THE HANDBOOK OF
MULTISENSORY PROCESSES

Edited by: Gemma A. Calvert
Charles Spence
Barry E. Stein

A BRADFORD BOOK
THE MIT PRESS
CAMBRIDGE, MASSACHUSETTS
LONDON, ENGLAND
CONTENTS

Introduction xi

1 PERCEPTUAL CONSEQUENCES OF MULTIPLE SENSORY SYSTEMS

1 The Cross-Modal Consequences of the Exogenous Spatial Orienting of Attention  CHARLES SPENCE AND JOHN MCDONALD  3

2 Modulations of Visual Perception by Sound  LADAN SHAMS,
YURISU KAMITANI, AND SHINSUKE SHIMOJO  27

3 Cross-Modal Interactions Evidenced by the Ventriloquism Effect in Humans and Monkeys  TIMOTHY M. WOODS AND
GREGG H. RECANZONI  35

4 Multisensory Integration of Dynamic Information  SALVADOR SOTO-FARACO AND ALAN KINGSTONE  49

5 Sweet and Sour Smells: Learned Synesthesia Between the Senses of Taste and Smell  RICHARD J. STEVENSON AND
ROBERT A. BOAKES  69

6 Cross-Modal Interactions in Speeded Classification  LAWRENCE E. MARKS  85

7 Multisensory Texture Perception  SUSAN J. LEDERMAN AND
ROBERTA L. KLATZKY  107

8 Cross-Modal Object Recognition  FIONA N. NEWELL  123
II IS SPEECH A SPECIAL CASE OF MULTISENSORY INTEGRATION?

10 From Multisensory Integration to Talking Heads and Language Learning DOMINIC W. MASSARO 153

11 Spatial and Temporal Constraints on Audiovisual Speech Perception KEVIN G. MUNHALL AND ERIC VATIKOTTIS-BATESON 177

12 Speech as a Supramodal or Amodal Phenomenon CAROL A. FOWLER 189

13 Audiovisual Speech Binding: Convergence or Association? LYNNE E. BERNESTEIN, EDWARD T. AUER, JR., AND JEAN K. MOORE 203

14 Multisensory Animal Communication SARAH R. PARTAN 225

III THE NEURAL MECHANISMS UNDERLYING THE INTEGRATION OF CROSS-MODAL CUES

15 Multisensory Integration in Single Neurons of the Midbrain BARRY E. STEIN, WAN JIANG, AND TERENCE R. STANFORD 243

16 Analysis and Modeling of Multisensory Enhancement in the Deep Superior Colliculus THOMAS J. ANASTASIO AND PAUL E. PATTON 265

17 The Resurrection of Multisensory Cortex in Primates: Connection Patterns That Integrate Modalities JON H. KAAS AND CHRISTINE E. COLLINS 285

18 Multisensory Convergence in Early Cortical Processing CHARLES E. SCHROEDER AND JOHN J. FOXE 295

19 Multisensory Neuronal Convergence of Taste, Somatosensory, Visual, Olfactory, and Auditory Inputs EDMUND T. ROLLS 311

20 Cross-Modal Memory in Primates: The Neural Basis of Learning About the Multisensory Properties of Objects and Events AMANDA PARKER AND ALEXANDER EASTON 333

21 Corticocortical Connectivity of Cross-Modal Circuits M. ALEX MEREDITH 343
IV MULTISENSORY MECHANISMS IN ORIENTATION

23 Auditory-Visual Interactions Subserving Primate Gaze Orienting
A. J. VAN OPISTAL AND D. P. MUNOZ 373

24 Modeling the Time Course of Multisensory Interaction in
Manual and Saccadic Responses  ADELE DIEDERICH AND
HANS COLONIUS 395

25 Multisensory Influences on Orientation and Movement Control
JAMES R. LACKNER AND PAUL DIZIO 409

26 Action as a Binding Key to Multisensory Integration
LEONARDO FOGASSI AND VITTORIO GALLES 425

27 Multisensory Neurons for the Control of Defensive Movements
MICHAEL S. A. GRAZIANO, CHARLES G. GROSS, CHARLOTTE S. R. TAYLOR, AND
TINN MOORE 443

28 Cortical Mechanisms of Tool Use Subserved by Multisensory Integration
HIDETOSHI ISHIKAWA, SHIGERU OBAWASHI, AND ATSUSHI IRIKI 453

29 Multisensory Representations of Space in the Posterior Parietal Cortex
YALE E. COHEN AND RICHARD A. ANDERSEN 463

V HUMAN BRAIN STUDIES OF MULTISENSORY PROCESSES

30 Hemodynamic Studies of Audiovisual Interactions
GEMMA A. CALVERT AND JAMES W. LEWIS 483

31 Multiple Electrophysiological Mechanisms of Audiovisual Integration in
Human Perception  ALEXANDRA FORT AND MARIE-HELENE GIARD 503

32 MEG Studies of Cross-Modal Integration and Plasticity
TOMMI RAJJ AND VEIKKO JOUSMAKI 515

33 Functional Imaging Evidence for Multisensory Spatial Representations
and Cross-Modal Attentional Interactions in the Human Brain
EMILIANO MACALUSO AND JON DRIVER 529

34 Electrophysiological Studies of Multisensory Attention
MARTIN EIMER 549

35 Neuroimaging Studies of Cross-Modal Integration for Emotion
JOHN O’DOHERTY, EDMUND T. ROLES, AND MORTEN KRINGSBACH 563
VI THE MATURATION AND PLASTICITY OF MULTISENSORY PROCESSES

37 Epigenetic Factors That Align Visual and Auditory Maps in the Ferret Midbrain
ANDREW J. KING, TIMOTHY P. DOUBELL, AND IRINI SKALIORA 599

38 Visual Instruction of the Auditory Space Map in the Midbrain
YORAM GUTFREUND AND ERIC I. KNUDSEN 613

39 The Development of Multisensory Integration
MARK T. WALLACE 625

40 Perceptual Development and the Origins of Multisensory Responsiveness
ROBERT LICKLITER AND LORRAINE E. BAHRICK 643

41 The Value of Multisensory Redundancy in the Development of Intersensory Perception
DAVID J. LEWKOWICZ AND KIMBERLY S. KRAEBEL 655

VII CROSS-MODAL PLASTICITY

42 Rewiring Cortex: Cross-Modal Plasticity and Its Implications for Cortical Development and Function
MRIGANKA SUR 681

43 Cross-Modal Consequences of Visual Deprivation in Animals
JOSEF P. RAUSCHEKER 695

44 Visual Cortical Involvement in Normal Tactile Perception
K. SATHIAN, S. C. PRATHER, AND M. ZHANG 703

45 Visual Cortex Engagement in Tactile Function in the Presence of Blindness
ESTEBAN A. FRIDMAN, PABLO CELNIK, AND LEONARDO G. COHEN 711

46 Compensatory Plasticity as a Consequence of Sensory Loss
BRIGITTE RÖDER AND FRANK RÖSLER 719

47 Audiovisual Speech Perception in Deaf Adults and Children Following Cochlear Implantation
TONYA R. BERGESON AND DAVID B. PISONI 749

48 Neuroimaging Studies of Cross-Modal Plasticity and Language Processing in Deaf People
RUTH CAMPBELL AND MAIREÁD MACSWEENY 773
VIII  PERSPECTIVES DERIVED FROM CLINICAL STUDIES

49  Multisensory Integration: Resolving Ambiguities for Human Postural Control  LESLIE ALLISON AND JOHN J. Jeka  785

50  Neuropsychological Evidence of Integrated Multisensory Representation of Space in Humans  ELISABETTA LADAVAS AND ALESSANDRO FARNE  799

51  Cross-Modal Integration and Spatial Attention in Relation to Tool Use and Mirror Use: Representing and Extending Multisensory Space Near the Hand  ANGELO MARAVITA AND JON DRIVER  819

52  Grapheme-Color Synesthesia: When 7 Is Yellow and D Is Blue  MIKE J. DIXON, DANIEL SMILEK, BRANDON WAGAR, AND PHILIP M. MERKLE  837

53  Behavioral and Brain Correlates of Multisensory Experience in Synesthesia  JASON B. MATTINGLEY AND ANNA N. RICH  851

54  Synesthesia, Cross-Activation, and the Foundations of Neuroepistemology  V. S. RAMACHANDRAN, E. M. HUBBARD, AND P. A. BUTCHER  867

Contributors  885
Index  889
The Cross-Modal Consequences of the Exogenous Spatial Orienting of Attention

CHARLES SPENCE AND JOHN McDONALD

Introduction

Our attention is often captured by the sudden and unpredictable sensory events that frequent the environments in which we live. For instance, we will normally turn our heads if someone suddenly calls our name at a crowded cocktail party. Similarly, if a mosquito lands on our arm, our eyes will be drawn immediately to the source of the unexpected tactile event. In these and many other such situations, objects that are initially processed in one sensory modality "grab" our attention in such a way as to enhance the sensory processing of stimuli presented in other modalities at the same spatial location. The cross-modal consequences of the involuntary orienting of our spatial attention is an area that has been extensively researched in recent years and is the subject of this review.

Research has demonstrated that the reflexive overt orienting of our attention conveys immediate cross-modal benefits: Not only do we see visual events more accurately at the fovea than in the periphery of our visual fields, but perhaps more surprisingly, we also hear and feel more acutely if we look—or even if we simply prepare to look—in the direction of nonvisual sensory stimulation (e.g., Driver & Grossenbacher, 1996; Gopher, 1973; Honoré, 1982; Honoré, Bourdeau'Hui, & Sparrow, 1989; Kato & Kashino, 2001; Rorden & Driver, 1999). While many researchers have focused their attention on the nature and consequences of these cross-modal shifts in specifically overt attention (i.e., involving shifts of the eyes, head, or body to better inspect an event of interest; e.g., Amlôt, Walker, Driver, & Spence, 2003; Jay & Sparks, 1990; Perrott, Saberi, Brown, & Srybel, 1990; Whittington, Hepp-Reymond, & Flood, 1981; Zanbarbieri, Beltram, & Versino, 1995; Zanbarbieri, Schmid, Prablanc, & Magenes, 1981; Zanbarbieri, Schmid, Magenes, & Prablanc, 1982), others have investigated the consequences of the covert shifts of attention that may occur prior to, or in the absence of, any overt orienting.

Covert shifts of attention take place very rapidly following the presentation of a peripheral sensory event, occurring prior to any shift of the sensory receptors themselves. Under normal circumstances, there is a close coupling between overt and covert orienting (e.g., Jonides, 1981a; Posner, 1978), with covert shifts of attention typically preceding any overt orienting response (e.g., Klein, Kingstone, & Pontefract, 1992; Rafael, Henik, & Smith, 1991; Rizzolatti, Riggio, Dascola, & Umiltà, 1987; Shepherd, Findlay, & Hockey, 1986), and both being controlled, at least in part, by the same neural structures (such as the superior colliculus; see Desimone, Wessinger, Thomas, & Schneider, 1992; Groh & Sparks, 1996a, 1996b; Robinson & Kertzman, 1995; Stein & Meredith, 1993; Stein, Wallace, & Meredith, 1995; Thompson & Masterton, 1978).

Psychologists have known for many years that the presentation of a spatially nonpredictive visual stimulus, or cue, can lead to a rapid but short-lasting facilitation of responses to visual targets subsequently presented at the cued location (or elsewhere on the cued side), even in the absence of any overt orienting toward the cue itself (e.g., Jonides, 1981b; Posner & Cohen, 1984). Similar intramodal cuing effects following the presentation of a nonpredictive cue have also been reported in subsequent years between auditory cue and target stimuli (e.g., McDonald & Ward, 1999; Spence & Driver, 1994), and more recently, between tactile cue and target stimuli as well (Spence & McGlone, 2001).

In most situations, these cuing effects appear to reflect the consequences of a transient shift of attention to the cued location rather than a passive sensory effect per se (see Posner & Cohen, 1984; but see also Tassinari, Aglioti, Chelazzi, Peru, & Berlucchi, 1994, for a sensory explanation of some facilitatory cuing effects reported in earlier visual cuing studies). Because the cues were spatially nonpredictive with regard to the likely target location in these early studies, researchers concluded that spatial attention can be
oriented involuntarily, and at least somewhat automatically, to the location of a cuing event. In line with the majority of previous research on this topic, we will refer to this as exogenous attentional orienting (as compared to the endogenous orienting that occurs following the presentation of a spatially predictive peripheral, or central symbolic, cue; e.g., see Driver & Spence, 2004; Klein & Shore, 2000; Spence & Driver, 1994; Wright & Ward, 1994).

Having provided evidence that exogenous shifts of attention to visual, auditory, and tactile cues can facilitate responses to targets presented subsequently in the same modality, the obvious question arises as to whether such shifts of covert attention can also facilitate responses to targets presented in sensory modalities other than that of the cue. Would, for example, an exogenous shift of attention to a sound on the left facilitate responses to subsequent visual targets appearing on that side? Any such cross-modal cuing effect might reflect the existence of a supramodal attentional mechanism, a finding that would have important implications at both the theoretical and applied levels (see, e.g., Spence, 2001). The evidence suggesting that the presentation of auditory or tactile cues can trigger an overt shift of visual attention in the cued direction provides at least prima facie evidence that this might be the case. However, as is often the case in the field of experimental psychology, proving (to the satisfaction of all) what is intuitively obvious to the majority of people has taken rather longer than one might have expected! Part of the problem with research in this area has often been the adoption of inappropriate experimental designs that have either used response measures that are relatively insensitive to the manipulation of attention or else do not satisfactorily rule out nonattentional explanations (such as simple detection latencies; see below). Meanwhile, other studies have incorporated experimental setups that failed to maximize the possibility of detecting any cuing effect present because the cue and target stimuli on ipsilaterally cued trials were presented from different spatial locations.1

A further problem has been an overreliance on particular experimental paradigms with relatively few attempts to understand why different groups of researchers have found different patterns of cross-modal cuing effects in their different experimental paradigms (see Ward, Prime, & McDonald, 2002, on this point). Over the past decade, contradictory and often seemingly incompatible findings have emerged from the laboratories of Lawrence Ward, John McDonald, and their colleagues, on the one hand, and Charles Spence, Jon Driver, and their colleagues, on the other (see, e.g., Spence & Driver, 1997a; Ward, 1994; Ward, McDonald, & Golestani, 1998; Ward, McDonald, & Lin, 2000). Although these differences were originally attributed to methodological problems with the particular studies involved, recent work has confirmed the validity and robustness of each pattern of results, at least within the particular experimental paradigms in which they were tested. Fortunately, as the various different research groups have argued over the "true" nature of the cross-modal links in covert spatial attention that exist between the different sensory modalities, a number of important methodological and theoretical advances have emerged from the debate in this area. What is more, and as we hope to show in this chapter, there is now convincing empirical evidence that the covert orienting of exogenous attention that is triggered by the presentation of auditory, visual, or tactile cue stimuli can facilitate the perception of target stimuli presented subsequently at the cued location, no matter what their sensory modality. In fact, cross-modal cuing effects have now been demonstrated behaviorally between all possible combinations of auditory, visual, and tactile cue and target stimuli under a subset of experimental testing conditions.

The way is now open, therefore, for researchers to start investigating a number of theoretically more interesting issues in this area, among them the following: (1) the precise nature of the relationship between mechanisms underlying exogenous shifts of attention to auditory, visual, and tactile stimuli (McDonald & Ward, 2003a); (2) the possible modulation of cross-modal cuing effects by top-down, or endogenous, attentional factors (e.g., McDonald & Ward, 1999; Spence, 2001); (3) the effects of posture change on cross-modal covert attentional orienting (e.g., Driver & Spence, 1998; Kennett, Spence, & Driver, 2002); and (4) the underlying reasons behind the different patterns of cross-modal cuing effects reported in different experimental paradigms (i.e., over and above any simple methodological limitations inherent in particular studies; Prime, McDonald, & Ward, 2003; Ward et al., 2002). Cognitive neuroscientists have also begun to investigate some of the neural underpinnings (and consequences) of the cross-modal orienting of covert exogenous spatial attention (e.g., Kennett, Eimer, Spence, & Driver, 2001; Macaluso, Frith, & Driver, 2000; McDonald, Teder-Sälejärvi, Heraldez, &

---

1 In this chapter, the terms ipsilaterally cued and contralaterally cued are used to refer to trials on which the target was presented on the same versus opposite side as the cue, respectively.
Early studies of the cross-modal consequences of covert spatial attention revealed that the presentation of spatially nonpredictive auditory cues also led to the facilitation of detection latencies for visual targets presented on the side of the cue (e.g., Buchtel & Butter, 1988; Farah et al., 1989; Klein et al., 1987). However, by contrast, the peripheral presentation of visual cues was found to have no spatially specific effect on auditory target detection latencies (e.g., Buchtel & Butter, 1988; Klein et al., 1987). These results provided some of the first empirical evidence for the existence of asymmetric audiovisual attentional cuing effects. In fact, the same asymmetric pattern of cuing effects has now been replicated in a number of more recent cross-modal cuing studies using both the simple detection task (e.g., Reuter-Lorenz & Rosenquist, 1996; Schmitt, Postma, & de Haan, 2000, 2001), as well as other kinds of discrimination tasks (e.g., Schmitt et al., 2000, 2001; Spence & Driver, 1997a; see below).

Subsequent research has, however, revealed a number of potential problems with the use of the simple speeded detection response measure in cross-modal attention research. First, simple auditory detection latencies often appear insensitive to the distribution of spatial attention, even when the shift of attention has been elicited by the presentation of an auditory cue (e.g., Buchtel, Butter, & Ayvasik, 1996; Spence & Driver, 1994). This may be because people can base their detection responses on very "early" tonotopic representations of the auditory targets, in which information about the spatial location of the stimulus is simply not made explicit (cf. McDonald & Ward, 1999; Rhodes, 1987; Spence & Driver, 1994). This contrasts with the earliest sensory representations of visual and somatosensory stimuli that are inherently spatiotopically organized, one retinotopically and the other somatotopically (though Spence & McGlone, 2001, recently suggested that tactile simple detection latencies may also be insensitive to the spatial distribution of attention).

Perhaps more important, many researchers have questioned whether, even when facilitatory effects are reported in such cross-modal cuing experiments, they actually reflect the beneficial effects of a shift of covert spatial attention on perceptual processing at the cued location. Instead, it has been argued, they may simply reflect the consequences of criterion shifts taking place at the cued (and/or the uncued) location, and/or speed-accuracy trade-offs. Participants might therefore respond more rapidly on ipsilaterally cued trials simply because less target-related information is required to respond (i.e., their criterion for responding to targets at the cued location might be lower than for targets at the uncued location; e.g., see Spence & Driver,
auditory targets at the shortest SOAs. This audiovisual asymmetry in cross-modal cuing effects was the opposite of that observed in the simple detection studies reported earlier, and also the opposite of that observed in subsequent studies using Spence and Driver’s (1997a, 1999) orthogonal cuing paradigm (see below).

Ward’s (1994) findings were initially met with some skepticism (e.g., Spence & Driver, 1997a), partly because of their apparent inconsistency with other results, and partly because of a potential methodological problem that affects this kind of spatial discrimination task. While the decision to use a spatial discrimination task is not, in and of itself intrinsically problematic, the majority of studies that have utilized such tasks (including Ward’s study) have resulted in experimental paradigms in which the dimension on which the cue was varied and the dimension on which participants were required to respond overlapped. Consequently, the cuing effects reported in such studies are open to an alternative nonattentional interpretation in terms of response priming, or stimulus-response compatibility effects. Specifically, the presentation of the lateralized cue may simply have primed the response associated with the side of the cue (i.e., a left response may have been primed by the presentation of a cue on the left, and a right response by the presentation of the cue on the right; see Simon, 1990, on this issue). Consequently, participants may have responded to ipsilaterally cued targets more rapidly than contralaterally cued targets in such experiments simply because the appropriate response had been primed in the former case and the inappropriate manual response had been primed in the latter case.

9A lowering of a participant’s criterion for responding to targets on the cued side would be expected to lead to a speeding up of response latencies, with a corresponding increase in errors (see, e.g., Klein et al., 1987, Experiment 5, for one such example of a speed-accuracy trade-off). This pattern of behavioral results can be contrasted with the more commonly held view of a perceptual facilitation due to attentional orienting that would lead to a speeding up of response times, with a concomitant reduction in the number of errors that are made (or, at the very minimum, no increase in errors). The problem for studies incorporating a simple detection response is that no measure of error rates is provided, and hence the criterion-shifting explanation cannot be ruled out (see Duncan, 1980; Müller & Findlay, 1987; Shaw, 1980; Sperling & Dosher, 1980).

4However, the peripheral presentation of a cue typically leads to at least two distinct effects on a participant’s performance. First, there is the spatial cuing effect, which is the focus of the present chapter. Second, there is an equally important and often behaviorally more dramatic alerting effect that is also triggered by the presentation of the cue (e.g., Niemi & Nätänen, 1981; Posner, 1978). Typically, alerting effects facilitate responses to all stimuli, regardless of their location, although this increased facilitation often comes at the cost of an increase in errors. Although Posner (1978) suggested that there may be a cross-modal asymmetry in alerting effects to stimuli of different sensory modalities, Ward (1994) demonstrated a clear cross-modal alerting effect from auditory (and visual) cues on visual (and auditory) targets (e.g., see Ward, 1994, Figs. 2 and 3, pp. 250 and 251). Similar cross-modal alerting effects have also been reported in the majority of subsequent cross-modal cuing studies. They will not be discussed further in the present chapter, in light of the inherently nonspatial nature of such effects.

5This problem of interpretation for spatial cuing studies that used overlapping cue and response dimensions is emphasized by the fact that a very similar experimental paradigm was used by researchers in the early 1970s to investigate a different research question related to stimulus-response compatibility effects. For example, Simon and Craft (1970) reported that participants made speeded left-right discrimination responses to visual targets presented on a screen more rapidly when the presentation of target was accompanied by a spatially nonpredictive auditory cue presented to the same side rather than to the opposite side over headphones. While the auditory and visual stimuli were presented simultaneously in Simon and Craft’s study, similar stimulus-response compatibility effects were also reported by Bernstein and Edelstein (1971) when the auxiliary auditory stimuli were presented up to 45 ms after the onset of the lateralized visual target.
Ward (1994) considered response priming as a possible explanation for his cross-modal cuing results but argued that any such response priming effect would have led to an auditory cuing effect on visual discriminations as well as the visual cuing effect on audition that he found. By contrast, Spence and Driver (1997a) raised the possibility that even Ward’s null result might be attributable to response priming, because the visual cues were presented much closer to fixation than the auditory targets. In particular, they suggested that even on ipsilaterally cued trials, the auditory cues and targets may have primed (or been associated with) opposite response tendencies. For example, an auditory cue on the left may have primed a left response, whereas a visual target on the left may have been coded in terms of a right response (at least initially) because the target was located on the right with respect to the cue (cf. Nicoletti & Umiltà, 1989; Umiltà & Nicoletti, 1985). Spence and Driver argued that the response tendencies generated on these ipsilaterally cued trials might therefore have been the same as on the contralaterally cued trials, thereby potentially canceling each other out and leaving no net cross-modal cuing effect.

Ward and his collaborators, however, have subsequently demonstrated that the cross-modal asymmetry reported in Ward’s (1994) original study still occurs when these potential problems are avoided. For example, Ward et al. (1998) found the same pattern of asymmetric cuing effects in subsequent experiments when the cue and target were presented from exactly the same lateral eccentricity. Moreover, in a different study, Ward et al. (2000) also replicated the asymmetry using an experimental procedure (implicit spatial discrimination; McDonald & Ward, 1999) that ruled out any response priming explanation more directly (see below).

Two important points therefore emerge from this early debate over Ward’s (1994) asymmetric cross-modal cuing results. First, response priming by the cue may be problematic on certain spatial discrimination tasks, and steps should be taken to avoid such problems, for example, by using techniques such as the orthogonal spatial cuing task or the implicit spatial discrimination task discussed below. Second, although a study may be methodologically confounded, this does not necessarily mean that the conclusions based on that study might not still turn out to be correct. Having demonstrated the robustness of the cross-modal cuing asymmetry first reported by Ward (1994) in subsequent studies, it therefore now becomes increasingly important to try to determine precisely why auditory cues have no effect on visual discrimination performance under at least certain experimental conditions.

Ward and his collaborators (Ward et al., 1998, 2000) have argued that auditory-on-visual cuing effects can be modulated by top-down control processes. In Ward’s (1994) original study, the cue modality was unpredictable and the cuing environment complex (five different possible cue types could be presented on any trial). Ward et al. (2000) therefore suggested that participants might not have fully processed the locations of the auditory cues under such circumstances. According to the account of Ward et al., auditory-on-visual cuing effects would be expected to emerge when tested under more simple experimental conditions, such as when the cue and target modalities are fixed and predictable (see also McDonald & Ward, 2003a; Mondor & Amirault, 1998; and Spence, 2001, for a discussion of the role of the complexity of the stimulus environment, both that of the cue and that of the target, on the pattern of cross-modal cuing effects observed). Indeed, Schmitt et al. (2000) demonstrated a cross-modal cuing effect from auditory cues on visual left-right spatial discrimination responses, as well as the effect of visual cues on auditory discrimination responses already documented by Ward, 1994) under conditions where the cue (and target) modalities were fixed throughout each block of experimental trials. Symmetric audiovisual cuing effects have also been reported in the more recently developed implicit spatial discrimination task (McDonald & Ward, 1997; 2003b, 2003c; see below).

Cross-modal cuing studies using the orthogonal spatial cuing paradigm

Ambiguities concerning the appropriate interpretation of these early cross-modal cuing studies led Spence and Driver (1994, 1997a) to develop a modified version of the spatial discrimination task. They eliminated any possibility that the cue could prime the appropriate response in their task by making participants respond on a dimension (or direction) that was orthogonal to that on which they were cued. In the majority of their studies, Spence and Driver presented spatially nonpredictive auditory and visual cues from either the left or the right of a central fixation point on each trial. However, the auditory and/or visual targets were presented from one of four locations situated directly above or below the cue location on either side (Fig. 1.1A provides a schematic outline of the experimental setup used in many of Spence and Driver’s studies). Participants in these experiments were required to make a speeded spatial discrimination response regarding the elevation (upper vs. lower) of the targets, regardless of the side and sensory modality of their presentation. Consequently, the left or right cues could not differentially affect responses on
ipsilaterally versus contralaterally cued trials, since any response bias elicited by a cue should affect responses on both types of trial equally.

In their first cross-modal cuing experiment, Spence and Driver (1997a; Experiment 1) found that the peripheral presentation of a spatially nonpredictive auditory cue led to a short-lasting facilitation of elevation discrimination responses latencies to auditory targets as well as to visual targets (Fig. 1.1B). This cross-modal cuing effect was observed despite the fact that participants were explicitly and repeatedly instructed to ignore the auditory cues as much as possible. What is more, they were given more than 1000 trials on which to try to overcome any tendency they might have had to orient covertly toward the cue. Performance was also somewhat more accurate on ipsilaterally cued trials than on contralaterally cued trials, thus allowing Spence and Driver to rule out a speed-accuracy tradeoff account of their findings.6

6Spence and Driver (1997a, Experiment 1) did not monitor the eye position of the participants in their original study. Therefore, one cannot rule out the possibility that their spatial cuing effect might reflect overt rather than covert
In their more recent research, Driver and Spence (1998) further investigated the spatial specificity of this specific cross-modal cuing effect by presenting auditory cues and visual targets from one of two locations on either side of fixation (one placed 13 degrees from central fixation, the other 39 degrees from fixation) (Fig. 1.2). Spence and Driver found that participants responded more rapidly and accurately to visual targets presented from LEDs directly above (or below) the auditorily cued loudspeaker (at SOAs of 100–150 ms), with performance falling off as the cue-target separation increased, even if the cue and target both fell within the same hemisphere. Thus, the presentation of an auditory cue led to a spatially specific cross-modal facilitation of visual elevation discrimination responses, akin to the spatially specific cuing effects reported previously in intramodal studies of covert orienting within both vision and audition (e.g., Rorden & Driver, 2001; Tassinari et al., 1987) and subsequently replicated in a number of other cross-modal cuing studies (Frassinetti, Bolognini, & Ladavas, 2002; Schmitt et al., 2001). Driver and Spence (1998) have also shown that visual targets by the auditorily cued loudspeaker are discriminated more rapidly and accurately than lights by the other (noncued) loudspeakers when participants deviate their gaze 26 degrees to either side of central fixation throughout a whole block of trials (while keeping their head fixed in a straight-ahead position), thus showing that these cross-modal cuing effects update so that when posture changes, covert exogenous attention is still directed to the correct environmental location.

Having demonstrated a cross-modal cuing effect from auditory cues on visual elevation discrimination responses, Spence and Driver (1997a) went on to investigate whether the peripheral presentation of a spatially nonpredictive visual cue would also lead to a facilitation of responses to ipsilaterally presented auditory targets (see Fig. 1.1C). In fact—and in direct contrast with the results reported by Ward (1994)—Spence and Driver found no such spatially specific effect of visual cues on auditory elevation discrimination responses (see Fig. 1.1D). Importantly, this null effect of visual cues on auditory target discrimination responses was replicated in a number of different experiments using the orthogonal spatial cuing design, hence demonstrating the robustness of this null effect (at least in the orthogonal cuing paradigm). For example, Spence and Driver (1997a, Experiment 6) assessed auditory elevation discrimination performance at a much wider range of SOAs following the onset of the visual cue in order to try and rule out the possibility that the time course of visual cuing effects on auditory target discrimination responses might simply be different from those reported following an auditory cue. However, no effect of visual cues on auditory elevation discrimination responses was found at any of the SOAs tested (in the range of 100–550 ms) when the possible confounding
effect of overt orienting had been ruled out (i.e., by ensuring continuous central fixation by means of an eye position monitor, and throwing out all trials on which participants either moved their eyes or else blinked).

Spence and Driver (1997a) also varied the nature of the visual cuing events that they used in order to try and demonstrate a cross-modal influence of visual cues on auditory discrimination responses. For example, they assessed the consequences of the presentation of visual cues consisting of visual offsets (Experiments 4–6), rather than the onset cues more commonly used in cross-modal cuing studies, and they also assessed the consequences of the presentation of cues consisting of the onset of a highly salient 3 x 3 array of high-luminance yellow LEDs on one side of fixation or the other (Experiments 3, 5, and 6). However, none of these manipulations elicited a significant cross-modal cuing effect. That is, no consistent evidence of a significant cross-modal spatial cuing effect from visual cues on auditory elevation discrimination responses was found in any of the four experiments that were conducted. Spence and Driver and others (e.g., Schmitt et al., 2000; Spence & Driver, 1999, 2000; Vroomen, Bertelson, & de Gelder, 2001) have also replicated the null effect of visual cues on auditory elevation discrimination responses in several further studies, including those that have incorporated more complex cuing environments (Fig. 1.3; cf. Ward et al., 2000). The asymmetric cross-modal cuing effects reported in Spence and Driver’s studies also mirrored the findings reported earlier by researchers using simple detection response measures (e.g., Buchtel & Butter, 1988; Klein et al., 1987; Schmitt et al., 2000, 2001).

Spence and Driver (1997a, 1999) interpreted the null effect of visual cues on auditory target elevation discrimination performance as supporting the existence of asymmetric cross-modal links in exogenous spatial attention between essentially distinct (i.e., separate) auditory and visual exogenous attentional systems (see Schmitt et al., 2000, for a similar theoretical standpoint). Spence and Driver argued that whereas the presentation of a spatially nonpredictive peripheral auditory cue would lead to a shift of both the auditory and visual attentional systems, the presentation of visual cues appeared only to trigger a shift of the visual spatial attention system, not of the auditory attention system.

Cross-modal cuing between audition, touch, and vision

In the years following their original cross-modal cuing study, Spence, Driver, and their colleagues went on to investigate the nature of any cross-modal links in exogenous spatial attention between other combinations of sensory modalities, using variations on their orthogonal spatial cuing design (e.g., Kennett et al., 2001, 2002; Spence, Nicholls, Gillespie, & Driver, 1998). For example, Spence et al. reported a series of experiments demonstrating that the peripheral presentation of a tactile cue to the index finger of either the left or right hand would elicit a cross-modal shift of both visual and auditory attention toward the position of the cued hand in space (Figs. 1.4A and B). Similarly, Spence et al. also reported that the peripheral presentation of either a spatially nonpredictive auditory or visual cue could facilitate a participant’s ability to discriminate continuous from pulsed vibrotactile stimuli presented to the ipsilateral hand (Figs. 1.4C and D). A similar pattern of symmetric cross-modal cuing effects between vision and touch has now been reported in a number of other studies (Chong & Mattingley, 2000; Kennett, 2000; Kennett & Driver, 1999; Kennett et al., 2001, 2002; see also Tan, Gray, Young, & Irawan, 2001), supporting the robustness of the cross-modal cuing effects found between touch and audition or vision.

Kennett et al. (2002) have also demonstrated subsequently that if participants hold their hands in a crossed hands posture (Fig. 1.5), then a vibrotactile cue
The results of a study by Spence et al. (1998) of cross-modal links in exogenous spatial attention between touch, vision, and audition. (A) Diagram showing the cross-modal facilitation of visual elevation responses following the presentation of a spatially nonpredictive tactile cue to either the left or the right hand (Spence et al., 1998; Experiment 3). Ipsilaterally cued trials are indicated by open squares connected by solid lines and contralaterally cued trials by open circles connected by dotted lines. Error rates are shown next to the associated RT value. (B) Tactile cues also facilitated elevation discrimination responses for auditory targets presented ipsilateral (solid squares connected by solid lines) as opposed to contralateral (solid circles connected by dotted lines) to the cue (Spence et al., 1998; Experiment 3). (C) In another experiment, Spence et al. (1998; Experiment 2) demonstrated that continuous versus pulsed discrimination latencies for tactile stimuli presented to one or the other hand were also facilitated by the presentation of a visual cue by the stimulated hand (ipsilateral trials, gray squares connected by solid lines) versus the unstimulated hand (gray circles connected by dotted lines). (D) Finally, Spence et al. (1998; Experiment 1) also demonstrated that continuous versus pulsed tactile discrimination latencies could be facilitated by the presentation of an auditory cue on the ipsilaterally cued side (gray squares connected by solid lines) as opposed to the contralaterally cued side (gray circles connected by dotted lines).

Presented to the left hand will facilitate visual discrimination responses for targets on the participant’s right (i.e., on the opposite side to the facilitation reported in the uncrossed posture, though in the correct external location). Cross-modal cuing effects therefore seem to result in the facilitation of information processing for all stimuli at the correct externally cued location, regardless of their modality and regardless of the person’s posture. These results suggest that covert shifts of attention operate on a multisensory representation of space that is updated as posture changes (see Spence et al., in press, for a fuller discussion of this issue).

Cross-modal cuing studies using the implicit spatial discrimination paradigm

While Spence and Driver were extending their orthogonal cuing paradigm beyond the audiovisual pairing that they had originally studied, John McDonald, Lawrence Ward, and their colleagues were highlighting a possible methodological problem with the paradigm. They noted that the vertical distance between targets in the orthogonal spatial cuing paradigm must be large enough for participants to be able to discriminate the elevation of the targets reliably. Consequently, the cue and target stimuli have often had to be presented from different positions not only on contralaterally cued trials, but on ipsilaterally cued trials as well. This is particularly noticeable for the combination of visual cues and auditory targets, given the difficulty that most humans have in discriminating the elevation from which auditory stimuli are presented. For example, in Spence and Driver’s (1997) experiments, auditory and visual targets were presented from approximately 24 degrees above or below the cued location. Ward et al. (2000, p. 1264; see also Prime et al., 2003; Spence, 2001) speculated that the spatial focusing of attention in response to a highly localizable visual cue might be too narrowly...
distributed to influence the processing of auditory targets in such tasks, whereas the spatial focusing of attention in response to a less localizable auditory cue might be broad enough to influence the processing of visual targets under comparable conditions (Fig. 1.6).7

In light of this possible limitation with the orthogonal spatial cuing paradigm, McDonald and Ward (1999) developed an alternative paradigm, called the implicit spatial discrimination task, in which the spatial component of the task was maintained while ensuring that the cue and target stimuli could be presented from exactly the same locations (hence maximizing the possibility of finding any cuing effects). In the implicit spatial discrimination task, participants are required to respond to targets presented from certain spatial locations (known as Go trials), but not to stimuli at one or more other spatial locations (known as No-Go trials). In the original version of the task, depicted in Figure 1.7, participants were instructed to respond to the onset of targets presented to the left and right of fixation, and to refrain from responding to the onset of targets presented at fixation itself (McDonald & Ward, 1999, 2003a; Ward et al., 2000).

In subsequent versions of the implicit spatial discrimination task, participants have been required to discriminate between peripheral targets on the basis of a modality-specific feature instead (such as the frequency of auditory targets or the color of visual targets) while still refraining from making any response to target stimuli presented at fixation (McDonald et al., 2001; McDonald & Ward, 2003a, 2003b). Nevertheless, the crucial feature of each version of the implicit spatial discrimination task was that the location of target stimuli was still relevant to the participant’s task, even though the participant’s responses themselves were explicitly based on some other nonspatial criterion (such as onset detection, frequency, or color). Thus, just as in the orthogonal spatial cuing paradigm, the implicit spatial discrimination task maintains a spatial component to the participant’s task while avoiding any problems associated with possible response priming by the cue. Later versions of the implicit spatial discrimination task also satisfy the definition of orthogonal cuing because the response dimension is independent of, and orthogonal to, the cuing dimension.

The implicit spatial discrimination task was originally used to investigate the intramodal consequences of the exogenous orienting of covert attention to auditory cues (McDonald & Ward, 1999), but it has since been used to gather both behavioral and electrophysiological data on the cross-modal consequences of exogenous shifts of attention. As mentioned earlier, Ward et al. (2000) used this task in an attempt to replicate Ward’s (1994) original audiovisual asymmetry. Separate groups of participants responded to auditory targets and to visual targets. As in Ward’s original study, the targets in each task were preceded on a trial-by-trial basis by either a visual cue, an auditory cue, both types of cues, or no cue (i.e., both cue modality and cue type were unpredictable). The results of this study replicated Ward’s original results in the crucial cross-modal conditions; namely, the presentation of peripheral visual cues led to a short-lasting facilitation of response latencies to auditory targets, but the presentation of peripheral auditory cues failed to have a significant influence on response latencies to the visual targets (Fig. 1.8). Importantly, eye position was carefully monitored for each participant and stimulus-response compatibility effects were eliminated.
McDonald and Ward (2003a) conducted a further series of reaction-time experiments in which the cue and target modalities were both fixed (unlike their earlier work, which involved unpredictable cue modalities). In their first experiment, a cue appeared from a single modality that was known in advance and a subsequent target appeared from a different sensory modality that was also known in advance. Thus, the conditions were similar to those in most unimodal spatial cuing studies except that the cue occurred in a sensory modality that was completely irrelevant to the participant’s task (see Schmitt et al., 2000, for similar left-right discrimination
experiments under conditions of predictable cue and target modality presentation). Participants were instructed to respond to peripheral targets while trying to refrain from responding to central targets. McDonald and Ward again found that visual cues facilitated response latencies on the implicit spatial discrimination task for auditory targets presented ipsilaterally to the cue, thereby providing additional evidence for the cross-modal cuing effect that has not as yet been found in orthogonal spatial cuing studies (e.g., Schmitt et al., 2000; Spence & Driver, 1997a, 1999; Vroomen et al., 2001). Moreover, McDonald and Ward now found that auditory cues facilitated implicit spatial discrimination latencies to ipsilateral visual targets as well (Fig. 1.9A). Once again, eye position was monitored; thus, the effects appeared to reflect exclusively the covert orienting of spatial attention.

One possible nonattentional explanation for the results obtained by McDonald and Ward is that the cues affected the participant’s criterion for responding to targets differentially on ipsilaterally cued as opposed to contralaterally cued trials. To rule out this possibility, McDonald and Ward (2003a, 2003b) conducted a further experiment in which participants discriminated between red and green visual targets that appeared to the left or right of fixation while withholding responses to visual targets appearing at fixation. A spatially nonpredictive auditory cue was presented prior to the appearance of every target. By requiring the participants to discriminate between different peripheral targets, McDonald and Ward were able to rule out a criterion-shifting explanation of their results by testing explicitly for a speed-accuracy trade-off in their data. Critically, response latencies were faster on ipsilaterally cued trials than on contralaterally cued trials, and there was no evidence of a speed-accuracy trade-off (Fig. 1.9B).

McDonald et al. (2001) used a similar experimental design to investigate the crucial missing link in exogenous cross-modal attention further. In their experiment, participants discriminated between low-frequency and high-frequency auditory targets that were presented to the left or right of fixation while ignoring auditory targets that were presented at fixation. Prior to the appearance of each target, a visual cue was presented randomly to the left or right of fixation (at cue-target SOAs of 100–300 ms). The behavioral results showed a clear cross-modal cuing effect: Responses latencies were significantly shorter on ipsilaterally cued trials than on contralaterally cued trials, and once again, there was no evidence of a speed-accuracy trade-off (Fig. 1.9C). This study therefore provided the first unequivocal behavioral evidence that exogenously orienting attention to a visual cue can modulate the processing of subsequent auditory targets.8

The fact that cross-modal cuing effects have now been demonstrated between all possible combinations

---

8Further support for the notion of Ward et al. (2000) that spatial relevance may enhance the magnitude of any cross-modal cuing effects reported comes from a comparison of the study of McDonald et al. with those reported by Mondor and
of auditory, visual, and tactile cue and target stimuli does not, however, mean that cross-modal cuing effects will occur in every task or under all testing conditions (McDonald & Ward, 2003a; Spence, 2001; Ward et al., 1998, 2000). For, as we have already seen, certain performance measures, such as simple auditory detection latencies, may be insensitive to the orientation of spatial attention. In addition, Ward and McDonald have argued that certain cross-modal cuing effects (specifically, from auditory cues to visual targets) may be harder to demonstrate in more complex cue-target environments. A number of results now also support the suggestion that cross-modal cuing effects appear to be both larger and more robust under conditions where “space” is somewhere relevant to the participant’s task (than in nonspatial tasks; see McDonald & Ward, 1999; Spence, 2001). Space can be made relevant either by requiring participants to make a spatial discrimination response regarding the target (as in the orthogonal spatial cuing task) or by instructing participants only to respond to targets coming from certain spatial locations (as in the implicit spatial discrimination task).

The two spatial discrimination paradigms reviewed thus far both manage to make space relevant to the participant’s task while avoiding stimulus-response compatibility issues such as response priming. Each of these tasks appears to have its own advantages. The implicit spatial discrimination task has the important methodological advantage that the cue and target stimuli can be presented from exactly the same location on ipsilaterally cued trials, hence maximizing any cross-modal cuing effects that might be present. However, as yet this task has only been used to investigate cross-modal links between audition and vision. Meanwhile, the orthogonal spatial cuing paradigm has been used to demonstrate cross-modal cuing effects between all possible combinations of auditory, visual, and tactile cue and target stimuli with the exception of visual cues and auditory targets. Therefore, the orthogonal spatial cuing paradigm may be a more appropriate paradigm for investigating the effects of peripheral cues on visual and tactile target performance (where the elevation separation between upper and lower targets need not be large), but one can run into problems when using this paradigm to investigate the effects of cross-modal attention on auditory discrimination performance, for which elevation separation between targets has to be much greater, owing to the poor elevation discrimination ability in audition in humans (Roffler & Butler, 1968).

Prime, McDonald, and Ward (2003; Ward et al., 2002) have begun to examine the visual-on-auditory cuing effect in closely matched implicit spatial discrimination and orthogonal spatial cuing experiments. In one such study, a visual cue was presented to the left or right of fixation along the horizontal meridian shortly before an auditory target was presented from one of five loudspeaker cones (at an SOA of either 100 ms or 500 ms). A pair of loudspeaker cones was positioned 15 degrees above and below the visual cue display on each side of fixation, and a single loudspeaker cone was positioned directly in front of the observer. In the implicit spatial discrimination condition, participants ignored targets appearing at fixation and pressed the same response button for lateralized targets appearing above and below fixation. In the orthogonal spatial cuing condition, participants ignored targets appearing at fixation and pressed different response buttons for the lateralized targets appearing above and below fixation (i.e., they were making speeded elevation discrimination responses). Response latencies were shown to be shorter for ipsilaterally cued trials than for contralaterally cued trials at the 100 ms SOA, but only for the implicit spatial discrimination condition (although a criterion-shifting explanation of these facilitatory effects cannot be ruled out). Interestingly, response latencies were actually significantly longer for ipsilaterally cued trials than for contralaterally cued trials in the orthogonal spatial cuing condition. The reason for this inhibitory cuing effect is unclear, although a similar effect was reported in another previous study by Ward et al. (1998). Nevertheless, this difference in cuing results depending on the participant’s task highlights the potential importance of processing requirements in the generation of cross-modal cuing effects.

**Psychophysical studies of cross-modal exogenous cuing effects**

The reaction time studies described so far have demonstrated that the exogenous orienting of attention to a stimulus in one sensory modality can facilitate

Amirault (1998, Experiment 1). Mondor and Amirault used an experimental design that was very similar to that of Ward et al., but without the spatial relevance component (i.e., participants had to respond to all targets). Whereas Ward et al. found cross-modal cuing effects in their experiment when space was relevant to the task, Mondor and Amirault did not (except under conditions where the cue was predictive with regard to the likely location of the upcoming target, and so endogenous attentional orienting mechanisms were presumably involved).

Spence (2001) suggested that this unusual result may be related to the use of a highly localizable white noise cue (as opposed to the hard-to-localize pure-tone cue used in the majority of other previous orthogonal spatial cuing studies; see also Fig. 1.6 on this point).
responses to stimuli in other sensory modalities that happen to be presented at the same location a short time afterward. However, it is very difficult to know from such results where exactly in the stream of information processing the facilitatory effect is occurring. In particular, the studies discussed so far do not answer the question of whether cross-modal cuing effects represent a facilitation of the preparation and/or execution of a participant's responses (i.e., without requiring any enhancement of the participant's perceptual experience), or whether they also demonstrate a genuine enhancement in perceptual processing (e.g., Hawkins et al., 1990; Luu, 1986; Watt, 1991).¹⁰

In order to learn more about the processes underlying the facilitatory effects demonstrated in cross-modal cuing studies, researchers have now started to use signal detection theory (see Green & Swets, 1966; Macmillan & Creelman, 1991, for an introduction) to separate the effects of cross-modal cuing on perceptual sensitivity (i.e., $d'$) from their effects on response-related processing (specifically criterion shifts). McDonald, Teder-Sälejärvi, Di Russo, and Hillyard (2000) reported one such study in which participants sat in front of two arrays of LEDs, one on either side of fixation (Fig. 1.10A). On each trial, a bright flash of red light was presented from four red LEDs on either the left or right of fixation. In half of the trials, this was preceded by a brief flash of green light from a single LED in the center of the LED array (the green light flash was absent in the remainder of trials). Participants had to judge whether or not the green target had been presented at the location of the red stimulus.¹¹ A spatially nonpredictive auditory cue was presented from one side or the other.

¹⁰This viewpoint is nicely illustrated by the following quote from Roger Watt's (1991) book, Understanding Vision: "My own opinion is that studies based on reaction times should be treated much as one would regard a region on a map which was marked with the legend 'Centaurus abide here.' The point is that present information is not to be relied on too heavily, but rather as an indication of something of interest and perhaps, adventure" (p. 213).

¹¹In this study, the red flash of light served both as a mask and as a postcue indicating the potential location of the target stimulus. The target could only occur at the postcued location; thus, there was no uncertainty about where the target may have occurred. By contrast, postcues were presented at multiple locations in some of the earlier signal detection studies of intramodal visual attention (e.g., Downing, 1988; Müller & Humphreys, 1991), thereby raising the possibility that the measures of perceptual sensitivity were confounded with difficulties in localizing the target itself (cf. Hawkins et al., 1990; Luck et al., 1994). The single-postcue technique avoids such confounds. Note also that by postcuing a single location, the technique also implicitly establishes some degree of spatial relevance to the task. Essentially, participants are asked whether a target appeared at a specific location rather than at any location. As we have seen already, the establishment of spatial relevance appears to be one important factor in producing reliable auditory and audiovisual cuing effects (cf. McDonald & Ward, 1999; Spence, 2001).

100–300 ms before the interval when the target might be presented. McDonald et al. found that the parameter β was lower for targets on ipsilaterally cued trials than on contralaterally cued trials, thus providing evidence for a spatially specific criterion shift elicited by the auditory cue. More important, however, McDonald et al. also found that both accuracy and perceptual sensitivity, as measured by $d'$, was higher for visual stimul
that appeared at the cued location than for visual stimuli that appeared contralateral to the cue (Fig. 1.10B). This result demonstrates that cross-modal cuing modulates stimulus processing at a perceptual level (see also Frassinetti et al., 2002, for similar results).

Kato and Kashino (2001) have also used signal detection theory to investigate the effects of spatially nonpredictive visual cues on auditory perceptual judgments (i.e., the reverse situation to that examined by McDonald, Teder-Sälejärvi, Di Russo, et al., 2000). Participants in their study were presented with a series of visual cues unpredictably from 30 degrees to either side of central fixation. The participants were instructed to make an eye movement in the direction of the visual cue and to respond to the direction of motion (left-to-right vs. right-to-left) of auditory broadband noise targets presented binaurally over headphones. Perceptual sensitivity (d') was higher for auditory targets presented on the cued side at SOAs that were too short for participants to have executed an eye movement toward the visual cue by the time that the auditory target was presented (i.e., at 0 ms and possibly also 100 ms SOA; see Rorden & Driver, 1999). As in the study by McDonald, Teder-Sälejärvi, Di Russo, et al. (2000) study, Kato and Kashino also reported a significantly higher criterion on ipsilaterally cued trials, which suggested that the cue also elicited a spatially specific shift in the participant's criterion for responding on ipsilaterally cued versus contralaterally cued trials.

Therefore, taken together, these two studies (Kato & Kashino, 2001; McDonald, Teder-Sälejärvi, Di Russo, et al., 2000) demonstrate that the peripheral presentation of either an auditory or a visual cue leads both to a lowering of the criterion for responding at the cued location and to a genuine improvement in perceptual sensitivity at the ipsilaterally cued relative to the contralaterally cued location. Given the benefits of the signal detection theory approach in determining the source of any facilitatory effects, it seems likely that we will see an increasing dependence on such measures of perceptual sensitivity in the coming years. In this regard, it would be interesting to extend the approach to investigate the perceptual consequences of the cross-modal cuing between audition and touch and between vision and touch (cf. Chong & Mattingley, 2000; Kennett et al., 2001, 2002; Spence et al., 1998).

Electrophysiological evidence for exogenous cross-modal spatial attention

A number of research groups have also started to collect information about the consequences of cross-modal attention from measures of electrical brain signals called event-related potentials (ERPs). Not only can this approach help to elucidate some of the brain structures where target-related information processing is being modulated cross-modally by the presentation of the cue, it can also help to demonstrate the genuinely perceptual nature of such cross-modal cuing effects (see Luck et al., 1994). The rationale for this approach and the neural underpinnings of ERPs are described in more detail in Chapter 34, by Martin Eimer, and so will not be reiterated here. Instead, we will simply focus on the results of a number of recent ERP studies that have provided dramatic evidence that the exogenous orienting of attention to a stimulus in one modality can modulate the neural sensory responses to stimuli in another modality. Such cross-modal ERP effects have been observed in reaction time studies with auditory cues and visual targets (McDonald & Ward, 2000), tactile cues and visual targets (Kennett et al., 2001), and visual cues and auditory targets (McDonald, Teder-Sälejärvi, Heraldez, et al., 2001). In each case, orienting attention in one sensory modality appeared to modulate neural activity in brain areas that are considered to be modality-specific for the target modality.

McDonald and Ward (2000) investigated the effect of the involuntary covert orienting of attention to auditory cues on the neural processing of subsequent visual target using the basic implicit spatial discrimination task. Manual responses to visual targets were faster on ipsilaterally cued trials than on contralaterally cued trials at shorter SOAs (100–200 ms) but not at longer SOAs (900–1100 ms). The cuing effects on the ERPs elicited by the visual targets followed a similar time course. Figure 1.11A shows the visually evoked ERPs for both ipsilaterally cued and contralaterally cued trials. At the short SOA, the ERPs were more negative on valid trials than on invalid trials between 200 and 400 ms after the appearance of the visual target. This negative difference extended to the contralateral occipital scalp (Fig. 1.11B), suggesting that the exogenous covert orienting of attention to the auditory cue modulated neural activity in modality-specific visual cortex. The absence of any ERP differences in the time range of the earlier P1 and N1 components indicated that the effect of cross-modal attentional cuing on activity in the visual cortex occurred after the initial sensory processing of the target had been completed. McDonald and Ward speculated that the enhanced occipital negativity might depend on feedback from higher, multisensory areas.

A more recent ERP study investigated the neural basis of cross-modal spatial attention in a task where attentional orienting to a nonpredictive auditory cue was found to improve visual target detectability.
(McDonald, Teder-Sälejärvi, Di Russo, & Hillyard, 2003). Figure 1.12A shows the ERPs to the left of visual field stimuli on both ipsilaterally cued and contralaterally cued trials. As in McDonald and Ward’s (2000) speeded discrimination task, the visually evoked ERPs were more negative on ipsilaterally cued trials than on contralaterally cued trials. This effect began 80–100 ms earlier than the effect reported by McDonald and Ward and consisted of at least two phases. The first phase was initially focused over the parietal scalp between 120 and 140 ms after stimulus onset, shifting to the occipital scalp between 150 and 170 ms after stimulus onset (Fig. 1.12B). The second phase was focused over the frontocentral scalp and extended to the contralateral occipital scalp between 240 and 260 ms after stimulus onset. Inverse dipole modeling suggested that the cerebral sources of the first phase were located in the superior temporal cortex and the fusiform gyrus of the occipital lobe (Fig. 1.12C). The superior temporal cortex has been identified as a site of multisensory convergence and multisensory integration (e.g., Calvert, Campbell, & Brammer, 2000). Interestingly, the superior temporal dipoles appeared to be active before the occipital dipoles were. This spatiotemporal sequence of dipole activity was in accord with McDonald and Ward’s (2000) speculation that cross-modal attention effects on visual cortical activity depend on feedback from higher, multisensory areas (cf. Driver & Spence, 2000). More specifically, an involuntary shift of attention to sound appears to modulate visual-evoked brain activity first in multimodal cortex and subsequently in modality-specific visual cortex.

Cross-modal effects have also been found in an experiment in which attention was cued to the left or right hand by a spatially nonpredictive tactile stimulus prior to the appearance of a visual target presented from near to the left or right hand (Kennett et al., 2001). ERPs were recorded from eight electrodes over the frontal, central, parietal, and lateral occipital scalp. There were no cue effects on the P1 component, but, as in the other studies that we have already discussed, the ERPs to visual targets were more negative on ipsilaterally cued trials than on contralaterally cued trials following the P1. A biphasic negativity was found at central electrodes, with the first phase occurring between 110 and 180 ms after stimulus onset and the second phase occurring between 220 and 300 ms after stimulus onset. An enhanced negativity was also observed at lateral occipital electrodes in the latency range of the N1 component. Thus, the results of Kennett et al. also support the view that exogenously orienting attention to a nonvisual cue modulates visually evoked activity in modality-specific visual cortex.

The most contentious issue in the literature on the cross-modal cuing of attention has been settled by recent
**ERPs to LVF Stimuli**

- Grand-averaged ERPs to left visual field stimuli at electrodes over the occipital scalp (electrodes PO7 and PO8). Negative voltages are plotted upward. Each tick on the $x$-axis represents 100 ms and on the $y$-axis represents a 2-μV voltage change. Mean ERP amplitudes were computed separately for ipsilaterally and contralaterally cued targets in three time ranges (120–140 ms, 150–170 ms, and 240–260 ms post-stimulus) that are spanned by the gray bars. (B) Scalp voltage topographies of the enhanced negativity in three time ranges. The maps were created by subtracting the ERPs to contralaterally cued targets from the ERPs to ipsilaterally cued targets and plotting the difference. The first two time ranges correspond to the early and late portions of the first phase of the enhanced negativity. The third time range corresponds to the second phase of the enhanced negativity. (C) Projections of calculated dipole sources onto corresponding brain sections of an individual participant. The dipoles were based on grand-averaged ERP data. Dipoles 1 and 2 (shown as circles) were located in the superior temporal lobe (Talairach coordinates of $x = \pm 43, y = -32, z = 9$). Dipoles 3 and 4 (shown as pentagons) were located in the fusiform gyrus of the occipital lobe ($x = \pm 33, y = -58, z = -5$). Dipoles 5 and 6 (shown as squares) were located in perisylvian parietal cortex near the post-central gyrus ($x = \pm 35, y = -25, z = 35$). The dipole model accounted for 97.6% of the variance in voltage topography over the 120–260 ms interval following stimulus onset.

**Maps of cuing effect for LVF Stimuli**

- 120–140 ms
- 150–170 ms
- 240–260 ms

---

**ERPs evidence.** Namely, McDonald, Teder-Sälejärvi, Heraldez, et al. (2001) have provided ERP evidence of the effect of auditory cues on visual target processing. As in the auditory-visual ERP study by McDonald and Ward (2000), ERPs were more negative on ipsilaterally cued trials than on contralaterally cued trials (Fig. 1.13A). The earliest portion of the enhanced negativity (120–140 ms after target onset) was focused over the ipsilateral parietal scalp, whereas a later portion of the enhanced negativity (220–260 ms after target onset) was focused over the midline frontocentral scalp (Fig. 1.13B). No dipole source modeling was performed, but it is conceivable that the neural source of the early enhanced negativity was generated outside of auditory cortex (perhaps in the parietal lobe), whereas the later enhanced negativity was generated in auditory cortex itself.
Conclusions

Interest in the cross-modal consequences of the covert orienting of exogenous spatial attention has grown rapidly in the past few years. An impressive array of behavioral and electrophysiological studies have now been published on the topic, and a number of important findings documented. Despite a number of early disagreements in this area, it now appears that a consensus view has been reached by researchers in the field. In particular, the empirical evidence currently available supports the view that the presentation of a spatially nonpredictive auditory, visual, and/or tactile cue can lead to a covert shift of exogenous spatial attention in the cued direction that will facilitate the processing of auditory, visual, and/or tactile targets subsequently presented at (or near to) the cued location. A number of studies have shown that this performance benefit is primarily focused at just the cued location and falls off as the target is moved farther and farther away from the cue, even when the cue and target stimuli are presented within the same hemifield (e.g., see Chong & Mattingley, 2000; Driver & Spence, 1998; Frassinetti et al., 2002; Kennett, 2000; Kennett & Driver, 1999; Schmitt et al., 2001). In everyday life, such exogenous covert orienting would probably be followed by an overt eye movement toward the cued location in order to fixate the event of interest. However, in the majority of studies reviewed in this chapter, the participants were instructed to keep their eyes on central fixation, thus resulting in a dissociation between covert and overt attentional orienting.

Several recently published cross-modal cuing studies have demonstrated that the cross-modal orienting of spatial attention can result in an enhancement of perceptual processing for targets presented at the cued location (Kato & Kashino, 2001; Frassinetti et al., 2002; McDonald, Teder-Sälejärvi, Di Russo, et al., 2000). Moreover, a number of researchers have now shown that cross-modal cuing effects also result in the modulation of neural activity in modality-specific regions of cortex associated with the target modality (Kennett et al., 2001; Macaluso et al., 2000; McDonald, Teder-Sälejärvi, Heraldez et al., 2001; McDonald et al., 2003; McDonald & Ward, 2000).

The suggestion that the existence of cross-modal cuing effects may reflect the consequences of a covert shift of a supramodal attentional system in the direction of the cue was first put forward 15 years ago by Martha Farah and her colleagues (Farah et al., 1989, pp. 469–470). However, this claim was based on a single experiment in which auditory cues were shown to facilitate simple detection latencies for visual targets presented on the cued side in patients suffering from parietal
brain damage. Farah et al. did not provide any evidence that visual or tactile cues could elicit an exogenous shift of spatial attention, and so failed to provide sufficient evidence to support their claim. Moreover, the majority of the subsequently published data (typically showing asymmetric cross-modal cuing effects) appeared to be inconsistent with the supramodal attention view, and so it soon fell out of favor.

In the years since the original study by Farah et al., researchers have argued for the existence of various different kinds of attentional control systems. For example, Mondor and Amirault (1998) proposed the existence of entirely modality-specific attentional systems (with any cross-modal cuing effects simply reflecting the engagement of endogenous attentional mechanisms). Meanwhile, Spence and Driver (1997a) argued for the existence of separate-but-linked attentional systems instead (see also Schmitt et al., 2000, for a similar view), and others have suggested some combination of modality-specific and supramodal attentional systems (e.g., Posner, 1990, pp. 202–203; Ward, 1994; see also Buchtel & Butler, 1988).

Given that cross-modal cuing effects have now been demonstrated between all (nine) possible combinations of auditory, visual, and tactile cue and target stimuli (at least under certain experimental conditions), the available evidence would appear to provide a reasonable degree of support for the original suggestion by Farah et al. (1989) that cross-modal covert attentional orienting effects may actually reflect the operation of a supramodal attentional system. It is important to realize, however, that the putative existence of a supramodal attentional spotlight does not, in and of itself, necessitate that cross-modal cuing effects should always be found between all possible pairs of sensory modalities, no matter what the paradigm used. For, as we have already seen, a variety of top-down factors, from the complexity of the cuing situation to the relevance of space to the participant’s task, can influence the magnitude of any cross-modal cuing effects observed (e.g., McDonald & Ward, 2000; Mondor & Amirault, 1998; Spence, 2001). Although the idea that top-down factors might modulate exogenous cuing effects reported is a relatively new one, it would be very surprising if they did not exert any influence, given the extensive literature on the contingent nature of attentional capture effects within purely visual attentional cuing paradigms (see Yantis, 1996, 2000, for reviews).

It is also important to note that differences between the senses in terms of the relative ease with which auditory, visual, and tactile stimuli can be localized will likely mean that the distribution of attentional facilitation will be somewhat different, depending on the modality (and hence the localizability) of the cuing event used (Spence, 2001; see also Fig. 1.6). Additionally, any differences in transduction latencies between the senses (see Spence & Driver, 1997a; Spence & Squire, 2003) will presumably also result in modality-specific differences in the time course of cross-modal cuing effects, depending on the particular cue-target modality combination being investigated. Therefore, our view is that while attention may well be controlled by a supramodal attentional mechanism, it will nevertheless exhibit some modality-specific features, given the differences in the spatiotemporal processing of stimuli in each of our senses.

Several of the seminal investigations into the nature of any cross-modal links in spatial attention presented peripheral spatial cues that were also informative (i.e., predictive) with regard to the likely target location (e.g., Buchtel & Butler, 1988; Butler et al., 1989). A number of researchers have subsequently argued that these early studies may have confounded the effects of exogenous and endogenous attentional orienting (e.g., Spence & Driver, 1996; see also Pashler, 1998). Spence and Driver went so far as to suggest that the findings of these particular studies had no necessary implications for the possible existence of any cross-modal links in spatial attention. More recently, however, researchers have been able to study each type of spatial attention in relative isolation, by using specifically designed experimental paradigms. For example, spatially nonpredictive peripheral cues have been used to investigate the cross-modal consequences of purely exogenous covert orienting (as highlighted in the majority of experiments reviewed in this chapter), while spatially predictive symbolic cues have been used to investigate the cross-modal consequences of purely endogenous covert spatial attention (e.g., Lloyd, Merat, McGlone, & Spence, 2003; Spence & Driver, 1996; Spence, Pavani, & Driver, 2000; see Driver & Spence, 2004, for a recent review). Now that a better understanding of each of these types of attention in isolation has been developed, it will be important for researchers to begin to investigate how the two forms of attention interact to control behavior, as they inevitably do in the majority of the more ecologically relevant situations outside of the experimenter’s laboratory. One need only think of the cocktail party example with which we started this chapter: the endogenous direction of our attention to a particular speaker can be overridden by the exogenous attentional capture that takes place when someone else suddenly calls our name. Studying the interaction between these top-down and bottom-up factors constantly competing to control the spatial distribution of attention will represent a particularly fruitful as well as challenging area for future cross-modal research.
REFERENCES


