

Spatial and temporal factors determine auditory–visual interactions in human saccadic eye movements

M. A. FRENS, A. J. VAN OPSTAL, and R. F. VAN DER WILLIGEN
University of Nijmegen, Nijmegen, The Netherlands

In this paper, we show that human saccadic eye movements toward a visual target are generated with a reduced latency when this target is spatially and temporally aligned with an irrelevant auditory nontarget. This effect gradually disappears if the temporal and/or spatial alignment of the visual and auditory stimuli are changed. When subjects are able to accurately localize the auditory stimulus in two dimensions, the spatial dependence of the reduction in latency depends on the actual radial distance between the auditory and the visual stimulus. If, however, only the azimuth of the sound source can be determined by the subjects, the horizontal target separation determines the strength of the interaction. Neither saccade accuracy nor saccade kinematics were affected in these paradigms. We propose that, in addition to an aspecific warning signal, the reduction of saccadic latency is due to interactions that take place at a multimodal stage of saccade programming, where the *perceived* positions of visual and auditory stimuli are represented in a common frame of reference. This hypothesis is in agreement with our finding that the saccades often are initially directed to the average position of the visual and the auditory target, provided that their spatial separation is not too large. Striking similarities with electrophysiological findings on multisensory interactions in the deep layers of the midbrain superior colliculus are discussed.

Humans, as well as other animals, are equipped with various specialized senses that provide them with information about their environment. Several of these sensory systems represent the spatial location of an object on the basis of the received sensory input. This information about stimulus location can already be present at the level of the sensory organ, as is the case in the visual and somatosensory systems, or it can be neurally derived on the basis of indirect cues, as in the auditory system. Many of the objects that surround an organism, however, provide it with sensory information through various modalities at the same time.

In the literature, there is accumulating evidence that multimodal information about an object's location can lead to a reduction of the response latency and to an improvement of localization accuracy. For example, it has been shown that a motor response toward a visual target can be made with a shorter latency when this target is accompanied by an auditory signal at the same location. Simon and Craft (1970) have investigated this effect for arm movements, and Lee, Chung, Kim, and Park (1991) report qualitatively similar findings for saccadic eye move-

ments. In both studies, this effect was not present when the visual and the auditory stimulus were presented at opposite sides of a central fixation point. Perrott, Saberi, Brown, and Strybel (1990) showed that the time to foveate and identify one of two visual symbols was markedly reduced when an auditory costimulus was spatially aligned with the visual target, not only if the targets were presented in the far periphery but even if the stimuli were presented within the subject's parafoveal visual field.

Stein, Hunnecutt, and Meredith (1988) reported that, under near-threshold conditions, cats are able to localize combined audiovisual targets more accurately than visual targets. In their study, cats were trained to make whole body movements toward dimly lit visual targets, thereby learning that the presence of an auditory stimulus was irrelevant. Nevertheless, the animals performed better when an audiovisual cue was presented in spatial alignment. Performance dropped dramatically when the auditory and the visual stimulus were spatially disparate.

It should be noted that such auditory–visual interactions pose far from trivial problems to the nervous system, since the different sense organs initially encode the outside world in very different ways. First, the visual world is encoded *retinocentrically*, whereas auditory cues are represented with respect to the pinnae. For humans, this results in a *craniocentric* code, since the pinnae are immobile with respect to the head. Second, the retina is a spatially organized structure, whereas the cochlea has a *tonotopic* organization. From various bin-

This work was supported by the University of Nijmegen and the Mucom II program (6615) of the European ESPRIT initiative (M.A.F. and A.J.V.O.). We acknowledge the valuable technical assistance of H. Kleijnen and T. Van Dreumel. Correspondence should be addressed to M. A. Frens, Department of Medical Physics and Biophysics, University of Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands (e-mail: maarten@mbfys.kun.nl).

aural and monaural cues, present in the acoustical signal at the eardrums, the azimuth and elevation of a sound source appear to be derived through separate neural pathways, involving binaural and monaural processes (see Blauert, 1983; Irvine, 1986). It is generally recognized that in order to respond to a multimodal (e.g., an audiovisual) target, the different modalities must, at some stage in the neural programming, be merged into a single frame of reference (e.g., Stein & Meredith, 1993).

We wondered which rules underlie the generation of multimodally evoked targeting movements in human subjects. We have chosen the saccadic eye movement system as a model system to study this problem, since it is a very precise natural orienting system in primates and much has been learned about the neural pathways that are involved in these movements. We therefore have investigated whether gradual changes in audiovisual spatial and temporal alignment result in systematic changes of relevant saccade parameters.

Some of the results described in this paper have been presented previously in abstract form (Van Opstal, Frens, & Van der Willigen, 1993).

EXPERIMENT 1

Experiment 1 was designed to study the spatial factors that rule multimodal interactions in saccadic eye movements.

Method

Subjects

Three male volunteers, 24, 27, and 36 years of age, participated in this experiment. All subjects were without any known uncorrected visual, auditory, or oculomotor deficits, except Subject J.O., who has a strong dominance of one eye and is basically monocular. From this subject, movements of the amblyopic eye were measured.

Experimental Setup

Experiments were performed in a completely dark, sound-attenuated room ($3 \times 3 \times 3$ m) in which acoustic reflections above 500 Hz were strongly reduced by means of sound-absorbing foam that covered walls, ceiling, and floor. The average background noise level was 30 dB (SPL). The subjects were comfortably seated in a chair with a head support that prevented them from making head movements. Stimulus presentation as well as data acquisition were controlled by a PC-386 equipped with a data acquisition board (Metrabyte Das16).

Auditory noise stimuli were generated by a white-noise generator (Hewlett-Packard HOI-3722a), band-pass filtered (150–20 kHz, Krohn-Hite 3343), amplified (Luxman 58A), and presented through a speaker (Philips, AD44725). The speaker was mounted on a two-joint robot arm that was controlled by a second computer (PC-486). The robot enabled rapid positioning of the speaker anywhere on the surface of a virtual sphere with a radius of 0.90 m, centered at the subject's head. Between trials, the speaker was first moved to a randomly chosen peripheral location. In this way, sounds produced by the two stepping motors did not provide the subjects with any cues about the speaker's position.

Visual targets were red LEDs (radius 0.3°), mounted on an acoustically transparent wire frame, which constituted a spherical surface just proximal to the range of the robot ($r = 0.85$ m). Viewing was binocular.

Movements of the right eye were measured in two dimensions by means of the scleral coil technique (Collewijn, Van der Mark, & Jansen, 1975). In short, the subject was seated in a rapidly oscillating horizontal and vertical magnetic field (30 and 50 kHz), generated through two orthogonal coils (3×3 m). The coils did not obstruct the visual field of the subjects, nor did they disturb the sound field.

A scleral search coil (Skalar Instruments, Delft) was placed on the subject's right eye. The magnetic induction voltage in this scleral coil was directly proportional to its orientation with respect to the magnetic fields. In this way, eye position could be accurately measured with a resolution of about 0.25° in all directions.

In order to decompose the signal from the scleral coil into a horizontal and a vertical component of eye position, it was passed through two phase-lock amplifiers (PAR 128A) that used the driving signals for the horizontal and the vertical field, respectively, as a reference signal. The resulting position signals were then low-pass filtered at 150 Hz, before being collected by the data acquisition board. The sampling rate was 500 Hz for both the horizontal and the vertical components of eye position. Each trial consisted of 2 sec of recording time, starting 400 msec before presentation of the peripheral stimuli.

To calibrate the recorded eye movements, we asked our subjects to foveate visual targets on the horizontal axis and on the vertical axis at eccentricities of 2° , 5° , 9° , 14° , 20° , 27° , and 35° from straight ahead. The signals that were thus obtained were used to calculate off-line linear regression lines between the target coordinates and horizontal and vertical eye position signals. This method provided accurate calibration for all directions.

Experimental Protocol

The subjects were first required to foveate a visual fixation spot straight ahead. After a random period of 0.5 to 2.0 sec, the fixation spot was extinguished and a visual target was presented in the periphery. Synchronous with the onset of the visual target, an auditory stimulus was presented in 80% of the trials. Duration of each stimulus was 500 msec. The subjects were instructed to redirect their gaze as quickly and accurately as possible toward the visual target and were explicitly told to ignore the auditory nontarget.

One of four different visual targets was presented at spherical polar coordinates $R = 27^\circ$, and $\phi \in [60^\circ, 120^\circ, 240^\circ, 300^\circ]$. In this coordinate system, R is the distance from the central fixation spot and ϕ is the direction of the target, where $\phi = 0^\circ$ is to the right and $\phi = 90^\circ$ is upward (Figure 1A). Possible positions of a synchronous auditory stimulus were at these same locations. Thus, combined visual and auditory stimuli could be presented in one of four spatial configurations: (1) spatially coincident (*coincident*, for short), (2) diametrically opposed to each other with respect to the fixation spot (*opposite*), (3) horizontally aligned but vertically opposite (*horizontally aligned*), and (4) vertically aligned but horizontally opposite (*vertically aligned*).

During each experiment, all possible visual/auditory stimulus combinations were presented in random order, randomly interleaved with visual-only trials, in which the auditory stimulus was not presented (20% of the trials). Each of the 20 different stimulus configurations was presented at least 8, and—if time permitted—12, times in one experimental session.

In separate experimental sessions, the intensities of the visual and auditory noise stimuli were set either at $0.15 \text{ cd} \cdot \text{m}^{-2}$ (measured with the Minolta LS 100 luminance meter) and 70 dB (SPL) (here denoted as *high intensities*) or at $0.015 \text{ cd} \cdot \text{m}^{-2}$ and 45 dB (SPL) (*low intensities*), respectively. Note that all intensities were well above detection threshold.

In a recent study (Frens & Van Opstal, 1994), we have shown that the composition of the auditory spectrum has a strong influence on a sound's localizability. Therefore, in two separate high-intensity experimental sessions, we also selected a different spectral content of the auditory stimulus. In one session, the spectrum

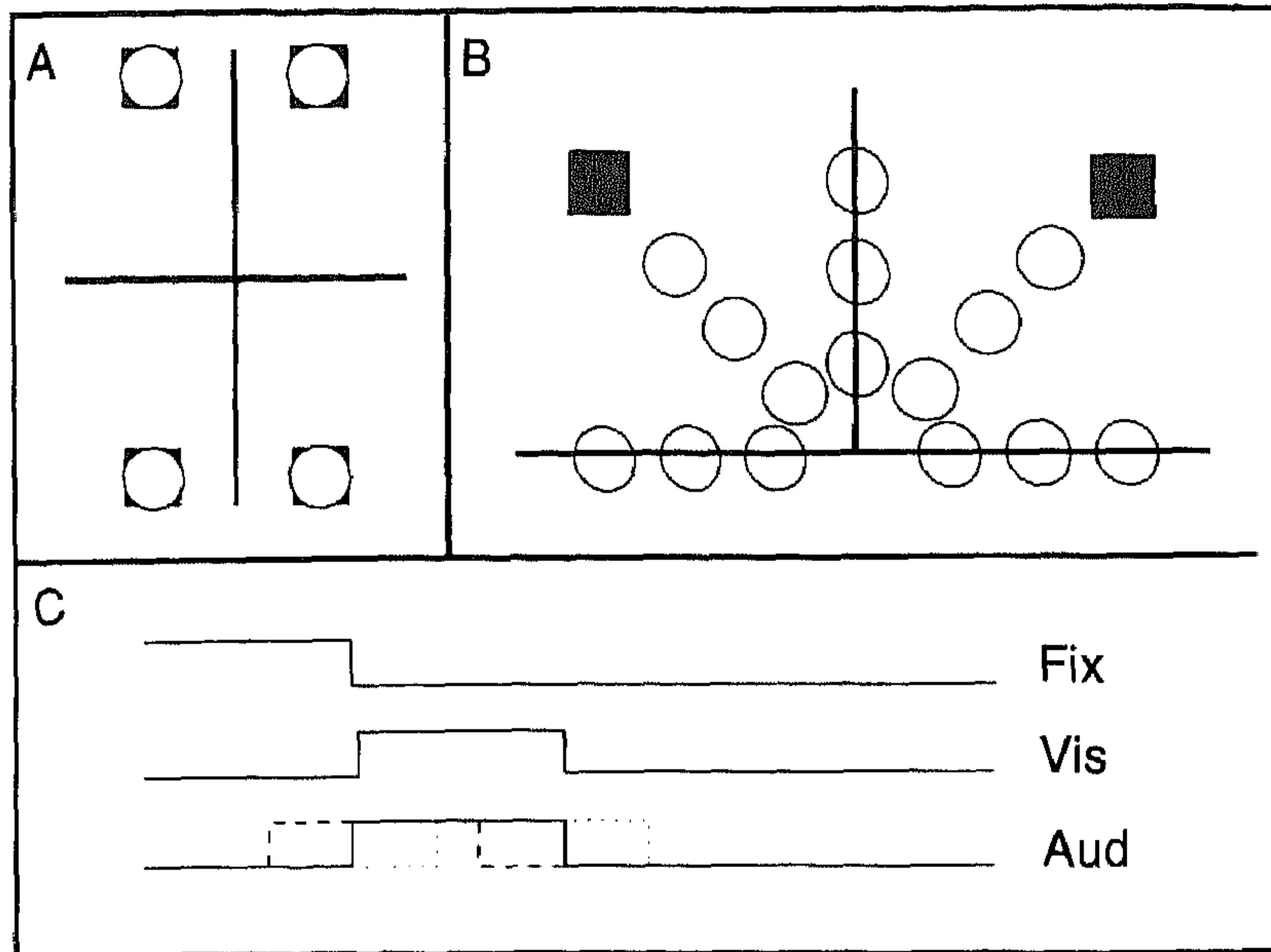


Figure 1. Schematic stimulus configuration of the three experiments described in this paper. (A) Experiment 1. Visual targets are represented as open circles; auditory stimuli are represented as filled squares. Note that the possible positions of visual and auditory stimuli coincide, but that each visual target could be presented in combination with any one of the four auditory stimuli or as a unimodal stimulus. (B) Experiment 2. A larger set of visual target positions was chosen in combination with one of two possible auditory stimuli. (C) Experiment 3. In all experiments, one of two possible visual targets was lit as soon as the central fixation spot was extinguished. The onset of the auditory stimulus (at either the same or the opposite position) could appear within a range of -50 (dashed signal) to $+100$ msec (dotted signal) with respect to the visual target onset. Neither timing nor target sizes have been drawn to scale.

was broad-band noise (see above); in the second session, it was a sharply peaked harmonic spectrum, having its most prominent component at 700 Hz (58 dB). Higher harmonics in this signal had an intensity that was at least 20 dB lower. This stimulus will be referred to as the *tone stimulus* in the rest of this paper.

At the end of an experimental session, the applied auditory stimuli were presented in a separate series, in which they served as targets. This auditory experiment aimed at measuring the latency, accuracy, and kinematic properties of auditory-evoked eye movements of our subjects. If time permitted, auditory-evoked saccades were also measured toward stimuli that were presented for 500 msec at random positions throughout the oculomotor range. These experiments served to assess the ability of our subjects to accurately localize acoustic targets.

Data Analysis

From the calibrated eye position signal, the onset and offset of saccadic eye movements were detected by the computer on the basis of velocity and mean acceleration criteria. All detection markings were visually checked by the experimenters. Subsequently, saccadic latency L (defined as the time interval between target onset and saccade onset, in milliseconds), overall saccade direction ϕ (in degrees), amplitude R (in degrees), and maximum velocity V_{\max} (in degrees/second) were determined from the calibrated eye position signals. Trials in which the primary saccade had a reaction time outside the 100–300-msec interval were discarded from further analysis. For the analysis of latencies, saccades with a direction that deviated more than 30° from the direction of the visual target were excluded.

The radial distance, ΔR , between the visual stimulus position, \vec{V} , and the auditory stimulus, \vec{A} , was defined as

$$\Delta R = \sqrt{(V_h - A_h)^2 + (V_v - A_v)^2}, \quad (1)$$

in which V_h , V_v , A_h , and A_v are the horizontal and vertical coordinates of the visual and the auditory target positions (in degrees), relative to the straight ahead fixation point.

Results

Effect on Saccade Trajectories and Kinematics

When the subjects made saccades toward well-lit unimodal visual targets, the trajectories of the movements were approximately straight in all four directions. When high-intensity noise stimuli were presented in combination with the visual targets, the trajectories of the saccades did not change with respect to the visually elicited saccades in any of our subjects. However, when the intensity of both stimuli was decreased (see Method section), the saccade trajectories depended strongly on the spatial configuration of the visual and auditory stimuli for 2 of the 3 subjects (M.F. and J.O.) in the following way (see Figure 2, for data from Subject J.O.): When the auditory noise stimulus was spatially coincident with the visual target, no systematic change with respect to the unimodal visual condition was obtained in the sac-

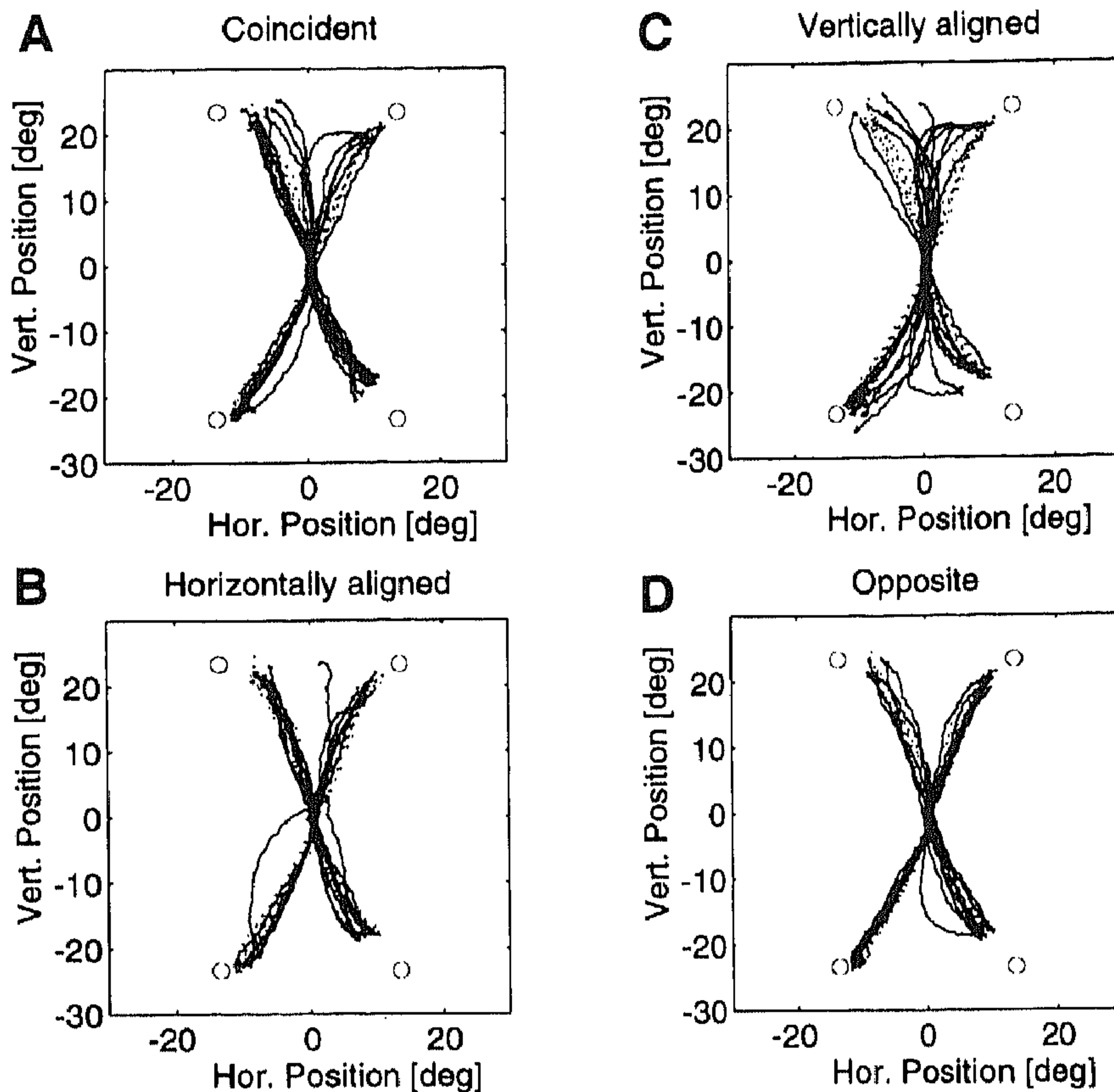


Figure 2. Experiment 1: Primary saccade trajectories (solid lines) of Subject J.O. under low-intensity bimodal stimulation. (A) Spatially coincident visual and auditory noise stimuli. (B) Horizontally aligned stimuli. (C) Vertically aligned stimuli. (D) Oppositely positioned stimuli. Open circles indicate the visual target positions (see Figure 1). For reference, in all panels, the trajectories of the unimodal visually elicited saccades are indicated by dotted lines. Note that all primary saccades undershoot the target position.

cade trajectories (Figure 2A). However, when the visual and auditory stimuli were presented vertically aligned ($\Delta\phi = 60^\circ$), the saccades typically started in a direction that was between the two stimuli. Subsequently, the movement curved in midflight toward the visual stimulus (Figure 2C). Within this population of saccades, no significant correlation between the initial direction and the latency of the responses was found ($p > .05$). Increasing the angle between the two stimuli to 120° (horizontally aligned) or 180° (opposite) resulted in straight saccades that were correctly directed toward the visual stimulus (Figures 2B and 2D). Whenever the trajectory of the saccade was not changed by the presence of the auditory target, the velocity profile and the duration of the movements were always indistinguishable from the unimodal visually driven saccades. The trajectories of Subject J.G. were straight and goal-directed under all conditions.

Effect on Latency

High-intensity stimuli. The reaction time results of this experiment are summarized in Figure 3. In Figure 3A, saccadic latencies of a representative subject (M.F.) are shown for visual targets in combination with a noise stimulus. Both stimuli have the highest applied

intensities (see Method section). One can see that the presentation of a spatially coincident auditory stimulus ("Coine" column) reduces the latency of the response to the visual targets by about 50 msec ($p < .01$) with respect to the values found for purely visual stimuli. Note that this figure shows the pooled data of the four visual targets. This pooling was allowed since no significant differences were obtained for the latency distributions of responses toward the four visual targets ($p > .1$) in any of the spatially similar multimodal configurations.

Increasing the distance of the auditory target with respect to the visual target reduces this latency facilitation. In the most extreme case tested, in which the auditory target was positioned opposite to the visual target at a distance $\Delta R = 54^\circ$ ("Opposite" column), the effect was absent. This reduction of facilitation was significant for all subjects tested. Spearman's rank order correlation coefficients (Press, Flannery, Teukolsky, & Vetterling, 1992) between ΔR and saccade latency for all 3 subjects are given in Table 1.

For all subjects, the latency distribution of coincident audiovisual targets (second column from left in Figure 3A) was similar to the latencies of unimodal auditory-evoked saccades (right column), which could, at least in

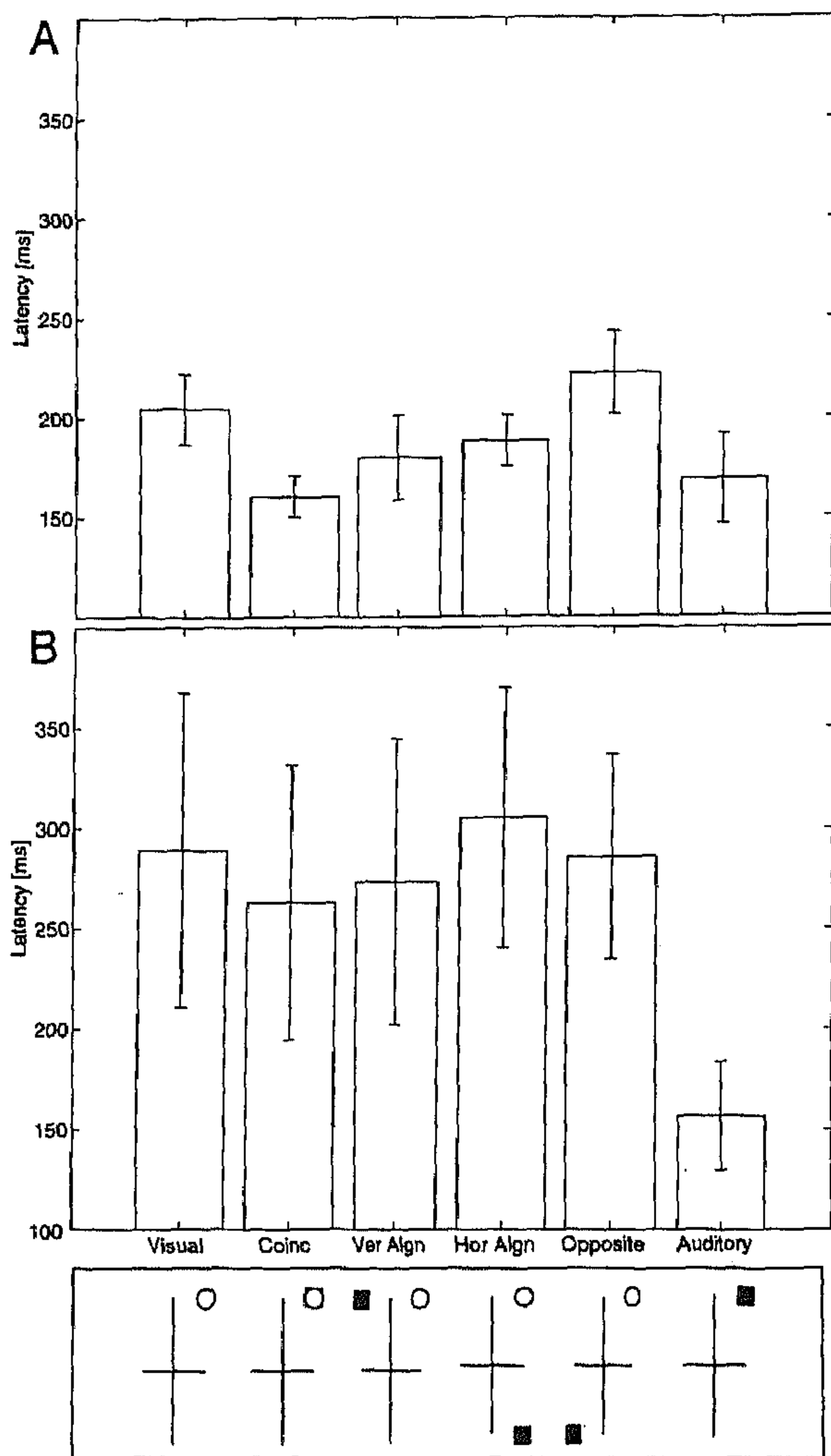


Figure 3. Experiment 1: Latencies (Subject M.F.). (A) Saccadic latencies (mean \pm SD) for the five different stimulus conditions in Experiment 1 (schematically indicated in the bottom panel for one of the visual targets) with high stimulus intensities and broad-band noise as the auditory nontarget (see Method section). Data are pooled over the four visual targets. Visual = unimodal visual stimulus; Coinc = coincident stimuli ($\Delta R = 0^\circ$); Ver Algn = stimuli are vertically aligned but horizontally opposite ($\Delta R = 27^\circ$); Hor Algn = stimuli are horizontally aligned but vertically opposite ($\Delta R = 39^\circ$); Opposite = stimuli are oppositely positioned with respect to the central fixation point ($\Delta R = 54^\circ$); Auditory = unimodal auditory stimulus. Note that the multimodal conditions are represented in ascending order of vectorial distance. The symbols at the bottom end of this figure exemplify each stimulus condition for the visual target at $(R, \Phi) = (27^\circ, 60^\circ)$. The visual target is represented as an open circle, whereas the auditory stimulus is a filled square. (B) Same format for low-intensity stimuli.

principle, be explained if, under this condition, the saccades were acoustically triggered. However, several arguments can be raised against this hypothesis. First, the accuracy of the responses to the coincident targets was much higher than was found for the auditory-guided saccades (compare, e.g., Figures 4A and 4B). In addition, the velocity profiles of audiovisual and visual saccades were stereotyped, in the sense that they obeyed the same

relations between amplitude and duration, maximum velocity or skewness. In contrast, auditory-guided saccades were generally slower and followed more curved trajectories. Finally, the results of the experiments in which low stimulus intensities were employed (see below) show that there can be substantial differences between the auditory and the coincident audiovisual latency distributions.

Low-intensity stimuli. In the low-intensity experiment, the latencies of the visually triggered saccades increased considerably with respect to those in the high-intensity condition (see Figure 3B, left-hand column). Nevertheless, the mean latency of auditory-evoked saccades remained approximately the same, which resulted in a much larger difference between the means of both unimodal latency distributions (for Subject M.F., high intensity, $\Delta L = 35$ msec, low intensity, $\Delta L = 132$ msec). Notwithstanding, the absolute effect of spatial alignment with the auditory stimulus was of the same order of magnitude as obtained in the high-intensity experiment (see also Table 1).

Note that, due to the decrease of intensities, the scatter in latencies of eye movements toward the visual targets increased, which reduced the correlations but kept the slope of the fitted linear relation intact. In contrast, the variability in the auditory saccades was not affected. This suggests that the visual stimuli were closer to the perceptual threshold than were their acoustical counterparts.

Tone stimuli. If the observed auditory-visual interactions could be attributed to a level where both modalities are represented in a common frame of reference, the localizability of the auditory target should influence the properties of the spatial interaction. Toward that end, we presented a tonal acoustic stimulus (see Method section). Figure 4 (data of Subject M.F.) shows that, compared with the broad-band sound (B), saccade accuracy to a 700-Hz tone was markedly reduced (C). In both panels, the trajectories of saccades toward the same four auditory target positions are shown. The data of Subject M.F. are representative for all subjects tested.

Table 1
Fit Parameters Between the Radial Interstimulus Distance, ΔR , and the Latency, L , of the Primary Saccades in Experiment 1

Condition	Subject	r	N	L_{\max}	σ_L	a	σ_a
High-intensity noise	M.F.	.72	172	156.5	3.2	0.96	0.09
	J.G.	.33	134	176.9	5.0	0.46	0.13
	J.O.	.53	176	177.2	3.0	0.62	0.08
Low-intensity noise	M.F.	.29	136	262.0	9.2	0.63	0.25
	J.G.	.33	171	264.1	6.8	0.75	0.18
	J.O.	.31	173	271.3	7.2	0.57	0.21
Tone	M.F.	.26	125	166.4	3.0	0.23	0.08
	J.G.	.11	131	174.8	4.2	0.16	0.10
	J.O.	.01	183	190.5	2.8	0.04	0.07

Note— r = Spearman's rank order correlation coefficient; L_{\max} and a = the offset and slope, respectively, of a straight line fit through the data; σ_L and σ_a = standard deviations of the parameters, which were determined by the bootstrap method (see Note 1). L_{\max} is given in milliseconds; a is given in milliseconds/degree.

