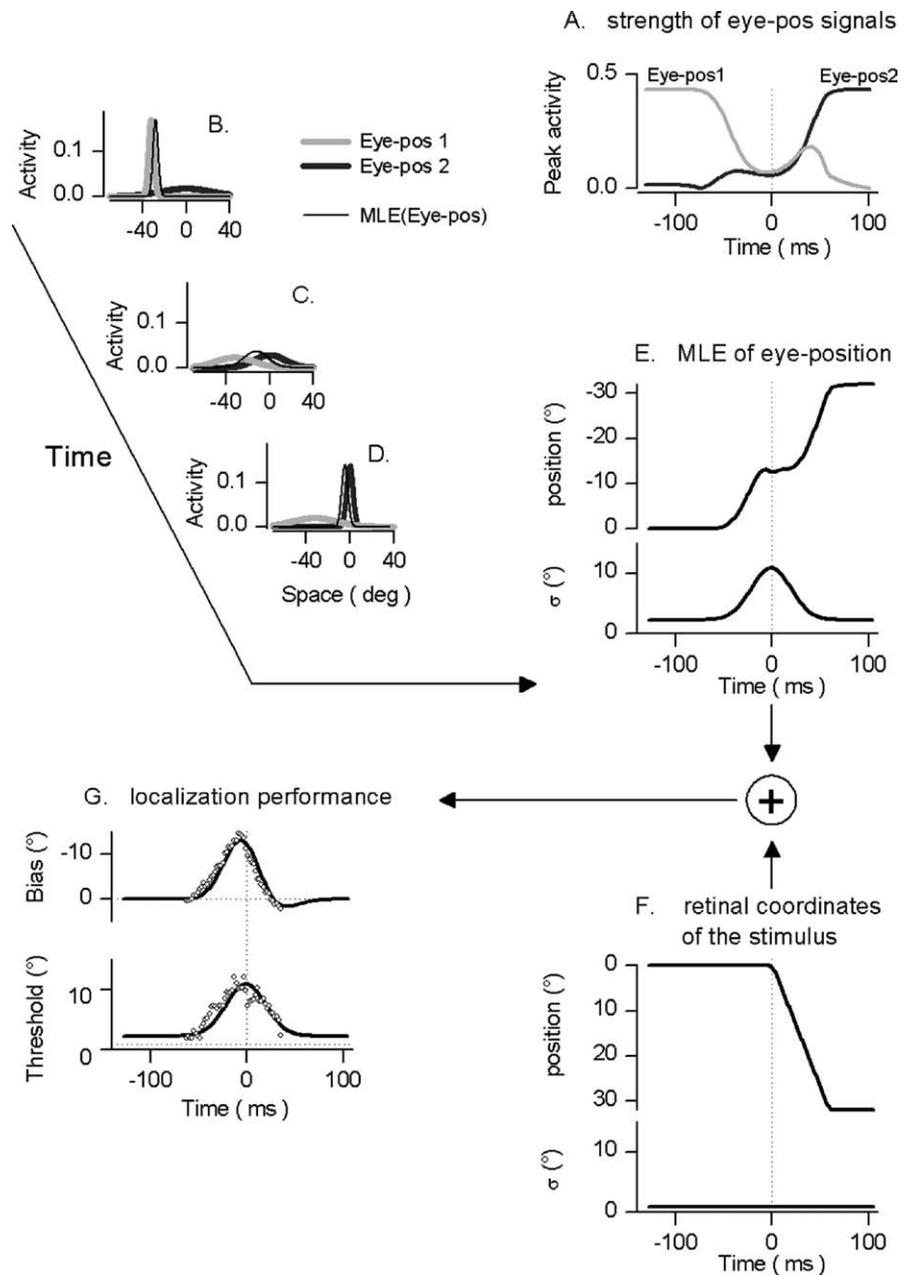


Figure 7. Model of perisaccadic visual localization. **A**, Time course of the maximum population activities at F_0 (light gray line) and F_1 (dark gray line). **B–D**, Spatial distribution of the two population activities $f_0(x,t)$ and $f_1(x,t)$ at three times (respectively, light gray and dark gray curves), together with the spatial distribution of the eye-position (eye-pos) signal $E(t)$ arising from the optimal integration of $f_0(x,t)$ and $f_1(x,t)$ (thin black curve). **E**, Time course of the maximum likelihood estimate of eye position $E(t)$ and its SD $\sigma(E(t))$. **F**, Retinal position of the stimulus $x(t)$ and its SD, that we assume to be uninfluenced by the eye movement. **G**, Observed time courses of bias and thresholds for visual localization (small circles), together with the fit from which the parameters of the model were derived (continuous lines). The abscissae in **A** and **E–G** indicate time relative to the actual saccadic onset, the latter being represented by the dashed vertical lines. Notice that ordinates in **E** and **F** (top plots, reporting estimated eye-position and stimuli retinal locations) are scaled differently than in **G** (top plot reporting bias values), but proportions are preserved.

free parameters. Consider, for example, the time course of activity at F_0 (Fig. 7A, light gray line): activity initially decreases but shows a “rebound” just after the saccadic onset. The observed time courses of bias and threshold jointly determine this variation: around the saccadic onset, the threshold [and therefore $\sigma(E(t))$] is maximum, forcing both gaze-related activities to be weak (Eq. 9). After some 25 ms from the saccadic onset, the threshold begins to recover to its usual value, forcing either the activity at F_0 or that at F_1 to increase (Eq. 9); meanwhile, the

localization bias is changing at a similar rate as does the retinal position of the stimuli: this makes $E(t)$ remain approximately constant (Eq. 4) and forces the activities at F_0 and at F_1 to have similar strengths (Eqs. 7, 8). The activity at F_0 is eventually allowed to decrease near the end of the saccade, when $\sigma(E(t))$ decreases and $E(t)$ progressively grows toward F_1 . Thus, whether the rebound has any physiological significance is yet to be determined, but it is a clear prediction of the present model.

Discussion

The general aim of this study was to investigate whether spatial mislocalization at the time of saccades may be consistent with statistically optimal Bayesian integration, by measuring spatial localization of perisaccadic visual, auditory, and bimodal audio-visual stimuli. We found that visual localization dramatically changes in the perisaccadic interval: not only does it become biased toward the saccadic target, but also far less precise. Auditory localization, however, was virtually unaffected by eye movements. Bimodal localization was affected in a way that was well predicted by optimal Bayesian integration of visual and auditory spatial cues, where localization depends less on visual signals at the time of saccades than during fixation.

For one of the four subjects tested extensively (G.A.), the pattern of perisaccadic visual mislocalization was atypical: unlike other subjects, for him perisaccadic visual stimuli were only slightly mislocalized, and the direction of the bias was opposite to that of the saccade. Although atypical, localization errors against the direction of the saccade have been observed previously for stimuli presented beyond the saccadic target (Morrone et al., 1997), leading to an apparent compression of the visual space. It may be that, in this atypical subject, compression of visual space was stronger than usual, affecting also stimuli presented close to the fixation spot. However, despite this idiosyncrasy, his results relative to the main findings of this study were in line with those of the other subjects: the precision of visual localization decreased in the perisaccadic interval relative to normal fixation, and the perisaccadic bimodal localization performance was well predicted by assuming optimal integration of visual and auditory cues.

Saccades have little influence on auditory localization

Auditory localization was almost unaffected by saccades. Perisaccadic auditory stimuli were localized with the same precision as during fixation, as Harris and Lieberman (1996) have previously shown for detection of auditory stimuli. Similarly, there was no statistically significant bias in location, and the direction of the bias varied across subjects. These results fairly agree with previous ones, showing that auditory localization during eye movements is unbiased (Vliegen et al., 2004) or that perisaccadic localization biases are small (Klingenhoefer and Bremmer, 2004).

Saccades influence visual space perception

In agreement with previous reports, in three of our primary subjects and in three additional observers, we found that visual localization was grossly biased around the time of saccades, with stimuli presented near the initial fixation spot shifted in the direction of the saccade; the effect was maximum at saccadic onset but commenced some 50 ms earlier, as observed by several research groups (Matin, 1972; Honda, 1989; Morrone et al., 1997; Ross et al., 1997). These errors have been proposed to result from a mismatch between the actual gaze position and the eye-position signal used by the visual system to recalibrate visual spatial maps across saccades (Pola, 2004), the latter being anticipated relatively

to the actual eye trace. As suggested by studies of saccadic adaptation and double-step saccades, the reason for this anticipation may be the presence of an intention-to-move signal (Dassonville et al., 1992; Bahcall and Kowler, 1999), a signal representing gaze at the intended postsaccadic location before the actual onset of the eye movement.

The current study is the first to specifically measure the precision of visual localization during saccades, using a bias-free technique (by measuring psychometric functions, we could derive independent estimates of random and systematic errors). The results from all tested subjects, showing a substantial decrease in precision, agree qualitatively with previous studies (Ross et al., 1997; Bockisch and Miller, 1999).

One possible cause of the degradation in spatial localization could be the detraction of attentional resources from the localization task at the time of oculomotor planning (Kowler et al., 1995). But it seems unlikely that attention could account for such a large (factor of ten) effect, because the effects of attention seem to be in the order of a factor of two or three, at most (Lee et al., 1999; Morrone et al., 2002).

Poor localization precision is more probably related directly to the systematic mislocalization, given that the two phenomena have similar temporal dynamics. Our model suggests that they have the very same cause: both can be accounted for by assuming that the actual shift of retinal images during saccades is compensated by a predictive, but sluggish and noisy, eye-position signal. Such a signal could arise from the optimal integration (maximum likelihood estimate) of two competing gaze-related signals, one representing gaze in the presaccadic direction and the other representing gaze in the postsaccadic one (possibly, the so-called intention-to-move signal).

Current computational work suggests that maximum likelihood estimation is a plausible rule to encode information at the population level (Jazayeri and Movshon, 2006). Neurophysiological evidence indicates that some information about the amplitude and direction of impending saccades reaches a subpopulation of visually driven neurons before actual saccadic onset and affects their function in the perisaccadic interval (Sommer and Wurtz, 2006). These neurons are said to demonstrate “predictive remapping” and often reside in parietal areas (Duhamel et al., 1992; Walker et al., 1995; Kusunoki et al., 1997; Umeno and Goldberg, 1997; Batista et al., 1999; Nakamura and Colby, 2000). In qualitative agreement with our simulations, the visual receptive fields of these neurons abruptly changes location just before the saccade, “jumping” from the presaccadic to its postsaccadic location, with no response at intermediate locations.

It is possible that the activity of these visually driven neurons encodes implicitly the gaze-related signal used to maintain stable the perceived positions of visual stimuli despite the shift of their retinal images. We suggest that, at the time of saccades, this gaze-related signal does not reproduce faithfully the actual displacement of the eyes [causing perisaccadic mislocalizations, as proposed previously by several authors (Matin, 1972; Dassonville et al., 1992; Miller, 1996; Morrone et al., 1997; Pola, 2004)] and is also far more noisy relative to normal fixation (causing the perisaccadic decrease of localization precision). Our simulations suggest that this may occur because, in this time window, gaze-related activities are weaker than usual. Several reasons could account for the gaze-related activity being weaker. For example, the fact that remapping does not proceed simultaneously in all cells would introduce a blurring of the response (Kusunoki and Goldberg, 2003). Another reason could be saccadic suppression

(Burr et al., 1994), strongest for the magnocellular system that projects predominantly to parietal cortex.

This particular model, the core parameters of which were derived from our psychophysical observations, rather than from independent physiological data, is not essential for the Bayesian approach to be useful, but it is one that seems to fit well with current physiology (Sommer and Wurtz, 2006).

Multimodal integration

Over the past few years, numerous experiments have shown that information from independent sensory channels is integrated with a near-optimal or Bayesian strategy, so the perceived size or position results from the weighted sum of all of the available sensory cues, with weights proportional to the reliability of each sense (Knill and Kersten, 1991; Ernst and Banks, 2002; Alais and Burr, 2004; Ernst and Bulthoff, 2004). Here, we test these predictions in an atypical situation: the conflict between multimodal cues and the degradation of sensory information were not produced by manipulating the stimuli but by taking advantage of the distortions of visual space that occur at the time of saccades. That the Bayesian framework works in these “natural” conditions provides very strong support for it.

Within this framework, our data provide firm evidence that the weight assigned to each sense is not fixed but is updated dynamically and rapidly so as to match the reliability of that sense at that particular moment. This is most clearly seen in Figure 6, where visual bias and sensitivity, and the Bayesian prediction of bimodal performance, varies smoothly over the interval 50 ms on either side of saccadic onset. This dynamic update of sensory weights implies that the nervous system has the ability to measure the error associated with each sensory estimate. It is still far from clear how neural mechanisms achieve this (Witten and Knudsen, 2005).

Conclusion

Seeing is usually believing: for about two-thirds of our waking lives, when we fixate, we perceive objects where vision tells us they are, which, more often than not, coincides with their actual position. In the remaining time, the visual system sends us erroneous spatial information, presumably because it is engaged in correcting the troublesome consequences of eye movements on retinal afferences. When this happens, we disbelieve visual information; if available, spatial cues from other senses become dominant. Here, we show that not only multimodal perception but also visual space perception can be modeled by assuming that, at any time, the perceptual system optimally integrates the available forms of information.

References

Alais D, Burr CD (2004) The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* 14:257–262.

Bahcall DO, Kowler E (1999) Illusory shifts in visual direction accompany adaptation of saccadic eye movements. *Nature* 400:864–866.

Batista AP, Buneo CA, Snyder LH, Andersen RA (1999) Reach plans in eye-centered coordinates. *Science* 285:257–260.

Bayes T (1764) An essay toward solving a problem in the Doctrine of Chances. *Proc R Soc Lond B Biol Sci* 370–418.

Bockisch CJ, Miller JM (1999) Different motor systems use similar damped extraretinal eye position information. *Vision Res* 39:1025–1038.

Burr DC, Morrone MC, Ross J (1994) Selective suppression of the magnocellular visual pathway during saccadic eye movements. *Nature* 371:511–513.

Clark JJ, Yuille AL (1990) Data fusion for sensory information processing systems. Boston: Kluwer.

Dassonville P, Schlag J, Schlag-Rey M (1992) Oculomotor localization relies on a damped representation of saccadic eye displacement in human and nonhuman primates. *Vis Neurosci* 9:261–269.

Deneve S, Latham PE, Pouget A (2001) Efficient computation and cue integration with noisy population codes. *Nat Neurosci* 4:826–831.

Duhamel JR, Colby CL, Goldberg ME (1992) The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* 255:90–92.

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

Ernst M, Bulthoff H (2004) Merging the senses into a robust percept. *Trends Cogn Sci* 8:162–169.

Harris LR, Lieberman L (1996) Auditory stimulus detection is not suppressed during saccadic eye movements. *Perception* 25:999–1004.

Honda H (1989) Perceptual localization of visual stimuli flashed during saccades. *Percept Psychophys* 162–174.

Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. *Nat Neurosci* 9:690–696.

Klingenhoefer S, Bremmer F (2004) Perisaccadic mislocalization of auditory targets. Paper presented at Fifth Annual Meeting of the International Multisensory Research Forum, Barcelona, Spain, June.

Knill DC, Kersten D (1991) Apparent surface curvature affects lightness perception. *Nature* 351:228–230.

Kowler E, Anderson E, Doshier B, Blaser E (1995) The role of attention in the programming of saccades. *Vision Res* 35:1897–1916.

Kusunoki M, Goldberg ME (2003) The time course of perisaccadic receptive field shifts in the lateral intraparietal area of the monkey. *J Neurophysiol* 89:1519–1527.

Kusunoki M, Colby CL, Duhamel J-R, Goldberg ME (1997) The role of the lateral intraparietal area in the control of visuospatial attention. In: *The association cortex: structure and function* (Fuster JM, ed), pp 191–206. New York: Academic.

Lee DK, Itti L, Koch C, Braun J (1999) Attention activates winner-take-all competition among visual filters. *Nat Neurosci* 2:375–381.

Mateeff S (1978) Saccadic eye movements and localization of visual stimuli. *Percept Psychophys* 24:215–224.

Matin L (1972) Visual psychophysics. In: *Handbook of sensory physiology* (Jameson D, Hurvich LM, eds), pp 311–380. Berlin: Springer.

Matin L, Pearce DG (1965) Visual perception of direction for stimuli flashed during voluntary saccadic eye movements. *Science* 148:1485–1488.

Miller JM (1996) Egocentric localization of a perisaccadic flash by manual pointing. *Vision Res* 36:837–851.

Morrone MC, Ross J, Burr DC (1997) Apparent position of visual targets during real and simulated saccadic eye movements. *J Neurosci* 17:7941–7953.

Morrone MC, Denti V, Spinelli D (2002) Color and luminance contrasts attract independent attention. *Curr Biol* 12:1134–1137.

Nakamura K, Colby CL (2000) Visual, saccade-related, and cognitive activation of single neurons in monkey extrastriate area V3A. *J Neurophysiol* 84:677–692.

Niemeier M, Crawford JD, Tweed DB (2003) Optimal transsaccadic integration explains distorted spatial perception. *Nature* 422:76–80.

Pola J (2004) Models of the mechanism underlying perceived location of a perisaccadic flash. *Vision Res* 44:2799–2813.

Robinson DA (1975) Oculomotor control signals. In: *Basic mechanisms of ocular motility and their clinical implications* (Bach-y-Rita P, ed), pp 337–374. Oxford: Pergamon.

Ross J, Morrone MC, Burr DC (1997) Compression of visual space before saccades. *Nature* 386:598–601.

Sommer MA, Wurtz RH (2006) Influence of the thalamus on spatial visual processing in frontal cortex. *Nature* 444:374–377.

Umeno MM, Goldberg ME (1997) Spatial processing in the monkey frontal eye field. I. Predictive visual responses. *J Neurophysiol* 78:1373–1383.

Vliegen J, Van Grootel TJ, Van Opstal AJ (2004) Dynamic sound localization during rapid eye-head gaze shifts. *J Neurosci* 24:9291–9302.

Walker MF, Fitzgibbon EJ, Goldberg ME (1995) Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements. *J Neurophysiol* 73:1988–2003.

Watson AB, Pelli DG (1983) Quest: a bayesian adaptive psychometric method. *Percept Psychophys* 113–120.

Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. *Nat Neurosci* 5:598–604.

Witten IB, Knudsen EI (2005) Why seeing is believing: merging auditory and visual worlds. *Neuron* 48:489–496.