
Cross-modal links in spatial attention

Jon Driver and Charles Spence

Phil. Trans. R. Soc. Lond. B 1998 **353**, 1319-1331
doi: 10.1098/rstb.1998.0286

References

Article cited in:

<http://rstb.royalsocietypublishing.org/content/353/1373/1319#related-urls>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

To subscribe to *Phil. Trans. R. Soc. Lond. B* go to: <http://rstb.royalsocietypublishing.org/subscriptions>



Cross-modal links in spatial attention

Jon Driver^{1*} and Charles Spence²

¹*Institute of Cognitive Neuroscience, Department of Psychology, Alexandra House, University College London, Gower Street, London WC1E 6BT, UK (j.driver@ucl.ac.uk)*

²*Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3LD, UK (charles.spence@psy.ox.ac.uk)*

A great deal is now known about the effects of spatial attention within individual sensory modalities, especially for vision and audition. However, there has been little previous study of possible cross-modal links in attention. Here, we review recent findings from our own experiments on this topic, which reveal extensive spatial links between the modalities. An irrelevant but salient event presented within touch, audition, or vision, can attract covert spatial attention in the other modalities (with the one exception that visual events do not attract auditory attention when saccades are prevented). By shifting receptors in one modality relative to another, the spatial coordinates of these cross-modal interactions can be examined. For instance, when a hand is placed in a new position, stimulation of it now draws visual attention to a correspondingly different location, although some aspects of attention do not spatially remap in this way. Crossmodal links are also evident in voluntary shifts of attention. When a person strongly expects a target in one modality (e.g. audition) to appear in a particular location, their judgements improve at that location not only for the expected modality but also for other modalities (e.g. vision), even if events in the latter modality are somewhat more likely elsewhere. Finally, some of our experiments suggest that information from different sensory modalities may be integrated preattentively, to produce the multimodal internal spatial representations in which attention can be directed. Such preattentive cross-modal integration can, in some cases, produce helpful illusions that increase the efficiency of selective attention in complex scenes.

Keywords: attention; cross-modal; touch; audition; proprioception; vision

1. INTRODUCTION

More than 40 years of intensive study have produced an extensive body of research concerning the dramatic effects of selective attention on perception and action (see, for example, Driver & Mattingley 1995; Pashler 1998, for recent reviews; plus the other papers in this issue). The pioneering psychological studies of the 1950s and early 1960s (see, for example, Cherry 1953; Broadbent 1958; Treisman 1964) were primarily concerned with attentional effects on audition. From the 1970s onward, the focus shifted toward the study of attentional effects within vision (see, for example, Eriksen & Hoffman 1972; Neisser & Becklen 1975; Posner 1978; Treisman & Gelade 1980). This trend continued through the main development of the 1980s, namely the increasing study of neural responses from single cells in behaving animals as a function of their attentional state (see, for example, Wurtz *et al.* 1982; Moran & Desimone 1985). The emphasis on vision has largely remained in more recent work on selective attention, which has used event-related potential (ERP), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI) methods (see, for example, Corbetta *et al.* 1993; Hillyard *et al.* 1995). Despite substantial progress with all these approaches, we would argue that most work has overlooked one crucial question; namely, the extent to which selective attention operates

cross-modally, rather than separately within each sensory modality.

It is customary for articles on selective attention to begin by noting that most everyday situations contain several stimuli, and that mechanisms of selective attention are therefore needed to pick out just the relevant sensory information for controlling current behaviour. However, the examples given typically refer to just a single sensory modality (e.g. to many objects in a visual scene all stimulating the retina at once; or to many sounds entering the ears simultaneously). Likewise, the experiments on attention that follow the opening remarks usually concern only a single sensory modality. Yet the fact is that most typical environments bombard all of our senses with numerous stimulations simultaneously. Even the textbook example of 'auditory' selective attention—listening to one conversation among many at a noisy party—turns out to be multimodal on closer inspection. Listeners typically depend not only on auditory information, but also on visual cues from the lips, face, and other bodily gestures of the speaker. Furthermore, note that in such situations listeners must coordinate their visual and auditory attention appropriately, so that information from a common relevant source gets selected together across the different senses (e.g. both the sights and sounds produced by just the relevant person speaking). Such coordination poses a considerable computational challenge, because the stimulus properties signalling a common source across the modalities (e.g. the various cues to location in audition

*Author for correspondence.

against vision) differ so greatly at the initial stages of sensory processing (e.g. vision is retinotopic, whereas audition is initially tonotopic and then head-centred).

In emphasizing the multimodal nature of everyday situations, by contrast with the unimodal nature of most previous attention experiments, we seek to highlight the point that many interesting and important questions are excluded when only unimodal situations are studied. These questions include some very fundamental issues about the psychological architecture of selective attention, such as: (i) the extent to which attention operates independently within each sensory modality in a strictly encapsulated modular fashion, as opposed to in a supramodal manner across all modalities; (ii) the nature of the spatial coordinates in which attention operates, and how these may depend on cross-modal integration; and (iii) the mechanisms by which attention is coordinated across modalities, allowing selection of a common source despite very different inputs to each sensory modality.

Finally, we note that many of the neural structures that have been particularly implicated in spatial attention (e.g. regions in the parietal and frontal lobes; and subcortical structures such as the superior colliculus) are known to be heavily involved in cross-modal interactions, and are thought by many investigators to subserve supramodal representations of space (see, for example, Stein 1992; Stein & Meredith 1993; Anderson *et al.* 1997). The study of cross-modal spatial interactions in such neural structures is a particularly active area in contemporary neuroscience. However, as yet the psychology of spatial attention has made little contact with this literature, beyond frequent assertions that the brain areas listed above play fundamental roles in spatial attention. We expect that the study of cross-modal spatial attention will prove to be a particularly fruitful area for interdisciplinary exchange. We also hope that its study will not only reveal new facts about selective attention, but may also provide a useful empirical handle on the mental representation of space. In this paper, we review recent studies with these aims in mind, concentrating primarily on just our own work with normal individuals, owing to length constraints (see Driver & Mattingley 1995; Driver *et al.* 1997; Mattingley *et al.* 1997; Driver 1998; for discussion of complementary work in brain-damaged patients with deficits of spatial attention, for whom many of the same cross-modal issues arise).

2. CROSS-MODAL LINKS IN COVERT SPATIAL ATTENTION: EXOGENOUS MECHANISMS

We begin with simple experiments that test whether shifts of spatial attention in one sensory modality tend to be accompanied by corresponding shifts in other modalities. Some standard terminology is helpful here. Overt shifts of attention involve the redirection of receptors (as in eye, head or hand movements) towards a region of interest. There is an interesting story to be told about cross-modal links in overt orienting (see, for example, Groh & Sparks 1996a; Jay & Sparks 1990; Yao & Peck 1997), but here we primarily restrict ourselves to considering internal covert mechanisms of spatial attention, when no receptor shifts are allowed. The spatial cueing paradigm (see, for example, Eriksen &

Hoffman 1972; Posner 1978) has been widely used to study covert spatial attention, particularly in vision. In this paradigm, the experimenter attempts to direct the subject's attention to a particular location before the appearance of a single target event, by means of a cue.

Many studies have now found that a variety of visual discriminations can be performed better at the cued location than elsewhere, even though subjects are not allowed to shift their eyes (Spence & Driver 1996). This result is attributed to the cue shifting the spatial distribution of covert attention (sometimes referred to as covert orienting). The stimulus-driven ('exogenous') effects of spatially nonpredictive peripheral cues (i.e. cues whose location does not predict which position is most likely for the target, but which nevertheless sometimes appear at that position) have been distinguished from the strategic ('endogenous') effects of informative symbolic cues, which do predict the likely location of the upcoming target, but in an indirect manner (e.g. a central arrow indicating that participants should shift their attention to one particular side). The facilitatory effects of an exogenous cue on the side of the subsequent target tend to be smaller and more short-lived than the endogenous effects of a spatial expectancy. Further qualitative differences have now been observed between these two forms of cueing, and it is suspected that different neural substrates are involved (Spence & Driver 1994, 1996, 1997). Accordingly, when assessing any cross-modal links in spatial attention, it is important to distinguish links that affect exogenous-reflexive attention from those concerning endogenous-voluntary attention.

We have adapted the standard spatial-cueing paradigm to provide a measure for the distribution of attention in audition and touch, as well as in vision (Spence & Driver 1994, 1996, 1997; Driver 1998; Spence *et al.* 1998a). Our basic method is as follows: subjects are first cued by various means toward the left or right, and are then required to make a speeded elevation discrimination (up as opposed to down) for a subsequent target event, regardless of the side on which it appears (see figure 1 for an example of such a task). We repeatedly find that performance in the elevation discrimination task is better (i.e. faster and/or more accurate) for a target that appears on the cued rather than uncued side. Given that no shift in peripheral receptors is allowed (because we clamp the head and hands, and monitor gaze direction with infrared gaze monitors), this result presumably reflects covert rather than overt orienting mechanisms. Note also that as the lateral position indicated by the cue (i.e. left or right) is entirely orthogonal to the discrimination that must be made (i.e. up against down), the effect of the cue cannot be caused merely by it biasing responses in favour of one decision or another (as might happen if, say, a left-right discrimination was required following a left or right cue; see Kustov & Robinson (1996), and Ward (1994), for examples of this confound). Instead, our cueing effects must reflect a genuine improvement in localization in the cued region (Spence & Driver 1994, 1996, 1997), owing to the cue attracting covert attention there.

Having first demonstrated that this simple up-down task with lateral cueing is sensitive to the distribution of covert spatial attention within each of the modalities of vision, audition and touch, we can then use the task to

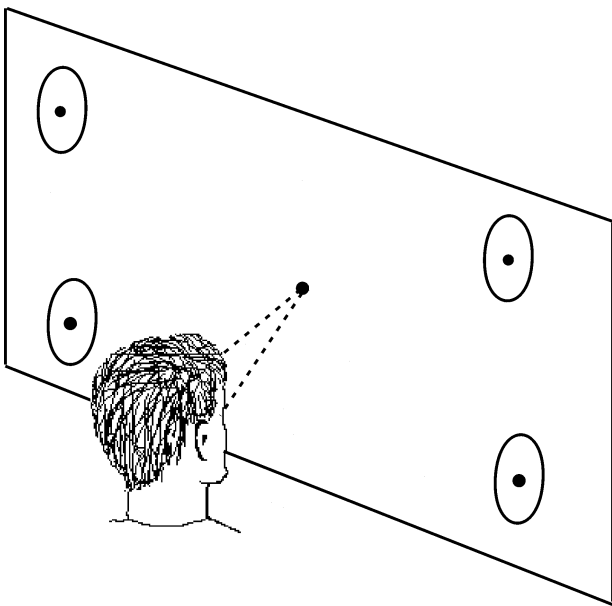


Figure 1. Schematic view of the position of possible target loudspeakers (shown by ellipses) and target lights (shown as black circles), plus central fixation light, in Spence & Driver's (1996, 1997) studies of audiovisual links in covert spatial attention. The subject's head is cartooned, and the direction of steady fixation is indicated with dotted lines. A single target was presented on each trial, in hearing or vision, and the subject made a speeded response discriminating whether this target came from the upper or lower row, regardless of its side, and regardless of which side attention had been cued toward (either by an instruction about which side to expect targets on; or by a spatially nonpredictive peripheral event at an intermediate elevation between the two possible target locations on one side).

examine any cross-modal links in covert attention. Here we first consider this cross-modal issue for the case of exogenous-reflexive attention. Will an irrelevant but salient event in one modality tend to attract covert attention toward it in other modalities? We find that the answer is clearly affirmative for most pairings of modalities (see, for example, Spence & Driver 1997; Spence *et al.* 1998a). A spatially nonpredictive, task-irrelevant abrupt sound on one side leads to better elevation judgements (on average, around 20–30 ms faster, and somewhat more accurate) for both visual and tactile events presented in the vicinity of the sound shortly (100–300 ms) after its onset. Thus, salient auditory events can evidently generate rapid cross-modal shifts of covert spatial attention. Similarly, spatially nonpredictive tactile events on one hand lead to better auditory and visual judgements on that side. So touch evidently generates cross-modal shifts of attention as well (Spence *et al.* 1998a). Finally, non-predictive visual flashes lead to better tactile judgements in their vicinity; yet to our initial surprise, we have repeatedly found (see, for example, Spence & Driver 1997) that peripheral visual cues do not affect auditory judgements (at least, not when eye-movements are prevented; see Rorden & Driver 1998). This finding holds up across numerous variations in the physical properties of the particular visual and auditory stimuli used, and so seems to reflect a general rule about how the modalities relate to each other, rather than being merely caused by

unintended differences in salience between the modalities for the particular stimuli used.

We are currently testing several possible accounts for the finding of no cross-modal exogenous influence from visual cues on auditory targets. Some of the potential explanations involve the neural properties of the colliculus, a subcortical structure that is thought by many investigators (Stein & Meredith 1993; Spence & Driver 1997) to be intimately involved in reflexive shifts of spatial attention, and in the cross-modal interactions which can influence such shifts. For now, we note only that the failure of visual events to attract auditory covert attention rules out the simplest explanation for all the many other cross-modal interactions we have found. It contradicts any proposal that exogenous covert orienting takes place within an entirely supramodal system (Farah *et al.* 1989). If a strictly supramodal system were shifted by sudden events in the periphery, then visual events which are demonstrably capable of attracting visual and tactile attention would necessarily attract auditory attention as well. Yet we find auditory attention to be unaffected by such visual events. Thus, although our experiments reveal extensive cross-modal links in the control of exogenous covert orienting (i.e. links between all possible pairings of cue modality with subsequent target modality, except for visual cues prior to auditory targets), the results also show that the underlying architecture is more complex than a strictly supramodal system.

3. SPATIAL COORDINATES OF EXOGENOUS CROSSMODAL ATTENTION

In all the experiments without saccades that we've described, the receptor systems for the various modalities were aligned in one particular 'default' posture. The subject's head and eyes were fixed straight ahead, with each hand resting on a table in its usual hemisphere (i.e. left hand on the left, right hand on the right). However, in daily life we can adopt many different postures, and the important point is that these spatially realign the receptors from the different modalities. For instance, every time you move your eyes, retinotopic visual inputs are realigned relative to somatotopic tactile space, and to head-centred auditory space. Likewise, each movement of your hand can realign its tactile coordinates relative to vision and audition. Such considerations raise the computational problem of how the nervous system can represent the 'common' position of a particular external source across the different senses. This seems challenging given that the spatial organization of the senses is so different at input levels (vision is retinotopic, touch somatotopic, and audition first tonotopic and then head-centred). Furthermore, the mapping of which particular receptors in one modality (e.g. tactile receptors on a finger) correspond spatially with those in another modality (e.g. particular retinal positions) changes every time a new posture is adopted, because the eye, head and body parts do not move as one.

The means by which the brain derives useful representations of stimulus location across the senses, despite such complexities, has been intensively investigated at the single-cell level by recent physiological studies (see, for example, Graziano & Gross 1993; Anderson *et al.* 1997). Some neural structures (e.g. the superior colliculus) are

known to code stimulation from several sensory modalities in approximate spatial register, such that cells tend to have spatially aligned receptive fields across the modalities (Stein & Meredith 1993). However, most of the initial single-cell recording studies on such cross-modal issues were done in anaesthetized animals, held in a fixed posture (typically with eyes, head and body all facing forward). A retinotopic visual map would of course appear to be in approximate alignment with a head-centred auditory map as long as the eyes faced forward within the head. Only when gaze is diverted would these spatial coordinate systems diverge. More recent single cell studies, often with behaving animals, have looked at spatial coding across modalities as the animals adopt different postures, which realign the receptors for different modalities. Several intriguing physiological results, from various different brain structures (e.g. colliculus, putamen, posterior parietal cortex, premotor cortex), have now shown that sensory responsiveness following stimulation at a particular location in one modality can depend on how the current posture spatially aligns the stimulated receptors with receptors from another modality (see, for example, Jay & Sparks 1984; Graziano & Gross 1993; Groh & Sparks 1996*b*; Anderson *et al.* 1997). For instance, the neural response to a sound or light can depend on how the eye is deviated in the head, and the response to tactile stimulation of a hand can depend on where that hand is located relative to the body or eye. Such findings may reveal how multimodal space is constructed at the level of individual neurons or cell populations. Our recent experiments have posed similar questions, but at the macroscopic functional level of where a person's attention gets directed in cross-modal situations, rather than at the microscopic level of how individual neurons behave.

4. VISUAL-TACTILE REMAPPING

Recall our earlier finding that tactile stimulation of the left hand leads to faster up-down discrimination in the left rather than right visual field shortly afterwards (and vice versa for stimulation of the right hand). Does this pattern depend on some fixed mapping between touch and vision, such that particular somatotopic activations in touch always lower thresholds for particular retinal positions (perhaps owing to which hemisphere the stimuli initially project to; Kinsbourne 1993)? Or are the cross-modal links in exogenous spatial attention more complex in spatial terms, such that visual attention gets drawn to the current location of a stimulated hand, and hence to different positions in the external world (corresponding to different visual receptors) depending on the posture adopted? The latter arrangement would seem more adaptive. If an insect is suddenly felt crawling on your left hand, you presumably need your visual attention to be drawn to the external location of the insect, rather than to some fixed default location in left hemispace (your left hand might be located on the right when the touch is first felt). However, directing attention to the appropriate external location in one modality, based on stimulation in another, is quite demanding computationally, given that receptors for the different modalities can be spatially realigned against each other with every change in posture.

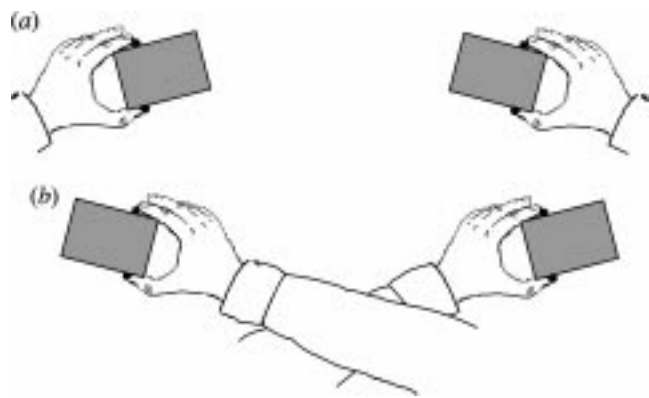


Figure 2. Illustration of relative hand position in (a) the uncrossed and (b) crossed conditions in a visual-tactile or tactile-visual cueing experiment. Subjects gripped a sponge cube in each hand with thumb and index finger. Both cubes incorporated one vibrotactile device at the thumb (next to a small light, illustrated here by a black circle), and a separate vibrator at the index finger (next to another small light). In one series of experiments, visual stimulation (from both the lights near one or other hand) served as a spatially nonpredictive cue, that was task-irrelevant. Each such cue was followed by a target vibration from any one of the four possible vibrators, and subjects made a speeded tactile discrimination of whether this vibration came from an upper (index finger) or lower (thumb) position, regardless of which hand was stimulated. In another series of experiments, the roles of cue and target were reversed; a target light was presented for a speeded up-down discrimination following tactile stimulation of one or other hand.

We have now done a number of studies where receptors in the three modalities of touch, vision and audition are shifted relative to one another, allowing us to determine the coordinates in which cross-modal exogenous attention operates, and thus how the modalities map into each other spatially. Figure 2 illustrates one such study. In figure 2*a*, the hands lie in an uncrossed posture. As described earlier, a visual flash in the left visual field leads to faster tactile discriminations with the left rather than right hand shortly (100–300 ms) after the flash; whereas a right flash gives an advantage to the right hand. What happens if the hands are placed in a crossed posture (see figure 2*b*)? We find (Spence & Driver 1998*c*) that the results now completely reverse in terms of which parts of the eye and body have to be stimulated to produce better performance; although they remain unchanged when considered in terms of external space. In other words, a visual flash on the right still leads to faster tactile discriminations on that side of space (but now with the left hand), whereas a flash on the left now leads to faster tactile discriminations with the right hand (in left hemispace). Evidently, the spatial mapping from particular retinal activations in vision, to somatotopic activations in touch, gets updated when the hands adopt different postures. This is presumably owing to an influence from proprioceptive signals specifying the current hand positions (as confirmed in a later study). Thus, a third modality (here proprioception) can apparently influence the attentional interactions between two other modalities (here vision and touch).

In situations with the reverse roles for cue and target (i.e. with a spatially nonpredictive tactile cue preceding a visual target for elevation discrimination), we have examined the spatial precision of the cross-modal mapping more closely. In particular, we have tested (Driver 1998; S. Kennett and J. Driver, unpublished data) whether cross-modal attentional effects can apply to specific locations within one hemifield, rather than merely affecting the cued hemisphere as a whole. In one such study, visual targets could be presented at any one of four possible eccentricities (outer left, inner left, inner right, outer right) for a speeded up-down elevation discrimination (there were thus eight possible visual target locations; four eccentricities, and two elevations at each). The uncrossed hands (i.e. left hand in left hemisphere, right hand in right hemisphere) could either be placed in alignment with the inner lights on each side, or with the outer lights. Stimulation of one hand led to faster visual elevation discriminations on that side at the visual eccentricity corresponding to the current hand position (i.e. cueing effects were bigger for the inner lights when the stimulated hand lay at the inner eccentricity on one side; but for the outer lights when the hands were placed further out).

This result once again demonstrates that the spatial mapping between modalities gets updated when different postures are adopted. Tactile stimulation can draw visual attention to the current location of the stimulated hand within a hemifield. The senses thus remain in useful register, with respect to each other and the outside world, despite changes in posture. However, to do so they must change the spatial mapping of which receptors in one modality (e.g. tactile receptors on the left hand) get linked to information from particular receptors in the other modalities (e.g. regions in the left or right visual field) to drive the attention shift. In the case of spatial cueing effects between touch and vision, this remapping presumably depends on proprioceptive (and/or visual) information about the current posture of the hands. The exact basis for the remapping can be tested. For instance, by occluding any view of the hands, or doing the study in total darkness, whether or not proprioception plays a critical role can be determined. In two studies we have confirmed that proprioception can indeed contribute to the spatial mapping between vision and touch that determines cross-modal exogenous cueing effects (Driver 1998).

5. WHEN REMAPPING FAILS

Thus far, even the reflexive aspects of cross-modal attention (i.e. shifts of attention that can be triggered within 100 ms of presenting a salient, but spatially nonpredictive, task-irrelevant event) have proved to be remarkably sophisticated, with the spatial mapping between the senses being sufficiently flexible to take into account how the current posture misaligns receptors across the modalities. However, a further covert orienting phenomenon reveals a more primitive level of the system. The studies described here all concerned the facilitatory effects found in discrimination tasks for targets presented near the location of a nonpredictive cue, shortly after its appearance. These effects typically emerge very rapidly but are short-lived (e.g. being found for targets presented

from around 50 ms after cue onset, but disappearing for targets presented 300 ms or more later; Spence & Driver 1994, 1997). It is well-known that an additional phenomenon often emerges at longer delays following a spatially nonpredictive cue, especially in speeded detection tasks. Although detection reaction times are typically faster for targets presented ipsilaterally rather than contralaterally to an immediately preceding cue, this pattern usually reverses at longer cue-target delays, so that targets on the cued side come to yield the slowest responses at delays of 300 ms or more. This robust finding has become known as 'inhibition of return' (IOR), reflecting Posner & Cohen's (1984) hypothesis that covert attention is first drawn to the cued position (thus producing the initial benefit there), and then moves on with a bias against returning to the location that has recently been attended (producing IOR). Most studies on IOR have been concerned only with vision, but we and others have recently found that a similar phenomenon can be found cross-modally (see, for example, Reuter-Lorenz *et al.* 1996; Tassinari & Campara 1996; Spence & Driver 1998*a,b*). For instance, visual detection can become slower at the location of a preceding auditory or tactile event than elsewhere, provided sufficient time has elapsed for the initial facilitatory effect in its vicinity to dissipate.

We have now begun to study the spatial coordinates in which cross-modal IOR arises. Surprisingly, our initial results suggest that tactile-visual IOR does not behave in the same manner as the earlier facilitatory cueing effect, when different postures are adopted (e.g. crossed compared with uncrossed hands, see figure 2). In other words, a tactile stimulation of the left hand can still produce IOR for a subsequent visual target in the left visual field, even when the hands are crossed, with the left hand lying in the right visual field. Unlike the initial facilitatory effects of cross-modal attention found near the cue shortly after its presentation, the later IOR phenomenon apparently does not remap spatially as a function of posture, thus revealing a more primitive level of the reflexive covert orienting system.

This primitive level, with no remapping, may also be revealed by eye-movements toward tactile events. Groh & Sparks (1996*a*) studied saccades towards tactile stimulation on an unseen hand. When the hands were crossed behind an occluding screen, saccades were accurate in their final landing position, but initially shot in the wrong direction (i.e. toward the uncrossed hemisphere, where the crossed hand would usually lie; e.g. rightwards for the right hand even when it was located on the left). This initially erroneous saccadic response may relate to our finding that cross-modal IOR likewise applies to the hemisphere where the stimulated hand would usually lie, rather than to its current position when crossed. Some relation seems plausible, given that several authors have previously proposed that IOR may be an after-effect of an initial saccadic program arising rapidly and reflexively in the superior colliculus (Rizzolatti *et al.* 1974; Wurtz *et al.* 1980; Spence & Driver 1998*a,b*). Note that while we did not allow any overt saccades to be executed in our own cueing studies of cross-modal IOR, this does not preclude some influence from internal saccade programmes generated automatically by peripheral events.

6. AUDIOVISUAL REMAPPING

Changing the position of the hands is by no means the only way to study whether spatial remapping between the senses can influence exogenous cross-modal attention, across changes in posture. For instance, similar issues can be addressed simply by deviating gaze, while holding the head and body in a constant position. In recent audiovisual studies (Driver 1998; Driver & Spence 1998) we have tested whether spatially nonpredictive auditory cues can attract covert visual attention to specific locations within a hemifield; and also how this depends on gaze posture. Visual targets could be presented at any one of four possible eccentricities (outer left, inner left, inner right, outer right) for a speeded up–down elevation discrimination (there were thus eight possible light positions; see figure 3*a*). A spatially nonpredictive sound cue could precede each visual target from any one of the four eccentricities. When the subject faced and fixated straight ahead, the auditory cue drew facilitatory covert visual attention to its specific eccentricity within the cued hemifield, rather than merely to that entire hemifield compared with the other (e.g. a sound on the far left was a better cue for outer left visual targets than for inner left visual targets; whereas a sound on the near left was a better cue for inner visual targets on that side). We next tested a situation where the head of the subject still remained fixed in a forward-facing position, but with his or her gaze now directed off to one side rather than straight ahead (e.g. the subject might now fixate toward the left, between the previous inner left and outer left locations; see figure 3*b*). We shifted all the possible locations for the visual targets laterally along with the deviated gaze, so that the retinal stimulation remained as before; but kept the possible sound locations where they had been, so that the auditory stimulation also remained as before with respect to the craniotopic midline (as the head did not move).

Note that owing to the deviation in gaze, the correspondence of particular retinal locations with particular sounds in external space was changed (e.g. visual targets that were inner right on the retina now corresponded in external space to an inner left sound, when the subject fixated just past the location of the inner left sound-source; see figure 3*b*). If the cross-modal attention effects from sound cues on visual targets that we had previously found with eyes fixating straight ahead were caused by a fixed ear–retina mapping, then particular sounds should still benefit the same retinal positions as previously. If instead the relation between the modalities gets spatially remapped when gaze is deviated, then a sound from a particular location should benefit visual targets that are close to it in external space. Hence the sound would benefit different visual targets, in retinal terms, depending on how the eye is deviated. The results clearly supported the latter remapping hypothesis. For instance, with eyes straight ahead (figure 3*a*), an inner right sound benefited an inner right retinal position more than an inner left sound did. In contrast, when subjects held fixation to the left of the inner left sound (figure 3*b*), the inner left sound now became more beneficial for the same inner right retinal position. Sounds drew visual attention in the ‘correct’ direction with respect to external space even when the eyes were

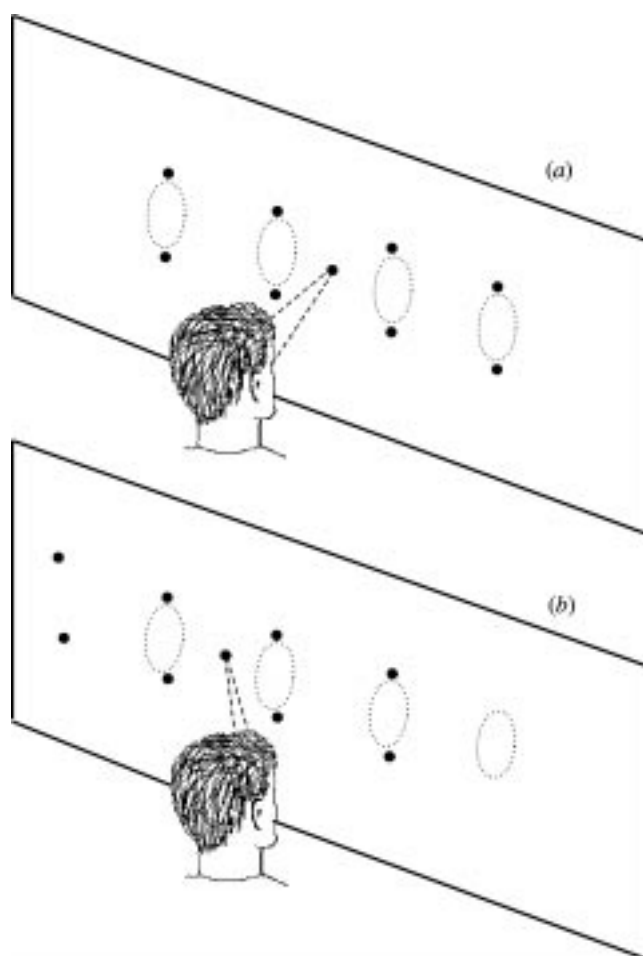


Figure 3. Illustration of possible cue loudspeakers and possible target lights in an audiovisual cueing experiment (the loudspeakers are each marked by dashed lines, to indicate that they could not be seen in the darkened chamber). The subject's head is cartooned, and the direction of steady fixation shown with dotted lines. There were four possible positions for the spatially nonpredictive cue sound (outer left, inner left, inner right, outer right). The target light could appear from any one of eight positions (four eccentricities, with two elevations at each). The task was speeded discrimination of the elevation (up compared with down) of each target light, regardless of its side, and regardless of where the immediately preceding sound had been. In (*a*) the subject fixates directly ahead, between the aligned loudspeakers and possible light positions. Visual discriminations were best at the same eccentricity and side as the immediately preceding sound. In (*b*) the subject fixates between the outer- and inner-left loudspeaker (and note that all visual events have been laterally translated along with gaze). Visual discriminations were again best for lights at the same external location as the immediately preceding sound, but these now occupied different retinal locations as compared with *a*. See text for full explanation.

deviated in the head. This entails that the mapping between auditory locations and retinal locations, which directed exogenous cross-modal attention, must have changed, to keep vision and audition in register as regards external space. These psychological findings on audiovisual attention effects in people can be related to recent physiological findings on single cell activity in the parietal lobe and superior colliculus of monkeys,

indicating that cells there change their coding of auditory location as a function of how the eyes are deviated in the head (Jay & Sparks 1984, 1990; Stricanne *et al.* 1996; Anderson *et al.* 1997).

To summarize the story so far, our studies with spatially nonpredictive peripheral cues have revealed many cross-modal links in the effects of exogenous covert attention. Indeed, such links were found for all pairings of modalities tested, except for auditory targets following visual cues (without saccades). The results also reveal that even these highly reflexive aspects of covert attention, triggered by salient but task-irrelevant events, are quite sophisticated in terms of the underlying spatial mappings between modalities. With the apparent exception of the mechanisms underlying cross-modal IOR, these mappings all change as a function of how the current posture realigns sensory receptors across the various modalities. Several of these findings indicate that the cross-modal interactions between two sensory modalities can be modulated by a third modality (e.g. proprioception), as when current hand position or eye position changes mappings between touch and vision, or between audition and vision.

7. CROSSMODAL LINKS IN ENDOGENOUS MECHANISMS OF COVERT SPATIAL ATTENTION

We now consider possible cross-modal links for the case of endogenous covert attention, which is directed in a voluntary manner on the basis of current spatial expectancies, rather than reflexively following a salient but task-irrelevant event (as in the preceding exogenous studies). Several pioneering authors (see, for example, Buchtel & Butter 1988; Butter *et al.* 1989) tested whether spatially informative peripheral cues in one modality, appearing at the probable location of a subsequent target in another modality, could influence performance in that second modality. Although positive findings from such studies are now common, we believe their method leads to an interpretative problem, and so may reveal little or nothing about the nature of any hard-wired links between the modalities. Informative cues by definition predict the likely location for events in the target modality. Hence any shift of attention may take place solely within only the latter modality, following interpretation of the informative cue. If so, the exact form of a predictive cue (be it visual, auditory, tactile, linguistic and so on) may scarcely matter, and so its modality in relation to the target modality becomes irrelevant to the underlying mechanisms.

Given these considerations, we have taken a different methodological approach when testing for any cross-modal links in endogenous spatial attention. Our general strategy has been to induce a strong spatial expectancy within only one sensory modality, and then test whether this has any influence on other modalities. For instance, Spence & Driver (1996) ran audiovisual experiments, with our usual elevation discrimination task (i.e. discriminating up compared with down, regardless of target side). In separate experiments, either audition or vision served as the more common primary modality (e.g. eight out of 11 targets in the pseudorandom sequence of trials would appear in this modality). Targets in another modality (termed secondary) were much rarer (e.g. only three out of 11 targets might appear in it). The subjects were given a strong spatial

expectancy about the likely side for targets in the primary modality (e.g. seven out of eight targets in that common modality would appear on one specified side throughout a particular block of 96 trials). In contrast, they had no reason to expect targets on that side in the intermingled secondary modality (in fact two out of three of targets in the rarer secondary modality appeared on the other side, to ensure that there was no strategic motivation to shift spatial attention to the same side across the two modalities).

The results revealed a substantial advantage in the efficiency of up-down discriminations for targets in the primary modality, on the side that was strongly expected for that modality. This is presumably owing to subjects directing their endogenous covert attention to the expected side within the primary modality (the spatial effect of such endogenous attention was typically two to three times larger than the exogenous effects described earlier). The more important result was that an advantage was also found on that same side for targets in the secondary modality, even though they were actually twice as likely on the other side! However, this spatial effect in the secondary modality was smaller (by about 50%) than that for the primary modality. This pattern of results held whether the primary modality was audition, with vision secondary, or vice versa. These findings suggest that when people have a strong spatial expectancy about the likely target side in only vision, or only audition, then their attention tends to shift in the same direction for the other modality as well, even when there is no strategic motivation for this to happen (indeed, even when the odds are somewhat stacked against it).

These findings demonstrate cross-modal links between audition and vision for covert endogenous attention. Note, however, that the results do not accord with a strictly supramodal system, contrary to Farah *et al.* (1989) and others; nor with averaging of probabilities across modalities by the subject, as the attentional effect was always larger in the primary modality (for which the strong spatial expectancy held) than in the secondary modality. Our findings therefore suggest that a strongly biased spatial distribution of endogenous attention in one modality tends to spread into other modalities as well, but at a reduced level. Recent evidence from audiovisual studies measuring neural activity via ERPs provides further support for this proposal (Hillyard *et al.* 1984; Eimer & Schröger 1998). Furthermore, we have recently extended the result to situations involving vision and touch (Spence *et al.* 1998*b*). We are currently testing various different accounts for how a spatial disposition might spread from the primary modality into the secondary modality, in attenuated form. For instance, on one account (Kinsbourne 1993) this may be owing to the relative activation of the two cortical hemispheres when attention is endogenously directed towards one hemifield in a particular modality.

Returning to our existing audiovisual study (Spence & Driver 1996), we next asked whether there were any circumstances in which endogenous covert attention could be simultaneously directed in opposite directions for audition and vision. Auditory and visual targets were now made equally common across blocks of 96 trials, but the subjects were told which side was more likely (by a ratio of 4:1) for targets within each modality. In common-side

blocks, the two modalities were both more likely on the same side. In the opposite-side blocks that are of particular interest here, visual targets were more likely on one side, but auditory targets on the other. In the latter blocks, we found that auditory performance was better on the likely side for auditory targets, whereas visual performance was better on the opposite side, which was more likely for visual targets. Measures of the variability in performance suggested that this was owing to a true ‘splitting’ of spatial attention between the two modalities, rather than to any strategy of committing all attentional resources to only a single modality on each trial. Thus, endogenous covert attention can apparently be directed in opposite directions concurrently for audition and vision. On the other hand, the advantage on the expected side for a particular modality in these opposite-side blocks was significantly smaller (reduced by over 60%) than that found for the common-side blocks, suggesting that covert endogenous attention is more effectively focused on a common location across the modalities, even though it can be spatially split between the modalities to some extent. In an earlier study (Driver & Spence 1994), we had reached very similar conclusions from an entirely different method, namely continuous dual-task performance in audition and vision.

Finally, we have recently compared opposite-side and common-side blocks in the manner described here for elevation discriminations in vision and touch (Spence *et al.* 1998*b*). We find that it is difficult to direct endogenous attention toward one side in touch, while simultaneously attending to the other side in vision (e.g. to the right hand in right hemisphere for touch, but to the left hemifield for vision). In contrast, directing visual and tactile endogenous attention to a common side does produce substantial advantages in elevation discrimination on that side for both modalities. When an uncrossed hand posture is compared with a crossed hand posture (see figure 2), the results reveal that it is easier to direct endogenous tactile and visual attention to a common external location, regardless of which particular visual and tactile receptors are involved. Thus, spatial remapping across different postures is evident in the cross-modal links for covert endogenous attention, as we had previously found for cross-modal exogenous attention.

8. CAN CROSS-MODAL INTEGRATION ARISE PREATTENTIVELY?

Thus far, our general approach has been to start by posing the simple question of whether a change in the spatial distribution of attention in one modality (be this an exogenous or endogenous change) tends to be accompanied by corresponding shifts in other modalities. Having found evidence for many such cross-modal links in spatial attention, we have then gone on to ask more detailed questions about the spatial mappings between the various sensory modalities that underly such cross-modal shifts. Specifically, we have investigated whether these mappings may change as a function of how the current posture realigns sensory receptors in the various modalities (e.g. when the eyes are deviated relative to the head, or the hands placed in crossed hemispaces, and so on). On this spatial mapping issue, we have already encountered several situations where the effect of one

modality upon another (e.g. of tactile or auditory cues upon response to visual targets) itself depends on inputs from a third modality (e.g. proprioceptive signals specifying the current location of unseen hands, or the current deviation of the eye in the head). Such results imply that cross-modal integration between the senses can itself contribute to cross-modal attentional interactions. For instance, proprioceptive information can evidently modify a tactile signal, so that the latter can indicate the true external location of a tactile event (rather than merely its somatotopic position on the body), and thus can direct visual and auditory attention to the appropriate external source of the tactile stimulation, whatever the current posture.

This general finding—that cross-modal integration can serve to direct cross-modal attention in an adaptive manner, by allowing the appropriate coding of space despite receptor misalignments—might be taken to suggest that cross-modal integration itself must precede attentional selection. The extent to which different types of sensory information can be integrated before attentional selection has, of course, attracted considerable research interest in purely visual attention research, under the framework of Treisman’s influential feature integration theory (see Treisman, this issue). It may seem natural to ask whether the integration of sensory information from separate modalities requires focused attention to a particular location, in the same manner as was originally proposed by Treisman and her colleagues (see, for example, Treisman & Gelade 1980) for the integration of separate dimensions within the visual modality (e.g. colour, motion, shape and so on).

We suspect that it would be naive to assume that integration between the senses is directly analogous to the integration of different properties in one sense. Furthermore, we believe that it would be highly maladaptive for the nervous system to be incapable of integrating information from separate modalities without first focusing attention on a particular location (as previously claimed for the integration of separate visual features). This would seem especially problematic when one considers how the suggested ‘location’ that attention should be focused on could possibly be coded in a common manner across the senses, without some cross-modal integration already having taken place. In our view, extensive cross-modal integration is first required to construct a suitable representational space in which attention might be effectively directed. From this perspective, cross-modal integration would often tend to precede attentional selection, rather than the reverse arrangement.

However, the experiments described so far do not provide very direct evidence on this issue. At first glance, some of our cases of cross-modal exogenous orienting might seem to entail a degree of cross-modal integration prior to attentional selection. For instance, in several situations we found that the effect of a cue in one modality on a target in a second modality depended in turn on an influence from a third modality concerning the cue’s location. This seems to entail that at least two of the modalities were integrated regarding the cue information, prior to the eventual influence of that cue on the subsequent target. One example of this is when proprioception influences the coding of a tactile signal, and thus changes which location visual

attention gets drawn to by that tactile event. However, one might plausibly argue that the cue itself was attended in such cases (after all, it was always a salient and abrupt event, designed by an experimenter to attract attention). If so, cueing results of this kind need not entail that any cross-modal integration arose in a strictly preattentive manner.

In several recent studies, we have adopted a somewhat different experimental approach than the usual cueing paradigms, in an effort to tackle the issue of preattentive cross-modal integration more directly. This series of experiments differs in that a strictly irrelevant stream of distractor events is now always presented concurrently with the stream of relevant target events. We measure the extent to which the distractors can be ignored; and in particular whether the efficiency of ignoring distractors is ever influenced by factors that depend on cross-modal integration. Our idea is that if the ease of selecting targets from their distractors can be shown to depend on cross-modal integration, then this implies that some cross-modal integration must have taken place before attentional selection was completed. We shall illustrate this approach with two examples.

9. TONIC TACTILE-PROPRIOCEPTIVE INTEGRATION CAN INFLUENCE TACTILE SELECTION

Driver & Grossenbacher (1996) presented subjects with concurrent vibrotactile stimulation to each hand. The task was to discriminate the vibrations (two short pulses, versus one longer continuous pulse) presented to only one of the two hands, which was specified as the relevant hand throughout a block of trials. The vibrations on the other hand were entirely irrelevant, and should be ignored. We varied whether concurrent target and distractor vibrations were the same or different. In this way we could assess the efficiency with which vibrotactile distractors were ignored, by measuring the extent to which they impaired judgements of the concurrent target when different from it.

The critical manipulation was the relative position of the two hands in external space. The hands were either placed close together in one hemispace (i.e. both being held on the far right of the body, or both on the far left) or they were widely separated (i.e. the left hand on the far left of the body, the right hand on the far right). Note that this manipulation of hand separation does not in any way alter the vibrotactile stimulation that was applied to the finger pads (indeed, control conditions confirmed that the different postures did not affect the discriminability of the vibrations delivered to each hand). Nevertheless, hand separation had a dramatic effect on the efficiency of tactile selective attention. Incongruent distractor vibrations on the irrelevant hand (as compared with congruent distractors) substantially impaired judgements of target vibrations on the relevant hand when the hands were held close together. However, this distractor interference was substantially reduced when the hands were placed far apart, even though the tactile stimulation remained just the same.

We take these results to show that tactile endogenous spatial attention does not operate on a purely somatotopic representation of space. The somatotopic position of relevant and irrelevant vibrations remained identical across the hand postures, so the results cannot be explained in these terms. Instead, endogenous tactile attention seems to operate on a representation of space that captures the current layout of

tactile receptors in external space; this representation changes as different postures are adopted, such as the various hand separations. Driver & Grossenbacher (1996) found their effects even for subjects wearing blindfolds, implying that somatotopic space gets recalibrated by proprioceptive inputs about current hand separation to produce the observed effect on tactile selection.

The effect of hand separation may seem counter-intuitive when one considers that only the vibrotactile stimulation was relevant to the required task, and that this stimulation was equivalent across the hand separation conditions. However, it makes better sense when one considers the usual function of touch in daily life. Touch does not serve merely to indicate the current stimulation on patches of skin, but instead uses this stimulation to derive felt properties of the external world. Tactile stimulation can only provide an indication of such external properties if it is integrated with proprioceptive and kinaesthetic information about the current spatial disposition of the tactile receptors as they move while exploring the world. Presumably, Driver & Grossenbacher's (1996) effect of hand separation on vibrotactile attention arises because tactile sensations are routinely recalibrated by proprioceptive inputs. In the case of their experiment, these inputs would indicate that the target and distractors vibrations were very far apart in the external world when the hands were widely separated, but were very close (so that they perhaps seemed more likely to stem from a common external source) when the hands were placed together.

From a theoretical perspective, the importance of this experiment is in demonstrating that some degree of cross-modal integration must have taken place (in this case, between proprioception and touch) before attentional selection was complete within one modality (in this case, within touch). How else could proprioceptive information about hand separation have influenced the ease with which tactile targets were attended and tactile distractors rejected? In this sense then, the study appears to demonstrate that some cross-modal integration can arise 'preattentively'. However, it should be noted that the separation of the hands was blocked in Driver & Grossenbacher's (1996) study, which means that fairly simple cross-modal interactions might be sufficient to explain the result. In particular, only a tonic influence from proprioception on touch (as opposed to phasic influences from kinaesthesia) would be required to explain the result. Proprioceptive inputs would only have to specify whether the hands were close together or far apart for an entire block of trials, and so this information did not need to be time-locked or integrated in any more specific fashion with particular tactile events. In other words, because the hands were held fixed, there were no transient kinaesthetic events to be integrated with concurrent tactile events stemming from a common external source (e.g. the shape of a felt object). This limitation of the study could be addressed in future experiments by applying similar techniques to dynamically moving hands, as they explore a real or virtual object in a haptic manner.

10. DYNAMIC AUDIOVISUAL INTEGRATION CAN INFLUENCE AUDITORY SELECTION

The final study that we shall describe provides an initial demonstration that even phasic cross-modal integration,

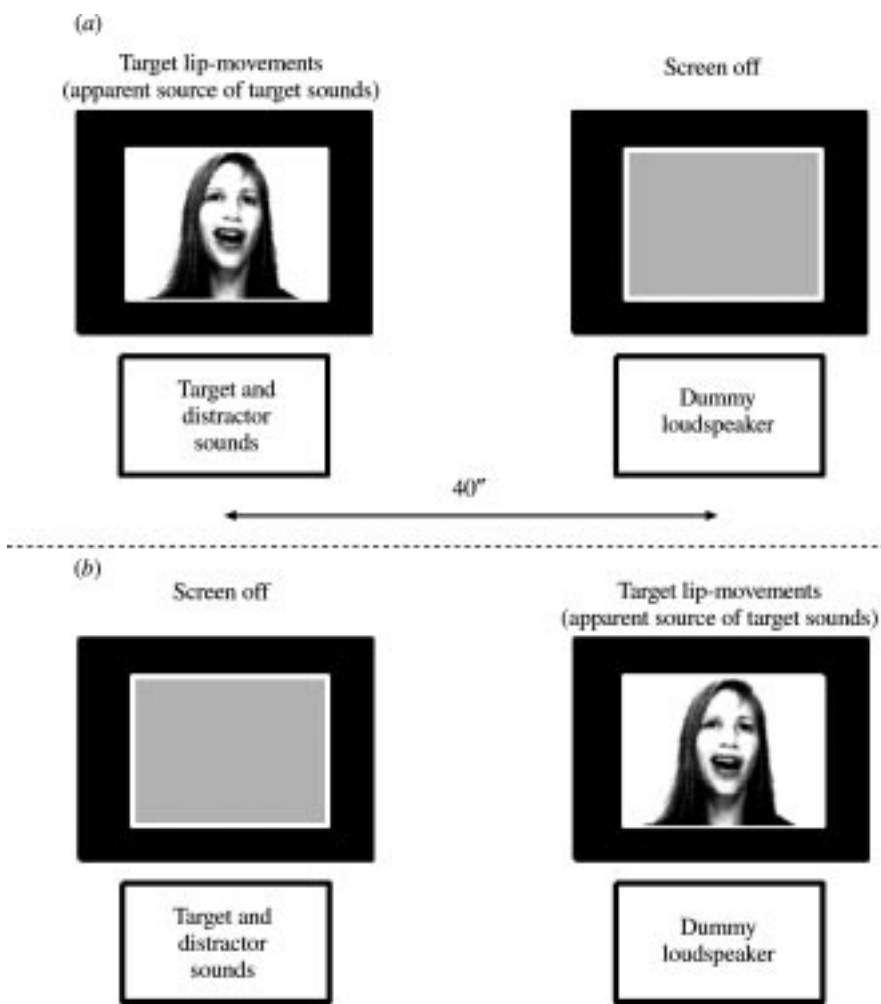


Figure 4. Schematic illustration of the apparatus used by Driver (1996). Relevant and irrelevant words were played from a single loudspeaker (shown on the left in the illustration, but equally likely to be on the right), and spoken in the same voice. A video monitor on one side (left in *a*, right in *b*) showed lip movements which matched the relevant auditory words that had to be repeated by the subject. Performance was better when the relevant lip movements were displaced away from the true monosound source (as in *b*) because this produced an illusory separation between relevant and irrelevant sounds owing to ventriloquism.

between concurrent dynamic events in separate modalities, can arise in an apparently 'preattentive' manner. This evidence stems from an audiovisual rather than a haptic situation. Driver (1996) presented subjects with two concurrent auditory messages. Each comprised three two-syllable words in random order (e.g. one message might comprise 'SUNSET-TULIP-HEADLINE' while the concurrent message might be 'MUSIC-FLOORING-PIGMENT'). As in the classic shadowing studies of auditory attention (Cherry 1953; Broadbent 1958; Treisman 1964), the task was to repeat one message (the target triplet of words) while ignoring the other concurrent message (the distractor triplet). Unlike standard shadowing situations, no auditory information specified which message was the target and which the distractor, because both messages were spoken by the same voice, on recordings played in synchrony from a single monosound source. Furthermore, both messages were random sequences of words. Which message was relevant was specified only visually. A screen displayed a continuous video of the face of the person who had recorded the two messages, showing her speaking the relevant words in perfect synchrony with the sounds she produced for that relevant message. The task was thus to repeat the words spoken on the video, with lip- and tongue-movements being available visually for the relevant message, but not for the irrelevant message.

It had previously been shown that adding visual lip-read information for just the relevant message can

improve selective shadowing (Reisberg 1978). This would be expected, because even people with normal audition rely to a considerable extent on lip-read information during speech perception, as it can provide further phonological information about the heard speech-sounds (McGurk & McDonald 1976; Massaro 1987). The novel twist to Driver's (1996) study was to manipulate where the visual lip-read information was situated relative to the monosound source that presented the relevant and irrelevant sounds (see figure 4).

It is well-known that people tend to mislocate sounds toward their apparent visual source; the so-called 'ventriloquist' illusion (for a review, see Bertelson 1998). For instance, film audiences invariably hear a spoken soundtrack as emanating from actors appearing on the screen, even though the true sound source (i.e. the loudspeakers presenting the soundtrack) is located elsewhere. Driver (1996) aimed to determine the possible implications of this well-known illusion for selective attention. He varied the position of visual lip-read information for the relevant message, either presenting this on the same side as the true monosound source for both auditory messages (directly above the active loudspeaker, see figure 4*a*), or displaced to the opposite side (directly above a 'dummy' loudspeaker which, unknown to the subjects, did not present any sounds; see figure 4*b*).

Driver (1996) found better selective shadowing of the relevant message in the latter condition, when the visual

lip-movements were displaced away from the true monosound source, rather than being presented at the same location as the sounds. He attributed this remarkable effect on objective performance to the subjective ventriloquist illusion. In the same-side condition, ventriloquism should scarcely affect performance, as the lips were presented very close to all of the speech sounds (figure 4a). However, in the displaced condition (figure 4b) any ventriloquism should pull the target sounds away from the distractor sounds, toward the side of the dummy loudspeaker where the lips appeared. Only the target sounds matched these lip movements, and so only they should migrate towards the lips (as ventriloquism arises only for matching visual and auditory events; Bertelson 1998). Hence the distractor sounds would be effectively 'left behind' at the true sound source, as ventriloquism pulled the relevant sounds toward the displaced lips.

Thus, any ventriloquism should lead to an illusory spatial separation between the target and sounds in the displaced condition (figure 4b). The 17% benefit in performance that Driver observed for this displaced condition, relative to the same-side condition, implies that selective shadowing can be objectively improved by an illusory spatial separation between target and distractor sounds, just as previously found for real spatial separations (see, for example, Cherry 1953; Broadbent 1958). In this respect, the result is reminiscent of the well-known 'binaural masking difference' found in auditory psychophysical studies. When a monaural target signal is made to sound as though it comes from a different position to a masking signal in the same ear, by presenting an additional mask to the other ear, then target detection improves (Moore 1982). Driver's (1996) results likewise suggest that auditory attention can be usefully focused on the phenomenal rather than the physical location of a target sound. However, the crucial difference is that, in his case, the illusory position of the target sound depended on a cross-modal influence from vision.

The implication of Driver's (1996) results is therefore that some cross-modal integration can take place between audition and vision before auditory spatial selection is fully completed. The beneficial effect of ventriloquism on selective shadowing performance in the displaced condition (see figure 4b) could only arise if just the target sounds were illusorily pulled toward the matching moving lips, leaving the distractor sounds behind, to produce the apparent spatial separation in audition that proved so useful for selective listening. This effective separation of the target and distractor sounds could only happen after the system had already worked out to some extent which sounds matched the synchronous lip movements (and as a result, should migrate toward them), and which sounds did not (and so should be left behind). As the illusory separation evidently benefited auditory selection, some of this cross-modal matching must have taken place before auditory selection was complete. In this sense, then, the results imply that cross-modal integration (i.e. the appropriate on-line matching of some speech sounds but not others with synchronous lip movements) can arise 'preattentively'. Endogenous auditory attention does not operate solely on auditory information, in a strictly modular

encapsulated fashion. Instead, attention operates on auditory information which has already been integrated to some extent with phasic visual information (note that only the dynamic correspondence between lip movements and concurrent speech-sounds indicated which auditory message matched the lips, and so should migrate towards them).

Driver's (1996) study provides an unusual case of an illusion (namely, ventriloquism) actually aiding veridical perception (i.e. identification of the target words), which might at first appear to be a peculiar state of affairs. However, it is important to realize that, in real-world settings, ventriloquism should usually improve veridical perception of auditory locations, rather than producing illusory mislocalizations. Speech sounds usually do emanate from the moving lips which match them, and vision tends to provide more accurate location information than audition, especially in noisy situations. Hence, in many cases it will be adaptive to weight visual evidence for the location of a sound source more strongly than any competing auditory evidence. This will only lead to illusory mislocations in unusual cases, such as in cinemas or during experiments, where the true sound source is artificially displaced from its apparent visual source.

Thus, by means of visual ventriloquism, the listener can compensate for the relatively poor coding of location by audition that arises in noisy environments, such as the textbook 'cocktail party' situation of many competing conversations in a crowded room. However, note that this cross-modal mechanism for enhancing auditory localization, by means of the better spatial resolution available in vision, could not possibly help us to listen selectively in noisy settings (i.e. could not benefit auditory attention) if it arose only after auditory spatial selection was already completed. To aid selective listening, cross-modal ventriloquism would have to arise 'preattentively', precisely as Driver's (1996) study suggests.

We suspect that a similar logic will apply for many of the other forms of cross-modal integration considered in this paper. In order for attention to be adaptively directed with respect to external space, considerable cross-modal integration is first necessary to construct a suitable internal representation of space. Even in apparently unimodal situations, such as those investigated by researchers who only consider visual attention, the representational space in which attention gets directed will probably turn out to be determined by other modalities as well. Diffuse proprioceptive, kinaesthetic and vestibular inputs are already known to influence the mental representation of visual space, and can do so even for the perceived position of a single spot of light (see Driver & Grossenbacher 1996). However, the possible implications of these cross-modal influences for the nature of the 'space' in which visual attention gets directed has scarcely ever been considered (although see Ladavas 1987).

While the study of spatial attention seems well-advanced when single sensory modalities are considered (see the other papers in this issue), it is still in its infancy as regards cross-modal issues. Nevertheless, we hope that the studies described here may persuade the reader that the cross-modal questions we have posed merit further investigation, and that the senses should no longer be approached in strict isolation by attention researchers.

This research was supported by the Medical Research Council (UK), and the Wellcome Trust. We thank S. Kennett, E. Maculoso, F. Pavani and C. Rorden for helpful comments. Illustrations by Kennett Enterprises.

REFERENCES

- Anderson, R. A., Snyder, L. H., Bradley, D. C. & Xing, J. 1997 Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *A. Rev. Neurosci.* **20**, 303–330.
- Bertelson, P. 1998 Starting from the ventriloquist: the perception of multimodal events. In *Advances in psychological science. I. Biological and cognitive aspects* (ed. M. Sabourin, F. I. M. Craik & M. Robert). Hove, UK: Psychological Press. (In the press.)
- Broadbent, D. E. 1958 *Perception and communication*. Elmsford, NJ: Pergamon.
- Buchtel, H. A. & Butter, C. M. 1988 Spatial attention 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.
- Cherry, E. C. 1953 Some experiments upon the recognition of speech with one and two ears. *J. Acoustic. Soc. Am.* **25**, 975–979.
- Corbetta, M., Miezin, F. M., Shulman, G. L. & Petersen, S. E. 1993 A PET study of visuospatial attention. *J. Neurosci.* **13**, 1202–1226.
- Driver, J. 1996 Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature* **381**, 66–68.
- Driver, J. 1998 Cross-modal links in spatial attention: the second EPS prize lecture. *Q. J. Exp. Psychol.* (In the press.)
- Driver, J. & Grossenbacher, P. G. 1996 Multimodal spatial constraints on tactile selective attention. In *Attention and performance. XVI. Information integration in perception and communication* (ed. T. Inui & J. L. McClelland), pp. 209–235. Cambridge, MA: MIT Press.
- Driver, J. & Mattingley, J. B. 1995 Selective attention in humans: normality and pathology. *Curr. Opin. Neurobiol.* **5**, 191–197.
- Driver, J. & Spence, C. J. 1994 Spatial synergies between auditory and visual attention. In *Attention and performance: conscious and nonconscious information processing*, vol. 15 (ed. C. Umiltà & M. Moscovitch), pp. 311–331. Cambridge, MA: MIT Press.
- Driver, J. & Spence, C. 1998 *Pöppel's paradox partially resolved in people*. (Submitted.)
- Driver, J., Mattingley, J. B., Rorden, C. & Davis, G. 1997 Extinction as a paradigm measure of attentional bias and restricted capacity following brain injury. In *Parietal lobe contributions to orientation in 3D space* (ed. P. Thier & H.-O. Karnath), pp. 401–429. Berlin: Springer.
- Eimer, M. & Schröger, E. 1998 ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology*. (In the press.)
- Eriksen, B. A. & Hoffman, J. E. 1972 Some characteristics of selective attention in visual perception determined by vocal reaction time. *Percept. Psychophys.* **11**, 169–171.
- 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.
- Graziano, M. S. A. & Gross, C. G. 1993 A bimodal map of space: somatosensory receptive fields in the macaque putamen with corresponding visual receptive fields. *Exp. Brain Res.* **97**, 96–109.
- Groh, J. M. & Sparks, D. L. 1996a Saccades to somatosensory targets. 1. Behavioral characteristics. *J. Neurophysiol.* **75**, 412–427.
- Groh, J. M. & Sparks, D. L. 1996b Saccades to somatosensory targets. 3. Eye-position dependent somatosensory activity in primate superior colliculus. *J. Neurophysiol.* **75**, 439–453.
- Hillyard, S. A., Simpson, G. V., Woods, D. L., Van Voorhis, S. & Munte, T. F. 1984 Event-related brain potentials and selective attention to different modalities. In *Cortical integration* (ed. F. Reinoso-Suarez & C. Ajmone-Marson), pp. 395–414. New York: Raven Press.
- Hillyard, S. A., Mangun, G. R., Woldorff, M. G. & Luck, S. J. 1995 Neural systems mediating selective attention. In *The cognitive neurosciences* (ed. M. S. Gazzaniga), pp. 665–681. Cambridge, MA: MIT Press.
- Jay, M. F. & Sparks, D. L. 1984 Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature* **309**, 345–347.
- Jay, M. F. & Sparks, D. L. 1990 Localization of auditory and visual targets for the initiation of saccadic eye movements. In *Comparative perception. I. Basic mechanisms* (ed. M. A. Berkley & W. C. Stebbins), pp. 351–374. New York: Wiley.
- Kinsbourne, M. 1993 Orientational bias model of unilateral neglect: evidence from attentional gradients within hemisphere. In *Unilateral neglect: clinical and experimental studies* (ed. I. H. Robertson & J. C. Marshall), pp. 63–86. Hillsdale, NJ: Lawrence Erlbaum.
- Kustov, A. A. & Robinson, D. L. 1996 Shared neural control of attentional shifts and eye movements. *Nature* **384**, 74–77.
- Ladavas, E. 1987 Is the hemispatial deficit produced by right parietal lobe damage associated with retinal or gravitational coordinates? *Brain* **110**, 167–180.
- McGurk, H. & MacDonald, J. 1976 Hearing lips and seeing voices. *Nature* **264**, 746–748.
- Massaro, D. M. 1987 *Speech perception by ear and eye*. Hillsdale, NJ: Lawrence Erlbaum.
- Mattingley, J. B., Driver, J., Beschin, N. & Robertson, I. H. 1997 Attentional competition between modalities: extinction between touch and vision after right hemisphere damage. *Neuropsychologia* **35**, 867–880.
- Moore, B. 1982 *Introduction to the psychology of hearing*. London: Academic.
- Moran, J. & Desimone, R. 1985 Selective attention gates visual processing in the extrastriate cortex. *Science* **229**, 782–784.
- Neisser, U. & Becklen, R. 1975 Selective looking: attending to visually specified events. *Cogn. Psychol.* **7**, 480–494.
- Pashler, H. E. 1998 *The psychology of attention*. Cambridge, MA: MIT Press.
- Posner, M. I. 1978 *Chronometric explorations of mind*. Hillsdale, NJ: Lawrence Erlbaum.
- Posner, M. I. & Cohen, Y. 1984 Components of visual orienting. In *Attention and performance: control of language processes*, vol. 10 (ed. H. Bouma & D. G. Bouwhuis), pp. 531–556. Hillsdale, NJ: Lawrence Erlbaum.
- Reisberg, D. 1978 Looking where you listen: visual cues and auditory attention. *Acta Psychol.* **42**, 331–341
- Reuter-Lorenz, P. A., Jha, A. P. & Rosenquist, J. N. 1996 What is inhibited in inhibition of return? *J. Exp. Psychol. Hum. Percept. Perf.* **22**, 367–378.
- Rizzolatti, G., Camarda, R., Grupp, L. A. & Pisa, M. 1974 Inhibitory effect of remote visual stimuli on visual responses of cat superior colliculus: spatial and temporal factors. *J. Neurophysiol.* **37**, 1262–1275.
- Rorden, C. & Driver, J. 1998 Does saccade preparation affect auditory attention? *Neuropsychologia*. (In the press.)
- Spence, C. J. & Driver, J. 1994 Covert spatial orienting in audition: exogenous and endogenous mechanisms facilitate sound localization. *J. Exp. Psychol. Hum. Percept. Perf.* **20**, 555–574.
- Spence, C. & Driver, J. 1996 Audiovisual links in endogenous covert spatial attention. *J. Exp. Psychol. Hum. Percept. Perf.* **22**, 1005–1030.

- Spence, C. & Driver, J. 1997 Audiovisual links in exogenous covert spatial orienting. *Percept. Psychophys.* **59**, 1–22.
- Spence, C. & Driver, J. 1998a Auditory and audiovisual inhibition of return. *Percept. Psychophys.* **60**, 125–139.
- Spence, C. & Driver, J. 1998b Inhibition of return following an auditory cue: the role of central reorienting events. *Exp. Brain Res.* **118**, 352–360.
- Spence, C. & Driver, J. 1998c Integrated visuo-tactile representation of space revealed by crossmodal attention effects in normal humans. (Submitted.)
- Spence, C., Nicholls, M. E. R., Gillespie, N. & Driver, J. 1998a Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision. *Percept. Psychophys.* (In the press.)
- Spence, C., Pavani, F. & Driver, J. 1998b Crossmodal links in spatial attention between vision and touch. (In preparation.)
- Stein, B. E. & Meredith, M. A. 1993 *The merging of the senses*. Cambridge, MA: MIT Press.
- Stein, J. F. 1992 The representation of egocentric space in the posterior parietal cortex. *Behav. Brain Sci.* **15**, 691–700.
- Stricanne, B., Anderson, R. A. & Mazzoni, P. 1996 Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *J. Neurophysiol.* **76**, 2071–2076.
- Tassinari, G. & Campara, D. 1996 Consequences of covert orienting to non-informative stimuli of different modalities: a unitary mechanism? *Neuropsychologia* **34**, 235–245.
- Treisman, A. 1964 The effects of irrelevant material on the efficiency of selective listening. *Am. J. Psychol.* **77**, 533–546.
- Treisman, A. M. & Gelade, G. 1980 A feature-integration theory of attention. *Cogn. Psychol.* **12**, 97–136.
- Ward, L. M. 1994 Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Can. J. Exp. Psychol.* **48**, 242–259.
- Wurtz, R. H., Goldberg, M. E. & Robinson, D. L. 1982 Brain mechanisms of visual attention. *Sci. Am.* **246**, 100–107.
- Wurtz, R. H., Richmond, B. J. & Judge, S. J. 1980 Vision during saccadic eye movements. III. Visual interactions in monkey superior colliculus. *J. Neurophysiol.* **43**, 1168–1181.
- Yao, L. & Peck, C. K. 1997 Saccadic eye movements to visual and auditory targets. *Exp. Brain Res.* **115**, 25–34.

BIOLOGICAL
SCIENCES



THE ROYAL
SOCIETY

PHILOSOPHICAL
TRANSACTIONS
OF

BIOLOGICAL
SCIENCES



THE ROYAL
SOCIETY

PHILOSOPHICAL
TRANSACTIONS
OF