Perception, 2001, volume 30, pages 795 – 810

DOI:10.1068/p3126

Identification of visual stimuli is improved by accompanying auditory stimuli: The role of eye movements and sound location

Melanie C Doyle, Robert J Snowden¶
School of Psychology, Cardiff University, Cardiff CF11 3YG, UK; e-mail: snowden@cardiff.ac.uk
Received 25 August 2000, in revised form 6 March 2001

Abstract. Can auditory signals influence the processing of visual information? The present study examined the effects of simple auditory signals (clicks and noise bursts) whose onset was simultaneous with that of the visual target, but which provided no information about the target. It was found that such a signal enhances performance in the visual task: the accessory sound reduced response times for target identification with no cost to accuracy. The spatial location of the sound (whether central to the display or at the target location) did not modify this facilitation. Furthermore, the same pattern of facilitation was evident whether the observer fixated centrally or moved their eyes to the target. The results were not altered by changes in the contrast (and therefore visibility) of the visual stimulus or by the perceived utility of the spatial location of the sound. We speculate that the auditory signal may promote attentional 'disengagement' and that, as a result, observers are able to process the visual target sooner when sound accompanies the display relative to when visual information is presented alone.

1 Introduction

Visual events, such as the dropping of a plate or the movement of someone's lips, are often associated with sound. It may therefore come as little surprise that there are neurons within the mammalian brain that appear to encode the cross-modal aspects of the perceptual scene (Stein and Meredith 1994). The auditory information that accompanies some visual events might facilitate detection of these events and may help to guide a person to a particular piece of visual information. For example, in a cockpit crowded with informative dials and gauges, auditory information might help a pilot to notice a particular dial or visual warning signal. In this paper we examine which aspects of the auditory information might be useful, and under what circumstances.

Similar questions have been explored previously. Todd (1912) reported that the reaction time (RT) to detect a visual target was speeded by as much as 80 ms by the simultaneous presence of an auditory signal (to which subjects were told not to respond). This is often termed the *redundant-signals effect*. Since this time, numerous studies have replicated and extended these findings (Bernstein 1970; Bernstein et al 1969; Bernstein and Edelstein 1971; Hershenson 1962; Morrell 1967, 1968; Nickerson 1973; Simon and Craft 1970). However, the interpretation of these results as evidence of intersensory facilitation is somewhat problematic. Firstly, the early studies did not include catch trials (Hershenson 1962; Todd 1912)—hence the participant was actually free to respond to either the visual or auditory stimulus. Given that under many circumstances RTs to auditory stimuli are faster than to visual ones, there is no need to postulate any cross-modal effect. Though it was argued that RTs to the combined stimuli were faster than to either stimulus alone, this effect may reflect statistical facilitation rather than any physiological mechanism (Hughes et al 1994; Nickerson 1973; Raab 1962). Following more careful consideration of these models, Gielen et al (1983) once again claimed that these effects reflected genuine cross-modal facilitation and could not be accounted for by statistical considerations. However, this later study once again did not have any catch trials.

¶ Author to whom correspondence should be addressed.

A method commonly used to address these criterion-based accounts of intersensory facilitation is to use a speeded choice-response to the target stimulus. If the auditory signal were to cause a change in the criterion for judging the visual target this would result in faster RTs accompanied by an increase in errors. A genuine change in the ability to process the visual stimulus would result in faster RTs accompanied by a reduction, or no change, in errors. In this series of experiments speeded choice-responses are employed.

A type of choice-RT paradigm was employed in the studies of Simon and Craft (1970; see also Bernstein et al 1969; Bernstein and Edelstein 1971). In these experiments the participant had to respond to visual targets on the left using one response and to those on the right using another response. It was found that auditory stimuli in the left ear facilitated a 'left' response and interfered with a 'right' response. Once again, however, the interpretation of such a result is not straightforward. The presentation of a tone in the left ear may simply have primed a 'left' response, so this may reflect a change in criterion at each of these two locations. In our experiments we took care to make sure that the location of the auditory stimulus did not provide any information (either supporting or conflicting) about the required response.

An example of this latter type of response (ie one in which sound-source location was independent of the visual response) comes from a later study of Perrott et al (1990). Observers made a choice-response on the basis of the shape of the visual target. The small size of the target meant that foveal vision was necessary to identify the target. Visual targets could be presented anywhere within a large visual field including locations to the side of and behind the head, where head movements were necessary to fixate the target, and where one might imagine that the auditory accessory might be highly beneficial in signaling the presence of this target. A train of clicks (the auditory accessory) accompanied the onset of the visual target and was played at either the location of the target (termed correlated), or at the fixation point (termed uncorrelated). Note that this click train gave no indication as to the identity of the visual target (and therefore no indication of the correct response).

As predicted, the correlated sound improved performance (relative to the uncorrelated sound) when the visual target was presented behind the head, however, it also improved performance at locations where the target was immediately visible, including those near the fixation mark. On the basis of evidence from this and related experiments Perrott et al (1990, page 225) suggest that

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

There is, however, one problem that might detract from such a position. There was a substantial effect of correlated compared with uncorrelated sound even when the visual target (and therefore the auditory signal) actually appeared at the point of fixation. Under these circumstances there was no need to move the eyes at all, hence explanations based on eye movements cannot account for this effect. Furthermore, identical stimuli were presented in the correlated and uncorrelated sound conditions so the difference in RTs must lie in the endogenous factors or strategies employed by the participant. The trials were blocked according to sound type: in the correlated condition the sound was always a useful indicator of target location but in the uncorrelated condition the sound was always presented at the centre of the display and, therefore, in the vast majority of trials it was not at the target location. Observers may have developed strategies for sound use based on its perceived utility and, as a result, may have come to rely on the location of the sound far more in the correlated sound condition than in the uncorrelated condition. Owing to these differing strategies, auditory location may have been more useful when correlated sound was employed despite the fact that, for targets at fixation, the stimulus conditions were identical in correlated and uncorrelated sound conditions.

Facilitation within the central portion of the visual field was re-examined in a later study (Perrott et al 1991) with a relatively cluttered display in which distractor items shared some characteristics with the target. Again, foveal vision was necessary to perform the visual discrimination task. Response times were faster when spatially congruent sound accompanied the target compared with when no sound was presented. This appears to demonstrate the utility of sounds located at the visual target location (though there was no detailed analysis of errors). However, as there was no condition where sound was presented away from the target's location, it does not tell us if the *location* of the sound was an effective component of this improvement, nor whether speeded or increased accuracy of overt eye or head movements are essential for the effect.

In the studies of Perrott et al (1990, 1991) both target-directed eye movements and head movements were employed: these movements may be viewed as an overt orienting of attention. However, the direction of spatial attention does not always have to follow the line of gaze, and it is possible to shift attention to a region of space without overt receptor movements to that location (Yantis 1998). There is considerable evidence that strong behavioural and neural links between overt orienting and attention may exist and even that "accurate saccades require shifts of perceptual attention to the target" [Kowler et al (1995, page 1913); for a review see Stelmach et al (1997)]. Facilitation of visual target acquisition by auditory information may be mediated by enhanced receptor movements to the target location, but might also occur when only covert orienting is employed.

However, it is also possible that the effect of the auditory signals might not have anything at all to do with overt/covert movements and that auditory facilitation may be relatively independent of sound-source location. The sound may simply act as a warning signal that improves RTs by inducing a faster button press—in this case faster responses may be accompanied by a cost to accuracy (Posner et al 1976)—or by improving performance for targets that occur at this location in time. Essentially this argument suggests that there is a degree of uncertainty about when the target will appear so information is gathered over a long temporal period, thus the information acquired contains a lot of 'noise' as well as 'signal'. By reducing the temporal uncertainty this period can be reduced so that less noise is gathered and the quality of information is improved—hence better performance.

The primary aim of this study is to examine the way in which auditory accessory signals may improve performance in a visual task. Our experiments focus upon three main questions. First, are overt eye movements necessary? Second, is the spatial aspect of the sound critical to facilitation? Finally, do other warning signals (eg visual ones) produce a similar effect?

2 Experiment 1

In order to test the above questions we felt it was necessary for us to replicate the findings of Perrott et al (1990) using target stimuli that closely resembled their 'highlighted' arrowhead stimuli (Perrott et al 1991, experiment 2). Critical to this experiment was that a saccadic eye movement was essential for the observer to be able to identify the target. This was achieved by reducing the size of the arrowhead so that foveal vision was essential to resolve the stimulus and complete the discrimination task. The arrowhead was contained within a red box (see figure 1a) whose presence was easy to detect when presented. The visual target was presented in isolation or target appearance was simultaneous with a sound at the target location, a sound at the fixation point, or a change in the colour of the fixation point.

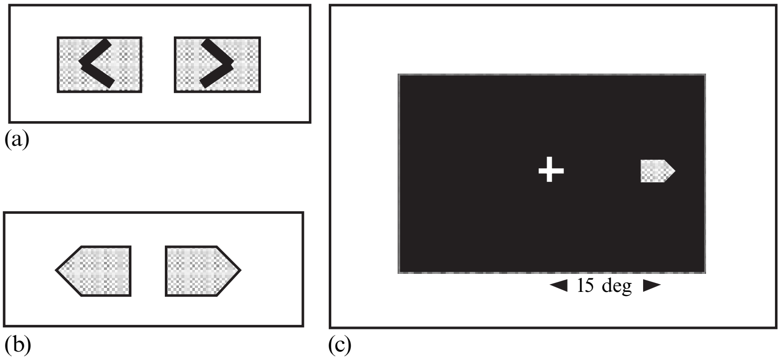

Figure 1. Illustration of the stimuli used. (a) Arrowhead stimuli. (b) Pentagon stimuli. (c) Display.

It was reasoned that these conditions would allow us to test the questions of interest. First, the effect of sound at the target location was compared with the vision-alone baseline to determine whether spatially congruent sound enhanced the processing of the visual target. Second, the effect of sound located at the target location was compared with that of a sound located at the fixation point to examine the impact of sound-source location upon facilitation. The possibility that any warning signal (including a visual one) might have the same effects was assessed by examining the results of the fixation-point-change (FP status) condition: a general alerting effect would result in better performance in the FP status than the visual baseline condition.

It was also of interest to consider what the role of auditory spatial information might be. If the spatial aspect of the accessory signals is important, one might imagine that 'attention' would be drawn to the location of the sound and that, relative to the visual baseline, performance would be worse in the central sound and FP status conditions. In all experiments, response times and errors were examined to determine whether facilitation was characterised by a change in response speed alone—thus indicating a genuine improvement in performance—or whether a speed/accuracy trade-off had occurred.

2.1 *Methods*

2.1.1 *Participants.* Ten undergraduates, aged between 18 and 36 years, participated in this experiment. All reported normal or corrected-to-normal vision and normal hearing.

2.1.2 *Apparatus and stimuli.* The laboratory was divided into a test chamber (3.67 m × 4.06 m × 2.34 m), the walls of which had been covered with sound-attenuating foam to reduce echoes, and a connecting room in which the equipment was housed. An array of loudspeakers was mounted on one wall of the test chamber and concealed by an acoustically transparent screen onto which the display was projected. Participants were seated in the test room at 2.36 m from the screen.

2.1.3 *Visual stimuli.* The visual display was generated and controlled with Microsoft Visual Basic (version 3.00 for Windows) on a Dell Pentium PC (Dimensions XPS P90) and projected onto the screen via a Sony LCD projector (VPL-3510). The stimuli were generated off-line in Microsoft Paintbrush (version 3.1) and imported to the visual display prior to each trial. The display was projected onto a black background (luminance 0.10 cd m^{-2}). This consisted of a red or white fixation cross (1.0 deg × 1.0 deg; luminance 1.00 cd m^{-2}) at 0 deg azimuth and elevation, and a solitary target. The target was a black arrowhead in a red rectangle (1.0 deg × 0.6 deg; luminance 1.00 cd m^{-2}) located 15 deg to the left or right of fixation (see figure 1). Eye movements to the target location were necessary to perform the discrimination task as foveal vision was needed to resolve these small targets.

2.1.4 *Auditory stimuli.* Auditory 'click-train' stimuli were generated with an analogue and digital I/O board (Data Translation: DT2821-F16SE) and produced with a SoundBlaster audiocard (version 1.0 0.02). Creative Wave Studio (version 1.0 0.02) was used to edit the stimuli so that click rate (10 Hz), the length of individual samples (15 ms), and the interval between these (85 ms) were relatively constant. Output was directed from the computer via an eight-channel amplifier (of local construction) to the loudspeaker array and the digital output component of the Data Translation card was used to activate the selected speaker. Speakers were aligned with visual stimuli (at +15 and 0 deg azimuth, 0 deg elevation) and concealed by the screen. The speakers were flat against the back of this flat screen. These meant that there was a small variation (3.3%) in the actual distance from the speaker to the observer for the central speaker compared with the peripheral speakers. Pilot tests were performed to ensure that participants could localise sound accurately: participants were able to differentiate right, left, and central locations for these sounds with almost 100% accuracy.

2.2 *Design*

In practice and test blocks, equal numbers of trials from each of the four conditions were interleaved in random order: the target was presented alone (visual), accompanied by a sound at the target location (spatial) or the fixation-point location (status), or by a change in the colour of the fixation point (FP status). This experiment was conducted within the same session as experiment 2 (which involved both overt and covert orienting to visual targets). Blocks of trials from the two experiments were counterbalanced across participants.

2.3 *Procedure*

Prior to the experiment, participants were given information about the display and instructed that it would be necessary to direct their eyes to the stimulus in order to perform the discrimination task. Participants completed two sets of practice trials (80 trials in each) and two blocks of test trials (200 trials in each) within the session. Initially the display consisted of a white fixation cross and 'trial start' button: trials were self-paced so that participants could initiate a break if necessary. The target and accompanying stimuli were presented 1.5 to 2.5 s after the start of the trial and continued until the observer responded or until a further 3 s had elapsed.

Response times and errors were recorded. Trials in which response times were less than 300 ms (anticipations) were discarded and those greater than 3000 ms (time-out errors) were also removed but were analysed separately. From the remaining data, median RTs (medians were used as RT data are skewed from normal) and errors were calculated; these were later analysed with a within-subjects analysis of variance. A posteriori Tukey tests were used to examine specific hypotheses as a priori tests could not be justified for all comparisons of interest and, though this test increased the possibility of a Type II error, a conservative view of the data was preferred.

2.4 *Results*

Group mean response times (of the individual medians) are shown in figure 2. A one-way ANOVA revealed a significant effect of auditory condition ($F_{3,27} = 5.27$, $p < 0.01$). Specific hypotheses were examined: despite trends in the appropriate direction neither the apparent effect of spatial sound (26 ms) nor that of status sound (20 ms) differed significantly ($p > 0.05$) from the vision-alone baseline. Spatially congruent sound produced very little extra facilitation relative to the status sound (6 ms, $p > 0.05$). The FP status produced RTs that were only slightly slower (13 ms) than the no sound baseline ($p > 0.05$). There were, however, significant differences between the auditory and the fixation change (FP status) conditions: a posteriori tests (Tukey HSD) revealed that both spatial and status sound facilitated response times relative to the FP status condition (39 ms, $p < 0.01$ and 33 ms, $p < 0.05$, respectively).

Mean errors in each condition ranged from 3.0% to 6.2% and the ANOVA revealed no significant effect of auditory condition ($F_{3,27} = 2.31$, $p = 0.10$).

2.5 *Discussion*

Experiment 1 was designed to replicate and extend the paradigm of Perrott et al (1990, 1991). The overall pattern of results was indeed consistent with that of Perrott et al (1991), and effects appeared to be similar in magnitude to those reported by Perrott et al (1991, experiment 2) for a solitary target located 14.8 deg from the fixation point. However, some of the crucial comparisons failed to reach statistical significance. Whilst the highlighted arrowhead target was similar to that used by Perrott et al (1991, experiment 2) it is possible that small differences in stimulus size or luminance across the two experiments, or reduced target uncertainty may account for the lack of significance of these individual comparisons. Also, observers' strategies for using sound may have reduced the impact of sound relative to earlier studies in which fewer auditory conditions were employed or different auditory conditions run in separate blocks (eg Perrott et al 1990, 1991). Nevertheless, our replication of the effect of sounds upon a visual-detection task for which foveal vision (and therefore fixation of the target) was necessary allows us to test the mechanisms behind this effect.

3 Experiment 2

In this experiment the arrowhead stimuli were replaced by solid five-sided red shapes, or pentagons (see figure 1b). It was possible to discriminate left and right pointing pentagons *with or without* overt orienting. The four auditory conditions examined mirrored those of the previous experiment: targets were presented alone or were accompanied by sound at fixation, sound at the target location, or a change in the fixation point. Additionally, two sets of instructions were used in separate blocks: observers were instructed to move their eyes rapidly to the target before responding (overt condition) or to maintain fixation on the central cross throughout (covert condition). The effect of the auditory accessory signal was compared across the overt and covert conditions to determine whether eye movements were necessary for facilitation to occur.

3.1 *Method*

New targets were employed and an eye-movement manipulation introduced. Otherwise, methods were identical to those employed in the previous experiment. As mentioned previously this experiment was run within the same session and interleaved with experiment 1. Therefore, the same observers took part in both experiments.

3.2 *Design*

The arrowhead stimuli were replaced by pentagon stimuli (see figure 1b) of similar size and luminance (0.75 to 1.25 deg by 0.6 deg; luminance 1.00 cd m^{-2}). It was, however, possible to discriminate these targets *without* foveal vision when they were located 15 deg to the left or right of fixation

3.3 *Procedure*

Observers were given information about the display followed by the appropriate eye-movement instructions. The eye-movement instructions were emphasised: in the overt orienting block participants were asked to fixate the central cross until the target appeared and then to move their eyes rapidly to the target before responding, whilst in the covert orienting condition they were instructed to fixate the central cross throughout the trial. Covert orienting was described as akin to seeing the stimulus 'out of the corner of one's eye'.

3.4 *Results*

Figure 2 illustrates the mean of the observers' median RTs for each combination of eye-movement and accessory signal conditions. A two-way ANOVA revealed a significant main effect of accessory signal ($F_{3,27} = 19.57$, $p < 0.0001$) and a marginal effect of eye-movement condition ($F_{1,9} = 3.59$, $p = 0.09$). The interaction of these two variables was not significant ($F_{3,27} = 0.29$, $p = 0.83$), which indicates that eye movements are not necessary for the facilitatory effects of accessory signals to occur.

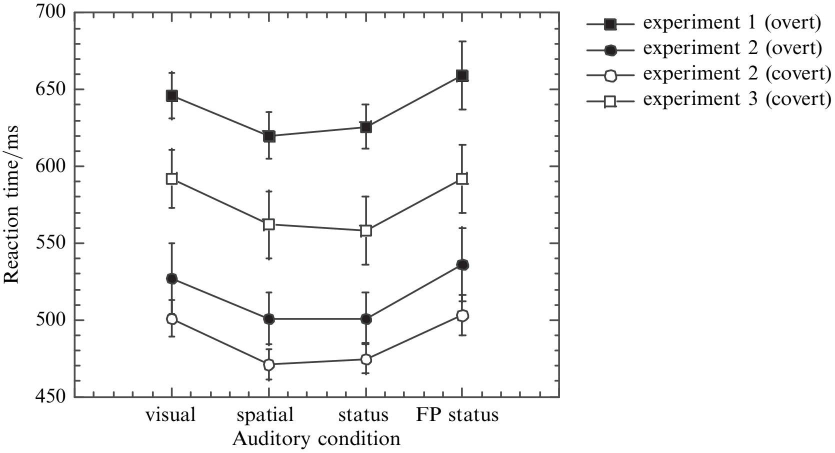

Figure 2. Data from experiments 1 – 3. Reaction time is plotted for the four conditions. Data from experiment 1 (where eye movements were necessary) are represented by the filled squares. Data from experiment 2 are represented by circles: filled symbols indicate trials in which the observers were instructed to move their eyes, and open symbols show trials in which they were instructed to maintain fixation. Data from experiment 3 (where eye movements were rendered useless) are represented by open squares. Error bars represent ±1 standard error of the mean.

Collapsed across eye-movement condition, the spatial sound condition produced faster RTs than did the vision-alone baseline (28.0 ms, $p < 0.01$), thus confirming the utility of auditory spatial accessory signals. Central sounds also produced facilitation relative to the visual baseline (25.7 ms, $p < 0.01$). Responses in the spatial sound condition were not significantly faster than those observed when central sound occurred (−2.3 ms, $p > 0.05$). Finally, the fixation-point-change condition was not significantly faster than the vision-alone baseline (−5.9 ms, $p > 0.05$).

Fewer than 5% errors occurred in any condition. The ANOVA showed no significant effect of accessory signal ($F_{3,27} = 1.23$, $p = 0.32$), and a marginal effect of eye-movement condition ($F_{1,9} = 3.94$, $p = 0.08$). The interaction of these two variables was not significant ($F_{3,27} = 0.73$, $p = 0.55$). The marginal effect of eye movement appears to be a tendency towards greater errors when eye movements are prohibited.

3.5 *Discussion*

Auditory accessory signals (whether spatial or central) produced facilitation relative to the visual baseline, whereas a visual accessory signal did not. This pattern of results was the same whether or not eye movements occurred. As there was no evidence of a speed/accuracy trade-off it appears that the facilitation provided by the auditory accessory stimulus reflects a genuine improvement in signal processing or decision making.

From these data it is tempting to conclude that eye movements are not necessary for the facilitatory effects of auditory accessory signals to occur and, therefore, that theories of faster and/or more accurate saccades under these conditions must be wrong or, at least, incomplete (eg Perrott et al 1990)—however, caution may be justified. We manipulated eye movements (or lack of them) by instructions to the observers.

Whilst our observers reported no problem in obeying these instructions, it is still possible that target-directed eye movements might have occurred (at least on some trials). We therefore repeated this experiment under conditions where any eye movement would be ineffective. We limited the duration of the visual signal (and auditory signal) to 150 ms. Under such conditions it is nearly impossible to initiate a saccade, let alone complete one, when a fixation point is continuously visible (Fischer and Weber 1993).

4 Experiment 3

In this experiment we aimed to replicate the essential conditions of experiment 2 when the eyes are fixating the central mark throughout a trial. The duration of the visual target was reduced to preclude overt orienting.

4.1 *Methods*

Apart from the following changes, all methods were the same as those employed in experiment 2. A left or right pointing pentagon served as the target; however, in this case, stimulus duration was reduced to 150 ms to preclude overt orienting. Due to software limitations the auditory stimulus had to be altered to limit and control the duration of the visual target and auditory accessory signal: instead of the click-train stimulus a 150 ms noise burst was employed. A pilot study revealed that both click-train and noise stimuli were localised with almost 100% accuracy. Observers were asked to fixate the central cross throughout each trial and were additionally informed that it would be impossible to make an eye movement to the target before it disappeared. Ten new naïve participants were recruited under the same conditions as experiment 1.

4.2 *Results*

Figure 2 shows the mean of median RTs in this study. The ANOVA revealed a significant effect of the accessory signal ($F_{3,27} = 11.96$, $p < 0.0001$). Again the crucial comparisons were examined with an a posteriori Tukey HSD test. Compared with the visual baseline, RTs were faster when spatial sound was employed (29.0 ms, $p < 0.01$) and when status sound was employed (33.3 ms, $p < 0.01$), but there was no significant change in RT for fixation-point change (-0.7 ms, $p > 0.05$). Latencies obtained with spatial sound were not significantly faster than those with status sound (-4.3 ms, $p > 0.05$).

Fewer than 3% errors were obtained in any condition and the ANOVA revealed no significant effect of the accessory signal ($F_{3,27} = 0.61$, $p = 0.62$) upon errors.

4.3 *Discussion*

Under conditions in which eye movements were excluded (or rendered useless) we obtained the same pattern of results as in the previous experiment, namely that the auditory accessory signals improved performance over the visual conditions and that the spatial aspect of the sound had no effect. Examination of figure 2 shows that the pattern of results is similar for experiments 1 – 3.

These results rule out differential eye movements (Perrott et al 1990) as an explanation of the effects of all auditory signals. Of course, it is still possible that, under the conditions where the visual stimulus is hidden or very hard to detect without appropriate eye movements, the spatial aspect of the auditory signal may well act to enhance target acquisition. The role of the detectability of the visual target will be addressed in the next experiment.

The fixation-point change was included to determine whether the accessory signal provided a temporal marker. It was suggested that the temporal marker afforded strategies that enhanced the signal-to-noise ratio for all targets relative to when no temporal marker was available. No such improvement was observed, which suggests that this is not a valid explanation of the facilitatory effects of auditory accessory signals.

In experiments 1 – 3 we found no difference between the effects of a sound at the target location and one placed centrally. This was somewhat surprising, as the sound could potentially play a role in target localisation, facilitating orienting to its location. One possibility is that the visual signal alone was perfectly capable of attracting attention to its location, hence the location of the auditory signals was redundant—this idea, of course, does not explain why the auditory signals still have any facilitatory effect. We therefore reasoned that if the visual information was poor then the utility of auditory spatial information may be increased and that, under these conditions, an effect of sound-source location might emerge. This would be evident as a significant improvement in performance in the spatial sound condition relative to a condition in which sound was presented centrally and therefore provided no information about target location.

5 Experiment 4

In this experiment we once again compared the effects of central and spatial sounds on the observer's ability to discriminate visual targets. The fixation-point-change condition was omitted from this experiment as we had established that this visual warning signal did not facilitate target discrimination. To manipulate the detectability of the visual stimulus we employed three levels of target luminance: high luminance corresponded to that used in the previous experiments, low luminance was chosen to be near the observers' thresholds (on the basis of a pilot study), and medium luminance was between these two values. Again, target duration was limited to prevent eye movements.

5.1 *Methods*

The methods were practically identical to those in the previous experiment, but target luminance was manipulated. Left and right pointing pentagon targets of high (1.00 cd m^{-2}), medium (0.53 cd m^{-2}), and low (0.20 cd m^{-2}) luminance were employed. Trials containing these three target types were interleaved within each block. Sixteen observers took part in this experiment and were recruited under the same conditions as experiment 1. After a practice block of 45 trials, each observer completed 2 blocks of 90 trials. Luminance and auditory conditions were combined to provide the nine types of trial and each of these was presented 10 times in each of the blocks. Right and left target types and locations were presented with equal frequency. Data from the two blocks were combined.

5.2 *Results*

The means of the median RTs are shown in figure 3. The ANOVA revealed a significant main effect of target luminance ($F_{2,30} = 65.29$, $p < 0.0001$). As expected, response times were reduced as target luminance increased. The main effect of accessory signals was also significant ($F_{2,30} = 20.14$, $p < 0.0001$), however, there was no significant interaction between these two variables ($F_{4,60} = 1.62$, $p = 0.18$).

Of crucial interest in this experiment was whether the central and spatial sounds produced different effects. To examine this we excluded the data from the vision-alone baseline and repeated the ANOVA. Once again the effects of luminance were significant ($F_{1,15} = 22.0$, $p < 0.001$), but now there was no effect of auditory condition ($F_{1,15} = 1.85$, $p = 0.19$) and no interaction ($F_{1,15} = 0.09$, $p = 0.77$).

An ANOVA of the percentage errors revealed a significant main effect of luminance level ($F_{2,30} = 30.48$, $p < 0.0001$; low 44.7%, medium 5.7%, high 2.2%), but no effect of accessory signals ($F_{2,30} = 0.50$, $p = 0.61$) or interaction ($F_{4,60} = 0.68$, $p = 0.61$). The ANOVA of the percentage of time-out errors (failure to respond within 3 s) revealed a significant main effect of luminance level ($F_{2,30} = 57.54$, $p < 0.0001$; low 34.8%, medium 7.4%, high 3.9%), but no effect of accessory signal ($F_{2,30} = 0.87$, $p = 0.43$) or interaction ($F_{4,60} = 0.78$, $p = 0.54$).

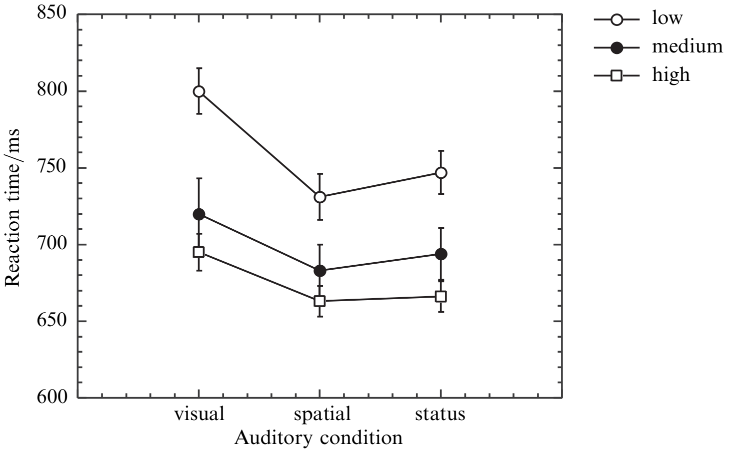

Figure 3. Data from experiment 4. Reaction time is plotted for the three conditions with target luminance (low—open circles, medium—filled circles, and high—open squares) as a parameter. Error bars represent ±1 standard error of the mean.

In sum, it is apparent that luminance had a dramatic effect upon performance to the extent that many (> 30%) of the targets at the lowest luminance level were completely missed, and those that were detected were often incorrectly classified (> 40%). Auditory stimuli had effects similar to those found in the previous experiments. Both spatially congruent and central sounds improved performance (producing faster RTs without greater errors) and did so by similar amounts. Finally, whilst there was no significant interaction of the effects of auditory information with luminance there was a noticeable trend towards greater overall facilitation with luminance (low = 61 ms, medium = 31 ms, high = 30 ms).(1)

5.3 *Discussion*

Despite lowering the quality of the visual information to near threshold levels we still obtained similar effects of both spatial and central sounds upon the identification of the visual targets. Under these poor quality visual conditions it seems most unlikely that the visual information alone would provide sufficient information about target location and render the auditory spatial information redundant. Results indicate that the spatial aspect of the auditory information may be redundant.

6 Experiment 5

Within the domain of spatial pre-cueing experiments it is well established that visual cues can facilitate orienting to the target location in one of two ways—by exogenous and endogenous selection (Theeuwes 1994). Exogenous selection is stimulus-driven, by such events as a stimulus onset, and is thought to be relatively automatic and effortless. Endogenous selection, on the other hand, is goal-driven by the aims of the observer and the cue is often presented in symbolic form such as an arrow indicating the likely location of a target. Such a distinction can also be made for cueing in other modalities (Spence and Driver 1994), and there is evidence that an auditory pre-cue can influence the processing of a visual stimulus in both an exogenous (Spence and Driver 1997) and endogenous (Spence and Driver 1996) manner.

So far we have failed to obtain any evidence that the spatial location of a simultaneous auditory accessory signal was important in aiding the processing of the visual

---

(1) In other studies we have further explored this issue. When we used purely spatial sounds and blocked luminance conditions, statistically significant greater auditory facilitation for low-luminance targets than high-luminance targets was obtained (see Doyle 1999).

target. However, in these preceding experiments we interleaved the various auditory and visual conditions within a block of trials. As a result, the spatial sound condition occurred on only a quarter of the trials in experiments 1 – 3 and in one third of trials in experiment 4. As the sound occurred infrequently and was not always presented in the same location as the target, observers may have perceived the spatial aspects of the auditory stimulus to be of limited use. They may therefore have developed strategies for responding to the visual target without using the sound. To investigate the issue of the strategic use of the spatial aspects of the visual cue we manipulated the validity of the spatial information provided by the auditory stimulus. Though our auditory accessory signal does not precede the visual target (and therefore cannot be defined as a cue) evidence from earlier studies (Perrott et al 1991) suggests that strategic, or endogenous, factors may have some impact upon the effects of simultaneous sound upon visual performance.

Sound now occurred on every trial (so it was always expected) and occurred either at the location of the visual target or on the *opposite* side. In separate blocks of trials we manipulated the validity (ie the proportion of trials with sound at the target location) from 80% to 50% to 20%. This manipulation produced situations in which the sound location could only influence performance through exogenous selection (50% validity of sound location), where both exogenous and endogenous cues could act together to aid in target localisation (80% validity of sound location) and finally, one in which the exogenous aspects were in opposition to the endogenous aspects (20% validity of sound location). It is important to note that the auditory signals occurred simultaneously with the visual target, unlike other recent studies that have looked at the cueing effects where the auditory stimulus precedes the visual target (eg Spence and Driver 1996, 1997).

6.1 *Methods*

The stimuli and methods were similar to those of experiment 4 save that all targets were of high contrast. Sixteen observers, naïve and recruited as in experiment 1, completed 3 blocks of 90 trials. In one block (80%) the target appeared on the same side as the sound on 80% of the trials, and therefore on the other side on 20% of the trials. On another block (50%) the target location and sound location were not correlated. On another block the target occurred on the side of the sound on 20% of the trials. Observers were informed of these contingencies before each block. The order of the three blocks was randomised for each observer.

6.2 *Results*

The means of the median RTs are shown in figure 4. An ANOVA revealed no effect of sound position ($F_{1,15} = 2.12$, $p = 0.17$), validity ratio ($F_{2,30} = 0.45$, $p = 0.64$), or interaction ($F_{2,30} = 0.20$, $p = 0.82$). An ANOVA of errors revealed no effect of sound position ($F_{1,15} = 0.00$, $p = 0.99$) or validity ratio ($F_{2,30} = 0.37$, $p = 0.69$), and no interaction ($F_{2,30} = 0.05$, $p = 0.96$).

6.3 *Discussion*

The results showed once again that the location of the sound source did not influence the magnitude of facilitation, despite the fact that the effect of a sound at the target location was compared to one at an opposite location. This was true under all the validity ratios tested, even when the sound was a good predictor of target location. The failure of auditory spatial information to influence response times in our previous experiments may, therefore, not be attributed to the effects of auditory validity upon observers' strategies.

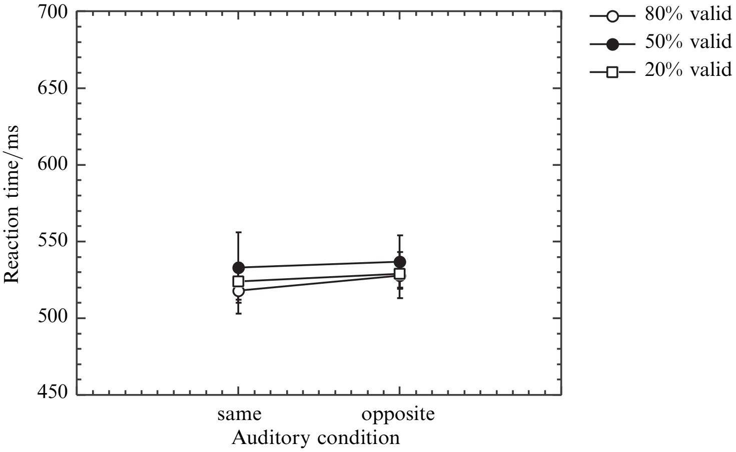

Figure 4. Data from experiment 5. Reaction time is plotted for the two conditions with target/sound location correspondence (80%—open circles, 50%—filled circles, and 20%—open squares) as a parameter. Error bars represent ±1 standard error of the mean.

7 General discussion

Our results can be summarised as follows.

(i) Auditory accessory signals enhanced performance in a visual-discrimination task: when sound occurred, reaction times to correctly identify the target were reduced without any loss in accuracy.

(ii) Auditory facilitation was evident whether target-directed eye movements were necessary (experiment 1), permitted (experiment 2), not permitted (experiment 2), or rendered impossible (experiment 3).

(iii) Visual accessory signals that were presented at an equivalent time to the auditory accessory signals did not produce such facilitation.

(iv) The location of the auditory signal had no significant impact upon the facilitation observed.

(v) These effects existed even when the quality of the visual information was very poor, and when the spatial location of the auditory signal was deliberately misleading (experiments 4 and 5).

Our conclusions appear to contradict those of Perrott et al (1990) in two ways. First, we found no role for a spatial aspect of auditory information in our study, whereas Perrott et al showed faster responding when the sound was at the location of the visual target rather than located at the centre of the display. However, as we noted in the introduction, this may have been due to the blocking of trials. A comparison of response times to visual targets and accessory sounds that both occurred at the central location revealed a difference between the block on which sound was always congruent with the target, and the block in which the sound was always at a central location (and therefore often not aligned with the target). As the stimuli were identical the difference in reaction times must arise from the differing strategies adopted in the two blocks. That a 'nonspatial' accessory auditory signal can improve visual performance may also be inferred from the data of Keuss et al (1990). They presented visual targets at a central location and auditory accessory signals through headphones and showed that response times decreased as the loudness of the auditory signal increased. Whilst no purely visual signal was tested, one can infer from the fact that low intensity sounds produced less facilitation than high intensity ones that such an effect would probably exist.

Second, Perrott et al (1990) claim that their results are due to enhanced eye movements to the audiovisual target. Whilst we do not dispute the notion that saccades to audiovisual targets are faster than to purely visual ones (see also Corneil and Munoz 1996; Engelken and Stevens 1989; Frens et al 1995; Hughes et al 1994; Nozawa et al 1994), two points seem pertinent. First, our data clearly demonstrate that somewhat similar effects (visual targets are identified more rapidly) are produced without eye movements. Second, it is necessary to establish whether the faster eye movements result from the spatial aspects of the sound or, as we have demonstrated here in a visual-identification task, whether the effects on saccadic latency can be non location-specific. In considering this second point there is considerable research that is germane. For example Konrad et al (1989) compared saccadic latencies to locations chosen at random from trial to trial, in the presence or absence of an auditory stimulus whose position was removed from that of the visual targets and was nonpredictive. Saccades were faster in the presence of this sound, showing that nonspatial sounds can increase the speed of saccades (unfortunately there are no data relating to response accuracy so this facilitation may reflect a criterion shift). A similar effect is apparent in the data of Ross and Ross (1981) in which an auditory signal was presented directly above the visual display (and was therefore neutral as to target position). Inspection of their figure 1 shows that auditory signals shorten saccadic latency by around 40 ms. Hence, there is previous evidence that nonspatial auditory information can shorten saccadic latency.

Returning to the earlier point—that we observed auditory facilitation in the absence of eye movements—one can ask whether the mechanisms that underlie auditory facilitation in the absence of eye movements might also underpin the effects with eye movements. One possibility is that covert attention might precede any overt receptor movement, and therefore the effects that we have documented are essentially the precursor of the effects observed in the eye-movement literature. There is a considerable body of evidence supporting the view that attention precedes an eye movement [at least when attention is summoned by stimulus-driven (exogenous) factors, whilst the evidence is somewhat contradictory for goal-driven (endogenous) factors; Stelmach et al (1997)]. Hence, our results could be explained if the auditory signal could somehow aid in shifting covert visual attention. Whilst there is a now good evidence that auditory signals can be used to guide visual attention in pre-cueing paradigms (eg Spence and Driver 1996, 1997) these studies cannot explain how a spatially neutral signal enhances performance.

There is another nonspatial effect that can aid in reducing saccadic latencies, and therefore has certain features in common with the results found in this study. The 'gap effect' refers to the demonstration that saccadic latency is reduced if the fixation point is removed a little before the saccade target is presented (Saslow 1967) and can, in some circumstances, induce 'express saccades'.(2) Furthermore similar 'gap effects' can produce faster response times to visual targets even when eye movements are not allowed (Mackaben and Nakayama 1993). The presence of a gap effect for covert orienting is relevant to the current pattern of results. Whilst there is still some controversy about the origins of this effect [see Fischer and Weber (1993) and related commentaries; Findlay and Walker (1999)], the most popular view is that removal of the fixation point leads to a state of 'disengagement' that allows faster movement of resources to the target location. In studies of saccade latency a similar 'gap' effect has been observed following presentation of an auditory stimulus and it has been suggested that the sound may act in a similar way to fixation-point offset [Kimmig (1986) as cited in

(2) There is a large literature on when and where express saccades might by induced and what they mean. A review of this issue is beyond the scope of this paper and we merely wish to draw attention to the fact that removal of the fixation point reduces saccade latency. The interested reader is referred to the review of Fischer and Weber (1993) and the related commentaries.

Fischer and Weber (1993)], a proposal that is supported by other data (Munoz and Corneil 1995; Ross and Ross 1981). Given the evidence that fixation-point offset may induce a gap effect for overt and covert orienting, and that nonspatial sound might produce a similar effect in overt orienting, we suggest that sound may also produce a gap effect for covert orienting. In essence, the auditory stimulus may promote disengagement of attention from the fixation mark and thus allow covert spatial attention to reach (or be reassigned) to the target location more rapidly.

That an auditory stimulus (in any location) can aid visual target acquisition seems in line with the previous findings reviewed above. However, one might also expect that sound at the target location might further enhance performance, perhaps by helping to guide attention to the correct area or by forming an audiovisual object that is more that the sum of its parts. There are several possible explanations for the absence of a spatial effect of sound. The first is that the observers know that the information needed to formulate the response is visual and so guide any required attentional movements via visual information alone (one has to presume that the 'disengagement' hypothesised above is not influenced by such a strategy). However, our attempts to manipulate the quality of the visual information (experiment 4) or to increase/decrease the perceived value of the position information given by the auditory signal (experiment 5) failed to uncover support for this notion. Second, there is considerable evidence that the perceived location of a sound may be strongly influenced by visual information—a ventriloquism effect (eg Bertelson and Aschersleben 1998; Driver 1996; Thomas 1941). As there was only a single visual item appearing simultaneously with the auditory signal, the status sound could have been drawn to the location of the visual target. Note that this doesn't necessarily mean that the perceived location of the sound would be at the visual target: the neural basis that controls our attentional movements may differ from those of our conscious report. Examination of the eye-movement literature yields only little support for this possibility. Lueck et al (1990) show that saccades to an auditory target are influenced by visual distractors such that the saccade falls to some weighted average of the two stimuli. However, such an effect did not occur when the two stimuli were in opposite hemifields, as was the case for our stimuli in experiment 5. Despite this, the notion that the spatial and nonspatial sounds produce similar effects due to a ventriloquism effect can not be ruled out on the basis of the present data.

Several brain areas contain 'multisensory' neurons that are capable of responding to stimuli from a number of modalities (Stein and Meredith 1994). For example, the deep layers of the superior colliculus contain neurons that are responsive to various combinations of visual, auditory, and somatosensory stimuli. The receptive fields of these neurons are arranged to form topographic maps of space that are aligned for all three modalities. When stimuli from different modalities co-occur at the same location, responses of these neurons may be significantly enhanced. The superior colliculus, which is involved in control of eye movements (eg Schiller 1998), and may also be involved in covert orienting (Kustov and Robinson 1996), contains a motor map of space. The motor map is aligned with sensory areas so that stimulation of a particular region of the sensory map will result in an eye movement that brings the fovea to that location (Stein and Meredith 1994). It has been reported that enhanced neural activity following sensory integration may influence orienting behaviour in cats (Stein et al 1988, 1989) and in humans (Costin et al 1991), and that the activity of multisensory neurons may underlie the effects of auditory stimuli upon visual saccade latencies (Frens et al 1995) and covert orienting. Given that such cells play a role in both attentional and eye movements, and that their response to a visual stimulus may be facilitated by auditory stimuli, these cells are obvious candidates for the possible neural substrate of the effects we have described.

In conclusion, we have demonstrated that auditory information may aid the acquisition of visual targets and have shown that these effects may occur even in the absence of eye movements. Surprisingly, the spatial aspects of the sound were of little value under these conditions. Sound may prove to be an inexpensive method of enhancing an operator's acquisition of visual information in visually cluttered environments, such as a cockpit. Auditory spatial information provided by spatially congruent sound appeared to provide little advantage over auditory information per se. This implies that it may be possible to avoid the costs of developing a spatially intelligent audio-visual display. However, the displays we have used here are simple and uncluttered and so the visual information is also salient. Our conclusions about the lack of the usefulness of spatial information may not apply to other conditions where the necessary visual information is not salient (Doyle 1999).

Acknowledgement. This research was partly funded by the Defence Research Agency, Farnborough, Hants.

References

Bernstein I H, 1970 "Can we see and hear at the same time? Some recent studies of intersensory facilitation of reaction time" *Acta Psychologica* **33** 21 – 35

Bernstein I H, Edelstein B A, 1971 "Effects of some variations in auditory input upon visual choice reaction time" *Journal of Experimental Psychology* **87** 241 – 247

Bernstein I H, Clark M H, Edelstein B A, 1969 "Effects of an auditory signal on visual reaction time" *Journal of Experimental Psychology* **80** 567 – 569

Bertelson P, Aschersleben G, 1998 "Automatic visual bias of perceived auditory location" *Psychonomic Bulletin & Review* **5** 482 – 489

Corneil B D, Munoz D P, 1996 "The influence of auditory and visual distractors on human orienting gaze shifts" *Journal of Neuroscience* **16** 8193 – 8207

Costin D, Neville H J, Meredith M A, Stein B E, 1991 "Rules of multisensory integration and attention: ERP and behavioural evidence in humans" *Society of Neuroscience Abstracts* **17** 656

Doyle M, 1999 *Facilitatory Effects of Response-irrelevant Sound upon Acquisition of Visual Targets* PhD thesis, Department of Psychology, Cardiff University, Cardiff, UK

Driver J, 1996 "Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading" *Nature* **381** 66 – 68

Engelken E J, Stevens K W, 1989 "Saccadic eye movements in response to visual, auditory and bisensory stimuli" *Aviation, Space and Environmental Medicine* **60** 763 – 768

Findlay J M, Walker R, 1999 "A model of saccade generation based on parallel processing and competitive inhibition" *Behavioral and Brain Sciences* **22** 661 – 721

Fischer B, Weber H, 1993 "Express saccades and visual attention" *Behavioral and Brain Sciences* **16** 553 – 610

Frens M A, Opstal A J van, Willigen R F van der, 1995 "Spatial and temporal factors determine auditory – visual interactions in human saccadic eye movements" *Perception & Psychophysics* **57** 802 – 816

Gielen S C A, Schmidt R A, Heuvel P J M van den, 1983 "On the nature of intersensory facilitation of reaction time" *Perception & Psychophysics* **34** 161 – 168

Hershenson M, 1962 "Reaction time as a measure of intersensory facilitation" *Journal of Experimental Psychology* **63** 289 – 293

Hughes H C, Reuter-Lorenz P A, Nozawa G, Fendrich R, 1994 "Visual – auditory interactions in sensorimotor processing: saccades versus manual responses" *Journal of Experimental Psychology: Human Perception and Performance* **20** 131 – 153

Keuss P G J, Zee F van der, Bree M B M van den, 1990 "Auditory accessory effects on visual processing" *Acta Psychologica* **75** 41 – 54

Kimmig H, 1986 *Express Sakkaden beim Menschen: die Rolle der Aufmerksamkeit in der Vorbereitungsphase zielgerichteter Blicksrünge* Dissertation, Fachbereich Medizin, Universität Freiburg, Freiburg, Germany

Konrad H R, Rea C, Olin B, Colliver J, 1989 "Simultaneous auditory stimuli shorten saccade latencies" *Laryngoscope* **99** 1230 – 1232

Kowler E, Anderson E, Dosher B, Blaser B, 1995 "The role of attention in the programming of saccades" *Vision Research* **35** 1897 – 1916

Kustov A A, Robinson D L, 1996 "Shared neural control of attentional shifts and eye movements" *Nature* **384** 74 – 77

Lueck C J, Crawford T J, Savage C J, Kennard C, 1990 "Auditory – visual interaction in the generation of saccades in man" *Experimental Brain Research* **82** 149 – 157
Mackaben M, Nakayama K, 1993 "Express attentional shifts" *Vision Research* **33** 85 – 90
Morrell L K, 1967 "Intersensory facilitation of reaction time" *Psychonomic Science* **8** 77 – 78
Morrell L K, 1968 "Cross-modality effects upon choice reaction time" *Psychonomic Science* **77** 14 – 18
Munoz D P, Corneil B D, 1995 "Evidence for interactions between target selection and visual fixation for saccade generation in humans" *Experimental Brain Research* **103** 168 – 173
Nickerson R S, 1973 "Intersensory facilitation of reaction time: Energy summation or preparation enhancement?" *Psychological Review* **80** 489 – 507
Nozawa G, Reuter-Lorenz P A, Hughes H C, 1994 "Parallel and serial processes in the human oculomotor system: bimodal integration and express saccades" *Biological Cybernetics* **72** 19 – 34
Perrott D R, Saberi K, Brown K, Strybel T Z, 1990 "Auditory psychomotor coordination and visual search performance" *Perception & Psychophysics* **48** 214 – 226
Perrott D R, Sadralodabai T, Saberi K, Strybel T Z, 1991 "Aurally aided visual search in the central visual field: effects of visual load and visual enhancement of the target" *Human Factors* **33** 389 – 400
Posner M I, Nissen M J, Klein R, 1976 "Visual dominance: an information processing account of its origins and significance" *Psychological Review* **83** 157 – 171
Raab D H, 1962 "Statistical facilitation of simple reaction time" *Transactions of the New York Academy of Sciences* **61** 193 – 199
Ross S M, Ross L E, 1981 "Saccade latency and warning signals: effects of auditory and visual stimulus onset and offset" *Perception & Psychophysics* **29** 429 – 437
Saslow M G, 1967 "Latency for saccadic eye movements" *Journal of the Optical Society of America* **57** 1030 – 1033
Schiller P H, 1998 "The neural control of visually guided eye movements", in *Cognitive Neuroscience of Attention: A Developmental Perspective* Ed. J E Richards (Hillsdale, NJ: Lawrence Erlbaum Associates) pp 3 – 50
Simon J R, Craft J L, 1970 "Effects of an irrelevant auditory stimulus on visual choice reaction time" *Journal of Experimental Psychology* **686** 272 – 274
Spence C, Driver J, 1996 "Audiovisual links in endogenous covert spatial attention" *Journal of Experimental Psychology: Human Perception and Performance* **22** 1005 – 1030
Spence C J, Driver J, 1994 "Covert spatial orienting in audition: exogenous and endogenous mechanisms" *Journal of Experimental Psychology: Human Perception and Performance* **20** 555 – 574
Spence C J, Driver J, 1997 "Audiovisual links in exogenous covert spatial orienting" *Perception & Psychophysics* **59** 1 – 22
Stein B E, Huneycutt W S, Meredith M A, 1988 "Neurons and behaviour: the same rules of multisensory integration apply" *Brain Research* **448** 355 – 358
Stein B E, Meredith M A, 1994 *The Merging of the Senses* Cognitive Neuroscience Series (Cambridge, MA: MIT Press)
Stein B E, Meredith M A, Huneycutt W S, McDade L, 1989 "Behavioural indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli" *Journal of Cognitive Neuroscience* **1** 12 – 24
Stelmach L B, Campsall J M, Herdman C M, 1997 "Attentional and ocular movements" *Journal of Experimental Psychology: Human Perception and Performance* **23** 823 – 844
Theeuwes J, 1994 "Endogenous and exogenous control of visual selection" *Perception* **23** 429 – 440
Thomas G J, 1941 "Experimental study of the influence of vision on sound localization" *Journal of Experimental Psychology* **28** 167 – 177
Todd J W, 1912 "Reaction time to multiple stimuli" *Archives of Psychology* **3** 1 – 65
Yantis S, 1998 "Control of visual attention", in *Attention* Ed. H Pashler (Hove, Sussex: Psychology Press) pp 223 – 256