

Auditory psychomotor coordination and visual search performance

DAVID R. PERROTT, KOUROSH SABERI, KATHLEEN BROWN, and THOMAS Z. STRYBEL
California State University, Los Angeles, California

In Experiments 1 and 2, the time to locate and identify a visual target (visual search performance in a two-alternative forced-choice paradigm) was measured as a function of the location of the target relative to the subject's initial line of gaze. In Experiment 1, tests were conducted within a 260° region on the horizontal plane at a fixed elevation (eye level). In Experiment 2, the position of the target was varied in both the horizontal (260°) and the vertical ($\pm 46^\circ$ from the initial line of gaze) planes. In both experiments, and for all locations tested, the time required to conduct a visual search was reduced substantially (175–1,200 msec) when a 10-Hz click train was presented from the same location as that occupied by the visual target. Significant differences in latencies were still evident when the visual target was located within 10° of the initial line of gaze (central visual field). In Experiment 3, we examined head and eye movements that occur as subjects attempt to locate a sound source. Concurrent movements of the head and eyes are commonly encountered during auditorily directed search behavior. In over half of the trials, eyelid closures were apparent as the subjects attempted to orient themselves toward the sound source. The results from these experiments support the hypothesis that the auditory spatial channel has a significant role in regulating visual gaze.

Statement of the Problem

The auditory system in human beings has only limited spatial resolving power; the ability to discriminate the location of a sound source, for example, is seldom better than 1°–2°.¹ Although an extensive literature exists on the topic of auditory spatial processes, little attention has been paid to evaluating the function of this system. In our search for a role for the auditory spatial system, we assumed that the function it serves must require no more than the limited resolution normally observed. We wish to suggest the following hypothesis: The primary function of the auditory spatial system may be to provide information that allows the individual to redirect the eyes in order to bring the fovea into line with an acoustically active object. Since the fovea, which is the most powerful information processing segment of the retina, extends over several degrees of visual angle, additional auditory spatial capacity may not have had any adaptive value. In the following section, we will attempt to present the arguments that led us to this conclusion.

Overview

In human beings, the eyes are located relatively close together at the front of the head. One cost of this arrangement is that people have available only a limited sample of the immediate environment. As noted by Gibson

(1950), at best we view the world through an oval “window” that extends approximately 180° laterally and 150° vertically in the frontal hemifield. Thus, at any moment, somewhat more than half of the immediate environment is unavailable for visual evaluation. But even within this “window,” visual information processing capacity is often quite limited. Only for a very narrow segment near the center of this field, immediately around the line of gaze, do we find really good spatial processing capacity. For images that fall even a few degrees outside of this area, acuity is substantially less. For events located 40° from the line of gaze, for example, spatial resolution is only 5% of that encountered for events within the foveal region. For the extreme peripheral portions of this field, approximately half of the available “window,” spatial resolution is even poorer (Wertheim, 1894).

It is within the context of the limitations of the visual channel that we believe one can most readily appreciate the advantageous position occupied by the auditory system. First, the ears can receive energy coming from any direction, regardless of the current line of gaze. Second, sound, unlike light, is relatively free to travel around objects, so that even if light from an object is partially blocked by foliage or other elements in the environment, acoustic information is still likely to be available. Third, the spatial resolving power of the auditory system is, relative to the visual mode, reasonably homogeneous for broad bandwidth signals (Perrott, 1988; Perrott & Saberi, 1989). Fourth, the sensory apparatus is embedded in the skull; with this arrangement, the spatial information available to the auditory system regarding events in the environment is always referenced to the current position of

This research was supported in part by grants from the National Science Foundation (BNS-8512317) and the National Institutes of Health (3S06-RR0801-1452). Correspondence may be addressed to David R. Perrott, Psychoacoustics Laboratory, California State University, Los Angeles, CA 90032.

the observer's head. In this context, even the absence of mobile pinna is probably an advantage.

All the features mentioned above are compatible with the notion that the auditory system may have a role in determining what portions of the immediate environment are brought into the visual field and, in effect, in determining what information will be present at the fovea. Shifts in gaze, in response to a sound, have been studied for a long time (see Sokolov, 1963). In fact, this component of the "orientation reflex" can be seen in human infants shortly after birth (Wertheimer, 1961). This observation is probably significant, for it suggests that the capacity to perform adjustments in gaze in response to a sound is essentially innate, or "hardwired."

While there is little question that the reorientation of an observer in response to an acoustic event is a common occurrence, there is no evidence as yet that such movements are precise enough to bring the event in line with the fovea. If the movement observed in response to a sound merely serves to bring the event within the general visual field, one might argue that the auditory system has a useful, but not critical, role in the process by which visual information is obtained. Under a broad variety of situations, human subjects do tend to give substantially greater weight to information arriving from the visual modality as opposed to any other sensory channel. This effect has been termed "visual dominance" (Posner, Nissen, & Klein, 1976). In the context of "visual dominance," one would predict that, once the target enters the visual field, any additional information needed for the movement of the eyes would be derived from the visual modality. If the latter description is true, then one could predict that the presence of spatial information from the auditory modality would be superfluous for visual targets that are already in the visual field. Such an outcome would, of course, be contrary to the experimental hypothesis.

General Paradigm

The primary task for the subjects in the first two experiments was to locate and identify a visual target. To ensure that they had in fact located and identified the target, one of two stimuli was employed on each trial (a two-alternative forced-choice, or 2AFC, procedure). To ensure that the subjects employed central vision on this task, the illumination level of the target was reduced to the point at which the discrimination between the two targets was at chance when they were positioned 5° from the subject's point of fixation. Detection performance, on the other hand, indicated that these targets could be detected, but not identified, out to at least 80° from the fovea on the lateral plane and 60° on the vertical plane (the functional visual field with these targets was somewhat less than the 90° lateral and 75° vertical limits of the visual field described by Gibson, 1950). In both experiments, a large region, never less than 260°, was examined within an experimental session. All points within the region selected were equally likely to be employed on the next

presentation. In effect, the experiments involved the localization and discrimination of visual targets under conditions of high spatial uncertainty. Response latency was the primary dependent variable in all of the experiments reported here.

EXPERIMENT 1

Method

Subjects. Five college students aged 21–26 served in this experiment. Two of the subjects had prior experience in psychoacoustic experiments, and three were experimentally naive. All were paid a fixed hourly rate (\$5/h). The subjects were informed that they could earn an additional \$5 if their reaction time (RT) performance was better than that of the other subjects in the same session. Since the bonus was paid for each of 14 sessions completed, a subject could earn an additional \$70. Our subjects appeared to be highly motivated to perform as rapidly as possible. Target identification performance of at least 95% correct was required of all subjects across all conditions examined. The subjects received feedback after each session, regarding their average RT, the number of errors made, and the performance of the other subjects on the same session.

Apparatus. The experiments were conducted in a large audiometric test chamber that had been modified for free-field testing. All interior surfaces were covered with acoustic foam wedges (Sonex). Sound field measurements indicated excellent reduction in reflected sounds for signals above 400 Hz. The subject was seated in a chair in the middle of the test chamber. The subject's head was positioned 1 m below a power shaft that entered the double-walled steel chamber through a hole in the center of the ceiling. Attached to this shaft was a sound boom, 1 m in radius, of local construction (see Figure 1). A loudspeaker (Sony, SRS-3) was mounted to the boom arm at eye level relative to the subject (0° elevation). Communication between this speaker and the audio system located in the external control room was made by a slip-ring connector. At the center of the boom-mounted speaker (Speaker 1), two red flat-surfaced incandescent lights were attached. These target lights subtended an angle of 0.5° (relative to the subject when seated at the center of the chamber). On the surface of one light was printed the letter "L," and on the other, the letter "R." An identical speaker (Speaker 2) was located in front of and at the eye level of the subject (0° elevation and azimuth). This speaker, which was not attached to the sound boom, remained in place throughout the various experimental conditions examined. A third light, placed at 0° azimuth and 3° elevation, was used to provide a fixation point for the subject prior to each stimulus presentation.

Since the sound boom was motorized, the lateral position of Speaker 1 and the two target lights mounted on it could be freely varied. Rotation speeds of 5°–130°/sec could be achieved. Motor noise was evident for rotation rates above 60°/sec but did not exceed 25 dB (SPL; 0.0002 dyn/cm²). Background white noise of 70 dB(A) was employed while the boom was in motion during the current experiment, even though there was no evidence that these sounds from the boom were correlated in any useful way with the current task. The sounds that were present, when preliminary tests were conducted without the masking noise, appeared to radiate from the power shaft directly above the subject. The rate and direction that the boom traveled and the position of the boom after the movement was terminated were under computer control. The positions of the boom-mounted speaker and target lights could be read to an accuracy of better than 0.1° azimuth. A more extensive description of this system has been given by Perrott and Tucker (1988). All other aspects of the experimental situation were also controlled by computer.

Procedure. The subjects were placed directly under the power shaft of the boom, with access to a response-key panel with three

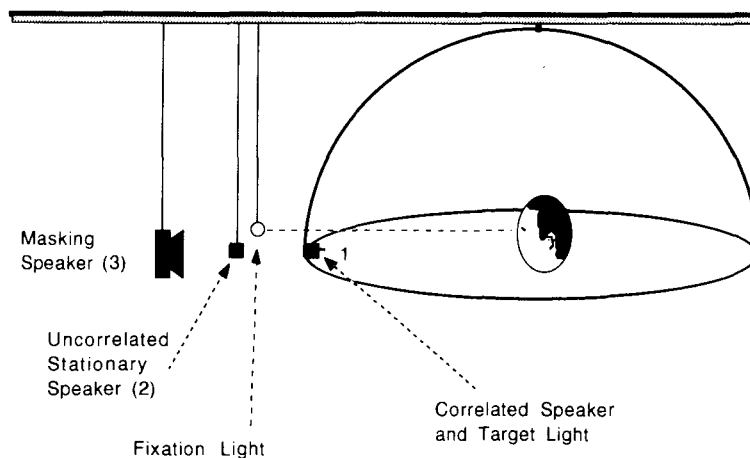


Figure 1. An illustration of the apparatus used in Experiment 1.

keys. One key, which was used by the subject to initiate each trial (the "start key"), was located on one end of the panel. The remaining keys, marked "L" and "R," were located respectively on the left and right sides of the other end of the response panel. At the beginning of each trial, the fixation light and the house lights were turned on (the latter were employed between trials to prevent dark adaptation over the long experimental sessions). The subjects were instructed to fixate on the light directly in front of them (the fixation light). When they were in position, they were instructed to hit the start key to begin the trial. All lights were immediately turned off. The masking speaker was activated, while the boom moved into position. With the boom in position, the masking speaker was deactivated, followed by a 500-msec silent interstimulus interval. At this point, a single speaker and a single target light were activated together. The auditory stimulus was a 60-dB(A), 10-Hz click train. The visual target was one of the two target lights mounted on the motorized boom (either the letter L or the letter R was visible). The start of the click train cued subjects to initiate search. The subject's task was to locate and identify the light and respond by pressing either the left key (if the letter L was lit) or the right key (if the letter R was lit). Each run consisted of 130 trials. An incorrect response rate of higher than 5% constituted a ratio defined as sufficiently large to cancel the run (no sessions were terminated, since performance was always better than 95%). The subject's response terminated the trial. The fixation light and the side light were reactivated, and the next trial began after the subject returned to the initial position (toward the fixation light) and pressed the start key.

Data were collected on target locations ranging from 130° to the subjects' right to 130° to their left. All locations in this region had an equal likelihood of being used on any trial with the following limitation. The 260° field was divided into 13 subregions of 20° each. Within a session, 10 samples were obtained from each subregion. However, all possible locations within each subregion were equally likely to be used within a session.² Sixty samples were collected for each subject at each subregion over a series of sessions. All tests on an experimental condition were completed by each subject prior to the beginning of testing on another condition.

Two conditions were examined in Experiment 1. In the first condition, the 10-Hz click train, which indicated that the subject could begin to search for the visual target, originated from Speaker 2 (the speaker that was always located directly in front of the subject). Thus, the location of the auditory stimulus was independent of the location of the visual target (spatially uncorrelated). The localiza-

tion and identification of the target light, in this condition, could only be based on visual information. In the second condition, the 10-Hz click train originated from Speaker 1 (the boom-mounted speaker). The location of the sound source was spatially correlated with the location of the visual target light. Information regarding the location, but not the identity of the visual target, was available to the subject at the beginning of the search phase of the trial. As mentioned above, each subject concluded either Condition 1 or Condition 2 prior to beginning testing on the other condition. At least one practice session was performed by each subject prior to beginning data collection on each condition. Three subjects served in the spatially correlated condition first, and the remaining subjects began with the spatially uncorrelated condition.

Results and Discussion

Figure 2 presents the mean RTs as a function of the location of the target. The positive and negative azimuth positions indicate that the target was, respectively, to the right or the left of the initial fixation point. In the lower right panel, Figure 2f, each point indicates the average for the 5 subjects employed in this experiment (each point is based on 300 observations). The mean plotted at an azimuth of -100°, for example, indicates the average latency required to locate and identify a target in a region extending from 90° to 110° to the subject's left. The open symbols depict the performance encountered when the source producing the 10-Hz click train was spatially uncorrelated with the location of the target. The performance obtained when the 10-Hz click train was emitted from the same location as that occupied by the visual target, the spatially correlated condition, is indicated by the filled symbols. The remaining five panels present the results for individual subjects.

Although substantial individual differences in this choice RT experiment are clearly evident (subject T.S. had RTs that were typically 200-300 msec longer than those for subject S.P., for example), the functions are nonetheless quite consistent across the subjects. All 5 subjects show U-shaped functions, indicating—not unexpectedly—that

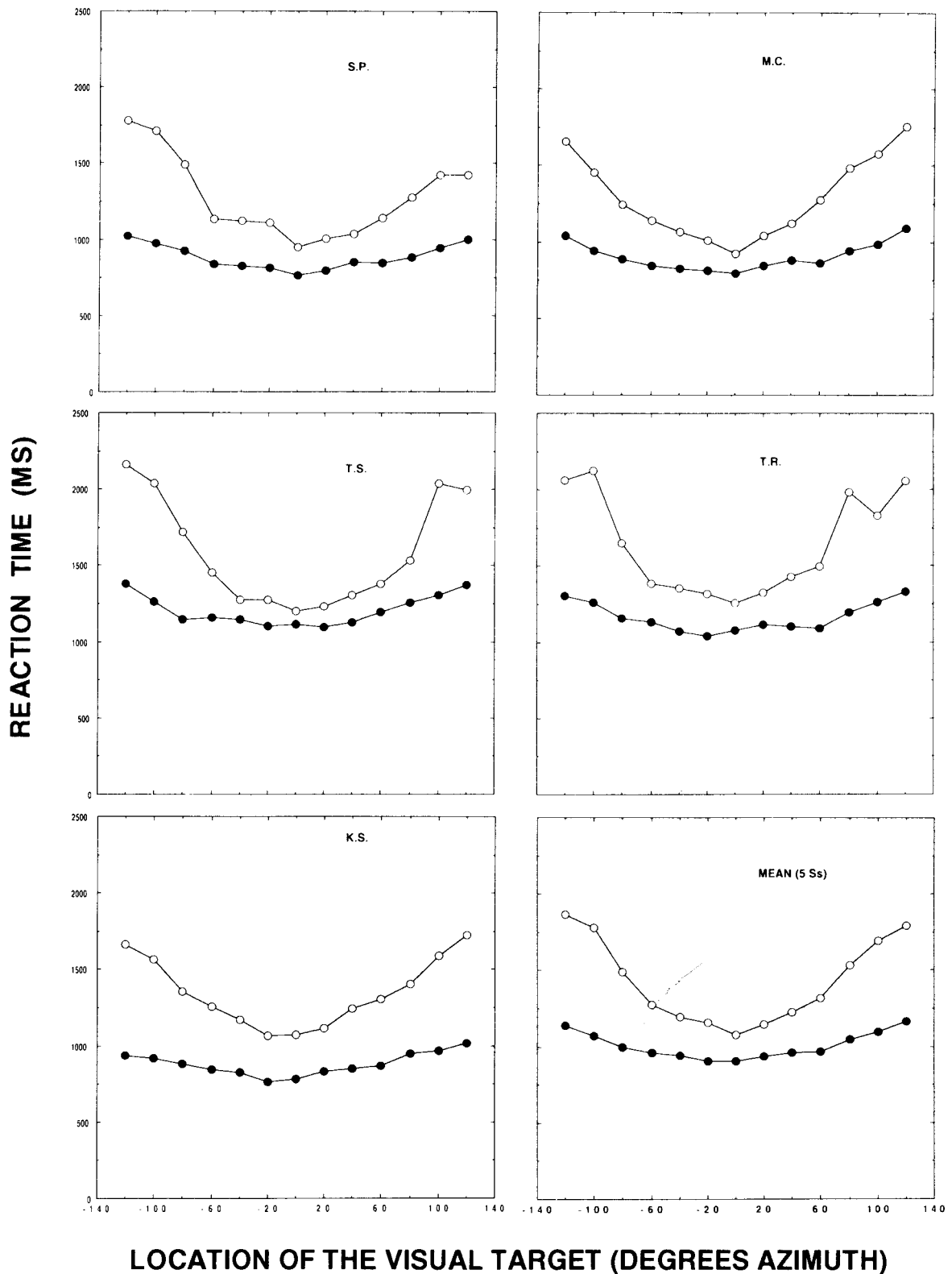


Figure 2. The mean latency to locate and identify the visual target, as a function of the relative location of the target at signal onset. Each point represents 60 observations obtained within a 20° region for an individual subject. The lower right panel presents the average results for all 5 subjects (each point is based on 300 observations). The open symbols indicate visual search performance in the absence of a spatially correlated sound; the filled symbols indicate performance in the presence of a spatially correlated sound.

the latencies are systematically shorter for "targets" located near the initial fixation point. For all subjects, performance is substantially better in the presence of a sound source that is spatially correlated with the visual target than when the sound is spatially uncorrelated. An analysis of variance performed on the data clearly supports these initial impressions. The effects of locus of the target [$F(12,48) = 48.04$, $p < .001$] and conditions [$F(1,4) = 413.746$, $p < .001$], as well as the interaction between these two variables [$F(12,48) = 22.39$, $p < .001$], are significant.³

Figure 3 presents the mean reduction in search time—in effect, the temporal advantage provided by the spatially correlated sound source. The largest effects were, not unexpectedly, obtained for events initially located outside the visual field (in our case, for targets located more than

80° from the initial fixation point). But substantial effects in excess of 150 msec are also evident for events that were initially located within the visual field. Probably what we found to be the most unexpected result was the observation that a statistically significant difference ($p < .01$) was apparent between the spatially correlated and uncorrelated conditions even when the visual target was located within 10° of the subject's initial fixation point.

That acoustic information can serve to reduce the visual search time for events initially located outside of the visual field was not particularly surprising. However, for an event that is located in the rear hemifield (more than 90° from the initial line of gaze), the latencies are only 200–300 msec longer than they are if the event occurs near the fovea—if the subject has spatial information from the auditory modality. In fact, RTs for targets located in

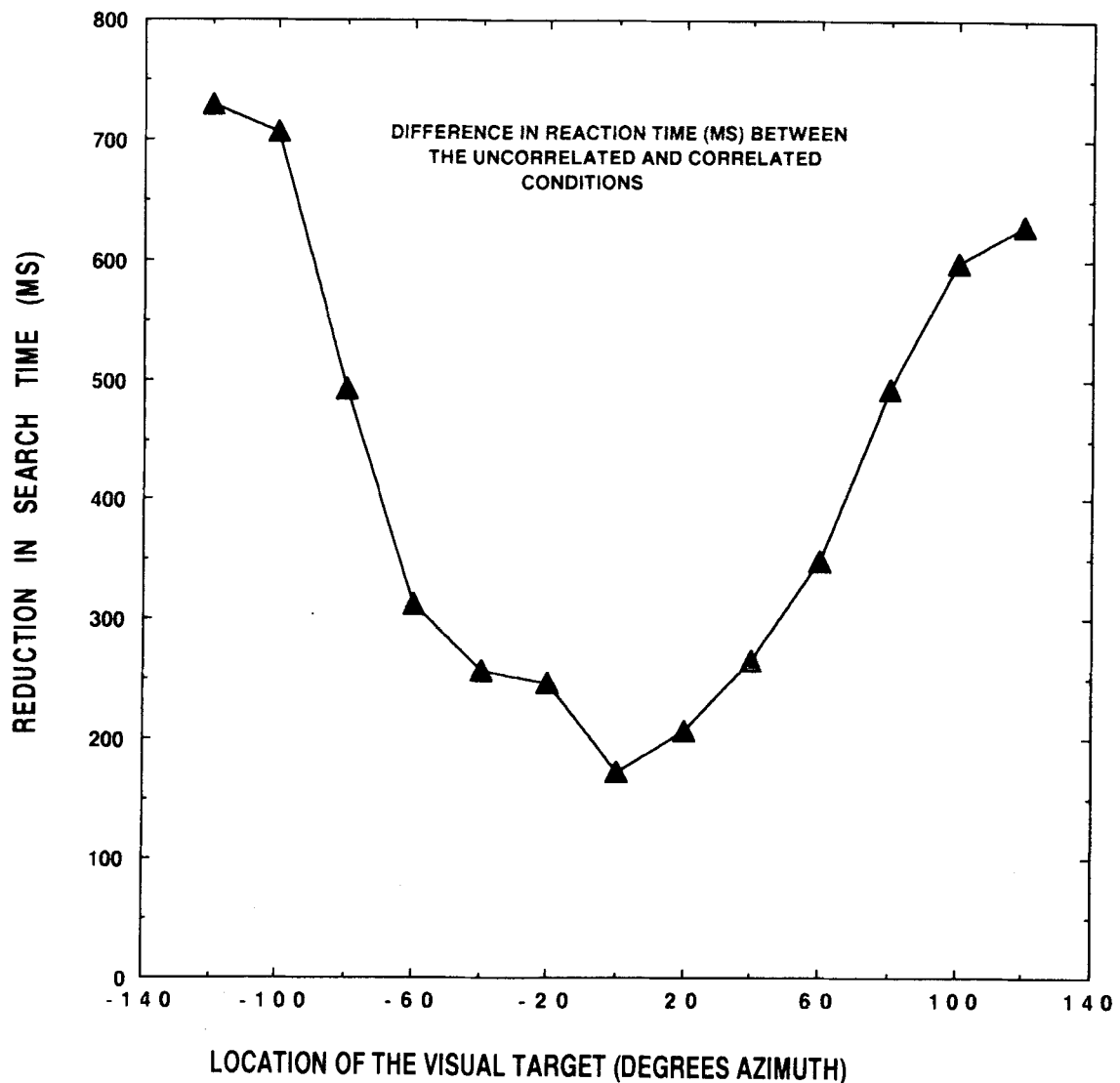


Figure 3. The mean difference in RT between performance observed in the absence of a spatially correlated sound and performance in the presence of a spatially correlated sound.

the rear hemifield in the presence of a spatially correlated sound are generally about the same as those obtained for targets located within 10° of the initial fixation point in the spatially uncorrelated condition. In the absence of this information—that is, when the subject must locate the target visually—an additional 500 msec or more are required for the subject to locate a target in the rear hemifield. If this were the only finding, we could still conclude that such a system could be quite meaningful to an organism that must contend with dynamic events in a three-dimensional space. The present results, however, indicate that the spatial information from the auditory modality also improves search performance by several hundred milliseconds even when the target is located within the visual field. Post hoc tests indicate a significant reduction ($p < .01$) in RT in all regions examined in the current experiment when spatially correlated sound was present.

The performance obtained with targets located within 10° of the initial line of gaze was quite unexpected. In this region, the subjects should have had excellent visual spatial resolving power. Eye movements necessary to bring the target onto the fovea should have been adequately supported by the existing information from the visual field. Yet the latencies obtained without the benefit of spatially correlated auditory information were 150–200 msec longer. There are a number of possible explanations for this interesting observation. First, our spatially uncorrelated condition may not have provided a fair test of the visual search capacity in the absence of auditory localization information. The presence of the sound from the initial fixation point, as used in the latter condition, may have triggered an oculomotor response (or inhibited a response, since the source was located at the fixation point) that was in conflict with the target location information that was available from the visual modality. If this were true, then one would have clear evidence for auditory dominance, at least in the regulation of gaze. While this hypothesis merits further attention, other observations from our laboratory suggest that superior performance can be observed in the central visual field even if the “spatial conflict” is eliminated.⁴

The second hypothesis centers on the notion that information from the auditory spatial channel becomes available to the oculomotor system more quickly than that from the visual channel. It has often been demonstrated that simple RTs to the onset of a light are slower, by about 40 msec, than RTs to the onset of a sound (e.g., Woodworth & Schlossberg, 1954). Differences in the relative signal level between our target light and our 10-Hz click train might account for another 40 msec. But such differences in RT (to a weak light versus a strong sound) are probably too small to explain the effect that we observed.⁵ The latter hypothesis assumes, of course, that the auditory spatial information is as effectively utilized by the oculomotor system as that acquired from the visual modality.

One final possibility is that aurally directed eye movements are simply more accurate than those dependent on information from the visual field. Eye movements can be extremely fast with rates in excess of $800^\circ/\text{sec}$. The time between successive saccades, on the other hand, generally exceeds 150 msec (Saslow, 1967). The difference that we see between the spatially correlated and uncorrelated conditions in the current experiment, with the target located within 10° of the initial line of gaze, could be predicted if an additional saccade was frequently required in the latter situation.

EXPERIMENT 2

Overview

In the natural setting, it is extremely likely that events will occur above and below the initial line of gaze. In fact, concurrent variations in elevation and azimuth would probably be far more common than the conditions examined in the first study, in which the relative elevation of the target was held constant. Aside from the obvious advantage of considering visual search behavior under more natural conditions, our interest in the introduction of concurrent variations in the elevation of the target was prompted by the recognition that such a manipulation might provide a clearer assessment of the potential role of the auditory modality in the regulation of gaze.

Given the intrinsic limitations of the visual modality, the presence of spatial uncertainty on the vertical dimension would probably have substantially more impact on visual search behavior in the absence of auditory spatial information than it would if such information were available. Under conditions of unaided visual search, if a subject fails to detect the target within the current visual field, a shift in gaze would be required. Successive regions would be examined until the target was detected. In effect, we can assume that visual search would proceed in a serial fashion. Thus the addition of variations in the relative elevation of the target would result in an increase in the number of potential regions that would have to be sampled.

In contrast, the auditory spatial channel can appreciate the relative location of an event, regardless of the current line of gaze. No time-consuming movements of the eyes or head are required to sample the field. Thus the addition of variations in elevation might have little or no impact on visual search performance if auditory spatial information is available.

Method

The subjects, apparatus, and procedures were the same as those employed in the first experiment, but with the following changes: Two sets of visual targets were employed in the same session. One pair of targets was positioned 46° above and the second pair was located 46° below the fixation point ($\pm 46^\circ$ elevation). In the condition in which the auditory signal was spatially correlated with the location of the visual target, a speaker was activated from both the same lateral position and the same elevation as the visual tar-

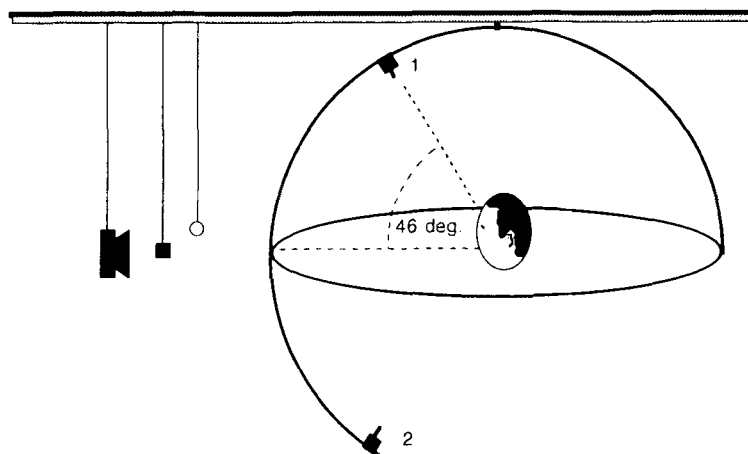


Figure 4. The apparatus employed in Experiment 2 consisted of two speakers and four visual targets (two above and two below the initial line of gaze), mounted on the boom. The speakers and the associated target light pairs were separated by 92° .

get. Figure 4 illustrates the general configuration employed. In the spatially uncorrelated condition, while the visual targets were free to vary over a 260° field in the lateral dimension and occupy one of two locations vertically ($\pm 46^\circ$), the click train was presented from a speaker located directly in front of the subject.

Results and Discussion

Figure 5 presents the mean RTs for the 5 subjects in this experiment. The open symbols present RTs collected under conditions in which the sound was spatially uncorrelated with the location of the visual target. The open squares indicate RTs obtained in the present experiment, when the target was free to vary in elevation ($\pm 46^\circ$) as well as azimuth. For purposes of comparison, the open circles are a replot of the results from the first study, in which no variation in elevation was permitted. The filled symbols present the performance obtained when spatially correlated sound was available during the visual search process. The filled squares indicate RTs obtained under conditions of elevation variation, and the filled circles are a replot of the same condition from the previous experiment. The vertical bars indicate the standard error of the mean for each point.

An analysis of variance performed on these data indicated that lateral location [$F(12,48) = 34.504, p < .001$], conditions [$F(1,4) = 341.196, p < .001$], and the interaction between these two variables [$F(12,48) = 15.192, p < .001$] were significant. The interaction is particularly evident in these data. In the uncorrelated condition, RTs increased by more than 700 msec as the target was displaced laterally. In contrast, in the correlated condition, RTs increased by only 100–200 msec. In fact, for targets located within 70° of the initial fixation point, no increase in RT was evident in the correlated condition. Overall, error rates were higher in this experiment than in the previous study (2.66% and 1.88% for the correlated sound and uncorrelated sound conditions, respectively).

The effects of the addition of variations in elevation are clearly evident in these data. Under conditions in which the sound is spatially uncorrelated (the open symbols), the introduction of a 92° variation in elevation produces a substantial increase in RT. Even for targets located within 40° (azimuth) of the initial fixation point, RTs increased 400–800 msec. A much smaller effect is evident under conditions in which the subjects had access to spatial information from the auditory modality (the filled symbols). Overall, RTs increased by 150–200 msec when a 92° variation in relative elevation was introduced.

The current results are in excellent agreement with the notion that, in the absence of spatially correlated sound, visual search must proceed in a serial fashion. The addition of variations in elevation directly increased the number of potential regions that would need to be sampled. Even for targets in the forward field ($\pm 40^\circ$ azimuth), the displacement of the target 46° in the vertical dimension from the initial line of gaze had a severe impact on performance. The results obtained in the spatially correlated sound condition are entirely different. Here a small but consistent increase in RT was evident across all lateral locations tested. The presence of this increase in RT in the correlated condition with the addition of the vertical dimension, however, may be particularly relevant to our understanding of the relation between the auditory spatial processes and the oculomotor system.

In the previous study, we noted that it was possible to locate and identify a visual target initially positioned outside of the visual field, if a spatially correlated sound source was present, in about the same time as a subject would need to perform an unaided visual search with a target located near the initial line of gaze. Such an effect seems counterintuitive, since in the former task a shift in gaze in excess of 80° would be required, and in the latter case, a shift of only a few degrees would be necessary. This finding could be explained if we assume that

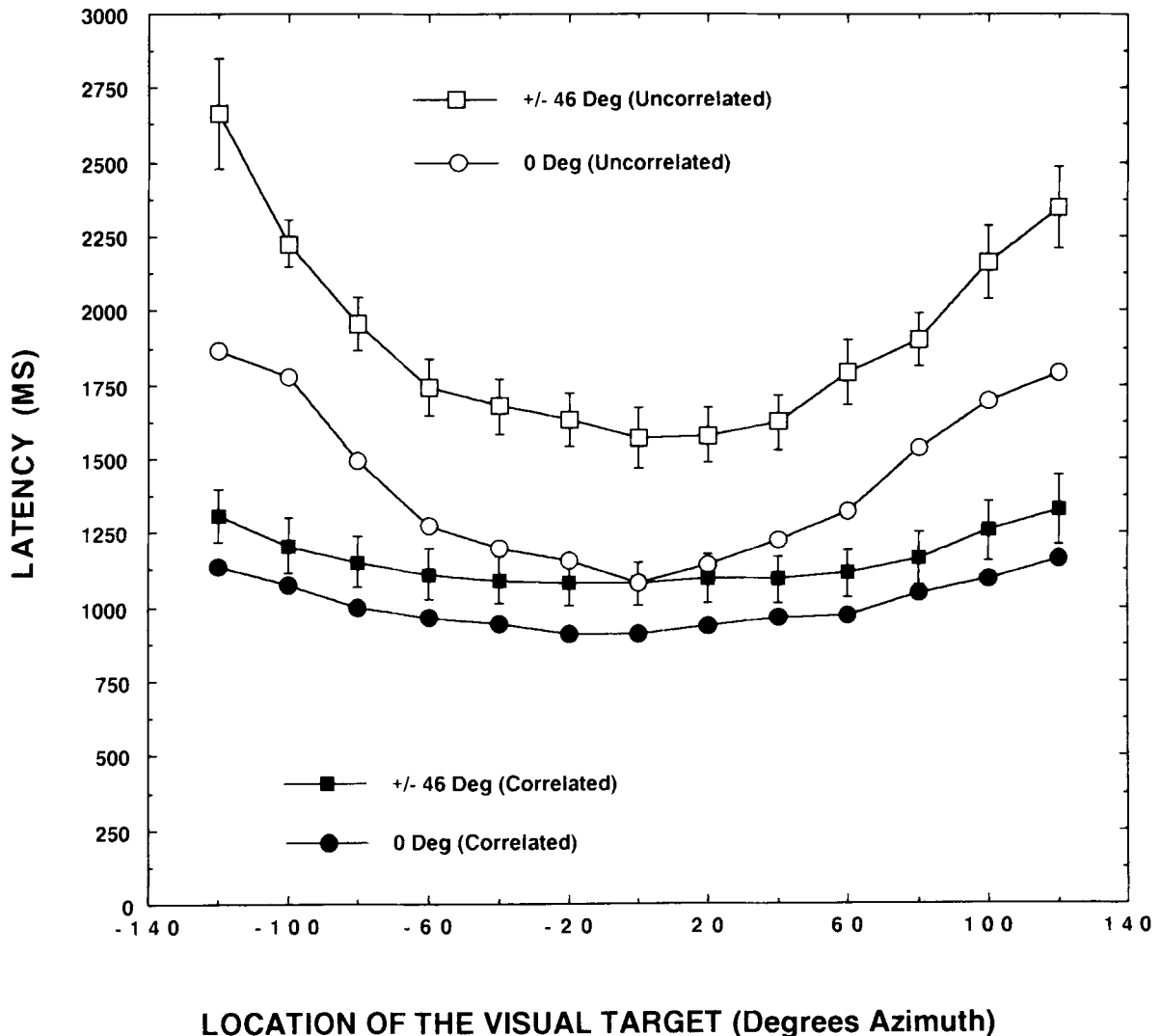


Figure 5. Each point represents the mean latency for 300 observations, as a function of the location of the target light relative to the fixation point at signal onset. The upper two curves (open symbols) present the performance obtained when spatially uncorrelated sounds were present, and the lower two curves (filled symbols) present the latencies observed in the same task when the sound source was spatially correlated. The square symbols indicate that the target varied in elevation ($\pm 46^\circ$ from the initial fixation point), and the circles indicate that no variation on the vertical dimension occurred.

gaze shifts in response to sound are simply more accurate than those dependent on information from the visual field. As noted earlier, shifts in gaze can be extremely fast. The time between successive saccades, on the other hand, is substantial (readily exceeding 200 msec). Thus, the superior performance obtained in the first experiment in the spatially correlated, as opposed to the spatially uncorrelated, sound condition could be explained if one assumed that fewer saccades are necessary when spatial information from the auditory channel is present. In effect, the accuracy with which the position of the target is first assessed would determine the relative probability that the first saccade would result in the image from the target falling sufficiently near the fovea to allow visual identifi-

cation of the event. Errors, when they occur, could be expected to have a disproportionate impact on the average RT, since any additional saccade would require another intersaccade interval.

Although auditory spatial acuity is relatively uniform, spatial resolution does degrade as sources are displaced laterally. For broad bandwidth stimuli, as employed in the current experiment, resolution for targets located in the extreme lateral field drops to about 30% of that obtained near 0° azimuth (Manligas & Perrott, 1988). Thus a substantial portion of the 200–300 msec increase in RT encountered with laterally displaced sources presented in Experiment 1 could have been due to the increased error present in the localization of the sound source.

Minimum audible angle thresholds on the vertical dimension, with broad bandwidth stimuli, are 2–4 times larger than those obtained near 0° azimuth on the lateral plane (see Perrott & Saberi, 1990). This resolution is quite similar to that encountered in the extreme lateral region of the listener's field. Thus, if the accuracy of the initial determination of the location of the target were a primary determinant of whether or not the first saccade would bring the target onto the fovea, performance requiring an assessment of the elevation of the target should be similar to that obtained for sources located in the extreme lateral field. In effect, the relative spatial resolution and not the distance of the target from the initial line of gaze would have a major impact on auditorily aided visual search performance. The 150–200 msec increase in RTs obtained in the spatially correlated sound condition in the current experiment is well within the limits suggested by this hypothesis.

EXPERIMENT 3

In the first two experiments reported here, the ability to adjust the line of gaze using auditory spatial information was explored indirectly by examining the time required to locate and identify a visual target. In the last experiment of this series, we wanted to obtain some direct information about the movements of the head and eyes in response to an acoustic signal in the largest test area possible (a full 360°). For practical reasons, we decided to focus on the initial and not the terminal movements that could be observed.

Method

Subjects. Three men and 4 women served in this experiment. Only 1 subject had prior experimental experience. All subjects had normal vision without glasses and no known hearing impairments. The data from 1 subject was dropped due to an unreported motor dysfunction that became evident during the course of testing.

Apparatus. The apparatus previously described was again used, but with the following additions: A video camera (NEC, TI-22A CCD) was positioned approximately 10° above the fixation light. The camera and illumination source were mounted on a track attached to the interior surfaces of the chamber. The output from the camera was fed into a video cassette recorder (Panasonic, AG-6300) located in the adjacent control room. The recorder had a frame rate of only 30/sec, which set the lower limit on the temporal differences that could be assessed (approximately 33 msec). The subject wore an adjustable headband, upon which had been attached a 1/4-in. black sequin.

Procedure. The procedures were essentially the same as those described for Experiment 1, but with the following changes made in the paradigm: First, only the spatially correlated sound condition was examined. Second, we expanded the test region from 260° to a full 360° (at an elevation of 0°). Since we wanted to record the eye and head movements on the videotape, the chamber remained illuminated during the session. A second chamber, made out of black parachute cloth, was constructed to fit within the perimeter of the boom. This arrangement eliminated most of the information that the subject might acquire regarding the location of the sound source from the visual modality. The task, then, had been reduced from a 2AFC paradigm to one in which the subjects were simply asked to look at the location from which the sound originated. A number

of relatively minor changes were made as well. For example, the 10-Hz click train was replaced with a broad bandwidth noise of approximately 50 dB(A). A small red light, within the field of the camera, was turned on simultaneously with the speaker. The latter change provided a convenient "marker" on the video recording, which indicated the video frame in which the auditory signal was initiated.

Data analysis. The video records obtained from each of the 600 trials were inspected one frame at a time. The latency of the movement of the eyes was measured as follows: A straightedge was held against the video image of the iris on the frame at which the light signaling the onset of the sound occurred. The videotape was then advanced until a detectable displacement of the iris was evident.⁶ The number of frames required was used to determine the latency of this response. A similar procedure was used to evaluate head movements, with the edge of the black sequin providing the referent image. During the course of the data collection, it became clear that yet another response was present. The subjects frequently closed their eyes while responding to the sound. We referred to this as eyelid closure rather than an eyeblink, since the closure, when it occurred, was usually sustained for most of the period that the subject's face remained in the camera's field of view. Although the latency of the eyelid closure was less readily determined than the head and eye movements, there was no ambiguity regarding the response itself.

Results and Discussion

Figure 6 presents, in four panels, frequency histograms of the latencies observed for eye movements (open circles), head movements (filled circles), and eyelid closures (open squares), as a function of the relative location of the sound source. The upper left panel (Figure 6A) presents the results obtained when the sound source was within 20° of the initial fixation point (right and left fields are combined in all of the following figures). Out of the 600 trials completed, only 9 involved cases in which no eye movement was observed. All but 2 of these occurred when the source was within 2° of the fixation point (the others were observed with the source at 3° and 5°). As can be seen from this panel, the modal latency for eye movement was approximately 275 msec. Within this region, head movements were seen on approximately 60% of the trials, and the modal latency for this response was approximately 375 msec. The lowest function indicates cases in which the eyelids were observed to close during the trial. There appears to be no consistent pattern to this response.

Tomlinson and Bahra (1986a), in an experiment with primates, also noted "blinks" during the course of measuring eye movements in a visual search task. They simply eliminated the trials on which these events occurred. The results in the upper right panel suggest why these eyelid closures should not be ignored. The distribution of latencies for eye and head movement are plotted along with the cases in which eyelid closure occurred. These observations were made with the source located between 20° and 60°. Probably what is most striking is the presence of a well-defined function for eyelid closures. The modal latencies for both head and eye movements have declined. Head movements are observed on approximately 90% of the trials for sources located in this region.

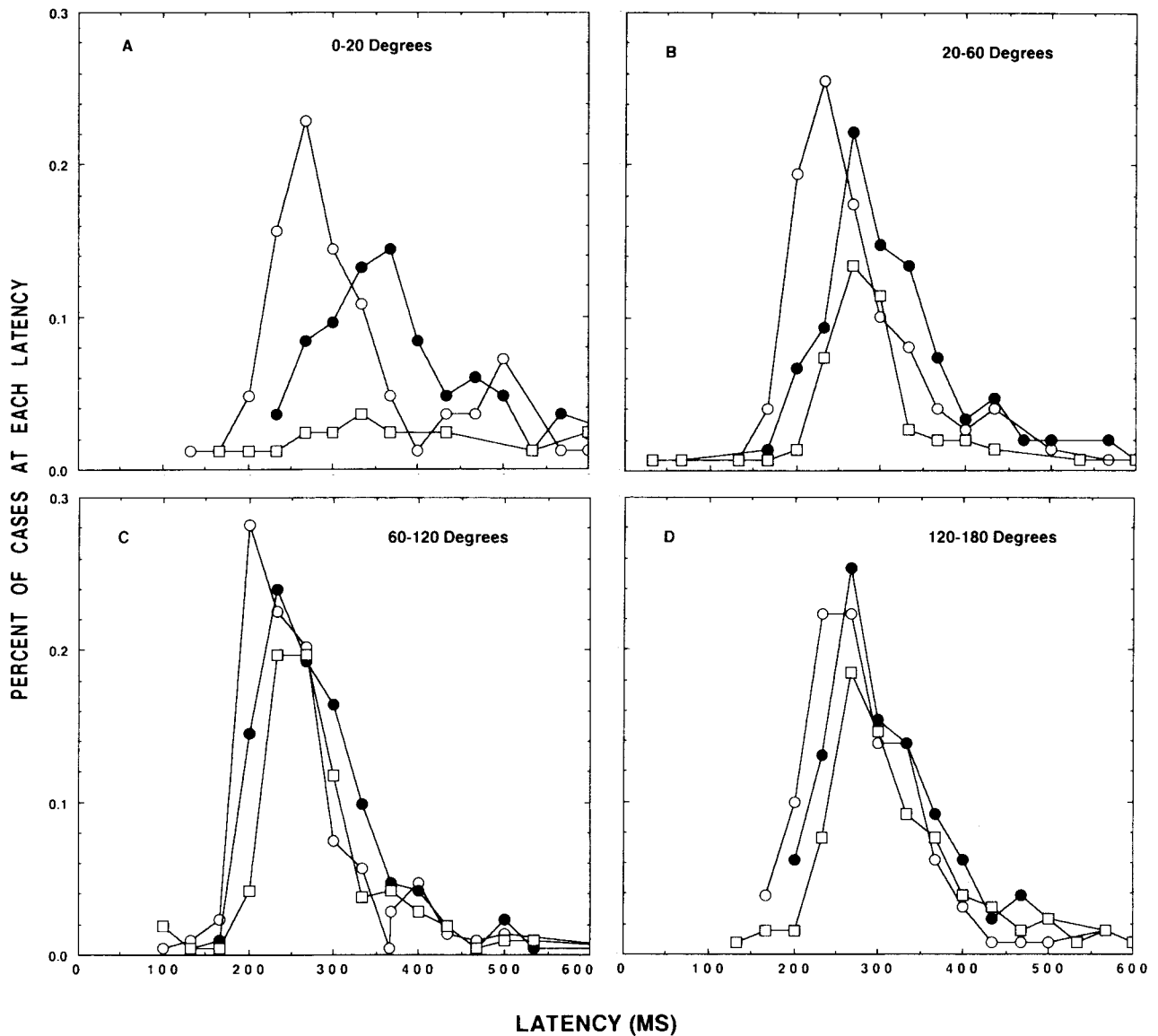


Figure 6. A frequency histogram of the following response latencies: eye movement (open circles); head movement (filled circles); and eyelid closure (open squares). Each panel indicates the location of the sound source relative to the initial fixation point.

Performance in the extreme lateral region (60° – 120°) indicates that head and eye movements occur on every trial; eyelid closures now occur on approximately 70% of the trials. Probably what is most obvious in this figure is that there is now only a small difference in the distribution of these three responses. Although eye movements are still generally seen earlier than head movements and eyelid closures, the difference is now less than 40 msec. The last panel is a summary of all the observations made in the rear (120° – 180°) hemifield. Movements of the head and eyes as well as eyelid closures tend to occur within the same 33-msec frame. The data on the latency of eye movements in this region must be considered somewhat incomplete however, since eyelid closures sometimes occurred before any eye movement could be detected.

GENERAL DISCUSSION

Our starting point for this series of experiments was an attempt to understand why the spatial resolving power of the human auditory system has an accuracy of only 1° – 2° . We speculated that this low level of acuity might, in fact, be indicative of the functional role of this system within the broad interplay of the various sensory and motor channels. For example, while the location of the eyes in the front of the human head brings with it a significant binocular advantage, it seems reasonable to suggest that this requires the processing of information from events that occur in the rear. Of course most features of the environment are silent, which means that the auditory system provides really very little information about the en-

environment outside of the visual field. Organisms, whether they vocalize or simply locomote, and objects in motion, such as falling rocks, are significant exceptions. Even though the auditory system might only provide information regarding the location of sound-generating objects, such objects are likely to be particularly relevant to the survival of the individual.

An extensive literature, of course, shows that shifts in gaze often follow the onset of a sound. Pavlov's description of the "what is it?" reflex (Pavlov, 1927) is probably familiar even to beginning psychology students. The question of whether this reflex results in a precise realignment of gaze that brings the source of the sound into the central region of the visual field (the fovea in human beings), which would require an accuracy of approximately 2° , or whether it merely results in bringing the event within any portion of the visual field, was the initial stimulus for the current series of experiments. We had assumed that there were potentially three separate problems, which involved the advantage that might be obtained in a visual search task if auditory spatial information regarding the location of the source was provided (1) when the source was initially located outside of the visual field; (2) when the source was located in the peripheral visual field (more than 10° from the initial line of gaze); and (3) when the source was located within a few degrees of the fovea.

The results of the first two experiments, for targets located in excess of 90° from the subject's initial fixation point, clearly show that a large reduction in the time required to locate and identify the target when auditory spatial information has been provided can be obtained. Without vertical uncertainty, the latencies for events located in excess of 90° dropped by nearly 700 msec when auditory information was provided. When subjects had to scan in both the vertical and the horizontal dimensions for the target, the advantage increased by an additional 500 msec. Any auditory information regarding the location of a remote target appears to benefit the visual search performance. Targets located outside the visual field that required a shift in gaze of 100° before identification could be made required only an additional 50 msec relative to a target located within 10° of the initial fixation point, if the first event was presented with spatially correlated sound and the second was solely dependent on visual information. In effect, when auditory information regarding the location of remote events exists, target identification proceeds at about the same level of efficiency as it does when the entire process occurs in the central visual field. Thus, human beings, with their limited field of view, have lost little capacity to identify sound-producing objects that are initially located outside the visual field.

Auditory spatial information is therefore of significant benefit for the eventual visual identification of events initially located outside the visual field. The second issue with which we were concerned was whether or not any advantage in visual search performance would be realized with spatially correlated sounds when the initial location of the target was within the peripheral visual field. Acuity declines rather dramatically as the image is dis-

placed off the fovea. Yet even in the most peripheral regions of the retina, spatial resolution is many times superior to that enjoyed by the auditory channel. Thus it seemed reasonable to predict that the visual system would have ample information with which to guide the oculomotor response so as to bring the image into the central field. Once again, a substantial advantage was evident, both when the search was restricted to the horizontal plane (Experiment 1) and when it included the vertical dimension as well (Experiment 2). For "targets" located more than 10° from the fixation point, the temporal advantage exceeded 200 msec in all cases. It is clear from these results that the spatial information from the auditory channel is not superfluous even for events in this region.

Finally, we considered the results obtained when the visual target fell within 10° of the initial fixation point. The smallest displacement from the initial line of gaze was 3° , since the fixation light was 3° above the horizontal plane of the boom. The spatial resolving power of the retina in this region should be 50–200 times that enjoyed by the auditory system. Clearly there is more than adequate information for the system to be able to reorient the eyes so as to acquire the target. As can be seen in Figure 2, each subject performed the task more quickly when auditory spatial information was available. The average advantage was nearly 175 msec. This finding was completely unexpected.

We believe that the present results clearly support our initial hypothesis regarding the functional role of the auditory spatial system; that it provides information directly to the oculomotor system such that acoustically active targets have a high probability of being brought into the foveal window for further evaluation. This system clearly has a high priority. As noted in the "orientation reaction" literature, acoustic events can readily "draw away" the gaze, even if the individual is engaged in a critical task. What was not expected was the clear indication that visual search performance, when guided by auditory spatial information, is superior to that observed when only visual information is available. The following comments are an attempt to provide an explanation for this observation.

If one were designing a servomotor system to perform shifts in the input field on the basis of the current input, it would be reasonable first to determine the exact movement necessary and then to perform the movement. But if the initial information were less than optimal, it might be useful to sample the data from the field during the course of the movement, making adjustments in the motor program on the basis of later arriving data. But the visual system does not seem to perform in the latter fashion. With a target in the visual field, but displaced sufficiently far from the fovea that a shift in gaze is necessary, the localization information available to perform the saccade appears to be restricted to the initial sample; that is, there is little evidence that the system utilizes information from the retina during the saccade. Davidson, Fox, and Dick (1973), for example, noted that visual input during a saccade is essentially eliminated (the saccade suppression effect). More recent research (e.g., Irwin, Yan-

tis, & Jonides, 1983; Rayner & Pollatsek, 1983) suggests that there is even little integration of information across successive fixations. The results of the last experiment in this paper may provide the strongest argument yet available against the notion that there is any substantial contribution by the retina in the regulation of a shift in gaze once the movement has started. For large shifts in gaze, in excess of 60° , the eyelids were closed in approximately 70% of the trials that we observed. For smaller displacements of 20° – 60° , eyelid closure occurred on approximately 35%–50% of the trials. Incidental observations in our laboratory confirm the presence of this eyelid closure even in the absence of an auditory stimulus.⁷ Aside from the obvious idea that eyelid closure would eliminate most of the useful information that might fall on the retina, eyeblinks (both voluntary and involuntary) have been shown to decrease the sensitivity of the retina to light even when the light was presented so as to bypass the lids and optics of the eye (Manning, Riggs, & Komenda, 1983). Thus, the visual system seems to have relatively little capacity to modulate the gaze movement on the basis of information arriving later from the retina.

The consequence of an error—that is, of the initial gaze adjustment's failure to bring the image onto the fovea—can be substantial. While eye-head movements can readily exceed hundreds of degrees per second, the delay between successive saccades can readily exceed 150 msec. Thus, if the auditorily directed movement required fewer corrective saccades, a substantial portion of the auditory advantage could be explained. In addition, there is at yet no reason to believe that the processing of auditory spatial information is suspended during either head or eye movements, and thus it is possible that a more efficient servomechanism may exist through this channel.

One additional factor must be considered. In our final experiment, it was clear that head movements were a significant factor underlying most adjustments in gaze that occurred during the search for a target. Even for events within 20° of the initial line of gaze, concurrent movements of the head and eyes were commonly observed. The results of this experiment are consistent with the observations reported by Guitton, Douglas, and Volle (1985), who used cats as subjects, as well as with two reports by Tomlinson and Bahra (1986a, 1986b), who used primates. In these studies, only visual information was provided (most of the time). They noted that as the displacement of the target increased, the tendency of both head and eye movements to occur together also increased. Tomlinson and Bahra (1986a) also observed that as the head accelerated, the eye movement decelerated. In effect, the eye becomes essentially stationary (relative to the head) over much of the period that the head is moving. Shifts in gaze during the search for a target are thus almost exclusively due to head movements (when the latter movements occur). As mentioned earlier, the auditory system is particularly well situated, since it is embedded in the mastoid bone, to assess the position of the head relative to an active sound source. In summary, we believe that one can make a strong case for the hypothesis that the auditory

spatial channel in human beings may have evolved primarily to serve as a guidance mechanism for the eye-head motor systems responsible for the regulation of eye position.⁸

REFERENCES

- CARR, H. A. (1935). *An introduction to space perception*. New York: Hafner.
- DAVIDSON, M. L., FOX, M.-J., & DICK, A. O. (1973). Effect of eye movements on backward masking and perceived location. *Perception & Psychophysics*, **14**, 110-116.
- DURLACH, H., & COLBURN, H. (1977). Binaural phenomena. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 4, pp. 365-466). New York: Academic Press.
- GIBSON, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- GUITTON, D., DOUGLAS, R. M., & VOLLE, M. (1985). Eye-head coordination in cats. *Journal of Neurophysiology*, **52**, 1030-1050.
- HOWARD, I. P., & TEMPLETON, W. B. (1966). *Human spatial orientation*. London: Wiley.
- IRWIN, D. E., YANTIS, S., & JONIDES, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics*, **34**, 49-57.
- MANLIGAS, C., & PERROTT, D. R. (1988, May). *Some evidence for foveation in the auditory modality*. Paper presented at the meeting of the Western Psychological Association, Long Beach, CA.
- MANNING, K. A., RIGGS, L. A., & KOMENDA, J. K. (1983). Reflex eyeblinks and visual suppression. *Perception & Psychophysics*, **34**, 250-256.
- MILLS, A. W. (1972). Auditory localization. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 2, pp. 303-348). New York: Academic Press.
- PAYLOV, I. P. (1927). *Conditioned reflexes* (G. V. Anrep, Trans.). New York: Oxford University Press.
- PERROTT, D. R. (1984). Binaural resolution of the size of an acoustic array: Some experiments with stereophonic arrays. *Journal of the Acoustical Society of America*, **76**, 1704-1712.
- PERROTT, D. R. (1988, October). Are there motion detectors in the auditory system? In *Localization of sounds by humans*. Symposium conducted by the Committee on Hearing, Bioacoustics and Biomechanics (National Academy of Sciences), Washington, DC.
- PERROTT, D. R., & MARLBOROUGH, K. (1989). Minimum audible movement angle: Marking the end points of the path traveled by a moving sound source. *Journal of the Acoustical Society of America*, **85**, 1773-1775.
- PERROTT, D. R., & PACHECO, S. (1989). Minimum audible angle thresholds for broadband noise as a function of the delay between the onset of the lead and lag signals. *Journal of the Acoustical Society of America*, **85**, 2669-2672.
- PERROTT, D. R., & SABERI, K. (1989, August). *Functional integration of auditory and visual sensory processes*. Paper presented at the Annual Meeting of the American Psychological Association, New Orleans.
- PERROTT, D. R., & SABERI, K. (1990). Minimum audible angle thresholds for sources varying in both elevation and azimuth. *Journal of the Acoustical Society of America*, **87**, 1728-1731.
- PERROTT, D. R., & TUCKER, J. (1988). Minimum audible movement angle as a function of signal frequency and the velocity of the source. *Journal of the Acoustical Society of America*, **83**, 1522-1527.
- PIERCE, A. (1901). *Studies in auditory and visual space perception*. New York: Longmans.
- POSNER, M. I., NISSEN, M. J., & KLEIN, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, **83**, 157-171.
- RAYNER, K., & POLLATSEK, A. (1983). Theoretical considerations on the function and circuitry of various rapid eye movements. In A. Fuchs & W. Becker (Eds.), *Progress in oculomotor research* (pp. 3-9). Amsterdam: Elsevier North-Holland.
- SASLOW, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, **57**, 1030-1033.

- SOKOLOV, Y. (1967). *Perception and the conditioned reflex* (S. W. Wayenfeld, Trans.). New York: Pergamon.
- STRYBEL, T. Z., & PERROTT, D. R. (1984). Auditory distance discrimination: The success and failure of the intensity discrimination hypothesis. *Journal of the Acoustical Society of America*, **76**, 318-320.
- TOMLINSON, R. D., & BAHRA, P. S. (1986a). Combined eye-head gaze shifts in the primate: I. Metrics. *Journal of Neurophysiology*, **56**, 1542-1557.
- TOMLINSON, R. D., & BAHRA, P. S. (1986b). Combined eye-head gaze shifts in the primate: II. Interactions between saccades and the vestibulo-ocular reflex. *Journal of Neurophysiology*, **56**, 1558-1569.
- WERTHEIM, T. (1894). Über die indirekte Sehscharfe. *Zeitschrift für Psychologie*, **7**, 172-187.
- WERTHEIMER, M. (1961). Psychomotor coordination of auditory and visual space at birth. *Science*, **134**, 1692.
- WOODWORTH, R. S., & SCHLOSSBERG, H. (1954). *Experimental psychology*. New York: Holt, Rinehart & Winston.

NOTES

1. Carr (1935) suggested that spatial systems must include the ability to resolve the attributes of size, shape, distance, motion, and location. Auditory spatial research with human subjects indicates the following: There is only a rudimentary capacity to resolve the size of an acoustic array and virtually no indication of a capacity to appreciate the shape of the array (Perrott, 1984). Auditory distance perception is notoriously poor. In fact, loudness, which we assume is not a spatial attribute per se, may account for much of the information employed in depth discrimination (Strybel & Perrott, 1984). Probably the discrimination of motion and location, when broad bandwidth stimuli are employed, represents the best developed spatial capacities in the auditory system (Perrott & Marlborough, 1989; Perrott & Pacheco, 1989). Resolution in both functions approximates 1° change in azimuth. It seems fair to point out that the latter functions are substantially less precise than the same functions obtained in the visual modality. For excellent reviews on this topic, see Pierce, 1901; Howard and Templeton, 1966; Mills, 1972; or Durlach and Colburn, 1978.
2. The subregions were divided, in turn, into 40 locations. Each location subtended approximately $\pm 0.25^\circ$ azimuth. Thus, there were 520 locations that the visual target could occupy on any given trial (40 locations $\times$ 13 subregions). The intent was to maintain a high degree of spatial uncertainty regarding the location of the target.
3. There is little evidence of a speed-accuracy tradeoff that could account for the superior performance observed in the spatially correlated condition. While the mean error rate was somewhat larger in the spatially correlated condition (1.49%) than in the uncorrelated condition

(0.98%), Subject S.P., for example, had an error rate in the uncorrelated condition that was 20% higher than her error rate in the correlated condition, yet she still had shorter reaction times when the spatially correlated sound was available.

4. In our initial attempt to collect data on this problem, we had employed a control condition in which only a visual target was present (no spatially uncorrelated sound was used). Performance was particularly variable in this condition, even when the target occurred in the central visual field. The reason for this variability became readily apparent. Lacking a clear signal indicating when to begin to look for the target, subjects sometimes began to search for the target prior to its onset. Such premature shifts in gaze were particularly costly when the target actually occurred near the initial fixation point. Any signal (auditory or visual) that might be used to mark the onset of the target would have the potential to initiate an orientation response, so that there does not appear to be any simple solution to the problem of the potential conflict between the spatial information from the cue to begin the search and the target itself, at least when the target stimulus is free to occur from nearly any direction.

5. Current research in our laboratory clearly indicates that more intense visual targets and less intense spatially correlated sounds can reduced the size of the advantage. However, assuming that the sounds are clearly audible, the decrease in the RT advantage is quite modest (generally less than 40 msec).

6. The identification of the "frame" in which the movement occurred was actually easy to assess. Given the combination that the eye movements themselves were quite rapid and our temporal resolution (approximately 33 msec/frame) was quite gross, a marked difference in the position of the video image of the iris between frames was generally evident.

7. We do not see eyelid closure every time a subject turns his head or every time a sound is initiated in the immediate field. If a subject is asked to look for a red square, for example, or to track a moving target, eyelid closures are not particularly evident. However, if the subject is asked to look at a target whose position is already known but is not currently in the line of gaze, closure of the eyelids will frequently be seen during the rapid shift in the line of gaze.

8. We have been careful to restrict our generalizations to human populations as much as possible. The size of the visual field and the organization of the retina are two factors that might influence how an organism employs spatial information from the auditory modality. Even the relative mobility of the eyes and head could be significant parameters. In the domestic cat, which has a relatively limited capacity to move the eyes, excellent mobility of the pinna is evident.

(Manuscript received September 20, 1989;
revision accepted for publication March 7, 1990.)