Supramodal and Modality-Specific Mechanisms for Stimulus-Driven Shifts of Auditory and Visual Attention

LAWRENCE M. WARD *University of British Columbia*

Abstract Two experiments are described in which visual and/or auditory location precues preceded visual or auditory targets. Observers were required to judge the location of the targets. Conditions were such that involuntary, stimulus-driven attention shifts were the only ones likely to occur and give rise to cueing effects. It was found that visual precues affected response time to localize both visual targets and auditory targets but auditory precues affected only the time to localize auditory targets. Moreover, when visual and auditory cues conflicted, visual cues dominated in the visual task but were dominated by auditory cues in the auditory task. These results imply that involuntary stimulus-driven attention shifts might be controlled by a modality-specific mechanism for visual tasks, whereas stimulus-driven shifts of auditory attention are controlled by a supramodal mechanism. This asymmetry in attention control is consistent with the idea that attentional dominance in a multimodal experimental task depends on the relative performance possible in the modalities involved; in this case visual localization is more precise than auditory and so auditory cues may be ineffective in cueing visual location, while visual cues are effective in both modalities.

Understanding the mechanisms by which attention to external stimuli is controlled is crucial to understanding how we obtain important information from those stimuli and how this information informs our behaviour. There has been significant progress in recent years in understanding how shifts of visual attention are controlled (Wright & Ward, 1994, provide a summary; also see other articles in this issue). This has led in turn to consideration of whether the mechanisms controlling visual attention are more general or whether they are specific to vision.

It is now established that precues to the location of a visual target lead to faster and more accurate responding to the target, both for near- and supra-threshold stimuli and for detection, localization, and discrimination tasks. Posner (1978, 1980) is usually credited with initiating the modern interest in the mechanisms that control the attention shifts that are presumed to give rise

to such cueing effects, and for providing much relevant data and theory. Interestingly, in an early article Posner, Davidson, and Nissen (1976) reported that information cues (an arrow at the centre of a stimulus display pointing toward or away from the location at which an impending target would appear), while effective in orienting attention in a visual simple reaction time task, were ineffective for auditory targets or tactile targets except when a discrimination task was used. Posner (1987) argued that this reflected " ... fundamental differences between the sensory modalities in terms of their mechanisms for activating attentional mechanisms" (p. 204). It is also possible that this difference reflects differences in task requirements, since localization is not necessary for detection (simple reaction time task) but may be necessary for discrimination and is demanded for localization tasks (see Egly & Homa, 1991).

Since that early work, several researchers have addressed the question of whether and how attention shifts to the location of auditory target stimuli can be initiated, and whether there is a supramodal mechanism that would allow cross-modal cueing of attention shifts; that is, visual cueing of shifts to the locations of auditory stimuli or vice versa. The results have been somewhat equivocal, both because of differences in the paradigms used (although most have been modeled on that of Posner and associates) and because of the inherent difficulties involved. Nonetheless they are suggestive that there might exist both supra- and intramodal mechanisms for the control of attentional orienting. The history of work on this problem is summarized in Table 1 and discussed below.

At least three studies have found cueing effects indicative of attention shifts for auditory tasks. Scharf, Canevet, Possamai, and Bonnel (1986) used a paradigm similar to that of Posner, but with both a visual and an auditory precue indicating the same location. When the cue was valid (indicated the actual location of the impending target), simple reaction time to the onset of a pure tone (which had to be discriminated from another tone to which no response was required) was about 100 ms faster than when the cue was invalid (indicated a location other than the one where the target appeared). Subjects knew that the precue was *informative*; that is, it was valid more often than invalid and more often than chance. Rhodes (1987) found that response time in an auditory localization task increased linearly with distance (up to 90°) between the location of the previous target and that of the present target to be localized. The difference between response times when a location was repeated and when locations were separated by 22.5° was about 70 ms, and increased to about 220 ms when locations were separated by 90°. The previous target provided an informative precue, since locations were repeated more often than chance. Finally, Bedard, El Massioui, Pillon, and Nandrino (1993) reported that an information cue (again an arrow in the centre of a screen) was effective in orienting attention in both an auditory localization

TABLE 1
Summary of previous research on cross-modality cue effects.

Study	Task[a]	Cue[b]	CTOA	Cue effect[c]
Posner et al. 1976	A SRT	information I	1 s	no
Scharf et al. 1986	A SRT go-nogo	A+V stimulus I	1 s	yes: 100 ms
Klein et al. 1987	A SRT / V SRT	V stimulus U / A stimulus U	0-500 ms / 0-500 ms	no / yes: 20 ms
Buchtel & Butter 1988	A SRT / V SRT	A/V stimulus I / A/V stimulus I	50-1000 ms / 50-1000 ms	no / yes: both 25 ms
Bedard et al. 1993	A SRT / A LOC	information I / information I	50-750 ms / 350-750 ms	yes: 20 ms @350 / yes: 100 ms
Rhodes 1987	A LOC	A previous stimulus I	2 s	yes: 70 to 220 ms
Butter et al. 1989	V SRT / T SRT	V/T stimulus I / V/T stimulus I	50-1000 ms / 50-1000 ms	yes: both 25-50 ms / yes: both 12-45 ms
Farah et al. 1989	V SRT	A/V stimulus U patients with right parietal lesions	50-1000 ms	yes: V 233 A 32 ms contra

[a]: A = Auditory; V = Visual; T = Tactile; SRT = Simple Reaction Time; LOC = LOCation
reaction time. [b]: A+V = both A and V cues at once; A/V = A or V cues in blocks;
I = Informative; U = Uninformative. [c]: Cue effect = reaction time on invalid trials minus
reaction time on valid trials.

task and an auditory simple reaction time task, although the latter effect was very small (about 20 ms). It thus appears that under the right conditions, attentional orienting can be demonstrated for auditory tasks. The implications of these experiments are, however, somewhat unclear, since the paradigms employed differed greatly. Because all paradigms used informative precues and relatively long Cue-Target Onset Asynchronies (CTOAs) (350 ms - 2 s), only goal-driven attention control mechanisms (those that control voluntary attention in response to an information cue) are implicated in these studies (see next paragraph).

Studies aimed specifically at uncovering a polysensory (Buchtel & Butter, 1988; Butter, Buchtel, & Santucci, 1989; Klein, Brennan, & Gilani, 1987) or supramodal (Farah, Wong, Monheit, & Morrow, 1989) attention shift mechanism have clarified the picture in some ways and muddied it in others. For example, Buchtel and Butter (1988) reported that both visual and auditory precues were effective in a visual simple reaction time task, but that neither was effective for an auditory simple reaction time task. Their cues were what Wright and Ward (1994) have called stimulus cues; that is, either abrupt-onset lights or sounds, that can cause involuntary, stimulus-driven attention shifts

(i.e., "capture" attention). However, Buchtel and Butter's (1988) cues were informative, so that there was probably a significant goal-driven component in the cue effects. This is confirmed by the fact that cue effects were found at CTOAs as long as 1000 ms. When such stimulus cues are *uninformative* (are not valid any more often than chance and thus are not predictive of target location), cue effects diminish to zero after 100 to 200 ms CTOA in visual tasks (Müller & Findlay, 1988; Shepard & Müller, 1989). On the other hand, informative stimulus cues can have effects that persist for as long as several seconds, presumably reflecting the action of both stimulus-driven (initial time course) and goal-driven (extended time course) mechanisms (see, e.g., Cheal, Lyon, & Gottlob, in press). Thus, although provocative, Buchtel and Butter's (1988) results cannot be unequivocally interpreted as being indicative of the presence of a stimulus-driven mechanism that operates across modalities, although they point to the existence of a goal-driven mechanism that does so at least for visual tasks.

An earlier study by Klein et al. (1987) does seem to indicate the existence of a cross-modal stimulus-driven attention control mechanism. They found that uninformative auditory precues were effective in a visual simple reaction time task but that uninformative visual precues had no effects on auditory simple reaction time. Interestingly, auditory precue effects on visual simple reaction time, although small (about 19 ms), were about the same at all CTOAs, from 0 ms (precue & target occurred simultaneously) to 500 ms. Klein et al. (1987) argued that this resulted from faster processing of the auditory precues than of the visual targets; that is, auditory localization was faster than visual simple reaction time. There is evidence (cited in Klein et al., 1987) that auditory localization and auditory overt orienting are faster than visual localization and visual overt orienting, respectively. However, I know of no evidence that (1) visual detection requires visual localization or visual orienting and (2) auditory localization is faster than visual detection. In fact, the data reported in this paper suggest that auditory localization can be considerably *slower* than visual detection (about 150 ms, see later). Since localization of the precue would seem to be necessary for it to be effective in orienting attention and thereby affecting simple reaction time, it is possible that the cue effects found arose in another way. Thus, although suggestive, Klein et al.'s (1987) results are not definitive (as indeed no single set of results can be).

Butter et al. (1989) did find evidence for cross-modal cueing of attention shifts between visual and tactile modalities in the same paradigm as that used by Buchtel and Butter (1988). Cue effects were asymmetric, being larger for both cue types in the visual task and for tactile cues in the tactile task and very small for visual cues in the tactile task. Their results suggested that mechanisms of attention are closely linked to orienting systems and that there is some system controlling spatial attention shifts that has polysensory inputs.

Farah et al. (1989) also addressed the question of the existence of a supramodal attention shift mechanism. They replicated Klein et al.'s (1987) finding that uninformative auditory precues were effective in shifting attention in a visual simple reaction time task. However, their subjects were all patients with right parietal lesions, and the effect of auditory cues occurred only for visual targets on the contra-lesional side (the left side), which is where the lesion would most affect visual orienting. Moreover, this effect was present at all CTOAs from 50 to 1000 ms, rather than diminishing sharply for the longer CTOAs as a stimulus-driven cue effect should. Thus, although there is some suggestion that a supramodal, involuntary, stimulus-driven attention shift mechanism might exist, none of the studies I have seen have unequivocally established their existence. It seems likely that such a mechanism does exist for goal-driven shifts, since several studies have reported cross-modal cueing effects with informative cues of several types.

The experiments reported here were designed to test whether, under conditions favourable for measuring involuntary stimulus-driven attention shifts, cueing effects for visual and auditory tasks would be modality-specific, or would instead indicate the existence of a supramodal mechanism for accomplishing attention shifts. The experiments reported here involve a localization task; experiments using simple reaction time and discrimination tasks are reported elsewhere. The results of a localization task are important for two reasons. First, auditory and visual localization are closely intertwined (viz, visual capture) and a supra-modal stimulus-driven attention shift mechanism seems more likely in such a context. Second, localization is arguably involved in all attention orienting, whether covert or overt, and should be studied directly.

In the present experiments a situation was created in which a stimulus could appear in one or the other of two locations, to the left or to the right of a fixation point. Response time to indicate the location of the target was measured. Brief, abrupt-onset, stimulus precues also could appear at either of the two target locations, or in a central location or not at all, and in one or the other, or both or neither, of two modalities, visual and auditory. The various combinations of cue conditions displayed in Table 2 were used in each of two experiments, one employing a visual localization task (i.e., only visual targets) and the other an auditory localization task (i.e., only auditory targets). All of the cue conditions were mixed within each experiment. This situation is highly ecologically valid since, when moving about in the world, humans and other animals encounter a variety of cues and targets in a variety of locations and modalities in an irregular distribution over time. Further, when such mixed cue experiments are run, a single strategy should be adopted for all trials, since it is impossible to anticipate the type of trial that will occur on a given occasion.

Whenever a visual or auditory cue (or both) was presented on the same side

TABLE 2
Design of the cue conditions of the experiments.

		Auditory Cue Position			
		Left	Right	Central	None
	Left	LL	LR	LC	LN
Visual Cue	Right	RL	RR	RC	RN
Position	Central	CL	CR	CC	CN
	None	NL	NR	NC	NN

VAcon = LL + RR; VC = LC + RC; VN = LN + RN;
VAincon = RL + LR; CA = CL + CR; NA = NL + RN

as a target (e.g., cue NL, target L) a *valid* trial resulted. Similarly, when a cue was on the opposite side from the target, an *invalid* trial resulted. In the condition designated VAincon in Table 2, in which a visual cue was presented on one side and an auditory cue on the other, one cue was valid and the other was invalid. The results are presented in terms of the validity of the same-modality cue (as target) in these conditions. In conditions in which the central cue in one modality was paired with either no cue or the central cue from the other, cue validity was neutral. On cued trials, valid and invalid trials occurred with equal frequency, so that cue location was not predictive of target location. Cues and targets occurred with equal frequency in each location, and data were collapsed across cue (see Table 2) and target locations. Under these conditions, subjects typically try to ignore the cues, and any cueing effects can be reasonably attributed to attention shift mechanisms not completely under subjects' control. In other words, these experiments were an attempt to determine the extent to which subjects' attention would be captured involuntarily by the stimulus cues.

METHOD
Subjects
Thirty university students, 15 in each experiment (7 males and 8 females in visual, 5 males and 10 females in auditory), were paid to participate. All had normal or corrected-to-normal vision and were aware of no hearing deficits. All could hear and localize the auditory targets and cues without effort. Two subjects who could not localize the auditory targets with above 90% accuracy were dropped and replaced (one performed near 90% and the other near 80%). Some subjects participated in both experiments and some in only one; results were not affected by how many experiments were done.

Stimuli and Apparatus
Stimulus and cue timing and presentation and data recording were accomplished by a HP Vectra ES/12 microcomputer. Visual cues and targets were presented on a NEC Multisync computer monitor screen. The screen continu-

ously displayed three square boxes in the centre vertically, one in the centre horizontally, and one at extreme left and one at extreme right, with centres 12° of visual angle from the centre of the central box. The central, fixation box measured 1.1° square and contained a 0.6° square "X". The peripheral boxes measured 2° square. Visual targets were "X"s that filled the peripheral boxes for 50 ms. Visual cues were 50 ms brightenings of the central or peripheral boxes accomplished by thickening their outlines and then thinning them. The central visual cue was used to correspond to the central localization of the neutral auditory cue (see next paragraph).

Auditory targets were 50 ms duration, .5 ms rise/fall time, 76 dB SPL (at the ears), 3000 Hz pure tones produced by a custom sound generator and played through speakers mounted as close as possible to each side of the monitor with cones at the same vertical position as the boxes present on the screen. The centres of the speaker cones were 12° from the centre of the nearest peripheral box, 48° from each other, and were located at 24° left or right azimuth with respect to the plane of the centre of the subject's head. The auditory cues were played from the same speakers but were 50 ms duration, .5 ms rise/fall time, 86 dB SPL, 964 Hz pure tones that were easily distinguished from the targets by all subjects. Central, or neutral, auditory cues were produced by playing the cue tone simultaneously from both speakers. Since the intensities at each speaker were identical, the resulting sound (92 dB SPL) was typically localized as coming from 0° azimuth, or directly in front of the subject's head, where the central box was positioned on the screen. For some subjects, neutral auditory cue location was perceived as ambiguous.

Subjects sat in darkness in a quiet room with their chins in a chin rest and eyes 55 cm from the screen and speakers. The controlling computer was in the same room and provided a relatively constant background fan noise of about 60 dBA. Errors were signaled by a beep from the computer speaker that sounded very different from either auditory cues or targets. Responses were made by pressing one of two large buttons that operated microswitches, one with the left hand for left side targets and the other with the right hand for right side targets.

Procedure

All trials in both experiments followed the same sequence of events. At the beginning of a trial, the screen displayed the three boxes and the central fixation X, which subjects were instructed to fixate at all times during a block of trials. Eye movements were not monitored since previous experiments have shown that under such instructional set eye movements occur relatively infrequently and are uncorrelated with results. Also, eye movements are irrelevant to detection of sounds, and most stimulus-driven visual cueing effects have occurred within 200 ms CTOA in previous studies (see Wright & Ward, 1994), an interval too short for eye movement initiation under such

circumstances. Two seconds after the response to the previous trial a cue (either visual, auditory, both, or neither) was presented and was followed by a target after a random one of four CTOAs: 100, 200, 550, and 1050 ms. After the response to the target, the next trial was initiated.

Trials were run in blocks of 128, which represented two trials each of the 64 different CTOA × Condition × Validity combinations explained earlier (see Table 2). Each block was shuffled to present the trials in a different irregular order. Trials on which either an incorrect response was made, or on which the response time was faster than 100 ms or slower than 1500 ms, were counted as errors. Such trials were rerun on an irregularly-placed later occasion in the same block. Total errors were recorded for each block, and subjects were required to make fewer than 12 errors per block on average. Thus perform- ance was required to be above 91% correct, and only response times for correct responses were recorded. Subjects practiced until they were performing the task with few errors and average response times in the range of 300 to 800 ms, which amounted to from 10 to 256 practice trials, depending on the subject. Each subject then completed 15 blocks in three sessions lasting about 45 min to one hour each, for a total of 1920 trials per subject, 30 for each of the 64 CTOA × Condition × Validity combinations. Subjects took a short break (at least 1 min) after each block and a longer break (at least 5 min) after each session. Some completed all three sessions in a single day and others completed the sessions on different days.

RESULTS AND DISCUSSION

Because response-time distributions in such tasks tend to be skewed, and these were no exception, medians were taken of the 30 response times available for each of the 64 conditions for each subject in each experiment. Figures 1 (visual task) and 2 (auditory task) display the averages across subjects of these median response times for the six cued conditions considered separately. Appropriate neutral cue conditions are plotted for each cued condition. Whenever cues were presented in a single modality, the neutral condition with only that modality cue presented centrally is plotted (e.g., NC for the NA condition). Whenever both modality cues were presented, CC is the appropri- ate neutral condition (e.g., for VAcon). The NN condition (no cues at all) is plotted in only one panel for each task.

The median response times were entered into a Condition (6 cued conditions only) × CTOA (4 CTOAs) × Validity (Valid, Invalid) repeated measures ANOVA for each task separately. Neutral cue conditions were not entered into these analyses because of the necessity of using the same neutral cue condition several times (e.g., CC for VAcon, VAincon, VC, & CA). Neutral cue conditions were analysed separately and showed highly reliable effects of Cuetype, CTOA, and their interaction in both experiments. However, these data are not discussed here; they are included only for completeness. A detailed
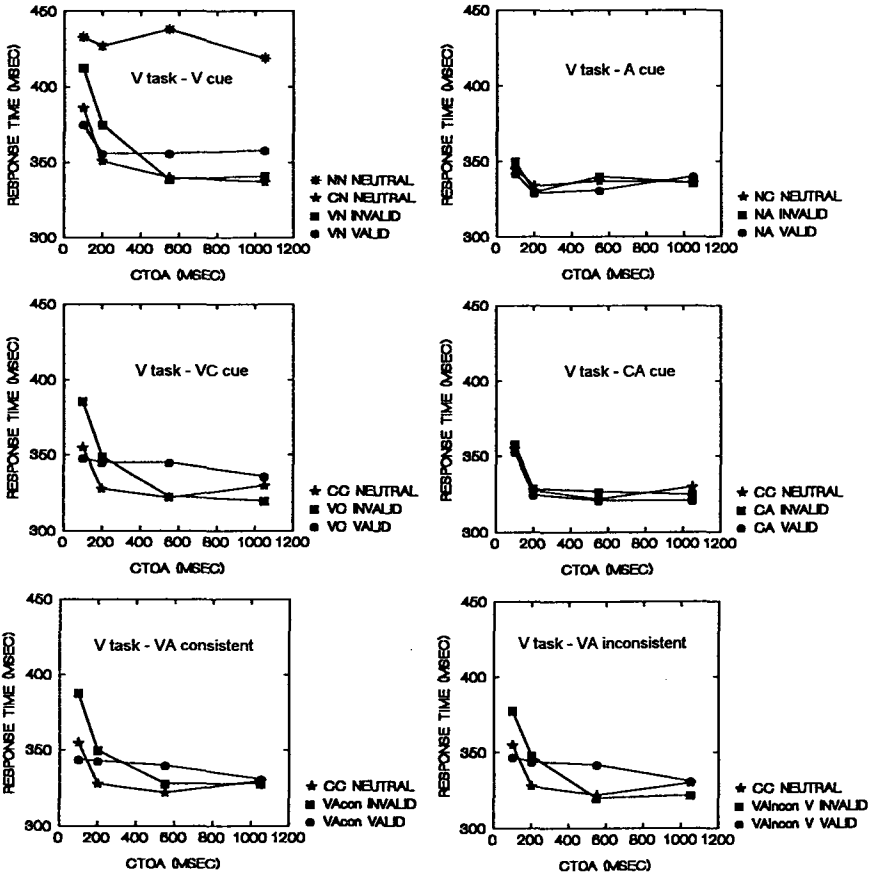
Fig. 1 Effects of visual and auditory cues on time to localize visual stimuli as a function of Cue-Target-Onset-Asynchrony (CTOA) for the six cued conditions. Neutral cue condition is CC for all cued conditions except VN (CN) and NA (NC). NN condition is displayed in VN condition graph.

cost-benefit analysis was not done for these data because of the difficulty of interpreting neutral cue conditions, especially the somewhat unconventional ones used here (cf., Jonides & Mack, 1984). Huynh-Feldt corrected degrees of freedom were used to determine p values when sphericity was violated for repeated measures factors with more than two levels. An α level of 0.05 was used for all significance determinations. Table 3 displays the average cue effects (response times on invalid trials minus those on valid trials) for each condition and CTOA for the visual and auditory tasks and indicates which ones were statistically reliable by Bonferroni t-tests. Experimentwise error (EW) was set at 0.10 for each cue condition in each task (the "experimental unit") because each Condition × Task combination is similar to the experimental unit
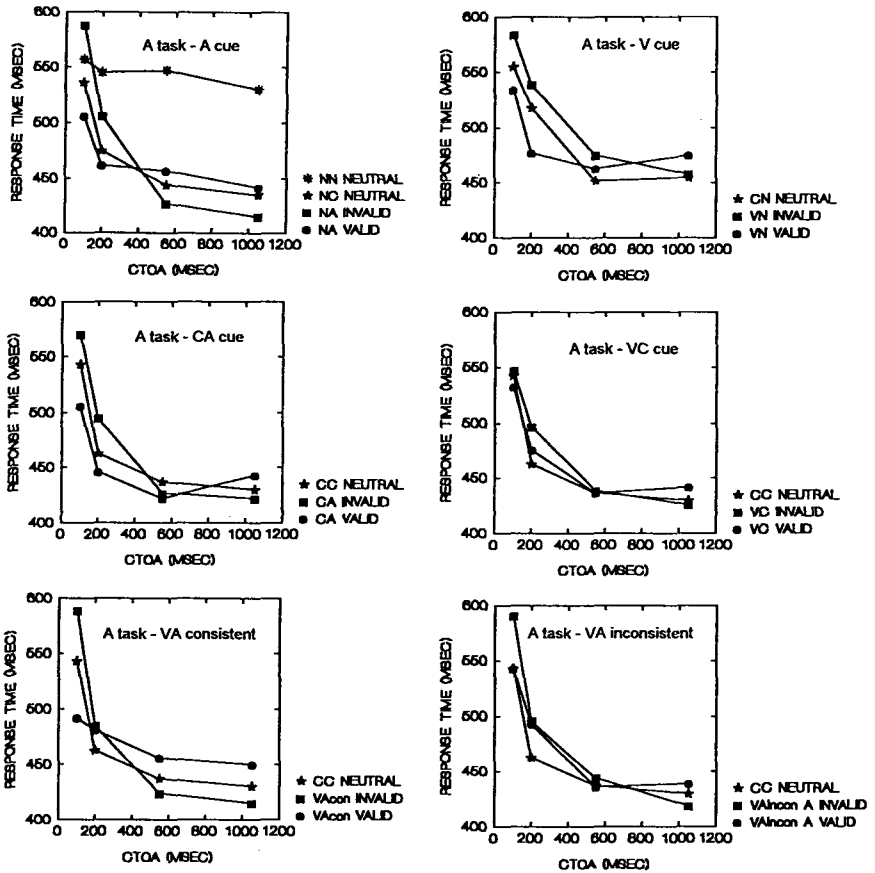
Fig. 2 Effects of visual and auditory cues on time to localize auditory stimuli as a function of Cue-Target-Onset-Asynchrony (CTOA) for the six cued conditions. Neutral cue condition is CC for all cued conditions except VN (CN) and NA (NC). NN condition is displayed in NA condition graph.

more typical of studies in this field (cf., Buchtel & Butter, 1988; Butter et al., 1989). The critical t value for 4 contrasts (i.e., cueing effect at each CTOA for a given Condition × Task combination) and 120 degrees of freedom (the highest tabled df < ∞) was used (t_{crit} = 2.27) along with the MS_e from the Condition × CTOA × Validity interaction for each task to calculate the critical differences for the cue effects. All such tests on cue effects (Table 3) were thus done two-tailed with α = 0.025 on each test (EW = 4 × 0.025 = 0.10).

Inspection of Figures 1 and 2 reveals the findings. The discussion will focus on cue effects and will ignore effects whose interpretations are obvious or obscure. In particular, the effect of CTOA on response time (which was statistically reliable for every condition) has been observed and discussed previously (e.g., Buchtel & Butter, 1988; Butter et al., 1989; Farah et al.,

TABLE 3
Cue effects (Invalid RT - Valid RT) for localization tasks

		CTOA		
Condition	100	200	550	1050
Visual Task				
VAcon	44*	7	−12	3
VAincon[a]	31*	5	−22*	−9
CA	5	4	6	5
VC	38*	4	−22*	−16*
NA	8	2	9	−4
VN	38*	19*	−17*	−17*
Auditory Task				
VAcon	97*	4	−32*	−35*
VAincon[b]	48*	3	8	−20
CA	64*	49*	5	−21
VC	14	21	2	−16
NA	83*	44*	−30*	−27*
VN	50*	62*	12	−17

* significant at α = 0.025 by Bonferroni two-tailed t-test as described in text; visual task critical difference = 14 ms; auditory task critical difference = 25 ms; [a] V Invalid RT - V Valid RT; [b] A Invalid RT - A Valid RT

1989) as have the alerting effects that are obvious for the neutral cue conditions (NN vs. CN, NC, & CC). The result of interest is the interaction of Condition × CTOA × Validity and it was significant and strong for both tasks, although, as will be seen, its interpretation is different for the different tasks.[1]

For the visual task, the results for the VN condition replicated previous results in similar paradigms for stimulus-driven cueing (see Wright & Ward, 1994). The cue effect is relatively large for the 100 ms CTOA (see also Table 3), and in these data both costs of the invalid cue and benefits of the valid cue can be seen (cost = VN invalid minus CN; benefit = CN minus VN valid). The cue effect is smaller for the 200 ms CTOA and reverses for the longer CTOAs,

1 The complete results of the statistical analyses follow. Visual task: The main effects of Condition [$F(5,70)$ = 30.55, p < .001] and CTOA [$F(3,42)$ = 33.73, p < .001] were significant, the main effect of Validity was not. Significant interactions were Condition × CTOA [$F(15,210)$ = 4.57, p < .001], CTOA × Validity [$F(3,42)$ = 11.04, p < .001] and Condition × CTOA × Validity [$F(15,210)$ = 10.98, p < .001]. Visual task neutral cue conditions: Main effects of Cuetype [$F(3,42)$ = 83.55, p < .001] and CTOA [$F(3,42)$ = 17.98, p < .001] and the Cuetype × CTOA interaction [$F(9,126)$ = 6.20, p < .001] were all signficant. Auditory task: The main effects of Condition [$F(5,70)$ = 8.89, p = .001], CTOA [$F(3,42)$ = 49.85, p < .001] and Validity [$F(1,14)$ = 8.17, p = .013] were all significant. The only significant interactions were CTOA × Validity [$F(3,42)$ = 36.50, p < .001] and Condition × CTOA × Validity [$F(15,210)$ = 4.27, p < .001]. Auditory task neutral cue conditions: Main effects of Cuetype [$F(3,42)$ = 17.97, p < .001] and CTOA [$F(3,42)$ = 47.82, p < .001] and the Cuetype × CTOA interaction [$F(9,126)$ = 5.36, p < .001] were all significant.

an indication of the possible existence of the phenomenon of inhibition-of-return (Posner & Cohen, 1984) in these data. In inhibition-of-return, targets that occur at locations that have been cued by an abrupt-onset stimulus cue are responded to more slowly if attention is directed to another location after first being directed to the cued location, for example, by a second abrupt-onset stimulus cue. In the current situation, subjects may have been directing their attention back toward the central fixation box shortly after having it involuntarily shifted to the location of the visual cue. Typically, when stimulus cues are uninformative, stimulus-driven cue effects peak at about 100 to 200 ms CTOA and diminish to zero or below (inhibition-of-return) for longer CTOAs. Results are similar for the VAcon and VC conditions, and the visual cue dominates at 100 ms CTOA when the cues are inconsistent (VAincon).

In contrast, the auditory cue alone had no effect on visual localization judgement times, as indicated by the NA and CA condition results. This result is inconsistent with those of previous studies (Buchtel & Butter, 1988; Farah et al., 1989; Klein et al., 1987), all of which found a significant effect of an auditory precue on visual simple reaction time. It is possible that the auditory cue used here was not sufficiently powerful to capture visual localization since pure tones are difficult to localize precisely near $0°$ azimuth. However, the cues used here were relatively intense (86 dB SPL), not easily confusable, sufficient to yield a large cue effect on auditory localization (next paragraph), and similar to the pure tone cues that have been shown to affect visual simple reaction time (Farah et al., 1989; Klein et al., 1987). Furthermore, in a pilot experiment ($n = 7$) using somewhat more intense broadband noise cues I have found no significant cueing effects at 100 or 200 ms CTOAs. In the pilot experiment the cue effects are of the same order as those reported here. It is also possible that visual localization is less sensitive to such cue effects than is visual simple reaction time. However, my own preliminary studies using the same cues in the same situation have yielded no significant auditory cue effects on visual simple reaction time or on auditory simple reaction time (Ward, 1993). A final possibility is that situational factors are responsible for the inconsistencies between these and earlier studies. The studies in which auditory cue effects on visual simple reaction time have been demonstrated have all used only one cue type in each experiment, whereas the present experiments involved a mix of auditory and visual cues. It is possible that different overall strategies are pursued by subjects in single-cue-modality and mixed-cue-modality situations, resulting in their being able to ignore auditory cues when cue types are mixed but not when they are blocked.

The pattern of results for the auditory localization task is somewhat different. It can best be characterized as effective cueing by both visual and auditory cues for auditory localization. This can be seen most clearly in the panels of Figure 2 depicting the results of the NA and VN conditions. In both

conditions, there are large cue effects for the 100 ms and 200 ms CTOAs, and these reverse for the longer CTOAs although the negative effects are not reliable for the visual cue condition. The effects are basically identical to those found in standard visual cueing paradigms, even to the suggested presence of both costs and benefits at 100 and 200 ms CTOAs (the neutral cue response times lie between the valid and invalid response times). The reversal of the effect at the longer CTOAs suggests that inhibition-of-return might also occur in the auditory domain, at least for localization judgements. When auditory and visual cues are consistent, a large cue effect also is found for the 100 ms CTOA (VAcon panel), along with a negative effect at the longest CTOAs. The lack of an effect at the 200 ms CTOA might indicate that the combined cue is processed more quickly than the single cues. When the combined cues are inconsistent (VAincon) the auditory cue dominates at 100 ms CTOA. However, the cue effect is only half as large as when the cues are consistent (only the cost remains), and there are no significant cue effects at longer CTOAs, presumably because of the conflicting visual cue. Similarly, when a central cue in one modality is presented with a directional cue in the other (VC & CA) there are smaller cue effects, although the effect for CA is significant at 100 ms and 200 ms CTOA. Since each cue modality alone has a reliable and similar effect on auditory localization, and conflicting cues yield diminished or no effect (a central cue conflicts with a directional cue much as two directional cues conflict), it seems clear that auditory localization is much more susceptible to involuntary attention shifts caused by visual cues than vice versa.

A reviewer suggested that all of the cue effects found in these experiments could possibly arise from a tendency of subjects, in response to the presentation of a cue, to prepare the response on the side corresponding to the cue. When the cue was valid, the required response would be the prepared one, resulting in a faster response than when the cue was invalid and the required response was unprepared. Simon (1969) demonstrated that such an effect is plausible. He found that subjects responded faster when they had to move a lever toward the same side as the ear in which an auditory stimulus occurred than when they had to move the lever in the opposite direction. Bedard et al. (1993) also found that effects of an information cue were larger for localization responses than for simple reaction time responses and attributed the difference to the effects of the cue on motor preparation time. However, I do not believe this can explain the present effects. First, a cue effect size difference similar to that of Bedard et al. (1993) can be shown for many discrimination tasks compared to simple reaction time; there are more differences than just motor preparation times between discrimination and simple reaction time tasks. Second, and more important, the auditory cue had no effect on visual localization. If the cue effects were caused by response preparation alone, there should have been a cue effect in this condition, since

responses to the visual stimulus were identical to those to the auditory stimulus and should have been equally prepared by an auditory cue. It seems more reasonable to conclude that the effects reported here arise from the effects of the precues on attentional orienting.

CONCLUSIONS

The results of the present experiments indicate that while stimulus-driven attention shifts on a visual localization task may be initiated only by visual cues, those on an auditory localization task can be initiated equally well by either visual or auditory cues. Moreover, on the visual task, auditory cues do not seem to compete successfully with visual cues for attention, while on auditory tasks, the auditory and visual cues seem to diminish or cancel out each other's effects. These results are inconsistent with the idea that stimulus-driven attention shifts caused by uninformative stimulus cues are modality-specific. On the contrary, those in the auditory modality, as least as measured by a localization task, conform to a supramodal model (cf. Butter et al., 1989; Farah et al., 1989; Klein et al., 1987).

The asymmetry between audition and vision revealed by these experiments would be expected from an approach that emphasizes the differences in localization accuracy in the two modalities, and the necessity of correctly localizing environmental events that have both auditory and visual signatures (like the approach of a large predator). It has been argued that the auditory modality is most useful for detection of important events, regardless of where they are located, while the visual modality is most useful for analysis of the significance of an event. Vision is also more useful for spatial localization because the visual scene is mapped topologically onto the retina and this mapping is preserved at higher visual processing areas. Thus, localization in visual space is both direct and very precise (typical resolution acuity in the fovea is about 1 min of arc which is 1/60 of a degree). However, auditory localization is accomplished by neural computations based largely on differences in time or intensity at the two ears and is thus less direct. It is also much less precise than visual localization: minimum audible angle for pure tones at 0° azimuth is about 2° and is far worse elsewhere; it is better but still poor for broad spectrum sounds. Moreover, there are other profound, and asymmetrical, interactions between visual and auditory localization. For instance, in visual capture, or the ventriloquism effect, auditory location is "captured" by a moving visual stimulus (like the movements of the ventriloquist's dummy's mouth) to a surprising extent (the distance between the actual sound source, the ventriloquist's mouth, and the apparent one, the dummy's mouth). The reverse effect, auditory capture, does not occur.

The present results extend this visual-auditory asymmetry to stimulus-driven attention shifts. This implies that the mechanisms that control such attention shifts are closely connected to early sensory processing, like the mechanisms

involved in stimulus localization. Previous suggestions that stimulus-driven shifts of visual attention are accomplished by a mechanism closely related to that involved in the control of saccadic eye movements are consistent with the present results (see Wright & Ward, 1994). It is interesting to speculate that a single, supramodal map of environmental events is constantly being updated (cf. Farah et al., 1989). However, the present results suggest that this map is based on a map of visual space, and that conflicts between the visual map and auditory computations are resolved either in favour of the visual map (visual capture) or by other computational compromises that are sufficient to eliminate the involuntary attention shifts that would create cue effects.

References
Bedard, M.A., El Massioui, F., Pillon, B., & Nandrino, J.L. (1993). Time for reorienting of attention: A premotor hypothesis of the underlying mechanism. *Neuropsychologia, 31,* 241-249.

Buchtel, H.A., & Butter, C.M. (1988). Spatial attentional shifts: Implications for the role of polysensory mechanisms. *Neuropsychologia, 26,* 499-509.

Butter, C.M., Buchtel, H.A., & Santucci, R. (1989). Spatial attentional shifts: Further evidence for the role of polysensory mechanisms using visual and tactile stimuli. *Neuropsychologia, 27,* 1231-1240.

Cheal, M.L., Lyon, D.R., & Gottlob, L.R. (in press). A framework for understanding the allocation of attention in location-precued discrimination. *Quarterly Journal of Experimental Psychology.*

Egly, R., & Homa, D. (1991). Reallocation of visual attention. *Journal of Experimental Psychology: Human Perception & Performance, 17,* 142-159.

Farah, M.J., Wong, A.B., Monheit, M.A., & Morrow, L.A. (1989). Parietal lobe mechanisms of spatial attention: Modality specific or supramodal? *Neuropsychologia, 27,* 461-470.

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin, 96,* 29-44.

Klein, R., Brennan, M., & Gilani, A. (1987). *Covert cross-modality orienting of attention in space.* Paper presented at the annual meeting of the Psychonomic Society, Seattle, WA.

Müller, H.J., & Findlay, J.M. (1988). The effect of visual attention on peripheral discrimination thresholds in single and multiple element displays. *Acta Psychologica, 69,* 129-155.

Posner, M.I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Erblaum.

Posner, M.I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3-25.

Posner, M.I., Davidson, B.J., & Nissen, M.J. (1976). The process of stimulus detection. Paper presented at the annual meeting of the Psychonomic Society, St. Louis. Cited in Posner, M.I. (1978). *Chronometric explorations of mind.* Hillsdale, NJ: Erlbaum.

Posner, M.I., & Cohen, Y. (1984). Components of visual attention. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention & Performance,* Vol. X. Hillsdale, NJ: Erlbaum.

Rhodes, G. (1987). Auditory attention and the representation of spatial information. *Perception & Psychophysics, 42,* 1-14.

Scharf, B., Canevet, G., Possamai, C-A., & Bonnel, A-M. (1986). Some effects of attention in hearing. Cited in Scharf, B. (1988). The role of listening in the measurement of hearing. *Advances in audiology, 5,* 13-26.

Shepard, M., & Müller, H.J. (1989). Movement versus focusing of visual attention. *Perception & Psychophysics, 46,* 146-154.

Simon, J.R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology, 81,* 174-176.

Ward, L.M. (1993). *Visual/auditory capture of auditory/visual attention.* Paper presented at the annual meeting of the Psychonomic Society, Washington, DC.

Wright, R.D., & Ward, L.M. (1994). The control of visual attention. Unpublished manuscript.

Sommaire

Mécanismes supramodaux et spécifiques de modalité pour les déplacements d'attention auditive et visuelle actionnés par un stimulus

Cette communication a pour objet de chercher à déterminer si les déplacements d'attention vers des emplacements spatiaux sont contrôlés par un mécanisme général, polymodal, ou par un mécanisme particulier à la vision. Les travaux antérieurs laissaient supposer un mécanisme spécifique de modalité, mais les recherches plus récentes penchent vers un mécanisme polymodal. Néanmoins, la plus grande partie des données recueillies ces derniers temps l'ont été au moyen de signaux préalables d'information ou dans des conditions qui ne permettent pas de répondre à la question de savoir si les déplacements d'attention en modalité croisée sont contrôlés par un mécanisme involontaire actionné par un stimulus. On décrit ici deux expériences dans lesquelles des signaux préalables d'emplacement visuel et/ou auditif précédaient des cibles visuelles ou auditives. Dans chaque modalité, les signaux étaient produits soit seuls, soit conjointement avec un signal de l'autre modalité (au même endroit ou à un endroit différent), ou encore n'étaient pas produits du tout dans chaque déroulement des essais, afin que la situation expérimentale soit semblable aux situations de la vie réelle où les signaux sont mélangés. On a demandé aux sujets d'indiquer l'emplacement de la cible dans chaque essai. Dans l'expérience de localisation visuelle, la cible consistait en l'apparition d'un «X» pendant 50 ms dans une de deux cases situées aux extrémités horizontales d'un écran d'ordinateur. Dans l'expérience de localisation auditive, la cible était l'émission d'une tonalité de 76 dB à 3 000 Hz, pendant 50 ms, provenant de l'un ou l'autre de deux haut-parleurs placés aussi près que possible des extrémités horizontales de l'écran. Les signaux à déclenchement brusque (illumination d'une case pendant 50 ms, ou émission d'une tonalité de 86 dB à 964 Hz pendant 50 ms) ne renseignaient pas sur l'emplacement de la cible, de sorte que pour des asynchronies d'apparition du signal et de la cible (CTOA, *cue-target-onset-asynchrony*) de l'ordre de 100 ms à 200 ms, les seuls déplacements d'attention susceptibles de se produire étaient des déplacements involontaires actionnés par un stimulus vers l'emplacement signalé. On a constaté que les signaux visuels préalables affectaient le temps de réaction nécessaire à la localisation des cibles visuelles aussi bien que des cibles auditives lorsque les asynchronies étaient plus courtes, mais que les signaux auditifs préalables affectaient seulement le temps de réaction nécessaire à la localisation des cibles auditives. Les signaux auditifs préalables n'avaient aucun effet sur le temps de réaction nécessaire à la localisation des cibles visuelles, quelle que soit l'asynchronie. De plus, lorsque les signaux visuels et auditifs étaient en conflit, les

signaux visuels dominaient dans la tâche visuelle, mais étaient dominés par les signaux auditifs dans la tâche auditive. La possibilité que ces résultats soient attribuables à des amorces de réaction données par les signaux préalables doit être écartée parce que les signaux auditifs préalables, qui ont pourtant eu un effet marqué sur la rapidité de localisation des cibles auditives, n'ont aucunement affecté la localisation des cibles visuelles à l'aide des réactions identiques. Les contradictions entre ces constatations et celles d'autres expériences peuvent se résoudre si l'on prend en considération les différences entre les situations entourant les signaux dans la présente étude, en particulier le mélange des signaux appartenant aux deux modalités. Dans l'ensemble, ces constatations et les autres résultats indiquent que les déplacements d'attention involontaires actionnés par un stimulus pourraient être contrôlés par un mécanisme spécifique de modalité pour les tâches visuelles, alors que les déplacements d'attention auditive involontaires actionnés par un stimulus peuvent être contrôlés par un mécanisme supramodal. Cette asymétrie dans le contrôle de l'attention est conforme à l'idée que la domination en matière d'attention, dans une tâche expérimentale à modalités multiples, dépend de la performance relative possible selon les modalités en question; dans le cas présent, la localisation visuelle est plus précise que la localisation auditive, et il se peut donc que les signaux auditifs manquent d'efficacité pour signaler l'emplacement visuel, alors que les signaux visuels sont efficaces dans les deux modalités.