Auditory-Visual Interactions Subserving Goal-Directed Saccades in a Complex Scene

B. D. CORNEIL,[1] M. VAN WANROOIJ,[2] D. P. MUNOZ,[1] AND A. J. VAN OPSTAL[2]

[1]*Centre for Neuroscience Studies, Department of Physiology, Queen's University, Kingston, Ontario K7L 3N6, Canada; and*
[2]*Department of Biophysics, University of Nijmegen, Geert Grooteplein 21, 6525 EZ Nijmegen, The Netherlands*

Corneil, B. D., M. Van Wanrooij, D. P. Munoz, and A. J. Van Opstal. Auditory-visual interactions subserving goal-directed saccades in a complex scene. *J Neurophysiol* 88: 438–454, 2002; 10.1152/jn.00699.2001. This study addresses the integration of auditory and visual stimuli subserving the generation of saccades in a complex scene. Previous studies have shown that saccadic reaction times (SRTs) to combined auditory-visual stimuli are reduced when compared with SRTs to either stimulus alone. However, these results have been typically obtained with high-intensity stimuli distributed over a limited number of positions in the horizontal plane. It is less clear how auditory-visual interactions influence saccades under more complex but arguably more natural conditions, when low-intensity stimuli are embedded in complex backgrounds and distributed throughout two-dimensional (2-D) space. To study this problem, human subjects made saccades to visual-only (V-saccades), auditory-only (A-saccades), or spatially coincident auditory-visual (AV-saccades) targets. In each trial, the low-intensity target was embedded within a complex auditory-visual background, and subjects were allowed over 3 s to search for and foveate the target at 1 of 24 possible locations within the 2-D oculomotor range. We varied systematically the onset times of the targets and the intensity of the auditory target relative to background [i.e., the signal-to-noise (S/N) ratio] to examine their effects on both SRT and saccadic accuracy. Subjects were often able to localize the target within one or two saccades, but in about 15% of the trials they generated scanning patterns that consisted of many saccades. The present study reports only the SRT and accuracy of the first saccade in each trial. In all subjects, A-saccades had shorter SRTs than V-saccades, but were more inaccurate than V-saccades when generated to auditory targets presented at low S/N ratios. AV-saccades were at least as accurate as V-saccades but were generated at SRTs typical of A-saccades. The properties of AV-saccades depended systematically on both stimulus timing and S/N ratio of the auditory target. Compared with unimodal A- and V-saccades, the improvements in SRT and accuracy of AV-saccades were greatest when the visual target was synchronous with or leading the auditory target, and when the S/N ratio of the auditory target was lowest. Further, the improvements in saccade accuracy were greater in elevation than in azimuth. A control experiment demonstrated that a portion of the improvements in SRT could be attributable to a warning-cue mechanism, but that the improvements in saccade accuracy depended on the spatial register of the stimuli. These results agree well with earlier electrophysiological results obtained from the midbrain superior colliculus (SC) of anesthetized preparations, and we argue that they demonstrate multisensory integration of auditory and visual signals in a complex, quasi-natural environment. A conceptual model incorporating the SC is presented to explain the observed data.

INTRODUCTION

Saccadic eye movements reorient gaze swiftly to a new target of interest. Much has been learned about the neural processes underlying the initiation of visually guided saccades (see Findlay and Walker 1999; Munoz et al. 2000 for review). Under natural conditions, the saccadic system is typically challenged by myriad possible targets to which gaze could be directed. Often, these potential targets emit multisensory signals that may provide different combinations of visual, auditory, and tactile inputs. The integration of multisensory signals from a single event into an orienting response is far from trivial as different sensory modalities are transduced uniquely and encoded initially in different frames of reference (see Sparks and Mays 1990 for review). The oculocentric frame of reference in which saccades are represented must be derived from retinotopic signals for visually guided saccades, and from head-centered space for aurally guided saccades. This latter transformation is particularly complex because the CNS constructs the head-centered space from different acoustic cues: sound azimuth is extracted from interaural timing and intensity disparities, and sound elevation from monaural spectral shape cues induced by the pinnae (see Blauert 1997; Irvine 1986 for review).

There is ample experimental evidence that a combined presentation of auditory and visual stimuli reduces saccadic reaction times (SRTs) (see Colonius and Arndt 2001 for a recent review). These reductions generally exceed the predictions of the so-called "race model," which entails that combined auditory and visual stimuli are processed independently but produce shorter SRTs so long as the unimodal distributions overlap, since subjects can react to either stimulus (Raab 1962). Exceeding the race model implies that the bimodal stimuli are neurally integrated prior to saccade initiation (Hughes et al. 1994; Nozawa et al. 1994). Observed SRT reductions range usually between 10 and 50 ms and diminish as the spatial and temporal separation of the stimuli increases (Colonius and Arndt 2001; Corneil and Munoz 1996; Frens et al. 1995; Harrington and Peck 1998; Hughes et al. 1998).

The neural correlates of multisensory integration have been studied extensively in anesthetized preparations and also depend on the spatial and temporal register of the stimuli (see Stein and Meredith 1993 for review). Another important property of neurons that display multisensory integration is that of "inverse effectiveness" (Meredith and Stein 1986), whereby smaller unimodal responses from near-threshold stimulus intensities are associated with conversely stronger amounts of multisensory integration. If similar mechanisms operate in awake preparations, then the behavioral benefits afforded by multisensory integration should also be greatest with low-intensity stimuli. Accordingly, improved orienting to low-intensity multisensory stimuli has been demonstrated in cats (Stein et al. 1989). So far, human studies using low-intensity stimuli have not demonstrated the dramatic behavioral benefits expected from inverse effectiveness: the SRT reductions afforded by pairing low-intensity stimuli usually approximate the SRT reductions afforded by pairing high-intensity stimuli (Frens et al. 1995; Hughes et al. 1994). Perhaps in these studies, the low intensities were not close enough to threshold, or the limited number of potential target locations may have allowed subjects to constrain their responses prior to stimulus onset. In addition, the auditory stimulus in some of these experiments did not serve as a potential target, but acted as a distractor that could have been ignored by the subject.

The purpose of the present study is to evaluate multisensory integration in human saccades in a complex experimental environment in which both the auditory and visual stimuli serve as potential targets. To this end, low-intensity unimodal or bimodal targets were distributed over 24 possible target locations within the two-dimensional (2-D) oculomotor range and embedded within an auditory-visual background (Fig. 1). Both the signal-to-noise (S/N) ratio of the auditory target relative to background, and the temporal register of the auditory and visual targets on bimodal trials were systematically varied. Attesting to the difficulty of this task, subjects generated saccade scan patterns that consisted of anywhere between 1 to over 10 saccades before localizing the target. This report focuses exclusively on the SRT and accuracy of the first saccade in each trial as indexes of how well the subjects initially localize the target(s). Accurate saccades at short SRTs imply a well-localized target, whereas inaccurate saccades at longer SRTs imply the opposite. Our results demonstrate that the behavioral benefits of auditory-visual integration vary systematically with the S/N ratio as predicted by inverse effectiveness, and that such benefits were greater in the elevation versus azimuth response component. Moreover, the observed effects depended in a systematic way on the relative timing of the auditory and visual stimuli. These behavioral data are in good agreement with the rules extracted from multisensory-evoked responses of cells in the mammalian superior colliculus (SC) (Stein and Meredith 1993).

Abstracts describing some of these data have been published (Corneil et al. 2001; Van Wanrooij et al. 2000).

METHODS

Subjects

Five male subjects (ages 23–43) participated in the experiments and provided their informed consent. Experimental procedures were approved by the local ethics committee of the University of Nijmegen.
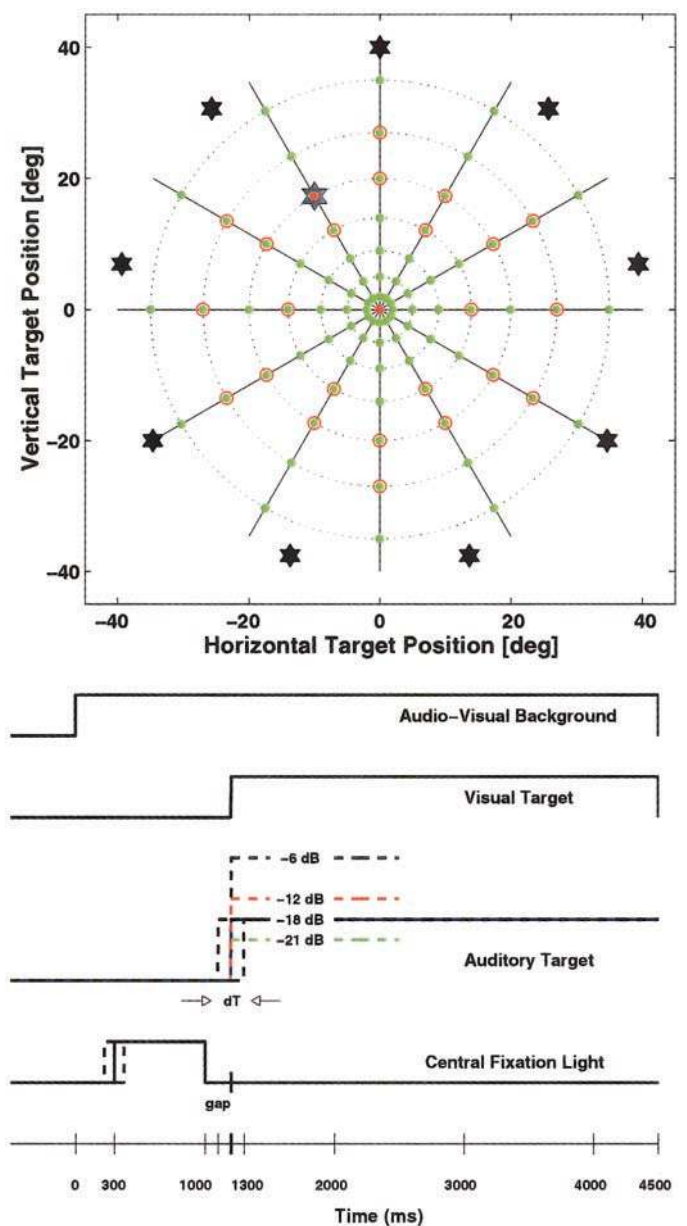

FIG. 1. Spatial (*top*) and temporal (*bottom*) depiction of the AV-multimodal experiment. *Top*: the auditory-visual background was produced by 9 background speakers (black stars) and 85 green light-emitting diodes (LEDs; filled green circles). Subjects were required to saccade from a central red fixation point (FP) to a peripheral auditory, visual, or bimodal auditory-visual target, which could appear at 1 of 24 possible locations (red outlines of green circles). In this particular example, the auditory-visual target is presented at $[R, \Phi] = [20,120]$ (red circle and blue star). *Bottom*: the auditory-visual background was presented at *time 0* and persisted for the entire trial. On unimodal trials, the visual or auditory target was presented 200 ms after the FP turned from red to green (gap). In bimodal trials, the visual target was presented at the same time, but the presentation of the auditory target varied between −100, 0 (as shown here), and 100 ms relative to the visual target (indicated by dT, and vertical dashed lines). The auditory target was presented at 1 of 4 possible intensities (indicated by the heights of the 4 horizontal dashed lines). See METHODS for further details.

All subjects were experienced with eye-movement recording protocols. *Subjects JO, BC, DM,* and *MW* are authors of this paper, although the latter three had no prior experience with sound localization studies. *Subject MZ* was naive as to the purpose of the study. All subjects had normal hearing, as determined by audiograms of both

ears that were obtained with a standard staircase procedure (10 tone pips, 0.5-octave separation, between 500 Hz and 11.3 kHz). With corrective glasses in the experimental setup (*subjects BC, DM,* and *MZ*), all subjects had normal binocular vision except for *JO,* who is amblyopic in his right (recorded) eye. The calibration procedure described below corrected for any nonlinearities from this subject.

Apparatus

Experiments were conducted in a completely dark and sound-attenuated room in which the inner walls, ceiling, and floor, as well as every large object present, were covered with black sound-absorbing acoustic foam that effectively eliminated echoes above 500 Hz. The overall background sound level within the room was approximately 30 dB SPL (A-weighted). The subject was seated comfortably on a chair with back and foot support, and the head was aligned with the center of the room. A customized neck rest, rigidly attached to the floor, prevented the head from moving. Eye movements were recorded with the scleral search coil technique (Collewijn et al. 1975). Horizontal and vertical eye position signals were demodulated by lock-in amplifiers (PAR 128A), amplified and low-pass filtered (cutoff 150 Hz), and sampled at 500 Hz per channel (Metrabyte DAS16H) before being stored on hard disk.

STIMULUS GENERATION. *Visual stimuli.* Visual stimuli were generated by 85 light-emitting diodes (LEDs) that were mounted on a thin wireframe that formed a hemispheric surface 85 cm in front of the subject (the "LED sky"). LEDs were positioned at visual angles that corresponded in a 2-D polar coordinate system to seven radial eccentricities $R \epsilon$ [2; 5; 9; 14; 20; 27; 35] deg with respect to the center of the LED sky, and 12 directions $\Phi \epsilon$ [0; 30; 60; . . . ; 330] deg, respectively (where $\Phi = 0$ deg is rightward, $\Phi = 90$ deg is upward, etc.; Fig. 1*A*). All LEDs could be turned green or red. The visual background was formed by turning all 85 LEDs green. The initial fixation point (FP) was presented by turning the central LED at [R, Φ] = [0, 0] red. The visual target was lit by turning one of the other green LEDs red. LED intensities were kept low to ensure that localization was difficult in the presence of the background (green LEDs: 0.25 cd/m^2; red LEDs: 0.18 cd/m^2). The LED sky was backed by an acoustically transparent thin black cloth.

Acoustic stimuli. The acoustic environment consisted of an auditory background sound and a target sound. All sound intensities were measured at the position of the subject's head with a calibrated sound amplifier and microphone (Brüel and Kjaer BK2610/BK4144), and are expressed in dB SPL (A-weighted). All auditory stimuli were generated digitally at 50 kHz (National Instruments DA board, DT2821) and tapered with a sine-squared onset and offset ramp of 5 ms duration. The signals were amplified by a Luxman A331 audio amplifier and band-pass filtered (0.2–20 kHz, Krohnhite) before being passed to the speakers.

The background sound was produced by a circular array of nine small speakers (Nellcor; response characteristics flat within 5 dB between 2 and 20 kHz, not corrected), mounted on the wireframe of the LED sky at an eccentricity of about 45° relative to center (Fig. 1), and consisted of broadband Gaussian white noise (bandwidth 0.2–20 kHz) that was presented at a fixed intensity of 60 dB. This acoustic environment was perceived by all subjects as a spatially diffuse sound that did not emanate from any specific location.

The auditory target sound consisted of periodic broadband noise (period 20 ms, sounding like a 50-Hz buzzer, clearly discernable from the Gaussian white noise background) that had a flat broadband characteristic between 0.2 and 20 kHz. The auditory target was emitted by a broadband lightweight speaker (Philips AD-44725; response characteristics flat within 12 dB between 0.5 and 15 kHz, not corrected) and was presented at a variable intensity relative to background (see following text). The speaker was mounted on a two-link robot, which consisted of a base with two nested L-shaped arms

controlled by a PC80486 computer that drove separate stepping engines (Berger-Lahr, type VRDM5) (see Hofman and Van Opstal 1998 for details). This setup enabled rapid (within 2 s) and accurate (within 0.5°) positioning of the speaker at a fixed distance of 90 cm from the subject at any location on a virtual sphere just behind the LED sky. Earlier studies (Frens and Van Opstal 1995) verified that the sounds produced by the stepping motors did not provide any consistent localization cues to the subject. Before every trial, the speaker was moved to a random location at least 20° away from the previous location before a final positioning movement was made. In this way, speaker displacement cues could not be related to final speaker location.

Paradigms

Every subject performed three types of experiments: a calibration experiment, the primary auditory-visual (AV) experiment, and an AV-control experiment. Every session began with one block of the calibration experiment without the AV-background, then two blocks of either the primary or control AV-experiment, with the AV-background.

CALIBRATION EXPERIMENT. In all experimental sessions, the subjects first performed a calibration experiment without the AV-background. Subjects were instructed to look from a central red FP to a randomly selected peripheral red LED target that was illuminated as soon as the FP was extinguished (1 block consisted of 72 targets: 12 directions × 6 eccentricities, $R \geq 5°$, each presented once), and press a hand-held button when the target was finally fixated.

The primary purpose of the calibration experiment was to provide the final fixation positions for off-line calibration of the eye coil signals (described below). However, the first-saccade data from the calibration experiment also established the SRT and accuracy of visually guided saccades in the absence of the AV-background, which was compared with visually guided saccades generated in the presence of the AV-background. Although it might seem paradoxical to analyze data from a calibration experiment, this reduced the amount of time per session. Further, only the first-saccade data were used for comparative purposes, whereas the calibration procedures used the final fixation position, regardless of the number of preceding saccades.

AV EXPERIMENT. The spatial and temporal layout of the AV-experiment is depicted in Fig. 1. At *time 0* in each trial, the AV-background was turned on. After a randomly selected interval of 100, 225, or 350 ms, the central red FP changed color from green to red, and the subject was required to fixate it. At time 1,000 ms, the central FP turned from red to green, and a peripheral target was presented 100–200 ms later (see following text). The subject was instructed to acquire the peripheral target as quickly and as accurately as possible. The location of the peripheral target was selected at random from 1 of 24 different positions. All 12 directions on the LED sky were equally likely, but for each direction only 2 of the following 3 eccentricities were selected: $R = 14, 20,$ or 27° (Fig. 1). Subjects made saccades to red visual targets (V-trials), auditory targets (A-trials), or to bimodal auditory-visual targets (AV-trials). The auditory target was presented at one of four different signal-to-noise intensity ratios (S/N ratio) relative to the fixed-intensity background: $-6, -12, -18,$ or -21 dB. For the unimodal V- or A-trials, the target was presented 200 ms after the FP turned green (i.e., at time 1,200 ms) and persisted for 3,300 ms, determining the maximal search time. In AV-trials, the auditory and visual targets were always spatially coincident. The red visual target was illuminated at time 1,200 ms; the auditory target was presented randomly at time 1,100 ms, 1,200 ms, or 1,300 ms (i.e., either synchronous, or 100 ms before or after the visual target). Note that any time a visual target was presented, it was presented 200 ms after the offset of the fixation point (i.e., a 200-ms "gap"). This was done so that our data would not have to be analyzed as a function of gap interval, given the known effects of gap interval on SRT (see Findlay and Walker 1999 for review).

For coding purposes, saccades were identified by their trial type, S/N ratio of the auditory target (if applicable), and stimulus asynchrony coded relative to stimulus onset (if applicable). Thus V-saccades denote data from unimodal V-trials. A_{-6}-saccades denote data from unimodal A-trials where the target intensity was set to -6 dB relative to the auditory background. $A_{-12}100V$-saccades denote data from AV-trials in which the auditory target (-12 dB relative to auditory background) led the visual target by 100 ms. Twelve different AV-trials were possible (3 temporal asynchronies $\times$4 S/N ratios). In total, 17 different trial types were tested at each target position (1 V-trial, 4 A-trials, 12 AV-trials), making a total of 408 different trials (17 trial types $\times$24 target positions) for one complete series. All trial types were randomly interleaved. Each experimental session contained 204 trials run in two blocks of 102 trials each. A subsequent session was typically run on another day and contained the remaining 204 trials to complete the series. Each subject completed at least three full series of AV-multimodal experiments (*DM, MZ,* and *MW*: 3 series; *BC*: 4 series; and *JO*: 5 series), yielding between 72 and 120 trials per trial type.

An oversight on our part replaced the $A_{-6}100V$-trials with V100A_{-6} trials. Although unfortunate, our conclusions were not affected by the lack of data from $A_{-6}100V$-saccades.

AV-CONTROL EXPERIMENT. It is known that the onset or offset of an auditory target can lower SRTs to visual targets, presumably by a warning effect that is independent of the spatial congruity of the auditory and visual stimuli (Ross and Ross 1980, 1981). To parse out the portions of the data set from the primary AV-experiment that were caused by this nonspecific warning effect, each subject was also tested in a separate control experiment. Three trial types from the primary AV-experiment (V-only, A_{-12}, and A_{-12}V) were mixed with a new type of bimodal stimulus in which the auditory target sound was generated by the nine background speakers. For this auditory control stimulus, the acoustic signal was a linear superposition of the Gaussian broadband white noise and the periodic buzzer stimulus. A pilot test indicated good audibility of this sound when its level was at -3 dB relative to background. The subjects perceived this control sound as emanating from a single point near straight ahead, although the exact location of this percept varied between subjects. Accordingly, when the auditory control stimulus was presented, the spatial coincidence of the visual and auditory targets was lost, and the subject's task became ambiguous because of a conflict between the location of the visual target and the perceived location of the auditory control stimulus. In this experiment, a total of 192 trials was measured (each trial type presented twice at each location, yielding 48 trials per stimulus type).

Data analysis

DATA CALIBRATION. Off-line calibration of horizontal and vertical eye position was achieved by training two three-layer neural networks with the back-propagation algorithm on the 72 fixation positions from the calibration experiment (when the button was pressed) and the target coordinates (see Goossens and Van Opstal 1997 for details). The absolute accuracy of the calibration was within 3% over the entire response range. The networks were subsequently applied to the raw data from the calibration (1st saccades only), AV-multimodal, and AV-control experiments to map the measured induction voltages onto the corresponding 2-D orientations of the eye. Target and response coordinates are expressed as azimuth (α) and elevation (ϵ) angles, determined by a double-pole coordinate system in which the origin coincides with the center of the head. In this reference frame, target azimuth, α_T, is defined as the angle between the target and the midsagittal plane. Target elevation, ϵ_T, is the angle between the target and the horizontal plane through the ears with the head in a straight-ahead orientation.

After calibration, saccades were detected by a custom-made program that set separate velocity and acceleration criteria for the iden-

tification of saccade onset and offset. All markings were checked by the experimenter and corrected if needed. SRTs below 80 ms were excluded since these saccades presumably were anticipatory (Corneil and Munoz 1996), and SRTs greater than 1,000 ms were excluded because of a presumed lack of subject alertness. SRTs are expressed with respect to the onset of the *first* target stimulus, regardless of the stimulus asynchrony. SRTs are plotted as cumulative percentage probabilities on a probit scale (i.e., inverted gaussian) as a function of the reciprocal SRT ($-1/$SRT) (see Carpenter and Williams 1995). In this format, a Gaussian distribution results in a straight line.

STATISTICS. For saccade accuracy, the optimal linear fit of the stimulus-response relation between saccade amplitude and target eccentricity was found by minimizing the sum-squared deviation of

$$\alpha_R = a + b \cdot \alpha_T \quad \text{and} \quad \epsilon_R = c + d \cdot \epsilon_T \qquad (1)$$

for the azimuth and elevation components, respectively. In *Eq. 1, a* and *c* are the response biases in degrees (offset of the fitted line), and *b* and *d* are the dimensionless response gains (slopes). Confidence levels for Pearson's correlation coefficients were obtained through the bootstrap method (100 regressions on randomly drawn realizations of the data set) (Press et al. 1992). The gain, bias, residual error (SD relative to the fitted line), mean absolute error (the mean sum-squared difference between the target and response coordinates), and the correlation coefficient extracted from *Eq. 1* describe different aspects of response behavior. Response gain and bias relate to spatial accuracy, whereas the residual error and the correlation coefficient relate to the variability and spatial resolution of the system, respectively. The absolute error depends on both the accuracy and the variability of the responses.

To statistically analyze the differences between two distributions, the one-dimensional (1-D) Kolmogorov-Smirnov (KS) statistic was determined (Press et al. 1992). This statistic is based on the cumulative probabilities constructed from the ranked data arrays and is particularly useful for limited numbers of data points for which the underlying probability distributions are unknown. In cases where a statistical comparison was made between two 2-D distributions of data (i.e., data described by both response accuracy and SRT; see Fig. 6), the 2-D KS statistic for the difference between distributions was determined (Press et al. 1992).

CALCULATION OF THE RACE MODEL PREDICTION. Previous studies of the SRT reduction afforded by presenting bimodal stimuli have utilized the concept of a race model to provide a prediction for the SRT distribution that would be expected if the subject reacted simply to whichever stimulus was perceived first (Colonius and Arndt 2001; Corneil and Munoz 1996; Harrington and Peck 1998; Hughes et al. 1994, 1998). This concept, which operates like a logical OR-gate, was originally developed to model manual reaction times and is alternatively referred to as statistical facilitation or probability summation (Gielen et al. 1983; Miller 1982; Raab 1962). The SRT distribution predicted by a race model, $R(\tau)$ (where τ is a given SRT), is derived from the normalized SRT distributions for saccades to the unimodal auditory or visual stimuli, $A(\tau)$ and $V(\tau)$, respectively, by the following equation

$$R(\tau) = A(\tau) \int_{\tau}^{\infty} V(t)dt + V(\tau) \int_{\tau}^{\infty} A(t)dt \qquad (2)$$

If the observed SRTs to bimodal stimuli are shorter than those predicted by the race model, the bimodal signals are assumed to have been neurally integrated prior to saccade initiation. We calculated the race model prediction for all bimodal trial types in the AV- experiment, using the unimodal SRT distributions for V-saccades and A-saccades at the appropriate S/N ratio in 10-ms bins. The unimodal distributions were shifted by 100 ms for A100V and V100A-saccades.
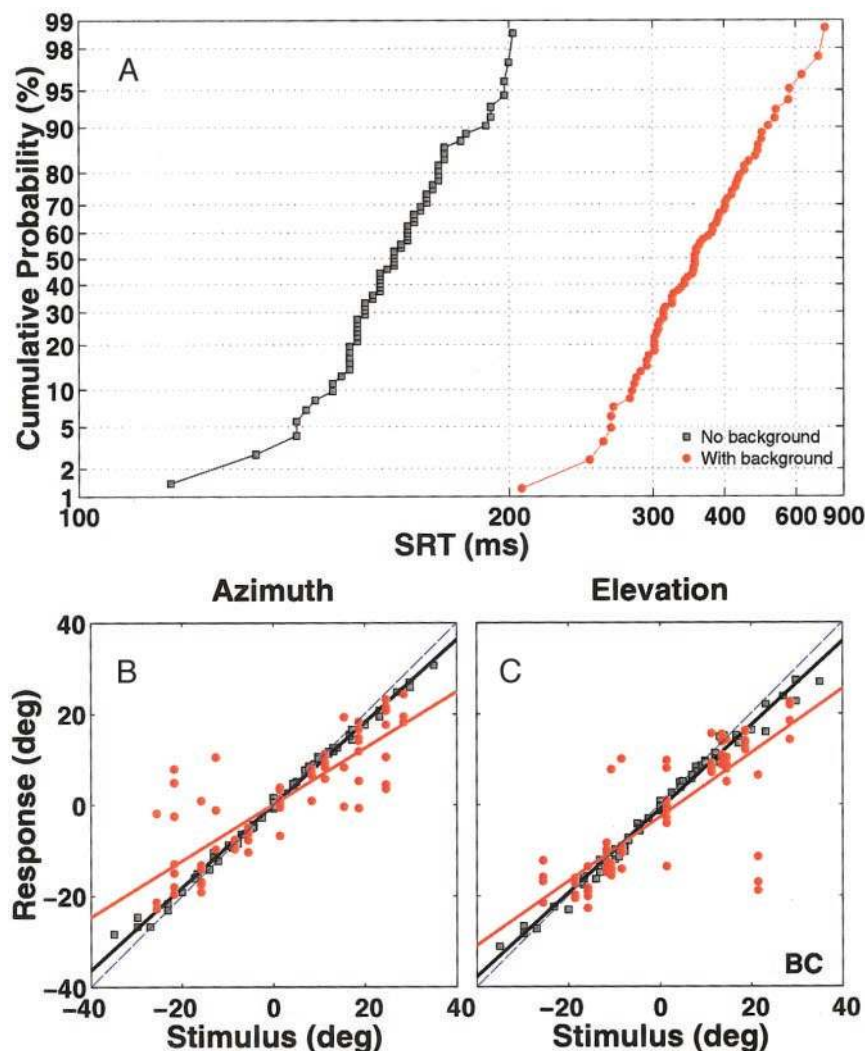
FIG. 2. Comparison of saccadic reaction times (SRTs; *A*) and accuracy (*B* and *C*) of V-saccades in the calibration experiment without the AV-background (gray squares) and in the AV-experiment with the AV-background (red circles) for *subject BC*. *A*: cumulative SRT probability distributions for the V-saccades in the 2 experiments. The cumulative probabilities have been plotted on a probit scale, and SRT is plotted on an inverted scale (see METHODS for further details). *B* and *C*: stimulus-response plots for the endpoints of V-saccades compared with target position in azimuth (*B*) and elevation (*C*). Blue dashed lines denote the unity lines; solid lines denote the linear regression lines.

RESULTS

Properties of unimodal V-saccades and A-saccades

The presence of the AV-background and the S/N ratio of the acoustic environment impacted the SRT and accuracy of unimodal V-saccades and A-saccades. V-saccades had longer SRTs in the presence of the AV-background, as evidenced by comparing the results from the AV-experiment (with the AV-background) to the results from the calibration experiment (without the AV-background; Fig. 2*A* and Table 1; $P < 10^{-8}$ for all subjects, 1-D KS test).[1] Response accuracy, quantified by the parameters of the linear regression analysis between saccade amplitude and target eccentricity (see METHODS), demonstrated that V-saccade accuracy in both azimuth (Fig. 2*B*) and elevation (Fig. 2*C*) was also compromised in the presence of the AV-background. These accuracy differences were significant across all subjects ($P < 0.05$ using the 1-D KS test; Table 1).

The SRT and accuracy of A-saccades depended on the S/N ratio of the auditory target relative to background. SRTs were

systematically longer and more variable for lower S/N ratios, as shown for a representative subject in Fig. 3*A* (Table 2 for all subjects). Interestingly, the accuracy of A-saccades decreased for the lower S/N ratios, but in a manner that differed for the azimuth and elevation response components. Targets were well localized in both azimuth and elevation at the highest S/N ratios (i.e., A_{-6}-saccades), although the residual error was greater in elevation (Fig. 3, *B* and *C*). At the high S/N ratios, the accuracy of A-saccades was in the same range as V-saccades (compare gain and error values in Tables 1 and 2). At the lowest S/N ratio (i.e., A_{-21}-saccades), saccade accuracy in azimuth decreased only slightly when compared with A_{-6}-saccades (Fig. 3*B*), yet was almost completely abolished in elevation (Fig. 3*C*). An analysis of the gain of the stimulus-response relationship (Fig. 4*A*) and the absolute response error (Fig. 4*B*) for the azimuth and elevation response components of A-saccades across S/N ratio confirmed the greater inaccuracy of aurally guided saccades in the elevation component at lower S/N ratios (i.e., for A_{-18} and A_{-21}-saccades) than in the azimuth component, which was only slightly compromised for A_{-21}-saccades (see also Table 2). These results confirm and extend earlier findings of auditory localization (Good and Gilkey 1996; Zwiers et al. 2001).

---

[1] This SRT difference cannot be due to the gap between FP offset and target onset, since the 200-ms gap in the AV-multimodal experiment would favor even shorter SRTs (see Fischer and Weber 1993 for review).

TABLE 1. *Median SRTs and linear regression results for V-saccades generated in the calibration experiment (without AV-background) and in the AV-experiment (with AV-background)*

Subject	SRT, ms	Azimuth				Elevation				n
		Gain	Bias	Corr	Err	Gain	Bias	Corr	Err	
		Calibration experiment without AV-background								
BC	162 ± 29	0.91	0.0	0.99	1.1	0.88	−0.9	0.99	1.8	643
DM	226 ± 47	0.94	0.3	0.99	1.1	0.92	−0.2	0.99	1.3	503
JO	196 ± 48	0.92	−1.1	0.98	2.8	0.91	0.5	0.98	3.1	716
MW	174 ± 55	0.88	0.4	0.99	2.1	0.88	−0.8	0.97	3.3	590
MZ	168 ± 20	0.93	−0.2	1.00	0.8	0.92	−0.1	0.99	0.9	431
		AV-multisensory experiment with AV-background								
BC	352 ± 113	0.62	1.0	0.84	6.3	0.71	−1.7	0.78	7.5	96
DM	419 ± 134	0.62	0.7	0.78	7.4	0.58	−1.2	0.73	11.0	72
JO	329 ± 95	0.66	−1.0	0.83	6.8	0.76	−0.7	0.85	6.9	120
MW	276 ± 82	0.76	0.9	0.86	6.4	0.70	0.9	0.77	8.1	72
MZ	319 ± 77	0.87	0.1	0.91	5.4	0.72	−0.9	0.80	7.5	60

Values in SRT are means ± SD; *n* is number of saccades. Note the much longer and more variable SRTs for V-saccades in the latter experiment, as well as the considerably lower accuracy, which is apparent from lower gain and correlation values and much higher residual errors (Bias and Err in deg; Gain and Corr are dimensionless). SRT, saccadic reaction time.

In summary, the increased SRT and decreased accuracy of V- and A-saccades (particularly at low S/N ratios) confirmed that the presence of the AV-background made the task much more difficult, although not impossible. A similar analysis on saccades generated to the different types of bimodal stimuli is now presented.

Properties of AV-saccades (no temporal asynchronies)

A representative example of the properties of AV-saccades is demonstrated in Fig. 5, in which V-, A_{-18}-, and A_{-18}V-saccades are compared. Note that the SRT distributions for A_{-18} and A_{-18}V-saccades are nearly superimposed (Fig. 5*A*), equaling but not exceeding the race model prediction based on the unimodal SRT distributions (black, solid line). The relationships of AV-saccades to the race model prediction are studied more thoroughly below. A comparison of saccade accuracy demonstrated that the residual errors of A_{-18}V-saccades were smaller than the residual errors for both V- and A_{-18}-saccades in the elevation (Fig. 5*C*), but not in the azimuth (Fig. 5*B*), component.

The SRT and accuracy of AV-saccades is contrasted more directly with unimodal saccades using a 2-D comparison of absolute localization error (combining both azimuth and elevation) versus SRT (see Fig. 6 for 1 subject). Each point in Fig. 6 stems from an individual saccade, and the ellipses circumscribe the mean values within one SD. Note that V-saccades were generated at longer SRTs but were more accurate than A_{-18}-saccades. However, the 2-D distribution of A_{-18}V-saccades is clearly distinct from either unimodal distribution, as the AV-saccades attained accuracies in the range of V-saccades, but at SRTs in the range of A_{-18}-saccades. The 2-D KS pair-wise statistic comparing the three distributions showed that all were significantly different ($P < 10^{-5}$ for all 3 comparisons). The results of this three-way statistical comparison across all subjects and at all S/N ratios is shown in Table 3. When the S/N ratio was low (−18 or −21 dB), the 2-D distributions for AV-saccades differed significantly from both unimodal A-saccades and V-saccades. For the higher S/N

ratios (−12 or −6 dB), the distributions for AV-saccades were often similar to the A-saccade distributions, but were always significantly different from V-saccades.

Another interesting observation from Fig. 6 is that the AV-saccades appeared to be distributed over a narrower accuracy-SRT range than A- and V-saccade distributions (compare the horizontal and vertical spans of the ellipses in Fig. 6). To quantify this point across all S/N ratios and subjects, we made two comparisons. First, we compared the SRT variance of AV-saccades to A-saccades (Fig. 7*A*) and demonstrated that the SRT variance for AV-saccades was similar to A-saccades at high S/N ratios, but had consistently lower variances at lower S/N ratios. Second, a comparison of the accuracy variance between AV-saccades and V-saccades showed that the accuracy variances for AV-saccades were consistently narrower than those for V-saccades, as the majority of data points lay below the diagonal in Fig. 7*B*. Thus, although auditory targets were barely detectable in elevation at low S/N ratios (Fig. 4 and Table 2), they were integrated effectively with the visual target to reduce both the mean and the variance of AV-saccade SRT and accuracy.

Taken together, the data suggest that the magnitude of multisensory interactions depended systematically on the S/N ratio of the auditory target relative to the background. At low S/N ratios, A-saccades were characterized by decreasing accuracy, and longer SRT and V-saccades were more accurate but were generated at much longer SRTs. When these two weak stimuli were combined, AV-saccades benefited from the "best of both worlds" in that they were as accurate as V-saccades and initiated at SRTs typical of A-saccades. Moreover, the variability of SRT and accuracy for AV-saccades was decreased compared with A-saccades and V-saccades, respectively, indicating more consistent responses.

AV interactions as a function of stimulus timing

In this section, we compare the properties of saccade responses to synchronous stimuli (i.e., AV-saccades) to those from asynchronous stimuli (i.e., A100V- or V100A-saccades).
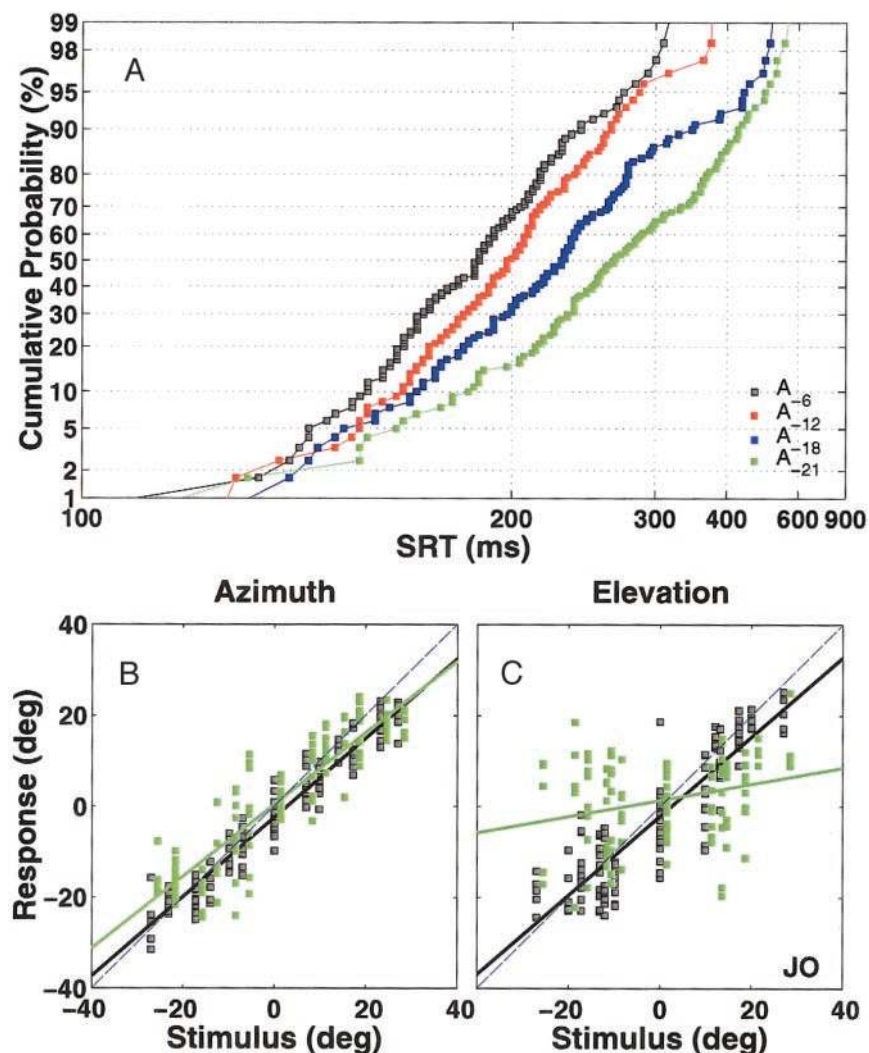
FIG. 3. Comparison of A-saccade SRT (*A*) and accuracy (*B* and *C*) in the AV-multimodal experiment with different signal-to-noise (S/N) ratios for *subject JO*. Same format as Fig. 2. Gray, red, blue, and green squares denote the data from A_{-6}, A_{-12}, A_{-18}, and A_{-21} saccades, respectively. Note in *A* that the distributions shift systematically to the right and cover a greater range for lower S/N ratios.

First, the SRT distributions for each stimulus asynchrony and S/N ratio combination were compared with the distributions predicted by the race model (see METHODS). To that end, the observed cumulative response distributions for bimodal stimuli were plotted as a function of the predicted cumulative race distributions (Fig. 8*A* for a representative subject). Such plots compare the relative differences between the observed and predicted cumulative distributions, regardless of absolute SRT. Note that the comparison plots for AV-saccades (solid lines in Fig. 8*A*) lay close to the unity line, implying that the observed SRT distributions were approximately equal to those predicted by the race model for all four S/N ratios, consistent with Fig. 5*A*. However, the comparison plots for V100A-saccades (dashed lines in Fig. 8*A*) lay well above the unity line, indicating that the observed SRT distributions were considerably shorter than those predicted by the race model. Conversely, the comparison plots for A100V-saccades (dashed-dotted lines in Fig. 8*A*) lay well below the unity line, meaning that the observed SRTs were much longer than predicted by the race model. This latter finding is quite striking since it implies that the delayed visual stimulus inhibits the SRTs for A100V-saccades compared with the SRTs for A-saccades.

It is not trivial to appreciate how these relationships with the race model change with the S/N ratio of the acoustic environment. To quantify this, we determined the area of the difference curve between the observed and predicted cumulative SRT distributions for those SRTs where the cumulative probabilities fell between 0.1 and 0.9. These calculated areas express the amount by which the observed data exceeded (positive values) or fell short (negative values) of the race model prediction regardless of the absolute SRTs, and are plotted for the same subject in Fig. 8*B*. Presented this way, it is clear that no systematic relationship emerged with the S/N ratio, hence the extracted areas were averaged across all S/N ratios (gray bars in Fig. 8*B*). AV-saccades did not deviate significantly from the race model ($P > 0.05$), whereas A100V-saccades had significantly longer SRTs by about 20% ($P < 0.02$) and V100A-saccades had significantly shorter SRTs by about 15% ($P < 0.02$) than those predicted by the race model. This pattern of SRT responses was found for three of five subjects. In the other two subjects, there was no significant difference between the observed and predicted SRT distributions for both AV-saccades and V100A-saccades. In these two subjects, the unimodal (shifted) SRT distributions did not overlap sufficiently, so that the race model prediction equaled the shorter unimodal SRT distribution (in this case for A-saccades). Regardless, averaging across all subjects revealed that the overall patterns were consistent (Fig. 8*C; P < 0.001* for the A100V-saccades,

TABLE 2. *Median SRT and linear regression results for A-saccades generated in the AV-experiment at different S/N ratios*

Subject	S/N, dB	SRT, ms	Azimuth				Elevation				n
			Gain	Bias	Corr	Err	Gain	Bias	Corr	Err	
BC	−6	178 ± 38	0.85	0.6	0.97	3.0	0.72	−1.3	0.86	6.3	83
	−12	190 ± 61	0.82	0.4	0.97	3.2	0.69	−0.6	0.85	5.9	83
	−18	212 ± 59	0.88	1.5	0.96	3.7	0.46	−0.5	0.57	9.1	81
	−21	264 ± 128	0.75	0.7	0.93	4.6	0.14	−0.7	0.24	9.3	83
DM	−6	201 ± 57	0.85	−1.3	0.96	4.0	0.55	4.6	0.75	7.1	72
	−12	226 ± 85	0.81	0.1	0.97	3.2	0.54	3.1	0.77	6.1	72
	−18	262 ± 75	0.90	1.4	0.96	3.8	0.36	3.0	0.57	7.9	71
	−21	360 ± 132	0.72	1.0	0.93	4.8	0.26	0.7	0.34	10.0	72
JO	−6	186 ± 47	0.87	−2.5	0.97	3.6	0.87	−2.1	0.87	6.6	120
	−12	199 ± 49	0.90	−1.7	0.97	3.5	0.70	−0.6	0.84	6.4	120
	−18	227 ± 84	0.97	0.1	0.96	4.3	0.55	−3.1	0.71	7.9	120
	−21	263 ± 106	0.79	1.5	0.90	6.0	0.18	1.6	0.26	8.8	120
MW	−6	153 ± 32	0.80	−1.7	0.94	4.7	0.69	0.5	0.79	7.1	60
	−12	170 ± 38	0.88	0.8	0.96	3.4	0.82	1.8	0.85	6.5	60
	−18	173 ± 47	0.84	2.4	0.96	4.3	0.31	−0.3	0.43	9.3	60
	−21	205 ± 116	0.76	3.6	0.91	5.2	0.19	3.2	0.33	8.9	60
MZ	−6	210 ± 53	0.84	−0.1	0.95	4.5	0.68	0.9	0.80	7.0	59
	−12	218 ± 57	0.89	0.7	0.97	3.4	0.74	−1.3	0.85	6.3	60
	−18	242 ± 96	0.91	−0.5	0.93	5.3	0.57	0.9	0.77	6.8	60
	−21	299 ± 170	0.75	0.1	0.87	6.6	0.29	−1.2	0.37	10.0	58

Values is SRT are means ± SD; n is number of saccades. Note increased SRT and residual error, and decreased elevation gain and correlation coefficient with decreasing signal-to-noise (S/N) ratio. SRT, saccadic reaction time.

$P > 0.05$ for AV-saccades, and $P < 0.05$ for V100A-saccades). Thus the relationships of the observed SRTs to those predicted by the race model depended on stimulus asynchrony.

To quantify the accuracy of bimodal saccades across stimulus asynchrony and S/N ratio, we first plotted the absolute azimuth and elevation localization errors as a function of S/N ratio for the different temporal asynchronies. Figure 9 shows data from one representative subject. Note that the accuracy of bimodal saccades almost always surpassed that of A-saccades, regardless of asynchrony or S/N ratio. In most cases, the accuracy of AV-saccades was also better than that of V-saccades. V100A-saccades tended to be among the most accurate, surpassing both AV- and A100V-saccades, particularly in elevation at low S/N ratios (Fig. 9*B*). Statistical analysis across all subjects confirmed that the elevation gain of bimodal saccades differed more from the gains obtained from V-saccades than A-saccades did at the lower S/N ratios (Table 4). However, this trend was not observed in the azimuth response component (Table 4).

A summary of the combined SRT-accuracy results for all bimodal stimulus conditions is shown in Fig. 10. These data were obtained by first normalizing the results for each stimulus condition with respect to the accuracy and SRT of V-saccades within each subject, and then averaging the normalized results for each condition across all subjects (note that data for $A_{-6}100V$-saccades are absent; see METHODS). All bimodal data in this accuracy-SRT plane are clearly distinct from the unimodal saccades, and there were obvious patterns depending on both the asynchrony and the S/N ratio. First, the normalized SRT and absolute localization error of A-saccades and bimodal saccades progressively increased with decreasing S/N ratios. Second, the position of the bimodal data in the accuracy-SRT plane depended strongly on the stimulus asynchrony. Relative to AV-saccades, A100V-saccades were more inaccurate and

had longer SRTs at the lower S/N ratios. This latter point is in agreement with our earlier analysis on the SRTs compared with the race model (Fig. 8) and again shows that the delayed visual stimulus slowed the SRT of A100V-saccades compared with unimodal A-saccades. In contrast, V100A-saccades were more accurate than AV-saccades, but had longer SRTs. Yet, V100A-saccades clearly surpassed the predictions of the race model (Fig. 8; recall that the unimodal distributions had to be shifted by 100 ms to determine the race model predictions). Overall, the best performances, indexed by the relative position of the bimodal saccades compared with the unimodal counterpart, were observed for AV- and V100A-saccades at the lowest S/N ratios.

AV-control experiment

We conducted a control experiment to test for the presence and influence of a generalized warning effect of the auditory target on both SRT and accuracy. Figure 11 shows the data pooled for all subjects from the AV-control experiment, which used an additional bimodal stimulus consisting of a visual target with a control auditory stimulus set up by the background speakers (recall that subjects perceived this sound as emanating from a fixed location near center). We emphasize two main points from this experiment. First, although the control auditory stimulus provided some warning cue information to shorten SRTs of AV-control saccades compared with V-saccades, the SRTs for spatially coincident $A_{-12}V$-saccades were still shorter (Fig. 11*A*). Thus, although one component (around 60 ms) of the shorter SRTs for $A_{-12}V$-saccades could be attributed to a warning effect, another component (accounting for an additional 65 ms) depends on the spatial alignment of the stimuli. Second, note that AV-control saccades were much more inaccurate than spatially coincident $A_{-12}V$-sac-
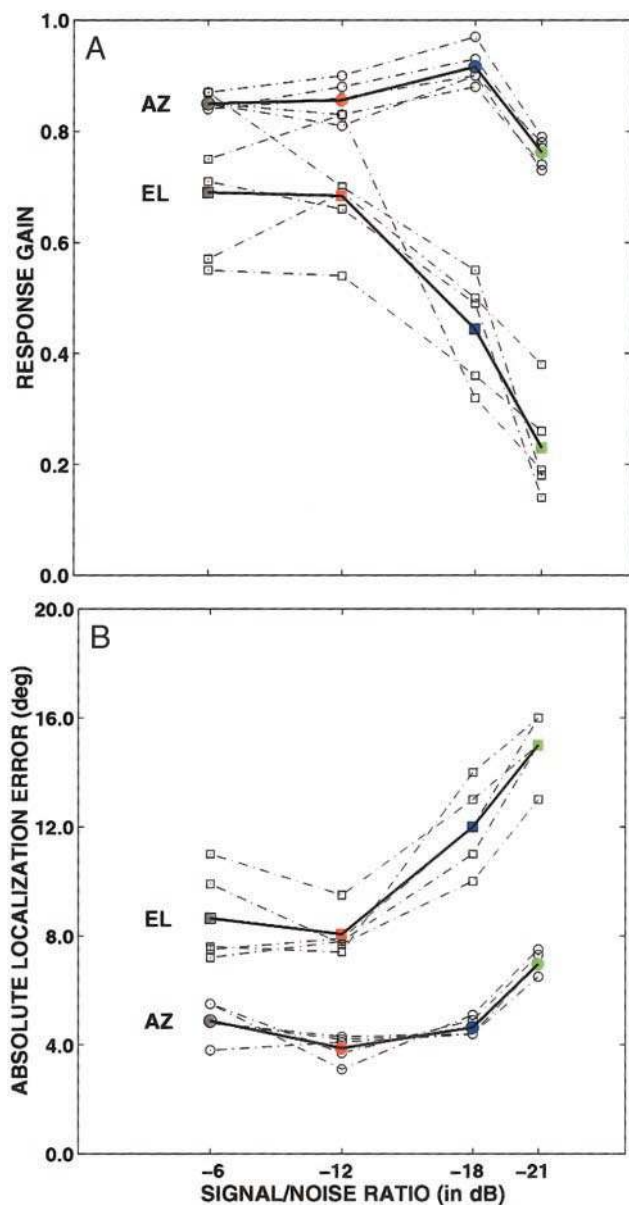
FIG. 4. Response gains (*A*) and absolute error (*B*) for azimuth (circles) and elevation (squares) for A-saccades as a function of S/N ratio for all subjects (open symbols) and the sample mean (solid symbols). Note the robustness of the azimuth response component, which degrades only slightly for A_{-21} saccades, whereas the elevation response component degrades precipitously at higher S/N ratios.

cades (Fig. 11, *B* and *C*) or either V- or A-saccades (Table 5). Thus, although the nonlocalizable auditory target conferred a beneficial warning effect on SRTs, it degraded saccade accuracy. These results were consistent across all subjects (Table 5), from which it was concluded that the combined benefits conferred by auditory-visual integration across SRT and accuracy depended on the spatial alignment of the stimuli.

DISCUSSION

This study investigated the first-saccade responses to visual, auditory, and bimodal stimuli distributed throughout the 2-D oculomotor range and embedded within a complex AV-background. We believe the timing and metrics of the first saccade

provide a measure for the speed and precision with which the oculomotor system can localize and orient to the stimuli. The properties of saccades to unimodal stimuli testify to the complexity of the task: the SRT and error of V-saccades increased greatly in the presence of the AV-background (Fig. 2, Table 1), and the SRT and error of A-saccades depended systematically on the S/N ratio of the acoustic scene, becoming prolonged and inaccurate, particularly in the elevation component, at lower S/N ratios (Figs. 3 and 4; Table 2). The properties of unimodal saccades provided wide ranges over which the benefits afforded by multisensory integration were realized. Specifically, saccades to bimodal stimuli were generated at SRTs typical of A-saccades, but at accuracies typical of V-saccades. These results depended critically on the temporal register of the stimuli and on the S/N ratio of the acoustic environment (Fig. 10). The control experiment demonstrated that the spatial register of the stimuli is also important (Fig. 11; Table 5), although this variable was not systematically manipulated. In this discussion, we argue that mechanisms other than neural integration of the auditory and visual signals cannot explain all aspects of our data. Our results are then related to behavioral and neurophysiological studies. Last, we propose a conceptual neural framework.

Consideration of mechanisms other than neural integration

We consider three mechanisms that could underlie the observed properties of bimodal saccades: race models, aurally assisted visual search, and auditory warning-cue effects. Each predicts specific patterns of SRTs and accuracy that differ substantially from those we observed. For example, race models state that subjects respond to whichever stimulus is perceived first, and derive SRT distributions from the unimodal data (*Eq. 2*) (Colonius and Arndt 2001; Corneil and Munoz 1996; Gielen et al. 1983; Harrington and Peck 1998; Hughes et al. 1994, 1998; Nozawa et al. 1994). Since the SRTs for A-saccades were much shorter than for V-saccades, race models predict that most saccades in bimodal trials would be initiated in response to the auditory target. However, if the subjects only reacted to the auditory target on bimodal trials, then the accuracy of bimodal saccades should equal the accuracy of A-saccades. This was never observed; bimodal saccades were always more accurate than A-saccades (Fig. 10). Even a trial-by-trial comparison of SRT and accuracy shows that individual AV-saccades combine properties typical of both A-saccades and V-saccades (Fig. 6).

Whereas the observed SRTs for AV-saccades agree nicely with those predicted by the race models, the observed SRTs for A100V- were longer and the SRTs for V100A-saccades were shorter than the race model predictions (Fig. 8), testifying to another inadequacy of a race model mechanism. At first, it might seem surprising that the observed SRTs for AV-saccades did not exceed the predicted SRTs, given the many examples of race model violations in the literature (Colonius and Arndt 2001; Harrington and Peck 1996; Hughes et al. 1994, 1998). However, many of these race model violations stem from simple experiments in which subjects orient to the target(s) without the presence of distracting stimuli. Complicating the experiments by employing distracting stimuli, or by instructing the subjects to orient to the auditory instead of the visual target, lead to observed SRTs to bimodal stimuli that do not exceed,
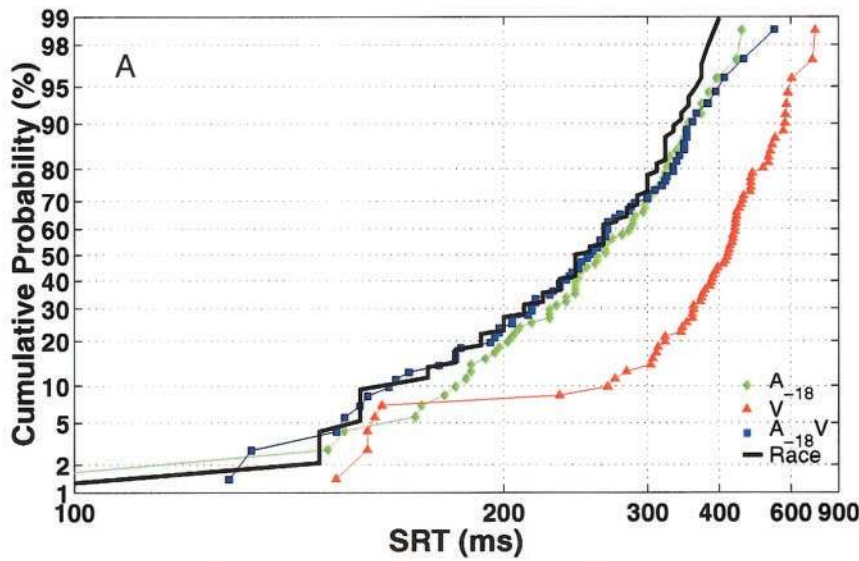
FIG. 5. Comparison of SRT (*A*) and saccade accuracy (*B* and *C*) in the AV-multimodal experiment for A_{-18}-saccades (green diamonds), V-saccades (red triangles), and bimodal A_{-18}V-saccades (blue squares) for *subject DM*. Same format as Fig. 2. The solid black line in *A* denotes the SRT distribution predicted by the race model.
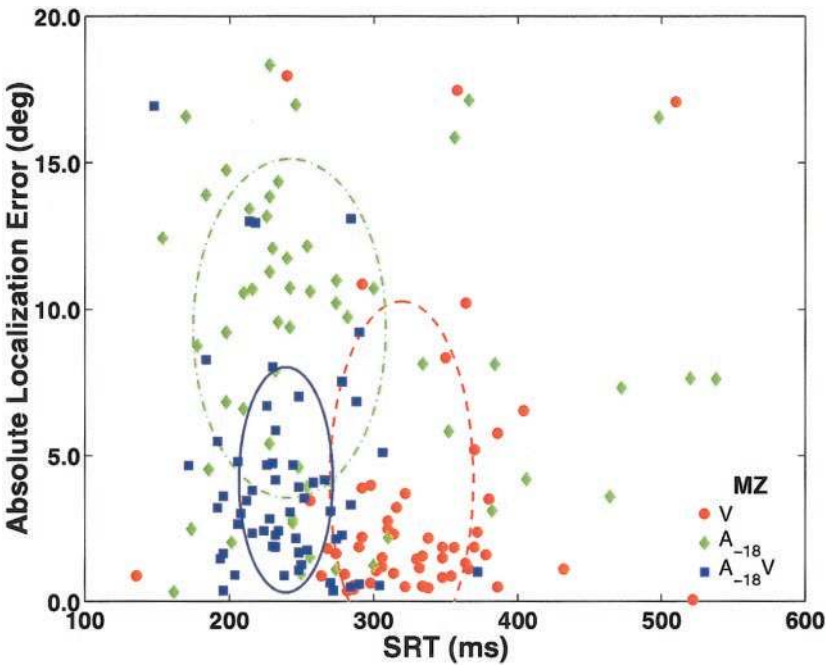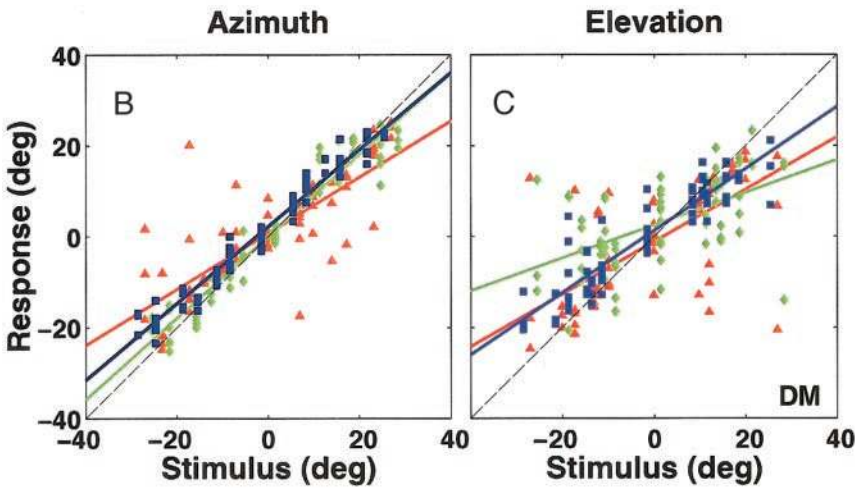

FIG. 6. Absolute localization error plotted as a function of SRT for *subject MZ*. Symbols denote observations from individual V-saccades (red circles), A_{-18}-saccades (green diamonds), and A_{-18}V-saccades (blue squares). Ellipses circumscribe 1 SD around the mean. Only data within 2 SDs of the SRT mean are shown.

TABLE 3. *Statistical summary of probabilities derived from a 2-D KS statistic comparing the distributions of SRT-accuracy saccade data*

Subject	S/N, dB	P(V = A)	P(V = AV)	P(A = AV)
BC	−6	*	*	n.s.
	−12	*	*	†
	−18	*	*	†
	−21	*	*	*
DM	−6	*	*	*
	−12	*	*	n.s.
	−18	*	*	‡
	−21	*	†	*
JO	−6	*	*	n.s.
	−12	*	*	n.s.
	−18	*	*	†
	−21	*	*	*
MW	−6	*	*	n.s.
	−12	*	*	†
	−18	*	*	*
	−21	*	‡	‡
MZ	−6	*	*	‡
	−12	*	*	n.s.
	−18	*	*	*
	−21	*	*	*

Note for all subjects the 2 lowest S/N ratios yielded significantly different distributions for all 3 comparisons. KS, Kolmogorov-Smirnov; n.s., not significant ($P > 0.05$); for other abbreviations, see Table 2. * $P < 0.00001$. † $0.00001 < P < 0.001$. ‡ $0.001 < P < 0.05$.

let alone meet, the SRTs predicted by the race model (Corneil and Munoz 1996; Hughes et al. 1994). More complex experimental paradigms, such as the one described here, presumably engage processes related to target selection and/or discrimination that elongate SRTs and demonstrate the insufficiency of a simple race model mechanism in accounting for the observed data. Below, we surmise on neural mechanisms that could account for the shorter SRT and improved accuracy of V100A-, but not A100V-saccades.

An "aurally assisted visual search" mechanism (Perrott et al. 1990, 1991) also cannot explain the combined patterns of SRT and accuracy. This mechanism proposes that the role of the auditory localization system is to bring the fovea into line with an auditory stimulus, constraining the area over which the visual system searches for a visual target, thereby expediting the time to locate and identify a visual target without necessitating auditory-visual integration. Importantly, while this mechanism considers processes beyond the generation of the first saccade and could explain the evolution of the scanning pattern, it holds that the first saccade to a bimodal stimulus is aurally guided. This mechanism therefore predicts that both the SRT and accuracy of AV-saccades should equal A-saccades, which differs from the observed data. As with the race model, the aurally assisted visual search mechanism cannot explain the improved accuracy of AV-saccades beyond the level typical of A-saccades. Another prediction of this mechanism is that A100V-saccades should be the most accurate and V100A-saccades the most inaccurate. This also differed drastically from the observed data (Fig. 10).

A third explanation of the observed data could be that the auditory system provides a nonlocalized "warning cue" to the subject to initiate the saccade, irrespective of the spatial register of the stimuli (Kingstone and Klein 1993; Ross and Ross 1980, 1981). While this mechanism could explain a partial reduction of SRTs of bimodal saccades (Fig. 11; Table 5), the AV-control experiment demonstrated that spatial alignment of the visual and auditory stimuli was crucial in mediating further improvements in SRT and accuracy, counter to a warning-cue mechanism (Fig. 11; Table 5). It is also hard to imagine how a warning cue mechanism could explain the larger improvements in accuracy at lower S/N ratios (Fig. 9) or why saccade accuracy and SRT varied systematically with the different temporal asynchronies (Fig. 10). While we would have liked to have systematically altered the spatial congruity between the AV-stimuli in this experiment, such an experiment is a major undertaking and is the focus of a separate and ongoing series of experiments.

In conclusion, all three mechanisms assume that bimodal saccades are driven in response to one modality, and therefore predict that their timing and metrics should be identical to either V- or A-saccades. Yet, in none of the 11 different stimulus configurations tested did bimodal saccades have an accuracy-SRT profile identical to V- or A- saccades (Fig. 10). The observed properties of bimodal saccades combine aspects of both A-saccades and V-saccades to achieve the "best-of-both-worlds," and accordingly the most parsimonious expla-
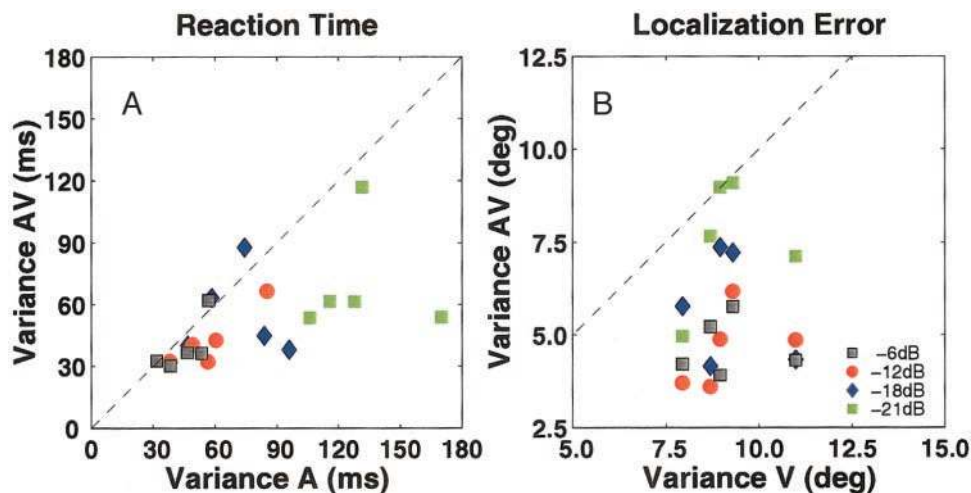

FIG. 7. Comparison of the variances in SRT distributions between A-saccades and AV-saccades (*A*) and of the absolute localization error between V-saccades and AV-saccades (*B*). Data are shown for all 5 subjects. The dashed lines denote the unity line, where the data would cluster if the variances were equal. Note, however, that most data points lie below these lines, indicating that the AV-saccades had narrower distributions.
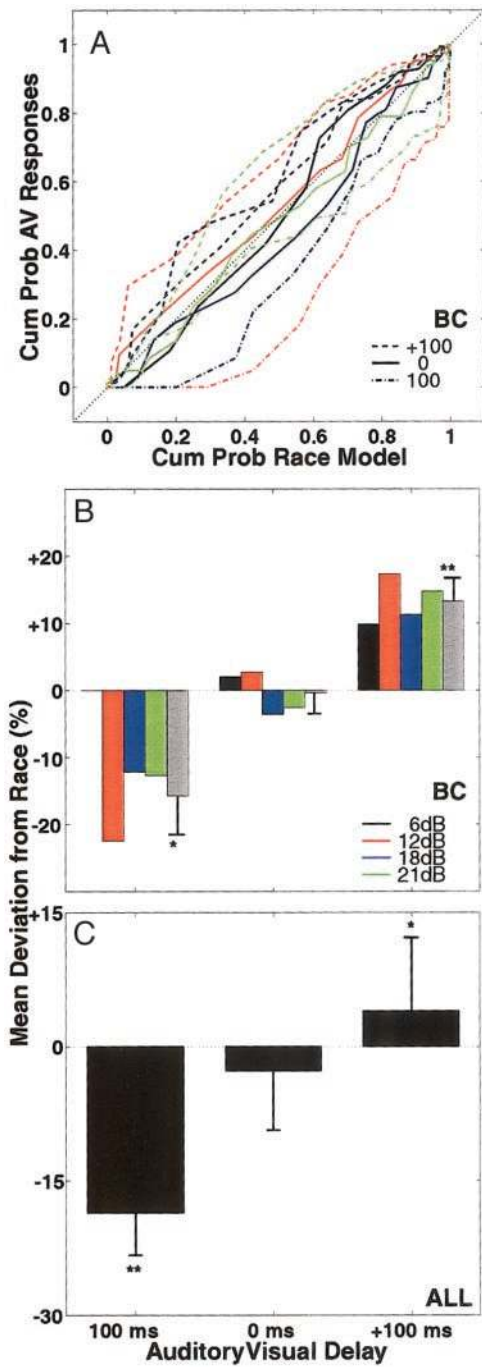
FIG. 8. Comparison of observed SRT data to that predicted by the race model, for bimodal saccades across stimulus asynchrony and S/N ratio. *A*: plot of observed cumulative SRT distribution vs. predicted cumulative SRT distribution for bimodal saccades from *subject BC*. SRTs from AV-saccades (solid lines) matched the predicted SRTs, since the lines lay near the unity line (black dotted line). Observed SRTs for V100A-saccades (dashed lines) surpassed the race model prediction since the data lay above the unity line. Observed SRTs for A100V-saccades (dashed-dotted lines) failed to match the race model prediction since the data lay below the unity line. Different S/N ratios are denoted by the different colors. *B*: quantification, for *subject BC*, of the departure of the observed SRT distributions from the predicted SRT distributions, segregated by stimulus asynchrony. The data were calculated by taking the area between the distribution line and the unity line. Gray bars denote the mean value across all S/N ratios. *C*: summary plot of distribution area averaged across all 5 subjects and all S/N ratios, segregated by asynchrony. Double or single asterisks in *B* and *C* indicate differences that were statistically different from zero at the $P < 0.001$ or $P < 0.05$ levels, respectively. Error bars represent the SD.
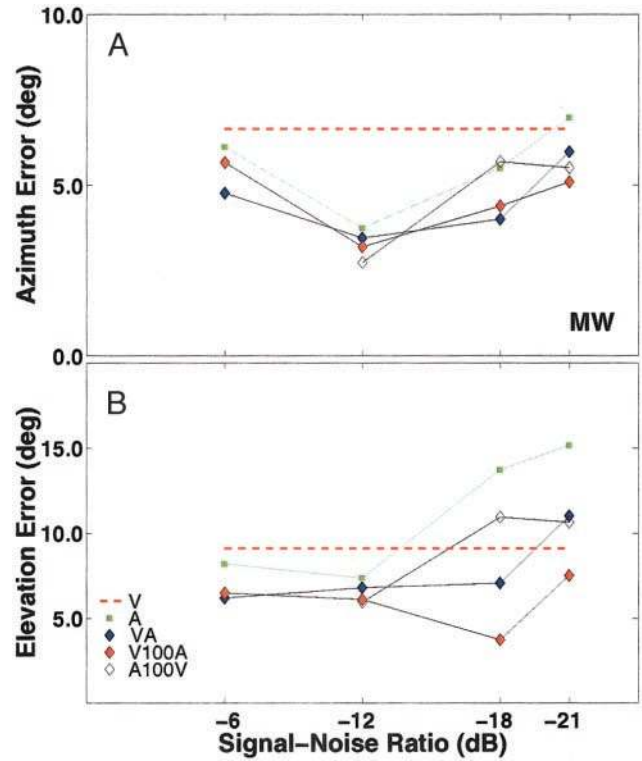

FIG. 9. Absolute azimuth (*A*) and elevation (*B*) localization error as a function of S/N ratio for *subject MW*. Data from different asynchronies are plotted in different series (see legend). Dashed red horizontal lines denote the data from V-saccades.

nation is that auditory and visual stimuli are integrated in a way that depends on their spatial and temporal register.

Rules for multisensory integration of bimodal signals and comparison to previous work

In the intermediate layers of the mammalian SC, many neurons respond to multimodal stimuli (Stein and Meredith 1993). Studies in anesthetized preparations show that the form and magnitude of multisensory interactions in these neurons

TABLE 4. *Statistical summary of probabilities comparing the gains of the azimuth or elevation stimulus-response relationship between A-saccades and bimodal A100V-, AV-, or V100A-saccades*

Response Component	S/N Ratio, dB	P (A100V = A)	P (AV = A)	P (V100A = A)
Azimuth	−6	n.a.	n.s.	n.s.
	−12	n.s.	n.s.	n.s.
	−18	n.s.	n.s.	n.s.
	−21	n.s.	n.s.	n.s.
Elevation	−6	n.a.	n.s.	n.s.
	−12	n.s.	n.s.	*
	−18	n.s.	†	‡
	−21	‡	‡	‡

For each response component, gains for A-saccades and bimodal saccades were subtracted from the gains for V-saccades for each completed series in each subject. This yielded a distribution of 18 data points for each S/N ratio. A 1-D KS test was then applied to compare the various distributions listed in the column headers. n.s., $P > 0.05$; n.a., not available; for other abbreviations, see Tables 2 and 3. * $0.01 < P < 0.05$. † $0.001 < P < 0.01$. ‡ $P < 0.001$.
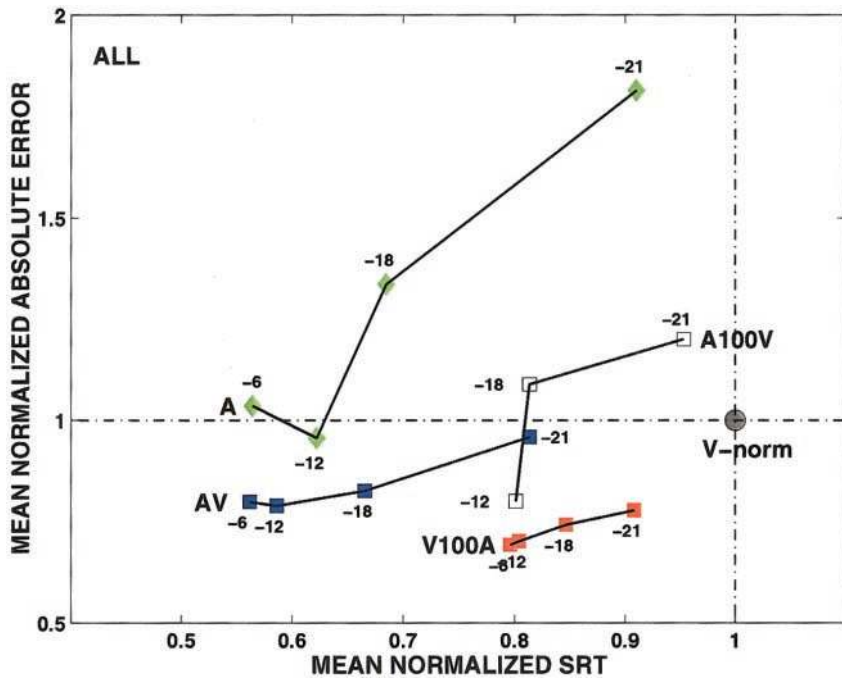
FIG. 10. Summary figure showing the mean normalized absolute localization error vs. the mean SRT averaged across all subjects. Values were first normalized for each subject to the absolute localization error and SRT for V-saccades (large gray circle and dotted-dashed lines), and then averaged across all subjects. Data from A-saccades (green diamonds), A100V-saccades (open squares), AV-saccades (blue squares), and V100A-saccades (red squares) are plotted in different series, with the different S/N ratios denoted on the graph. Note the degradation in SRT and accuracy with decreasing S/N ratio. Also, compare the position of the bimodal saccades series with the temporal asynchronies. The largest improvements in accuracy-SRT performance compared with unimodal saccades were realized at the lower S/N ratios.
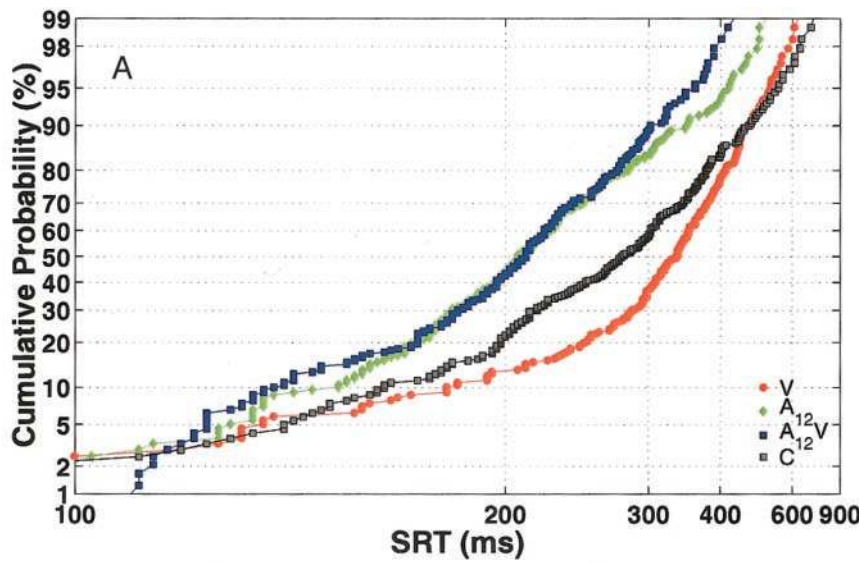

FIG. 11. Comparison of SRT (A) and accuracy (B and C) in the AV-control experiment. Data pooled from all subjects. Same format as Fig. 2. Filled gray squares denote the data from AV-control saccades. Blue squares denote the data for $A_{-12}V$-saccades.
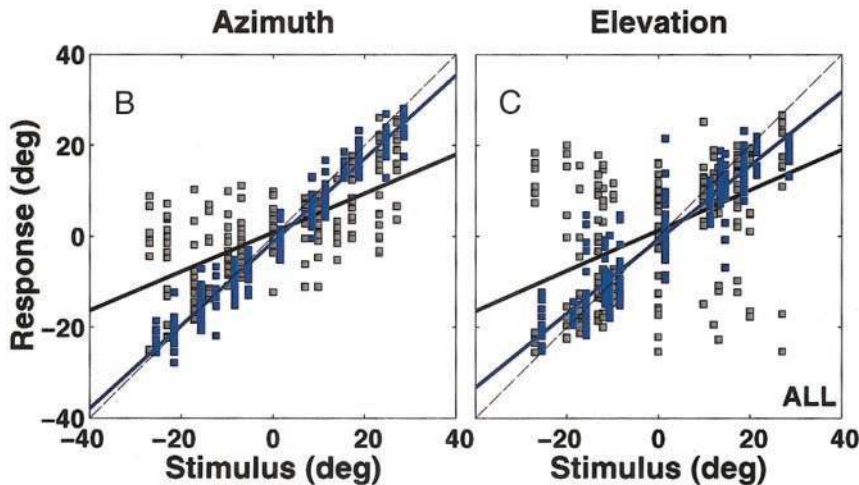
TABLE 5.   *Regression results from AV-control experiment*

Subject	Parameter	Azimuth				Elevation			
		V	A_{-12}	$A_{-12}V$	AV-cont	V	A_{-12}	$A_{-12}V$	AV-cont
BC	Corr	0.91	0.96	0.98	0.69	0.87	0.69	0.89	0.26
	Gain	0.83	0.84	0.90	0.35	0.68	0.47	0.75	0.27
	Bias	−0.4	0.9	1.0	−0.6	−0.2	1.3	−0.2	7.6
	Error	6.1	4.3	3.4	12.0	4.3	11.0	6.2	17.0
DM	Corr	0.97	0.97	0.98	0.89	0.95	0.83	0.91	0.83
	Gain	0.85	0.84	0.89	0.68	0.90	0.58	0.74	0.71
	Bias	0.0	−0.3	0.0	0.0	0.5	1.3	1.1	1.1
	Error	4.2	3.7	3.2	7.3	4.2	8.2	6.0	8.3
JO	Corr	0.82	0.96	0.97	0.15	0.84	0.91	0.94	0.61
	Gain	0.83	1.10	1.02	0.11	0.68	0.70	0.85	0.43
	Bias	−0.1	−2.5	−1.2	3.2	1.0	−0.2	−0.9	4.1
	Error	8.6	4.1	3.4	5.8	6.9	4.5	3.9	8.6
MW	Corr	0.73	0.94	0.98	0.58	0.70	0.88	0.88	0.30
	Gain	0.63	0.93	0.95	0.37	0.60	0.79	0.86	0.29
	Bias	−0.7	1.2	0.6	0.2	1.0	4.7	2.2	0.4
	Error	11.0	4.9	2.9	12.0	11.0	8.7	7.0	16.0
MZ	Corr	0.92	0.92	0.99	0.81	0.71	0.89	0.93	0.53
	Gain	0.87	0.86	0.93	0.64	0.62	0.77	0.87	0.51
	Bias	−1.2	0.8	0.1	1.3	−1.9	2.2	−0.1	−6.8
	Error	5.7	4.8	2.8	9.0	9.9	7.0	4.3	15.0
Median SRT, ms		332 ± 93	206 ± 65	210 ± 56	275 ± 100				

Values in Median SRT are means ± SD. Saccade accuracy was better when the bimodal stimuli were spatially aligned ($A_{-12}V$) than when the auditory target was nonlocalizable (AV-cont). V-saccades had the longest SRTs (bottom row, averaged across subjects). SRT, saccadic reaction time.

depend on the temporal and spatial alignment of the stimuli. Further, SC neurons obey the principle of inverse effectiveness, whereby the magnitude of multisensory interactions are largest when the multisensory stimuli are presented at near-threshold intensities (Meredith and Stein 1986). Studies of SC activity in awake preparations have confirmed these basic rules (Bell et al. 2001; Frens and Van Opstal 1998; Peck 1996; Peck et al. 1995; Wallace et al. 1998). However, linking these rules to behavior is not always straightforward. For example, mean SRTs in humans to high-intensity audio-visual stimuli are typically 10–50 ms shorter than SRTs to unimodal stimuli (Colonius and Arndt 2001; Engelken and Stevens 1989; Frens et al. 1995; Goldring et al. 1996; Harrington and Peck 1998; Hughes et al. 1994, 1998; Munoz and Corneil 1995; Nozawa et al. 1994). Studies with lower intensity stimuli have found that the SRT reductions to low-intensity bimodal stimuli are in the same range (Frens et al. 1995; Hughes et al. 1994; Nozawa et al. 1994), contrary to what would have been predicted given inverse effectiveness (Meredith and Stein 1986). Is it possible that inverse effectiveness is masked by other neural processes operating only in behavioral experiments? If so, could these processes also confound SRT and accuracy of bimodal saccades?

In light of these questions, we highlight several limitations or confounds of previous behavioral studies. First, auditory stimuli have been typically constrained to the horizontal plane, meaning that only sound-source azimuth needed to be extracted. Our extension to the 2-D oculomotor range, as well as the manipulation of the acoustic S/N ratio, provided the opportunity to observe major differences in the sensitivity of azimuth and elevation perception. The ability of the auditory system to extract stimulus elevation degraded at higher S/N ratios than stimulus azimuth (Fig. 4), consistent with recent studies (Good and Gilkey 1996; Zwiers et al. 2001). This effect relates presumably to the different mechanisms the CNS uses to extract sound-source azimuth and elevation from the acoustic cues (see Blauert 1997 for review). Consequently, the accuracy improvements afforded by presenting bimodal targets at low S/N ratios were greater in elevation than in azimuth (Fig. 9, Table 4). Previous studies may have underestimated the contributions of multisensory integration to saccade accuracy by constraining targets to the horizontal plane.

A second limitation of previous studies is the use of a limited number of potential target positions. This could allow subjects to use prior knowledge about potential target positions to prepare movements before target presentation, which, if left unaccounted for, would also lead to underestimations of the contributions of multisensory integration to saccade accuracy. The present setup used 24 potential target locations, making such a strategy highly unlikely.

Third, subject instructions and experimental context affect the temporal expression of multisensory integration (i.e., SRT). For example, requiring a subject to orient specifically to one modality while ignoring the other yields SRT distributions that violate the race model when the instructed target is visual, but not when the instructed target is auditory (Corneil and Munoz 1996; Hughes et al. 1994). In general, requiring subjects to discriminate between modalities prolongs SRTs (Corneil and Munoz 1996) and could confound the estimation of the contributions of multisensory integration to SRT. In the present experiments, subjects could orient to both the auditory and visual stimulus, so this was not a concern. Overall, the setup employed in our experiments allows for a behavioral assessment of the consequences of multisensory integration over both spatial and temporal domains, while being removed from con-

founds, such as the three discussed here, that affected the interpretation of previous studies.

A few behavioral studies have manipulated the temporal alignment between auditory and visual stimuli to address the temporal window over which stimuli may interact (Colonius and Arndt 2001; Corneil and Munoz 1996; Engelken and Stevens 1989; Frens et al. 1995). For the saccadic system, the temporal window is about ±100 ms, presumably allowing AV-integration in spite of differences in retinal versus cochlear transduction times (∼50 ms and 2–10 ms, respectively) (Gouras 1967; Kraus and McGee 1992) and the speed of sound versus light over a large range of stimulus distances. In the SC of awake, behaving primates, auditory response latencies usually range around 30 ms (Bell et al. 2001; Jay and Sparks 1987) and visual response latencies around 60 ms (see Munoz et al. 2000 for review), suggesting that the more complex transformation of auditory responses into oculocentric coordinates does not greatly affect the relative arrival times at the SC. As shown in Fig. 10, the combination of SRT and accuracy of AV-saccades surpassed that for either unimodal A- and V-saccades, and we argued above that a race model could not account for these data. However, Fig. 10 also shows that the temporal window permitting excitatory interactions is not symmetrical around synchronously presented stimuli. V100A-saccades were initiated at SRTs that surpassed the race model prediction and were more accurate than any other saccade type. Conversely, A100V-saccades were initiated at SRTs that fell well short of the race model prediction (and were even slower than A-saccades) and were more inaccurate than AV-saccades and V-saccades at low S/N ratios (but were still more accurate than A-saccades). These findings suggest a nonlinearity in the interactions of delayed visual or auditory signals. Apparently, a delayed auditory signal facilitates saccade initiation and sharpens the accuracy of a developing visually guided saccade, but a delayed visual signal inhibits saccade initiation and worsens the accuracy of a developing aurally guided saccade. Although surprising, this nonlinearity bears some resemblance to multisensory recordings in the SC of anesthetized cats, wherein response enhancements are observed if the visual stimulus leads the auditory stimulus but response depressions are observed if the auditory stimulus leads the visual stimulus (Meredith et al. 1987). Understanding the neural mechanism(s) responsible for this nonlinearity requires neuronal recordings from awake, behaving preparations.

Conceptual model of auditory-visual interactions in a complex scene

Figure 12 presents a conceptual model to explain how activity within the SC might evolve prior to A-, V-, and AV-saccades. We assume that visual and auditory signals are initially processed separately and converge on the SC, inducing modality-specific profiles of SC activity. At high intensities, aurally induced profiles arrive earlier than visually induced profiles, but with lower firing rates and a broader tuning (dashed lines and empty profiles in Fig. 12, B and C) (Bell et al. 2001; Frens and Van Opstal 1998; Jay and Sparks 1987; Peck et al. 1995; Wallace et al. 1996, 1998). These profiles continue to develop until a threshold is exceeded, here modeled by an integrated number of spikes. Achieving threshold silences the activity of omnipause neurons (OPNs), permitting
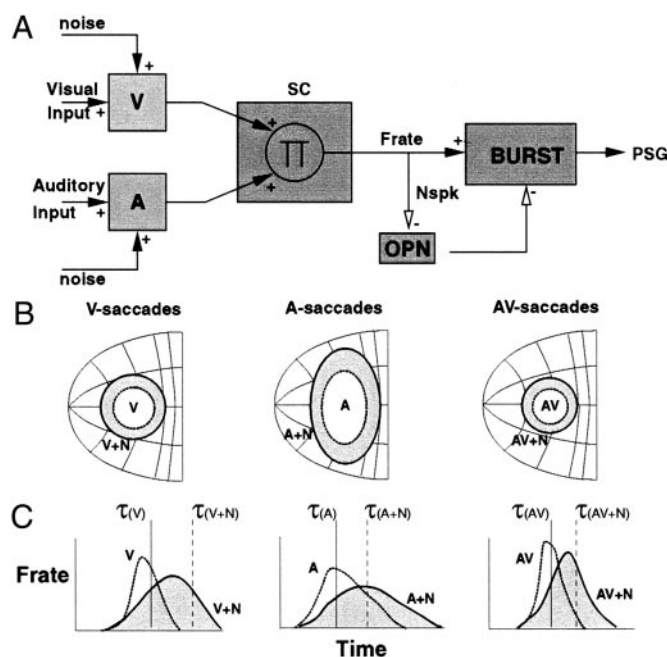

FIG. 12. A: conceptual model of AV-interactions in the superior colliculus (SC). Auditory and visual information is initially processed separately. The signals are nonlinearly integrated in the SC (Π), and relayed to the downstream circuitry (OPn, onmipause neurons; burst, burst generator; PSG, pulse step generator) when a saccade threshold is reached. B and C: sketch predicted spatial and temporal profiles of SC activity, respectively, prior to the generation of V-, A-, or AV-saccades, either in the presence of noise (indicated by +N; solid lines and shaded profiles) or not (dashed lines and empty profiles). In B, the width of the profiles determines the accuracy of the responses (e.g., note the ellipsoid shape for A-saccades, indicating more variable responses in elevation). In C, we assume that saccadic threshold is reached when an integrated number of spikes is exceeded, here denoted at time τ. Varying levels of noise (depending either on S/N ratio and/or the number of distractors) are assumed to reduce SC firing rates and smear the population activity in both the auditory and visual channels, resulting in prolonged SRTs and more inaccurate saccades. Provided SC activity induced by the auditory and visual channels overlap temporally and spatially, nonlinear interactions within the SC sharpens the population profile, leading to increased accuracy. See DISCUSSION for further details.

the activation of the burst and pulse-step generators that results in saccade generation. Obviously, SRT relates to the time the SC threshold is surpassed (denoted by time τ in Fig. 12C). We assume that saccade accuracy is determined by the center of gravity of the SC activity profile at this time, so that, over multiple trials, a sharper profile leads to more accurate saccades. Although speculative, these assumptions are consistent with the premotor processing prior to visually or aurally guided saccades to single, high-intensity stimuli (see Findlay and Walker 1999; Munoz et al. 2000 for review).

To our knowledge, there are no data from behaving animals on the SC activity patterns in the presence of the AV-background that address the effects of manipulations of S/N ratio and stimulus asynchronies. Our conceptual model makes predictions about these activity profiles that could be readily tested in future investigations. In the presence of the AV-background or low S/N ratios, unimodal V- or A-saccades have longer SRTs and/or are more inaccurate, respectively (Figs. 2–4; Tables 1 and 2), presumably related to the introduction of "noise" from competitive interactions within or prior to the SC. As a result, unimodal activity profiles within the SC are broader (shaded shapes in Fig. 12B) and take longer to achieve

threshold (Fig. 12*C*). In such a "noisy" environment, the pairing of AV-stimuli permits nonlinear excitatory interactions that sharpen and increase the firing rate of the SC activity profile, culminating in more accurate saccades generated at shorter and more consistent SRTs. However, such excitatory interactions can only occur over restricted spatial and temporal windows. Inhibitory interactions are observed if the bimodal stimuli are not aligned in space (AV-control experiment, Fig. 11 and Table 5), or time (i.e., for A100V-saccades, Fig. 10). While such interactions could be mediated by an intrinsic inhibitory network within the SC (Kadunce et al. 1997; Meredith and Ramoa 1998; Munoz and Istvan 1998), determining the exact mechanism(s) requires recording in awake, behaving preparations.

Conclusions

The experimental architecture presented in this paper provides a novel and illuminating way to investigate the behavioral significance of multisensory integration in the saccadic system. The complexity of the background and the manipulations of the S/N ratio and temporal register of the stimuli present a formidable challenge to the saccadic system. Importantly, such features mimic those occurring in everyday life, wherein auditory-visual stimuli may be presented over a wide range of distances from the subject or embedded within a complex auditory-visual background. Our results demonstrate the importance of multisensory integration to facilitate orienting in a complex quasi-natural environment.

REFERENCES

Bell AH, Corneil BD, Meredith MA, and Munoz DP. The influence of stimulus properties on multisensory processing in the awake primate superior colliculus. *Can J Exp Psychol* 55: 123–132, 2001.

Blauert J. *Spatial Hearing: The Psychophysics of Human Sound Localization.* Cambridge, MA: MIT, 1997.

Carpenter RH and Williams ML. Neural computation of log likelihood in control of saccadic eye movements. *Nature* 377: 59–62, 1995.

Collewijn H, Van Der Mark F, and Jansen TC. Precise recording of human eye movements. *Vision Res* 15: 447–450, 1975.

Colonius H and Arndt P. A two-stage model for visual-auditory interaction in saccadic latencies. *Percept Psychophys* 63: 126–147, 2001.

Corneil BD and Munoz DP. The influence of auditory and visual distractors on human orienting gaze shifts. *J Neurosci* 16: 8193–8207, 1996.

Corneil BD, Van Wanrooij M, Munoz DP, and Van Opstal AJ. Auditory-visual interactions subserving goal directed saccades in a complex scene. *Soc Neurosci Abstr* 71: 26, 2001.

Engelken EJ and Stevens KW. Saccadic eye movements in response to visual, auditory and bisensory stimuli. *Aviat Space Environ Med* 60: 762–768, 1989.

Findlay JM and Walker R. A model of saccade generation based on parallel processing and competitive inhibition. *Behav Brain Sci* 22: 674–721, 1999.

Fischer B and Weber H. Express saccades and visual attention. *Behav Brain Sci* 16: 553–610, 1993.

Frens MA and Van Opstal AJ. A quantitative study of auditory- evoked saccadic eye movements in two dimensions. *Exp Brain Res* 107: 103–117, 1995.

Frens MA and Van Opstal AJ. Visual-auditory interactions modulate saccade-related activity in monkey superior colliculus. *Brain Res Bull* 46: 211–224, 1998.

Frens MA, Van Opstal AJ, and Van Der Willigen RF. Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept Psychophys* 57: 802–816, 1995.

Gielen CCAM, Schmidt RA, and Van den Heuvel PJM. On the nature of intersensory facilitation of reaction time. *Percept Psychophys* 34: 161–168, 1983.

Goldring JE, Dorris MC, Corneil BD, Ballantyne PA, and Munoz DP. Combined eye-head gaze shifts to visual and auditory targets in humans. *Exp Brain Res* 111: 68–78, 1996.

Good MD and Gilkey RH. Sound localization in noise: the effect of signal-to-noise ratio. *J Acoust Soc Am* 99: 1108–1117, 1996.

Goossens HH and Van Opstal AJ. Human eye-head coordination in two dimensions under different sensorimotor conditions. *Exp Brain Res* 114: 542–560, 1997.

Gouras P. The effects of light-adaptation on rod and cone receptive field organization of monkey ganglion cells. *J Physiol (Lond)* 192: 747–760, 1967.

Harrington LK and Peck CK. Spatial disparity affects visual-auditory interactions in human sensorimotor processing. *Exp Brain Res* 122: 247–252, 1998.

Hofman PM and Van Opstal AJ. Spectro-temporal factors in two-dimensional human sound localization. *J Acoust Soc Am* 103: 2634–2648, 1998.

Hughes HC, Nelson MD, and Aronchick DM. Spatial characteristics of visual-auditory summation in human saccades. *Vision Res* 38: 3955–3963, 1998.

Hughes HC, Reuter-Lorenz PA, Nozawa G, and Fendrich R. Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *J Exp Psychol Hum Percept Perform* 20: 131–153, 1994.

Irvine DRF. *Progress in Sensory Physiology. The Auditory Brainstem,* edited by Ottoson D. Berlin: Springer Verlag, 1986, vol. 7.

Jay MF and Sparks DL. Sensorimotor integration in the primate superior colliculus. I. Motor convergence. *J Neurophysiol* 57: 22–34, 1987.

Kadunce DC, Vaughan JW, Wallace MT, Benedek G, and Stein BE. Mechanisms of within- and cross-modality suppression in the superior colliculus. *J Neurophysiol* 78: 2834–2847, 1997.

Kingstone A and Klein RM. Visual offsets facilitate saccadic latency: does predisengagement of visuospatial attention mediate this gap effect? *J Exp Psychol Hum Percept Perform* 19: 1251–1265, 1993.

Kraus N and McGee T. Electrophysiology of the human auditory system. In: *The Mammalian Auditory Pathway: Neurophysiology,* edited by Popper AV and Fay RR. New York: Springer, 1992, p. 335–403.

Meredith MA, Nemitz JW, and Stein BE. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 10: 3215–3229, 1987.

Meredith MA and Ramoa AS. Intrinsic circuitry of the superior colliculus: pharmacophysiological indentification of horizontally orienting inhibitory interneurons. *J Neurophysiol* 79: 1597–1602, 1998.

Meredith MA and Stein BE. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol* 56: 640–662, 1986.

Miller JO. Divided attention: evidence for coactivation with redundant signals. *Cog Psychol* 14: 247–279, 1982.

Munoz DP and Corneil BD. Evidence for interactions between target selection and visual fixation for saccade generation in humans. *Exp Brain Res* 103: 168–173, 1995.

Munoz DP, Dorris MC, Paré M, and Everling S. On your mark, get set: brainstem circuitry underlying saccadic initiation. *Can J Physiol Pharmacol* 78: 934–944, 2000.

Munoz DP and Istvan PJ. Lateral inhibitory interactions in the intermediate layers of the monkey superior colliculus. *J Neurophysiol* 79: 1193–1209, 1998.

Nozawa G, Reuter-Lorenz PA, and Hughes HC. Parallel and serial processes in the human oculomotor system: bimodal integration and express saccades. *Biol Cybern* 72: 19–34, 1994.

Peck CK. Visual-auditory integration in the cat superior colliculus: implications for neuronal control of the orienting response. *Prog Brain Res* 112: 167–277, 1996.

PECK CK, BARO JA, AND WARDER SM. Effects of eye position on saccadic eye movements and on the neuronal responses to auditory and visual stimuli in cat superior colliculus. *Exp Brain Res* 103: 227–242, 1995.

PERROTT DR, SABERI K, BROWN K, AND STRYBEL TZ. Auditory psychomotor coordination and visual search performance. *Percept Psychophys* 48: 214–226, 1990.

PERROTT DR, SADRALODABAI T, SABERI K, AND STRYBEL TZ. Aurally aided visual search in the central visual field: effects of visual load and visual enhancement of the target. *Hum Factors* 33: 389–400, 1991.

PRESS WH, FLANNERY BP, TEUKOLSKY SA, AND VETTERING WT. *Numerical Recipes in C* (2nd ed.). Cambridge, MA: Cambridge, 1992.

RAAB DH. Statistical facilitation of simple reaction times. *Trans NY Acad Sci* 24: 574–590, 1962.

ROSS LE AND ROSS SM. Saccade latency and warning signals: stimulus onset, offset and change as warning events. *Percept Psychophys* 27: 251–257, 1980.

ROSS SM AND ROSS LE. Saccade latency and warning signals: effects of auditory and visual onset and offset. *Percept Psychophys* 29: 429–437, 1981.

SPARKS DL AND MAYS LE. Signal transformations required for the generation of saccadic eye movements. *Annu Rev Neurosci* 13: 309–336, 1990.

STEIN BE AND MEREDITH MA. *The Merging of the Senses*. Cambridge, MA: MIT, 1993.

STEIN BE, MEREDITH MA, HUNEYCUTT WS, AND MCDADE L. Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *J Cog Neurosci* 1: 12–24, 1989.

VAN WANROOIJ M, CORNEIL BD, MUNOZ DP, AND VAN OPSTAL AJ. Auditory-visual integration in human saccades elicited in a noisy environment (Abstract). *2nd Annual Multisensory Research Conference,* 2000, p. 56.

WALLACE MT, MEREDITH MA, AND STEIN BE. Multisensory integration in the superior colliculus of the alert cat. *J Neurophysiol* 80: 1006–1010, 1998.

WALLACE MT, WILKINSON LK, AND STEIN BE. Representation and integration of multiple sensory inputs in primate superior colliculus. *J Neurophysiol* 76: 1246–1266, 1996.

ZWIERS MP, VAN OPSTAL AJ, AND CRUYSBERG JR. A spatial hearing deficit in early-blind humans. *J Neurosci* 21: 1–5, 2001.