

Auditory attention and the representation of spatial information

GILLIAN RHODES

Stanford University, Stanford, California

Three experiments were carried out to investigate the spatial properties of auditory attention. If auditory spatial information is represented analogically, then shifts of auditory attention should be constrained by the spatial structure of the representation. In particular, the time taken to shift auditory attention should increase with the distance moved. As predicted, the time for attention shifts was a linearly increasing function of the angular distance moved, for distances up to 90°. Moreover, the rate at which attention was shifted appeared to depend on the capacity available, which in turn depended on task difficulty (Experiment 1 vs. Experiment 2) and memory load (Experiment 3). Explanations based on general expectancies or response priming were ruled out. The results suggest that auditory spatial information, like visual spatial information, is represented analogically, and that this structure constrains the way that attention can be moved within the representation.

An important function of perception is to provide information about the locations of objects. Such spatial information appears to be represented analogically. Shepard and his colleagues have demonstrated that transformations in mental space occur in real time and mirror equivalent transformations carried out in actual space (Cooper & Shepard, 1973a, 1973b; Shepard & Cooper, 1982; Shepard & Metzler, 1971). This one-to-one correspondence between transformations in mental space and in actual space is a key property of analogical representations, which Shepard (1981) has formalized as the psychophysical complementarity principle. In an analogical, or map-like representation, spatial relations and distance information are directly represented by the structure of the representation. In a nonanalogical representation, for example, a list of the (x, y, z) coordinates of objects, relative position or distance between points must be computed. Nonanalogical representations may therefore be less efficient for representing spatial information (see Marr, 1982, for a discussion of the appropriateness of different types of representation for different types of information and processing requirements).¹

Shifts of attention in visual space appear to be constrained by the analogical structure of spatial representations. Tsal (1983) has demonstrated that the time taken to shift attention increases linearly with the distance moved. Downing and Pinker (1985) have shown that reaction time to detect a probe increases with its distance from

the point of attention in three-dimensional (3-D) as well as 2-D space.² Moreover, Shulman, Remington, and McLean (1979) found that when attention is shifted from one place to another detection of probes presented along the hypothesized path, at appropriate times during the shift, was facilitated. These results suggest that movements of visual attention occur continuously and in real time, as would be expected if they were occurring in an analogical representation. Thus, in vision, the analogical nature of the spatial representation constrains the behavior of attention within the representation.

An important theoretical issue is whether the analogical or topographical nature of spatial representations is confined to visual information or whether it is a general property of spatial representations. There is evidence that tactile as well as visual spatial information is represented analogically (Carpenter & Eisenberg, 1978; Marmor & Zaback, 1976). For both the visual and tactile modalities, representations at the surface receptors and in the primary sensory cortex are topographical or analogical maps. It is therefore not surprising that the same organization would be preserved in more abstract spatial maps.

Are representations of auditory space also analogical? Unlike vision and touch, spatial information is not mapped topographically at the auditory receptor surface. In audition, it is pitch that is mapped topographically and that may be more directly analogous to visual spatial position than is auditory spatial position (Kubovy, 1981). In audition, spatial information must be computed from interaural time and intensity differences. At the early stages of analysis, therefore, the code is nonanalogical. However, single-unit recordings of responses to auditory stimuli by cells in the superior colliculus indicate that spatial information is coded topographically or analogically at higher levels of the auditory system in birds (Knudsen, 1982, 1984) and mammals (Drager & Hubel, 1975; Palmer & King, 1982). Moreover, modality-specific spa-

I thank Cathryn Downing, Edward Kessler, Ann O'Leary, Michael Pavel, and especially Roger Shepard for helpful discussions about this work, and Al Bregman and two anonymous reviews for comments on an earlier version of the manuscript. I also thank Mac Feliciano and Gilbert Mesa for helping run subjects, and Edward Kessler, Ian McLean, and Gilbert Mesa for programming assistance. This work was supported by NSF Grant BNS 80-05517 to Roger Shepard. Requests for reprints should be sent to G. Rhodes, Department of Psychology, University of Canterbury, Private Bag, Christchurch, New Zealand.

tial maps were found to be superimposed, so that neighboring visual, auditory, and bimodally responsive neurons responded optimally to the same regions of space. Clearly this organization is ideally suited to intermodal integration of spatial information.

If auditory spatial information is also organized topographically or analogically in humans, then shifts of attention within the representation should be constrained by the spatial properties of such a representation, as they are for visual attention. The experiments reported here were designed to test the hypothesis that shifts of auditory attention would be a linearly increasing function of the angular distance moved, a prediction based on the spatial properties of a topographical or analogical representation.

A localization task was used in preference to a detection task. Detection tasks have been used to study visual attentional shifts (Posner, 1978; Posner, Snyder, & Davidson, 1980) because they allow response-level explanations to be ruled out (the response is always the same). Visual-detection reaction times have been found to increase linearly with distance from the point of attention. However, in audition, one would not expect any such dependence of detection reaction times on distance from the point of attention. This is because the auditory receptors do not map spatial positions topographically (as they do in vision). Detection occurs early in perceptual processing, and at the early levels of the auditory system spatial position is coded digitally (by interaural time, phase, and intensity differences). One would not therefore expect auditory detection times to depend on the spatial position attended. Nor has such a dependence been found (Posner, 1978). Therefore, although a detection task may be preferable on methodological grounds, it is unsuitable for investigating spatial properties of auditory attention. It is only in higher level auditory representations that space could be encoded analogically, and only at those levels would one expect to find spatial constraints on shifts of attention.

EXPERIMENT 1

Method

Subjects. Eight female and 8 male Stanford undergraduates participated in the experiment.

Stimuli and Apparatus. The stimuli were 100-msec 1500-Hz sine waves. Stimulus presentation was controlled by an Apple II microcomputer. The tones were played over nine 7.5-cm-diam speakers, located at equal intervals (22.5°) around the circumference of a semicircle, radius 50 cm, in front of the subject. The end speakers were directly out from the subject's ears, and the middle speaker was directly in front of the subject.

The speakers were labeled 1 through 9. This numbering went from left to right for half the subjects and from right to left for the other half. The task was to name the location of each tone (by giving its number) as quickly and accurately as possible. The subject's voice activated a relay which stopped a clock in the computer. Responses were recorded by the experimenter, and reaction times were recorded by the microcomputer. The experiment was carried out in a sound-attenuating chamber, as were Experiments 2 and 3.

Procedure. At the beginning of the experiment, the subject was shown the nine speakers and told the labels for them. He/she was instructed to say the number of each speaker that was played as quickly and accurately as possible without making any irrelevant vocalizations (e.g., "umm") before responding. The subject was blindfolded for the remainder of the experiment. Head movements were prevented by using a chinrest.

The subject started each trial by pressing a key. Then the 100-msec tone occurred in one of the nine possible locations, and the subject responded by saying the appropriate location as quickly and accurately as possible. Locations were chosen at random by the microcomputer, except that on approximately 29% of trials the last location played was repeated. The subject was told that repetitions were likely and was encouraged to use this information to help him/her respond as quickly as possible. It was assumed that each tone would capture the subject's attention. The incidence of repetitions was raised above chance level (11.1%) to 29%, to encourage subjects to keep their attention on the location of each sound until the next one occurred. Immediately after responding in one trial, the subject initiated the next, new trial.

There were 11 blocks of 45 trials each. The first block was a practice block with accuracy feedback given after every trial. No accuracy feedback was given during the experimental blocks. At the end of each block, the subject was told his/her mean reaction time for the block. All 11 blocks were given in a single session that lasted approximately 1 h.

Results and Discussion

Reaction times. Reaction times for incorrect trials were not included in the analysis. Nor were reaction times from trials immediately following incorrect trials, because it was not clear where subjects were attending at the beginning of such trials (some errors presumably result from failures to focus attention or incorrect labeling of the point of attention). Reaction times longer than 3 sec or shorter than 500 msec were also excluded from the analysis. For the remaining reaction times, a further cutoff point of 3 standard deviations above the mean reaction time for each subject and each distance was established.

In this experiment, the angular distance of a sound from where the subject was attending (i.e., the location of the previous sound) was partly confounded with the location of the sounds. For instance, a distance of 22.5° could have been between any two speakers, whereas a distance of 180° had to be between the end speakers (one and nine). As distance increased, the set of possible speakers decreased. Because people are not equally good at localizing sounds at different eccentricities (Mills, 1958, 1972), the effect of position had to be partialled out before the effect of distance could be assessed.

Consistent with Mills's results, there was a significant effect of location on reaction times to localize sounds at attended positions [$F(8,120) = 2.47, p < .02$]. This effect was partialled out of the reaction times using the following procedure. For each subject, the mean reaction time³ to localize sounds at each of the nine locations, when being attended to, was calculated ($\bar{X}_i, i = 1, \dots, 9$) (see Table 1), provided that at least three data points were available for the location. The number of subjects for whom correction factors could not be calculated at the various locations were as follows: $-90^\circ, 5; -67.5^\circ, 3; -45^\circ, 4; -22.5^\circ, 0; 0^\circ, 0; 22.5^\circ, 0; 45^\circ, 1; 67.5^\circ, 1;$

Table 1
Experiment 1: Mean Reaction Times (in Milliseconds) and
Percent Errors to Localize Sounds at Each Location
When it is Being Attended To

Location	Reaction Time		Percent Errors	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
-90.0°	872.5	48.4	61.6	6.5
-67.5°	944.8	49.6	56.4	4.9
-45.0°	967.4	49.0	43.7	5.4
-22.5°	928.7	47.1	13.0	3.0
0.0°	858.7	50.4	8.8	2.2
22.5°	931.4	44.4	19.9	4.2
45.0°	956.1	50.4	35.0	5.1
67.5°	985.9	65.5	38.2	3.7
90.0°	905.0	56.0	46.2	6.8

Note—*M* = mean, *SE* = standard error. Locations are measured in degrees of eccentricity to the left (negative numbers) or right (positive numbers) of straight ahead.

90°, 3. The grand mean for all locations was also calculated (\bar{X}_T). A correction term was obtained for each location by subtracting the mean for that location from the overall mean ($\bar{X}_T - \bar{X}_i$, $i = 1, \dots, 9$). The correction term was added to all reaction times for sounds at that location. This correction procedure should ensure that any effect of the distance of a sound from the previous sound would be independent of the location of the sounds.

Figure 1 shows mean (corrected) reaction times as a function of the angular distance from the point of focal attention. Reaction times increased from 917.9 msec at a distance of 0° up to 1,267.0 msec at a distance of 135°. The increase was monotonic (except for a small, nonsignificant decrease at 90°), and over this range of distances, the data was well fit by the linear function

$$RT = 957.7 + 2.5D, \quad (1)$$

where RT is reaction time measured in milliseconds and *D* is the angular distance from the point of attention measured in degrees. This function accounted for 91% of the variance. The contrast weights -3, -2, -1, 0, 1, 2, 3 were applied to the mean reaction times to test for a linear trend. There was a highly significant linear trend [$F(1,90) = 145.44$, $p \ll .001$]. However, the residual was also significant [$F(5,90) = 2.98$, $p < .05$], indicating that the linear trend did not completely characterize the data. The contrast weights -5, 0, 3, 4, 3, 0, -5, were used to test for an orthogonal quadratic trend. The quadratic effect was significant [$F(1,90) = 9.23$, $p < .01$]. There was no significant residual effect over and above the linear and quadratic effects [$F(4,90) = 1.43$, n.s.].

Reaction times to sounds at the largest distance from the point of focal attention (180°) were relatively fast. These

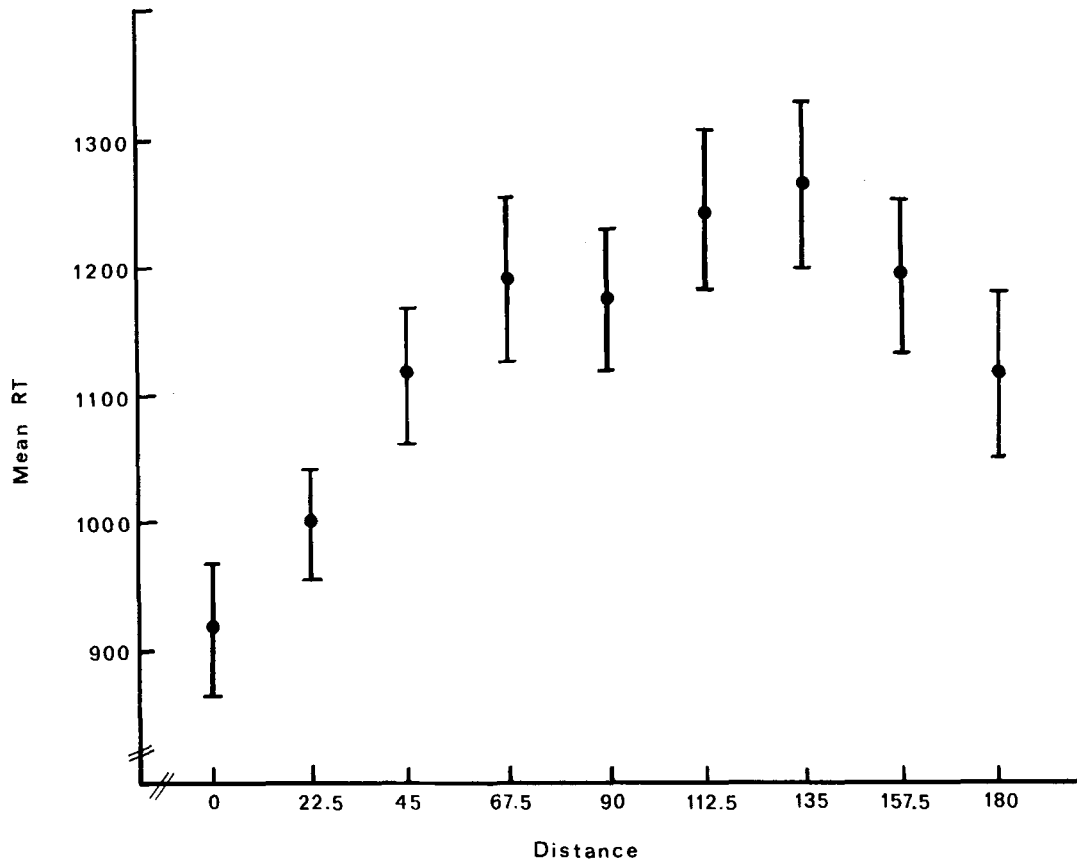


Figure 1. Experiment 1: Mean reaction times (in milliseconds) to localize sounds as a function of angular distance (degrees) from the focus of attention. Standard error bars are shown.

fast reaction times may have occurred because the responses were to the end speakers of the array, each of which had only a single adjacent speaker,⁴ thereby reducing the number of likely alternative locations from 2 to 1. Another possible explanation for the fast reaction times for 180° shifts is that there is a special relationship between diametrically opposed locations that makes it relatively easy to shift attention between them. The latter hypothesis was tested in Experiment 2. Although both hypotheses appear plausible, examination of the error data (see Table 1) suggested that the fast reaction times may have simply been due to a speed-accuracy tradeoff, because the localization error rates for the most eccentric locations (which are involved in 180° shifts) were very high.

The error data (see Table 1) indicated that subjects could not consistently discriminate between the two most eccentric locations on each side (i.e., 67.5° vs. 90°). Therefore, the data were reanalyzed excluding trials on which sounds were presented at, or immediately following sounds presented at, the two most eccentric locations on either side of the speaker array (see Figure 2). Mean reaction times increased monotonically from 919.3 msec at a distance of 0° to 1,294.1 msec at a distance of 90°. The increase was linear, and the function

$$RT = 903.7 + 4.3D \quad (2)$$

accounted for 99% of the variance. The contrast weights -2, -1, 0, 1, and 2 were applied to the mean reaction times to test for a linear trend. The linear trend was highly significant [$F(1,60) = 141.40, p \ll .001$], and there was no significant residual [$F(3,60) = .52$]. Thus, the relationship between reaction time and distance of a sound from the point of focal attention was completely characterized by Linear Function 2. The inverse of the slope of this function can be used as an estimate of the rate at which attention was shifted from one part of the representation to another. A rate of 233°/sec was obtained.

Errors. The percentage of trials on which sounds were incorrectly localized was quite high in this experiment ($M = 36\%$, $SD = 9\%$). The percentage of localization errors increased from a minimum of 9% of the trials on which the sounds were directly in front of the listener to 54% for the most eccentric sounds (90°) (see Table 1). These data confound location with distance from the point of attention. It would be better to examine errors as a function of location, when the subject was attending to that location. However, there were essentially no errors (3%) for locations that the subject was attending to. The increase

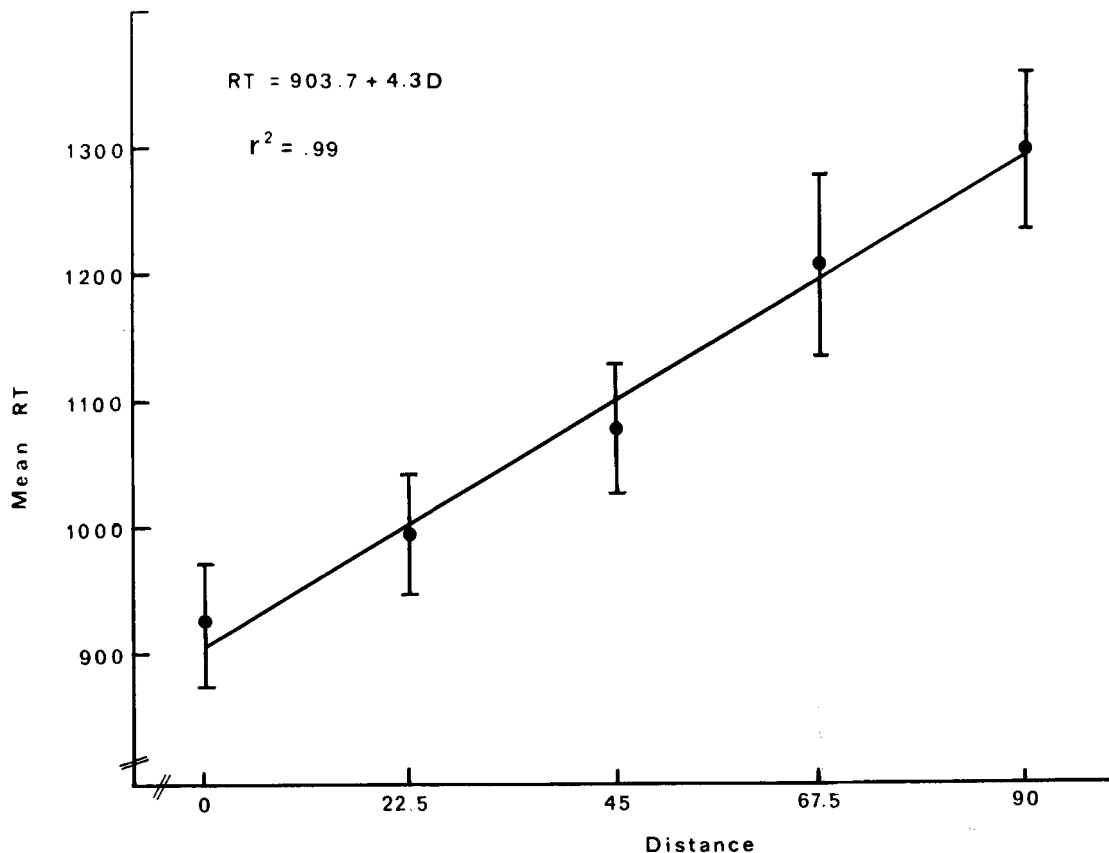


Figure 2. Experiment 1: Mean reaction times (in milliseconds) to localize sounds as a function of angular distance, D (degrees), from the locus of attention (excluding the two most eccentric locations on each side of the speaker array). Standard error bars and the best-fitting straight line through the points are shown.

in localization errors with increasing eccentricity was linear, and the function

$$E = 8.78 + 0.54L, \quad (3)$$

where E is percent errors and L is location measured in degrees of eccentricity (symmetric left-right positions averaged), accounted for 95% of the variance. The contrast weights $-2, -1, 0, 1,$ and 2 were applied to test the linearity of the increase. There was a significant linear trend [$F(1,60) = 113.58, p \ll .001$] and no significant residual [$F(3,60) = 1.92, n.s.$]. Thus, Linear Function 3 completely characterized the data. These results are consistent with Mills's (1958, 1972) finding that localization is most accurate for sounds that are straight ahead, and that it decreases with the increasing eccentricity of a sound. Given that subjects confused only adjacent locations,⁵ chance performance was 50% correct for the most eccentric locations (which have only one adjacent location) and 33% correct for the other locations. Localization of the most eccentric sounds (90° ; 46.3% correct) was no better than chance [$t(15) = -.14, n.s.$]. Localization was better than chance for the other eccentricities: $M = 91.2\%$ (0°), $M = 83.8\%$ (22.5°), $M = 61.2\%$ (45°), $M = 52.5\%$ (67.5°) (all $t_s > 6.57$).

Figure 3 shows percent errors as a function of distance when sounds presented at or immediately following sounds presented at the two most eccentric locations on each side (67.5° and 90°) were excluded (cf. Figure 2 for reaction times). The linear function

$$E = 1.55 + .46D \quad (4)$$

accounted for 98% of the variance. The linear trend, tested with the contrast weights $-2, -1, 0, 1,$ and 2 , was significant [$F(1,60) = 58.51, p \ll .001$], and the residual effect was not significant [$F(3,60) = .36, n.s.$]. Thus, like the reaction time data, errors increased linearly with the distance of the sound from the point of attention.

There was a large, significant, positive correlation [$r(6) = .97$] between the error rates and the corresponding mean reaction times. This suggests that the increase in reaction times with distance from the focus of attention was not due to a speed-accuracy tradeoff.

EXPERIMENT 2

It is possible that the increase in reaction times with distance from the point of focal attention in Experiment 1 resulted from the differential probability of occurrence

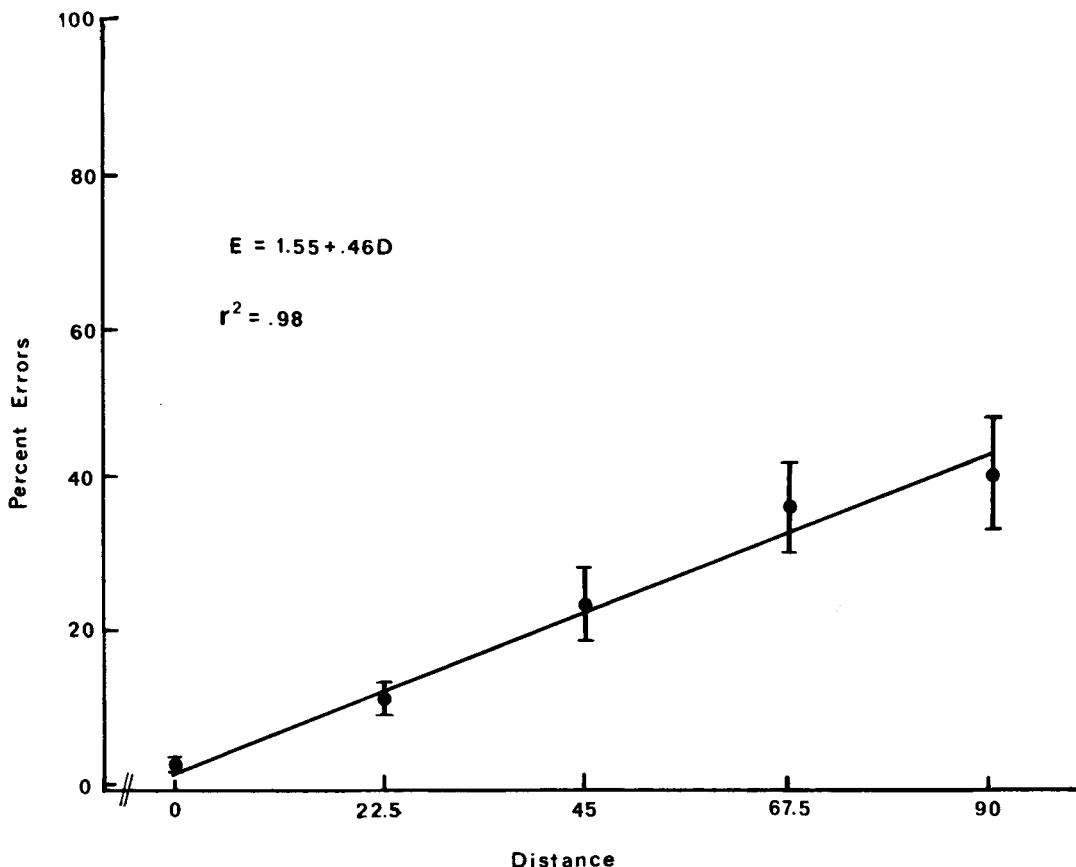


Figure 3. Experiment 1: Percent errors (E) in localizing sounds as a function of angular distance, D (degrees), from the focus of attention (excluding the two most eccentric locations on each side of the speaker array). Standard error bars and the best-fitting straight line are shown.

of trials at different distances. Because sound locations were chosen randomly there were progressively fewer trials at greater distances. Therefore, it is possible that the reaction time function was due to general expectancies about the likelihood of different-sized shifts, rather than to any active shifting of attention. Experiment 2 was designed to replicate the results of Experiment 1, using a paradigm in which location and distance were completely independent.

A circular array of speakers that surrounded the subject was used. With this arrangement, shifts of different distances were equally likely, except for 180° shifts, which were twice as likely as the others (because one could go halfway around the circle in either direction), and shifts of 0°, whose probability was deliberately raised (see below). According to the expectancy hypothesis, more 180° shifts should lead to shorter reaction times, whereas the attention-shifting model would predict relatively long reaction times for 180° shifts.

Another advantage of this arrangement over that used in Experiment 1 is that large shifts are no more likely than small shifts to involve eccentric locations, which are difficult to localize. The use of a circular array of speakers does, however, have one disadvantage. It is difficult to distinguish between symmetrically placed sound sources in front and back (Stevens & Newman, 1936). It was because of this difficulty that a semicircular speaker arrangement was used in Experiment 1. However, subsequent pilot testing indicated that front-back confusions could be kept to an acceptably low level if a complex sound involving many harmonics was used. The speakers were also arranged so that symmetric front and back arrangements were avoided.

Method

Subjects. Twelve women and 8 men were subjects in the experiment. These included Stanford undergraduates, who participated to fulfill a course requirement for Introductory Psychology, and others, who were paid for their time.

Stimuli and Apparatus. Eight speakers were arranged in a circle, radius 77 cm, around the subject. The speakers were placed at 8 of the 12 possible clock positions. Each of the 20 subjects had a different speaker arrangement. Speaker arrangements were chosen according to the following scheme. Four basic arrangements, each consisting of the 12, 3, 6, and 9 o'clock positions plus four of the other possible positions, were used. The other four positions were chosen so that there was one from each quadrant and so that there were no arrangements with symmetric front-back positions filled (except for 6 and 12). This produced the four basic arrangements mentioned above. From each one of these, four more arrangements were created by shifting the speakers at 12, 3, 6, and 9 o'clock, respectively, into adjacent unfilled positions. This resulted in 20 different circular arrangements of eight speakers.

The stimuli were 100-msec squarewave tones with a fundamental frequency of 2000 Hz. Square waves were used because they were easier to localize than sine waves (Experiment 1). The subjects labeled the speakers vocally according to their clock positions.

Procedure. The procedure was the same as for Experiment 1, except for the following differences, which were designed to help increase the subjects' accuracy. Practice blocks were given, with accuracy feedback after every trial, until a criterion of 2 consecu-

tive blocks with more than 70% correct responses or a maximum of 5 practice blocks was reached. The total number of blocks (including practice) was 11. Instead of giving reaction time feedback as in Experiment 1, accuracy feedback (number of errors made in the block) was given after each experimental block.

The same instructions were given as in Experiment 1 except that the subjects were explicitly told to keep their attention focused on the location of each sound until they heard the next one. They were told that this would help them to respond quickly and accurately because the sound would occur in the same place on consecutive trials over 40% of the time. Repetitions actually occurred on approximately 42% of the trials (cf. 29% in Experiment 1).

Results and Discussion

Reaction Times. As in Experiment 1, reaction times for incorrect trials or trials immediately following incorrect trials were excluded from the analysis. Reaction times exceeding 3,000 msec or 3 standard deviations above the subject's mean for the remaining data for a given distance were also excluded.

Figure 4 shows mean reaction times as a function of distance from the point of focal attention. Mean reaction times increased linearly over the range 0°-90°. The linear function

$$RT = 778.0 + 2.6D \quad (5)$$

accounted for 96% of the variance. There was a significant linear trend (assessed using the contrast weights -3, -1, 1, 3) [$F(1,57) = 91.86, p \ll .001$] and no significant residual [$F(2,57) = 2.06, n.s.$]. Therefore, the Linear Function 5 completely described the relationship between reaction times and distance from the point of attention. This replicates the results of Experiment 1 in a paradigm in which distance and location are unconfounded. In addition, the probability of the occurrence of different size shifts was equated (except at 0° and 180°), so the effect could not be due to general expectancies about the likelihood of the occurrence of sounds at different distances from the point of attention.

From Equation 5 it can be seen that attention was shifted at a rate of 385°/sec (the inverse of the slope) over the range 0°-90°, which is faster than the rate obtained in Experiment 1 (233°/sec). This is probably because the task was easier (as indicated by faster reaction times and lower error rates than in Experiment 1), thereby leaving more processing resources available for controlling attentional shifts.

For distances of 90°-180°, reaction times were essentially constant at 989.4 msec. Therefore, the linear increase in reaction times with increasing distance from the point of focal attention appears to be restricted to distances of 90° or less. There was no evidence in Experiment 2 for a decrease in reaction times for sounds 180° from the point of attention. This suggests that the decrease in reaction times for 180° shifts found in Experiment 1 did not reflect any special facility for 180° shifts.

Errors. The mean percentage of trials on which sounds were incorrectly localized was 17% ($SD = 2\%$). This was lower than in Experiment 1. More accurate performance

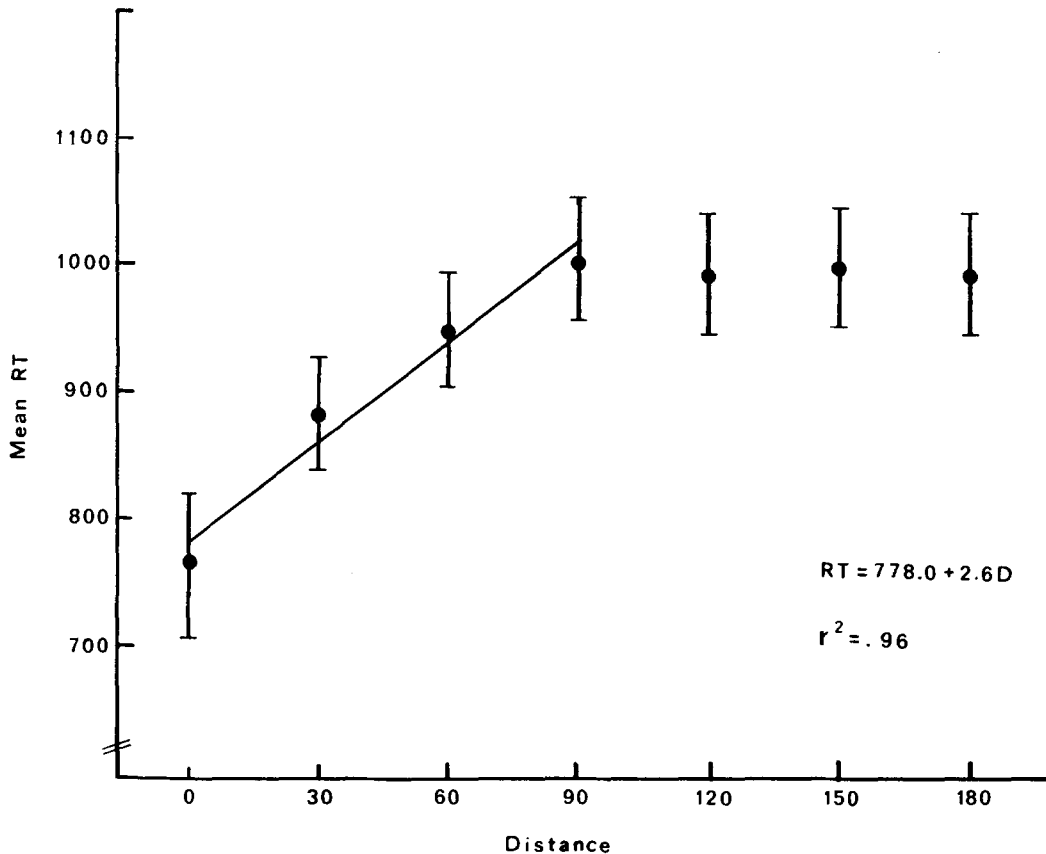


Figure 4. Experiment 2: Mean reaction times (in milliseconds) to localize sounds as a function of angular distance, D (degrees), from the focus of attention. Standard error bars and the best-fitting straight line over the range 0° - 90° are shown.

and faster reaction times suggest that localization was easier than in Experiment 1. Thus, the changes in procedure made to improve localization performance (complex tones were used, subjects had more practice, accuracy feedback was given, and a greater angular separation between the speakers was used) appear to have been successful.

Figure 5 shows the percentage of sounds at each distance that were incorrectly localized. There was a significant effect of distance on the error percentages [$F(6, 114) = 3.90, p < .002$]. There were more incorrect localizations at intermediate distances (90° and 120°) than at smaller or larger distances. There was no evidence of a speed-accuracy tradeoff.

Localization errors were more frequent for sounds at the side than for those in front of or behind the listener (see Table 2). This is consistent with Mill's (1958, 1972) finding that accuracy of localizing sounds (measured in terms of the minimal angle at which the left-right relationship of two sounds could be correctly discriminated more than 50% of the time) decreased as the eccentricity of the sound increased.

The results of Experiments 1 and 2 indicate that the time taken to localize a sound increased linearly with its distance from the listener's point of focal attention, up to a distance of 90° . This is consistent with the view that auditory spatial information is topographically organized,

and that this structure constrains the way in which attention is moved through the representation. For distances up to 90° , the time taken to move attention through the representation depends directly on the distance to be moved. For distances greater than 90° , however, this is no longer the case. The reaction time function asymptotes at around 90° . It may be that for these larger distances, it is no longer efficient to carry out controlled shifts of attention, which presumably require capacity. Lansman, Farr, and Hunt (1984) have shown that attention based on cuing an expected location, as in the present experiments—but not attention based on stimulus probabilities—requires capacity. Consistent with their results, attention was shifted more slowly in Experiment 1 ($233^{\circ}/\text{sec}$), which had a more difficult localization task, than in Experiment 2 ($385^{\circ}/\text{sec}$), which had an easier localization task. When a sound is relatively far (greater than 90°) from the listener's focus of attention, it may be faster for attention to pop up in the region of the distant sound than for it to shift over a large intervening space.

EXPERIMENT 3

It is possible that the increase in reaction times with distance from the point of attention observed in Experiments 1 and 2 was due to the use of a response set (num-

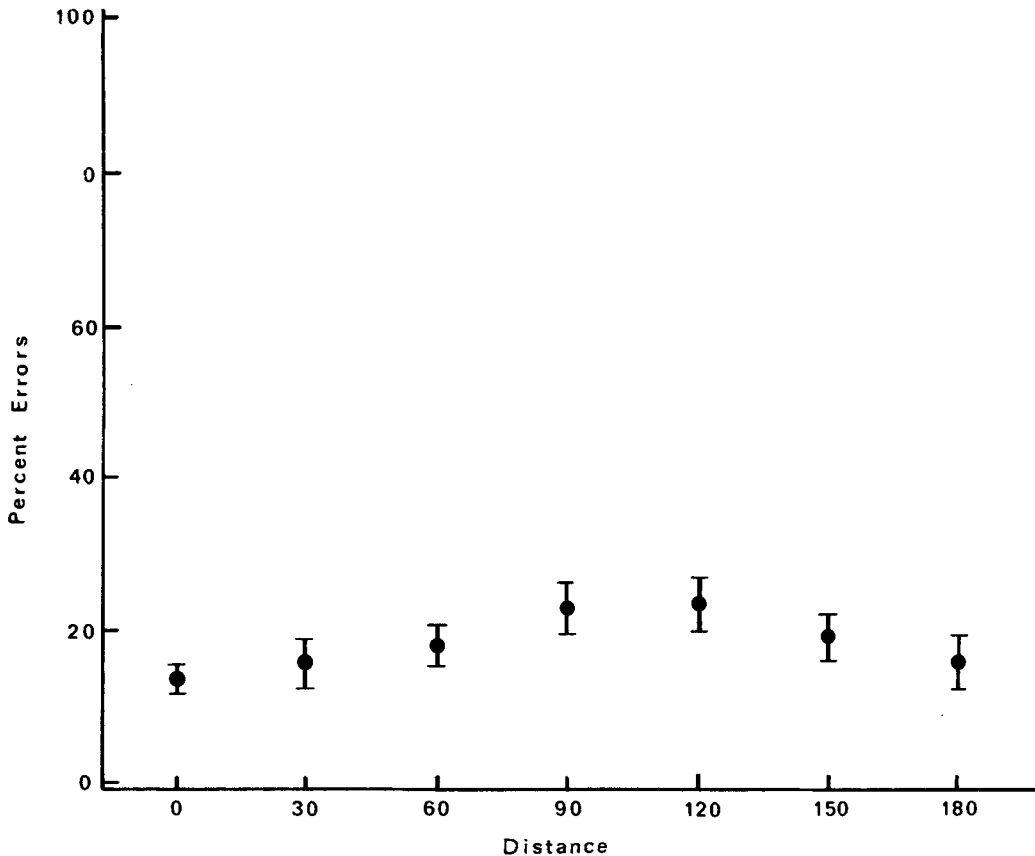


Figure 5. Experiment 2: Percent errors in localizing sounds as a function of angular distance (degrees) from the focus of attention. Standard error bars are shown.

bers) that was already ordered in memory. Marcel and Forrin (1974) found that the time taken to name a digit from the set 2-9 increased with its ordinal distance from the preceding digit named. It should be noted that the size of the digit effect is at least an order of magnitude smaller than the distance effect found in Experiments 1 and 2. This difference suggests that the use of digits as the response set is unlikely to completely account for the distance effect.

Digit effects can be avoided by the use of an arbitrary response set. Such a procedure also makes it possible to determine the effect of a memory load on the rate at which attention can be shifted, because the arbitrary response set must be held in working memory. If analogical shifts of attention require capacity, then a concurrent memory load should reduce the capacity available for attentional shifts and slow their rate. In Experiment 3, it was predicted that attention would shift more slowly when an arbitrary response set (names), which must be held in memory, was used than when an overlearned response set (numbers) was used. It might also take longer to retrieve the correct response when arbitrary responses are used. This would affect the intercept of the function relating reaction time and distance from the point of focal attention, but not the slope, which reflects the rate at which attention was shifted.

Method

Subjects. Ten female and 10 male Stanford undergraduates participated in the experiment.

Stimuli and Apparatus. The stimuli were 100-msec-duration 1500-Hz sine waves. Five speakers were placed at $\pm 45^\circ$, $\pm 22.5^\circ$, and 0° on a semicircle, radius 50 cm, in front of the subject. The more eccentric locations, which are difficult to localize, were not

Table 2
Experiment 2: Mean Reaction Times (in Milliseconds)
and Percent Errors to Localize Sounds at Each
Location When it is Being Attended To

Location	Reaction Time			Percent Errors		
	<i>M</i>	<i>SE</i>	<i>n</i>	<i>M</i>	<i>SE</i>	<i>n</i>
0°	737.9	59.8	16	5.2	1.7	16
30°	749.8	56.1	12	2.8	1.5	12
60°	739.0	97.3	12	30.6	8.9	12
90°	848.0	72.1	16	25.1	5.1	16
120°	906.1	72.8	12	37.4	6.1	12
150°	757.3	74.5	12	5.6	2.1	12
180°	808.9	70.8	16	3.2	1.2	16
210°	757.4	62.3	12	5.3	1.5	12
240°	842.7	71.7	12	29.1	7.0	12
270°	702.1	59.5	16	10.3	3.0	16
300°	794.9	94.2	12	21.5	5.4	12
330°	664.4	66.7	12	3.3	2.4	12

Note—*M* = mean, *SE* = standard error, *n* = number of subjects tested at each location. Locations are measured in degrees of eccentricity from straight ahead.

used. The waveform and speaker separation were the same as those used in Experiment 1 so that the results of this experiment would be comparable to those obtained in Experiment 1.

Procedure. The procedure on each trial was the same as for Experiment 1. The subjects were instructed to localize each tone that was played and to name the speaker it came from as quickly and accurately as they could. Attention was to be focused on the location of the previous tone until the next tone was played. The subjects were told that this would help them to respond quickly and accurately because the next tone was likely to be played over the same player. The repetition rate was approximately 40%.

In the name condition, Smith, Jones, Brown, Hart, and White were used as the speakers labels. For each subject, these names were assigned randomly to the five speakers. In the number condition, the subjects labeled the speakers using the numbers 1 through 5. Half the time, the speakers were labeled 1 through 5 from left to right, and half the time, from right to left.

There were 14 blocks of 45 trials each. The first 4 blocks were for training and practice. During the first 2 blocks, the subject was taught the names or numbers for the five speakers. For ease of reference, I will use "name" to refer to either the speaker name or number, depending on which condition the subject was in. As each tone was played, the experimenter named it and the subject repeated the name. During Block 3, the subject attempted to name each speaker as it was played. Accuracy was emphasized, and the subject was encouraged to take as long as he/she needed to name the speaker on each trial. Accuracy feedback was given on every trial. At the end of Block 3, the subject was told his/her average reaction time. At the beginning of Block 4, the subject was instructed to respond as quickly and accurately as possible. The subject was also instructed to attend to the location of the previous tone until the next tone occurred. On Blocks 4-14, accuracy feedback was given after a trial only if the subject made an error on that trial. Block 4 was treated as a practice block. The data from Blocks 5-14 (10 blocks) were used in the analyses. At the end of every block, subjects were told their mean reaction times and the number of errors they had made for the block.

Results and Discussion

Reaction times. As in the previous experiments, only reaction times from correct trials following a correct trial were used in the analyses. All reaction times below 300 msec or above 3,000 msec were excluded, as were reaction times more than 3 standard deviations above the mean of the remaining data for the appropriate distance. There was a significant effect of location [$F(4,72) = 4.14$, $p < .005$; see Table 3], which was partialled out using the same procedure as in Experiment 1.

Table 3
Experiment 3: Mean Reaction Times (in Milliseconds) and Percent Errors to Localize Sounds at Each Location When it is Being Attended To

Location	Reaction Time		Percent Errors	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
-45.0°	715.6	22.9	7.1	1.0
-22.5°	729.7	22.9	9.5	1.3
0.0°	705.2	21.5	7.8	1.0
22.5°	727.9	20.8	7.6	1.4
45.0°	699.8	20.4	11.4	1.2

Note—*M* = mean, *SE* = standard error. Locations are measured in degrees of eccentricity to the left (negative numbers) or right (positive numbers) of straight ahead.

Mean corrected reaction times were subjected to an analysis of variance with response condition (numbers or names) as a between-subjects factor and distance from the point of attention as a within-subject factor. Figure 6 shows mean reaction times as a function of distance from the point of attention. There was a significant effect of distance [$F(4,72) = 83.10$, $p \ll .001$], with longer reaction times for greater distances (except at 90°). There was no effect of response condition [$F(1,18) = 3.50$, $p > .08$], but this factor interacted with distance [$F(4,72) = 11.15$, $p \ll .001$], so that reaction times increased more steeply with distance when names rather than numbers were used as the response (see Figure 6). There were significant linear and quadratic trends for both the name condition [linear $F(1,36) = 164.52$, $p \ll .001$; quadratic $F(1,36) = 49.78$, $p < .001$; leftover $F(2,36) = .19$, n.s.] and the number condition [linear $F(1,36) = 103.44$, $p \ll .001$; quadratic $F(1,36) = 15.71$, $p < .001$; leftover $F(2,36) = 3.10$, n.s.]. For both the name and number response conditions, reaction times increased monotonically up to 67.5°. Over this range, the linear function

$$RT = 726.9 + 2.2D \quad (6)$$

accounted for 98% of the variance in the number condition, and the linear function

$$RT = 735.2 + 4.4D \quad (7)$$

accounted for 93% of the variance in the name condition. The intercepts of these two functions were very similar, indicating that when a location was being attended to the time taken to localize a sound there did not depend on whether the required response was from a newly learned arbitrary response set (names) or from an overlearned, ordered response set (numbers). This suggests either that the retrieval time was identical for names and numbers or that retrieval had already occurred because the subject was expecting the sound to recur at the same location.

Although there was no difference between response conditions for a distance of zero and no main effect of response condition, response condition did influence the slope of the function relating reaction times and distance, and therefore the rate at which attention was shifted. Using the inverse of the slope of these functions as an estimate of the rate at which attention was shifted gave rates of 455°/sec and 227°/sec for the number and name response conditions, respectively. Thus, the rate at which attention shifted when there was a memory load (name responses), was about half what it was when there was no such load (number responses). These rates were calculated for shifts of attention over the range of 0°-67.5°, where the increase in reaction times with distance from the point of attention was linear. It could be argued that for a distance of 0°, where the sound is presented in a location that the subject is already attending to and for which the correct response may have already been retrieved, should be excluded. If the 0° point is removed, the relationship between reaction time and distance from

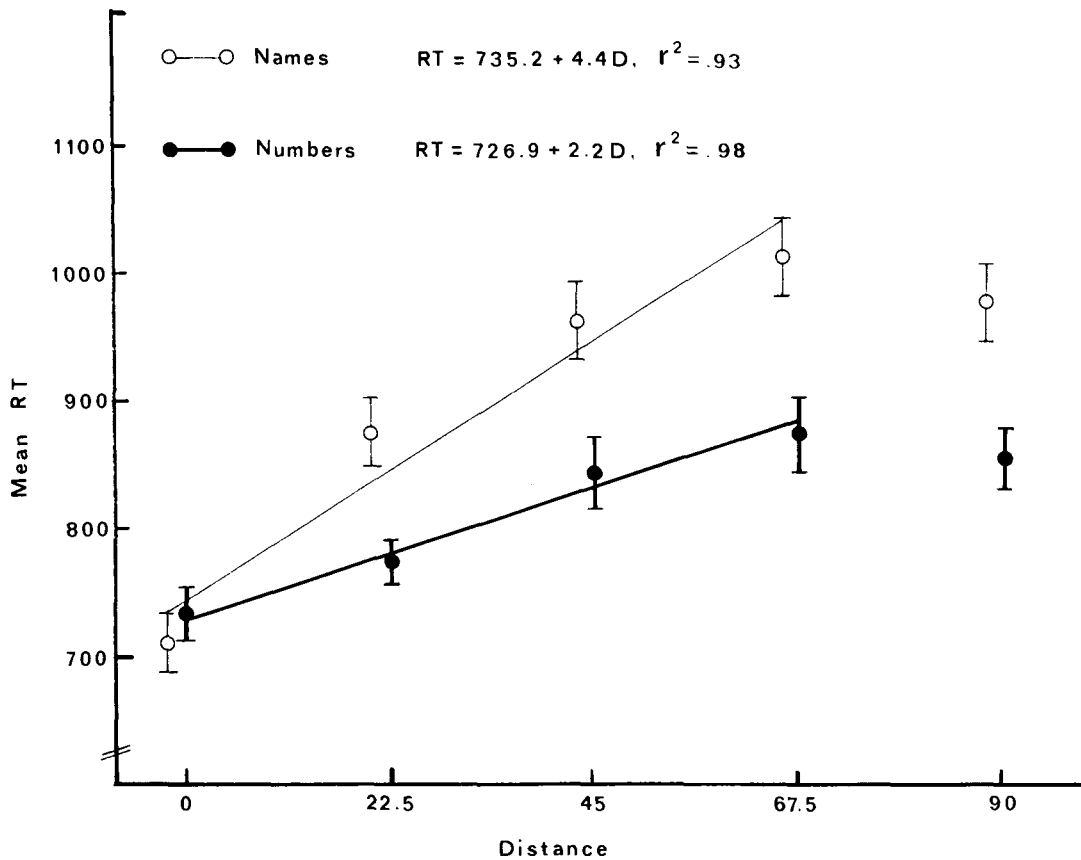


Figure 6. Experiment 3: Mean reaction times (in milliseconds) to localize sounds as a function of angular distance, D (degrees), from the focus of attention for each response type (names or numbers). Standard error bars and the best-fitting straight line over the range 0° - 67.5° are shown.

the point of attention over the range of 22.5° - 67.5° can be described by the linear functions

$$RT = 724.1 + 2.3D \quad (6a)$$

$$RT = 809.5 + 3.0D. \quad (7a)$$

These account for 95% and 97% of the variance of the number and name data, respectively. The rate at which attention was shifted in the two conditions was $435^\circ/\text{sec}$ (numbers) and $333^\circ/\text{sec}$ (names). Thus, attention was shifted more slowly in the name condition where there was a concurrent memory load, even if the 0° point was excluded.

In both the name and number conditions, reaction times were relatively fast for 90° shifts, that is, for shifts between the endpoints of the array (cf. 180° shifts in Experiment 1). It is not clear why this should occur. One possibility is that the endpoints may be easier to localize because there is only one adjacent speaker instead of two. This is consistent with the fact that there was no advantage for 90° shifts between $+45^\circ$ and -45° in Experiments 1 and 2, where these were not endpoints. However, the correction procedure should have eliminated any advantage for localizing sounds at the end of the array. Another possibility is that it is particularly easy to shift

attention from one end of the array to the other. Although there is no obvious reason why this should be so, it is at least consistent with the relatively fast reaction time to shift attention between the endpoints of the array (180° shifts) found in Experiment 1. A fast shift between the endpoints of the array would not be partialled out by the correction procedure because that procedure was based on reaction times to locations at the point of focal attention (i.e., when no shift of attention was necessary).

It could be argued that different slopes were obtained for the name (memory load) and number (no memory load) response conditions because subjects serially retrieved the responses (mentally) for positions between the point of focal attention and the target sound. This argument requires that the name responses take longer to retrieve than the number responses. There was no evidence for this. Reaction times to locate sounds at the point of attention did not depend on the response type. However, as noted above, it is possible that in this case the response had already been retrieved. If the response retrieval model were correct one would also expect the slope of the function in the number condition to be similar to that in Marcel and Forrin's (1974) digit-naming experiment. However, the slopes (which would depend on retrieval time) differed by an order of magnitude. Most

importantly, a serial response retrieval or priming explanation would predict that the time required for a shift of attention should depend on the number of intervening locations rather than on the angular separation of the initial and target locations. Rates of attention shifting as a function of number of intervening locations were calculated for Experiments 1, 2, and 3 (number responses only). The rates obtained were 10.6, 8.4, and 21.3 locations/sec (number responses only) for Experiments 1-3, respectively. According to the response priming explanation, these rates should be identical, except for differences correlated with the difficulty of the task (which would affect the resources available for response retrieval or priming). However, these rates were not identical and they were not correlated with task difficulty. The localization task was much easier in Experiment 2 than in Experiment 1 (compare reaction times and errors), yet it had a slower rate. Examination of reaction time and error data suggests that the order of difficulty of the tasks was: Experiment 1 (most difficult), Experiment 2, and Experiment 3 (least difficult). The rates as a function of angular distance (233°/sec, 385°/sec, 455°/sec for Experiments 1, 2, and 3, respectively) are consistent with this ordering, whereas the rates as a function of location are not.

Errors. The overall percentage of trials on which sounds were incorrectly localized was quite low ($M = 8\%$, $SD = 3\%$). An analysis of variance was carried out on the percentage of errors at each eccentricity (see Table 3), with response condition as a between-subjects factor and eccentricity as a within-subject factor. There was no significant effect of response condition [$F(1,18) = 2.27$, n.s.; $M = 11\%$, names; $M = 8\%$, numbers] or eccentricity [$F(2,36) = 1.64$, n.s.; $M = 8\%$, 0° ; $M = 9\%$, 22.5° ; $M = 12\%$, 45°] and no interaction between the two [$F(2,36) = 1.02$, n.s.].

Figure 7 shows the percentage of errors (averaged across response conditions) as a function of distance from the focus of attention. The percentage of errors increased with distance up to 67.5° and dropped (nonsignificantly; difference = 4%, $HSD = 7\%$) at 90° . The linear function

$$E = 2.32 + .28D \quad (8)$$

accounted for 99% of the variance of the error data over the range 0° - 67.5° . Thus, errors showed the same relationship to distance as did reaction times (see Figure 6), indicating that there was no speed-accuracy tradeoff. An analysis of variance indicated a significant effect of distance [$F(4,72) = 2.89$, $p < .03$] but no effect of response

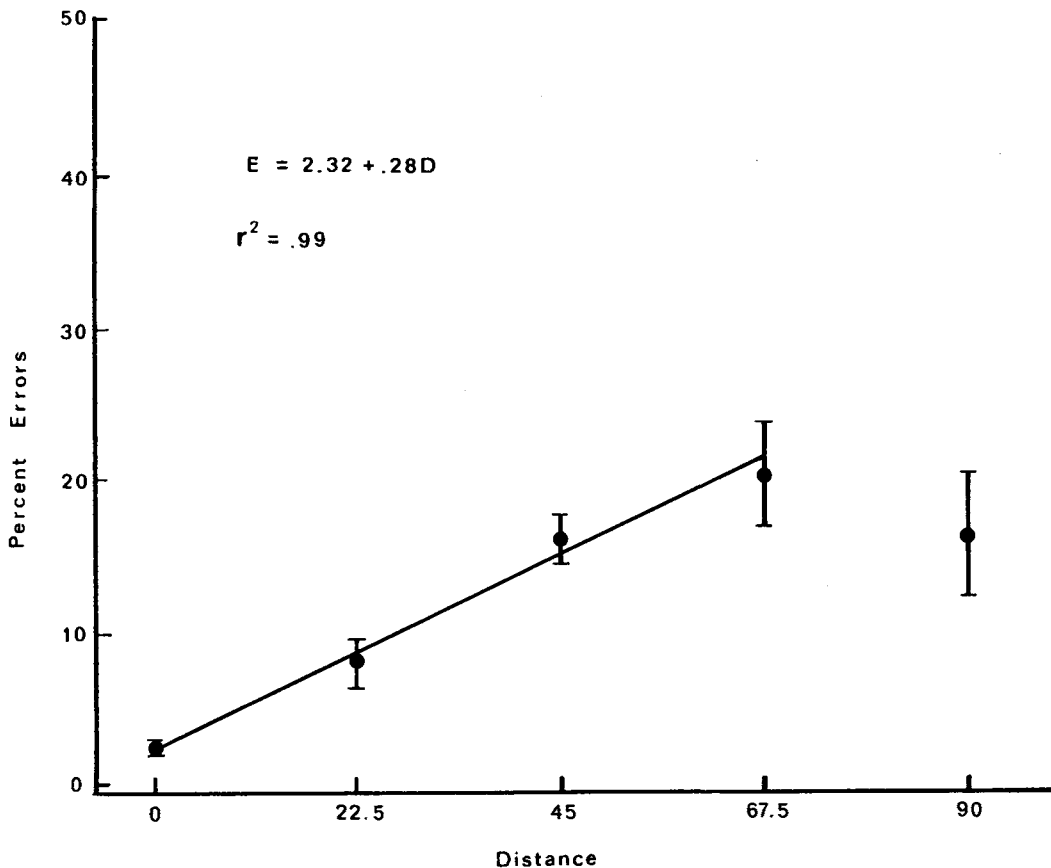


Figure 7. Experiment 3: Percent errors (E) in localizing sounds as a function of angular distance, D (degrees), from the focus of attention. Standard error bars and the best-fitting straight line over the range 0° - 67.5° are shown. Note the difference in scale of ordinate from that of Figures 3 and 5.

condition [$F(1,18) = .71$, n.s.]. Nor was there any interaction between response condition and distance [$F(4,72) = 1.15$, n.s.].

GENERAL DISCUSSION

The results of the three experiments reported here show that the time taken to localize a sound increases linearly with its distance from the point of attention for distances up to 90° (when end effects are not present). These results parallel those obtained for shifts of visual attention and are consistent with the view that auditory spatial information, like visual spatial information, may be represented topographically or analogically. The analogical nature of the representation appears to constrain the way attention can be shifted within the structure. In particular, it takes time to move from one part of the representation to another.

The time taken to shift attention depends on the angular separation between the initial and target locations and not on the number of intervening locations. This result rules out serial response priming or retrieval explanations of the linear relationship between response time and distance from the point of focal attention. Moreover, the linear relationship cannot be accounted for by subjects' expectancies about where sounds are likely to occur (Experiment 2). These experiments therefore provide the first evidence that auditory attention exhibits spatial properties. The results are consistent with the hypothesis that representations of auditory space are analogical and that the spatial structure of such representations constrains the behavior of auditory attention.

The analogical nature of auditory spatial representations is especially interesting because, unlike the visual case, there are no constraints from the topographical structure of the receptor surface that could lead to this type of organization. Rather, the need to integrate spatial information from a number of modalities (together with the special status of visual information in human information processing) might be responsible for the analogical representation of auditory spatial information. The attentional effects on sound localization reported here may be occurring in a high-level representation that contains multimodal spatial information (O'Leary & Rhodes, 1984). This hypothesis could be tested directly by examining cross-modal attentional cuing effects.

The time taken to shift attention did not continue to increase for distances greater than 90° . Why should this be so? One possibility, noted earlier, is that the capacity required for a large shift is greater (or the time required is longer) than that required to simply reallocate attention discretely. Thus attention may simply pop up anew in the distant location.

Another possible explanation of the 90° asymptote for reaction times is suggested by the functional utility of analogical shifts of attention in tracking continuously moving objects (Shepard, 1981). It should not, therefore, be surprising that such shifts fail to occur when two sounds

are separated by a large distance but are close in time, because they will be unlikely to have the same source. This may be analogous to apparent motion in which the distance between two lights determines the interstimulus interval required for apparent motion of a single object to be perceived (Korte, 1915; Shepard & Zare, 1983). If the time interval is too long relative to the distance between the lights, then two separate lights flashing on and off will be seen. Similarly, the distance over which analogical shifts of attention occur may be a function of the time relations between successive stimuli. These time relations would determine whether a stimulus was perceived as a single object moving in space (for which such shifts of attention would be appropriate) or as two distinct objects in different locations (in which case attention would pop up discretely at the new location). This hypothesis is consistent with reports of auditory apparent motion in which stimuli are organized into sound objects (Briggs & Perrott, 1972) on the basis of space-time relations, with reports of auditory stream segregation in which the organization is based on analogical pitch-time relations (Bregman, 1978, 1981; Bregman & Campbell, 1971; Bregman & Rudnick, 1975), and with Kubovy's (1981) arguments that pitch is analogous to a visual spatial extent dimension.

The 90° limit for spatial constraints on shifts of attention found in the present experiments may have been due to the particular interstimulus intervals (ISIs) that occurred. The ISIs were subject-controlled. Subjects typically initiated each trial (stimulus) immediately after responding to the last sound. Thus, the ISIs were similar to the response times (around 1 sec), which were short enough to have affected the auditory system's decision about whether motion had occurred. It is interesting to note that in Experiment 3, in which reaction times, and therefore ISIs, were shorter, the curve appears to flatten at 67.5° rather than at 90° . This is consistent with the "time" explanation. However, it is also possible that the flattening is due to an endpoint effect of the type mentioned earlier. Experiments are needed in which time and distance between successive sounds are parametrically varied in order to determine whether these parameters are lawfully related.⁶

An alternative possibility is that the 90° limit is imposed by constraints from the representation of visual information. If auditory spatial information is topographically organized and integrated with visual spatial information in a multimodal representation, then the auditory map may be constrained by properties of the visual representation. This appears to be the case in owls (Knudsen, 1982). In vision, humans have a view that subtends almost 180° , and visual representations cover essentially this range. It is possible that when attention is focused on a location, one might have a 180° "view" centered on that location. If a perceptual "object" (either visual or auditory) at the attended location moves to a new location within this 180° window, then its path should be followed by an analogical shift of attention. If it moves out of the window, then

a discrete shift should occur. This model predicts that a shift of up to 90° in either direction from the attended location would be followed by an analogical shift of attention, whereas for larger movements attention would just pop up in the new place.

To summarize, these experiments provide the first evidence that auditory attention exhibits spatial constraints. The time taken to localize a sound increased linearly as a function of its distance from the point of attention (up to 90° shifts, provided that end effects were not present). This spatial constraint is consistent with the hypothesis that for sounds to be localized, attention is shifted through an analogical or topographical representation. Alternative explanations based on general expectancies, location probabilities, or response-level effects (e.g., serial retrieval or priming of responses) were ruled out. Further experiments are required to confirm that the spatial constraints on attention are indeed analogical, that is, that attention traverses intervening space when it shifts to a target location.

REFERENCES

- BREGMAN, A. S. (1978). The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, NJ: Erlbaum.
- BREGMAN, A. S. (1981). Asking the what-for question in auditory perception. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum.
- BREGMAN, A. S., & CAMPBELL, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, **89**, 244-249.
- BREGMAN, A. S., & RUDNICKY, A. I. (1975). Auditory stream segregation: Stream or streams? *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 263-267.
- BRIGGS, R. M., & PERROTT, D. R. (1972). Auditory apparent movement under dichotic listening conditions. *Journal of Experimental Psychology*, **92**, 83-91.
- CARPENTER, P. A., & EISENBERG, P. (1978). Mental rotation and frame of reference in blind and sighted individuals. *Perception & Psychophysics*, **23**, 117-124.
- COOPER, L. A., & SHEPARD, R. N. (1973a). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.
- COOPER, L. A., & SHEPARD, R. N. (1973b). The time required to prepare for a rotated stimulus. *Memory & Cognition*, **1**, 246-250.
- DOWNING, C. J., & PINKER, S. (1985). The spatial structure of visual attention. In M. I. Posner, & O. S. M. Marin (Eds.), *Attention and performance XI*. Hillsdale, NJ: Erlbaum.
- DRAGER, U. C., & HUBEL, D. H. (1975). Responses to visual stimulation and relationship between visual, auditory and somatosensory inputs in mouse superior colliculus. *Journal of Neurophysiology*, **38**, 690-713.
- KNUDSEN, E. I. (1982). Auditory and visual maps of space in the optic tectum of the owl. *The Journal of Neuroscience*, **2**, 1177-1194.
- KNUDSEN, E. I. (1984). Synthesis of a neural map of auditory space in the owl. In G. M. Edelman, W. M. Cowan, & W. E. Gall (Eds.), *Dynamic aspects of neocortical function*. New York: Wiley.
- KORTE, A. (1915). Kinematoscopische Untersuchungen. *Zeitschrift für Psychologie der Sinnesorgane*, **72**, 193-296.
- KUBOVY, M. (1981). Concurrent-pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum.
- LANSMAN, M., FARR, S., & HUNT, E. (1984). Expectancy and dual task interference. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 195-204.
- MARCEL, T., & FORRIN, B. (1974). Naming latency and the repetition of stimulus categories. *Journal of Experimental Psychology*, **103**, 450-460.
- MARMOR, G. S., & ZABACK, L. A. (1976). Mental rotation by the blind: Does mental rotation depend on visual imagery? *Journal of Experimental Psychology: Human Perception & Performance*, **2**, 515-521.
- MARR, D. (1982). *Vision*. San Francisco: W. H. Freeman.
- MILLS, A. W. (1958). On the minimum audible angle. *Journal of the Acoustical Society of America*, **30**, 237-246.
- MILLS, A. W. (1972). Auditory localization. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 2). New York: Academic Press.
- O'LEARY, A., & RHODES, G. (1984). Cross-modal effects on visual and auditory object perception. *Perception & Psychophysics*, **35**, 565-569.
- PALMER, A. R., & KING, A. J. (1982). The representation of auditory space in the mammalian superior colliculus. *Nature*, **299**, 248-249.
- POSNER, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.
- POSNER, M. I., SNYDER, C. R. R., & DAVIDSON, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, **109**, 160-174.
- SHEPARD, R. N. (1981). Psychophysical complementarity. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum.
- SHEPARD, R. N., & COOPER, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- SHEPARD, R. N., & METZLER, J. (1971). Mental rotation of three-dimensional objects. *Science*, **171**, 701-703.
- SHEPARD, R. N., & ZARE, S. L. (1983). Path-guided apparent motion. *Science*, **220**, 632-634.
- SHULMAN, G. L., REMINGTON, R. W., & MCLEAN, J. P. (1979). Moving attention through visual space. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 522-526.
- STEVENS, S. S., & NEWMAN, E. B. (1936). The localization of actual sources of sound. *American Journal of Psychology*, **48**, 297-306.
- TSAL, Y. (1983). Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 523-530.

NOTES

1. I am not claiming that all spatial representations are analogical, or that no tasks exist for which a nonanalogical representation is more efficient. For example, representations in long-term memory may be nonanalogical. Rather, I am suggesting that perceptual representations of spatial information are analogical and that this has important functions, such as facilitating intermodal integration of information and continuous tracking of the position of an intermittently perceptible object (see later discussion).
2. Downing and Pinker (1985) interpret their results in terms of a gradient rather than a shift of attention. These two models have not yet been compared experimentally, and testable predictions to distinguish between them appear to be difficult to formulate. Computer simulations of the two models might be helpful in generating such predictions.
3. Mean, rather than median, reaction times were used in all analyses. However, the same pattern of results is obtained if medians are used.
4. The correction procedure was designed to eliminate such an effect. However, because there were a lot of localization errors for the most eccentric positions (see Table 1), the relevant correction factors were based on relatively few reaction times and in some cases could not be calculated at all.
5. On average, 92.7% ($SD = 5.0$) of each subject's mislocalizations involved naming the adjacent location to where the tone was presented. Of these erroneously localized tones, 80.2% ($SD = 19.7$) were localized closer to the midline than they actually were. Only 19.8% were judged to be more peripheral than they really were.
6. An apparent inconsistency with the time-distance hypothesis is that to know whether spatial constraints should operate on the attention-shifting mechanism, the auditory system must know whether the tones

were too far apart to support a continuous-motion interpretation. This decision itself depends upon knowing the location of the tones, which is the very thing that the attention mechanism is required to determine. However, the exact locations of the tones may not be necessary for this decision, only whether they are "near" or "far" from one another (where "near" and "far" are parameters whose values depend on the ISI). If the second tone is "far" from the first tone, then the discrete mechanism is invoked. The "near"/"far" decision could be made automatically and unconsciously (cf. the collicular visual system for deciding

where objects are), without requiring central capacity. Controlled attention, however, may be required for a conscious decision about the exact location of a sound. In this case, information processing capacity (i.e., conscious attention) would be moved to the location of the sound in order to localize it precisely in a higher level cortical representation containing explicit location information (i.e. names of locations).

(Manuscript received July 30, 1986;
revision accepted for publication February 14, 1987.)