RULES ARE MADE TO BE BROKEN: MULITSENSORY INTERACTIONS AT TWO STAGES OF CORTICAL PROCESSING

by

IAN CHRISTOPHER FIEBELKORN

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2011

This manuscript has been read and accepted for the
Graduate Faculty in Psychology in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Date	Sophie Molholm, Ph.D. Chair of Examining Committee
Date	John J. Foxe, Ph.D. Chair of Examining Committee
Date	Maureen O'Connor, Ph.D. Executive Officer

Supervisory Committee:

Josh Wallman, Ph.D.

Jennifer A. Mangels, Ph.D.

Steven A. Hillyard, Ph.D.

Peter Lakatos, Ph.D.

THE CITY UNIVERSITY OF NEW YORK

Abstract

RULES ARE MADE TO BE BROKEN: MULTISENSORY INTERACTIONS AT TWO STAGES OF CORTICAL PROCESSING

by

Ian Christopher Fiebelkorn

Advisers: Sophie Molholm, Ph.D., and John J. Foxe, Ph.D.

Research over the past few decades has illuminated the multisensory brain. While information from the various senses is first processed in segregated channels, this segregation is more the exception than the norm. It has now been convincingly demonstrated that the senses can begin to interact at the onset of processing in early sensory cortices (e.g., Foxe et al., 2000; Foxe & Schroeder, 2005; Lakatos, Chen, O'Connell, Mills & Schroeder, 2007; Lakatos, Karmos, Mehta, Ulbert & Schroeder, 2008; Lakatos et al., 2009; Molholm et al., 2002; Murray et al., 2005). These multisensory interactions continue as environmental stimuli proceed to be processed in higher-order cortical areas, but the rules and outcomes change. The following experiments were designed to investigate the neuroanatomic and neurophysiologic underpinnings of multisensory interactions at two stages of processing: (1) an earlier stage at the onset of cortical processing, where multisensory interactions contribute to detection and selection, and (2) a later stage of cortical processing, where multisensory features are combined into a coherent object. We also focus on the rules that govern these interactions. Basic rules for multisensory integration were first established in the cat superior colliculus (Meredith & Stein, 1983; Meredith & Stein, 1986; Meredith, Nemitz & Stein, 1987). These rules state that multisensory integration is more likely when (1) the

unisensory components arise from approximately the same location (i.e., the spatial rule), (2) the unisensory components occur at approximately the same time (i.e., the temporal rule), and (3) the unisensory components elicit weak responses when they are presented in isolation (i.e., the rule of inverse effectiveness). While these seminal rules have provided useful guidelines, more recent research has shown that they are not applicable to all multisensory interactions (e.g., Murray et al., 2005; Stein, London, Wilkonson & Price, 1996; Teder-Sälejärvi, Di Russo, McDonald & Hillyard, 2005; Van der Burg et al., 2008a). Here we provide further evidence that the rules for multisensory integration, as well as its outcomes, depend on several factors, including the stage of cortical processing and the observer's strategic goals.

Dedication

This dissertation is dedicated to my parents, Richard and Deborah Fiebelkorn, whose love, support, and encouragement have made all the difference. They are truly the loveliest and most inspirational people I know.

Acknowledgments

I would first like to thank my brilliant mentors, Sophie Molholm and John Foxe, who brought me to New York. Their enthusiasm and support kept me going when times were tough, and their teaching has provided me with an incredibly strong and varied scientific base. Any successes I have going forward will be theirs. I would also like to thank the post-doctoral fellows who have made significant contributions to my graduate education: Simon Kelly, Edmund Lalor, and Manuel Gomez-Ramirez. I am eternally grateful for your continued friendship and support. Many thanks also to the other friends who have made the past five years in New York nothing short of amazing: Ilana Abeles, Drew Avery, Daniella Blanco, Stephan Bickel, Daniel Belyusar, John Butler, Jennifer Cashion, Filip and Virginia De Sanctis, Edel Flynn, Hans-Peter Frey, Andie Iatrou, Clare Kelly, Rosemary Lalor, Josh and Andrea Lucan, Manuel Mercier, Jeremy Murphy, Lars Ross, Natalie Russo, Pejman Sehatpour, and Adam Snyder. A special thanks to my brother and sister, Eric Fiebelkorn and Claire Flynn. I love you dearly.

Finally, I would like to declare my deepest love and devotion to Lauren Mine, my best friend and partner in crime. I am incredibly excited about our future together, wherever it takes us.

Table of Contents

List of Tables

List of Figures

GENERAL INTRODUCTION

The traffic light changes and you step into the crosswalk, your mind turned inward, contemplating the intricacies of multisensory integration. Suddenly, your thoughts are interrupted by the blast of a car horn. As you turn toward oncoming traffic, you identify the taxi screeching to a halt as the source of the interruption, and your nostrils are filled with the smells of exhaust fumes and burning rubber.

The nervous system has unique sensory epithelia that sample various components of the surrounding environment. Although these components (such as sound waves and light) are first processed in segregated channels, combining incoming information from the different sensory epithelia improves the likelihood of accurate detection, localization, and identification. The body of work presented here investigates multisensory integration at two stages of cortical processing: (1) an earlier stage in low-level cortical areas, where multisensory interactions contribute to detection and sensory selection (see **Chapters 1 and 2**), and (2) a somewhat later stage in higher-level cortical areas, where multisensory interactions contribute to feature integration (see **Chapters 3 and 4**). Our findings highlight the mechanistic differences across these two stages, as well as the variable rules for multisensory interactions, which depend on the behavioral, neuroanatomic, and neurophysiologic context.

A potential pathway for multisensory interactions during the initial stages of cortical processing. It was once thought that the various components of a multisensory stimulus were processed extensively in unisensory cortex before being combined in so-called higher-order multisensory association areas, such as the superior temporal sulcus (see Calvert, Hansen, Iversen & Brammer, 2001). Recent evidence, however, has demonstrated that the sensory systems can also interact in early sensory cortices near the onset of stimulus processing (e.g., Foxe et al., 2000; Foxe & Schroeder, 2005; Lakatos et al., 2007; Lakatos et al., 2008; Lakatos et al., 2009; Molholm et al., 2002; Murray et al., 2005). This broader view of multisensory integration gained considerable traction when it was confirmed that such interactions among early sensory cortices were anatomically plausible. Several studies have now described corticocortical projections between the sensory cortices, dispelling the myth of purely unisensory cortex (Cappe & Barone, 2005; Falchier, Clavagnier, Barone & Kennedy, 2002; Falchier et al., 2010; Rockland & Ojima, 2003; Smiley & Falchier, 2009).

For obvious reasons, these direct neural pathways between the sensory cortices are often cited as a possible anatomic substrate for multisensory interactions during the earliest stages of cortical processing (e.g., Molholm et al., 2002; Van der Burg, Olivers, Bronkhorst & Theeuwes, 2008a). Van der Burg et al. (2008a), for example, suggested that corticocortical projections from auditory cortex to visual cortex might be responsible for the auditory facilitation of visual-target detection, where a co-occurring sound improves detection within a cluttered, continuously changing visual scene (such as the street of New York). Auditory-to-visual projections are hypothesized to increase subthreshold activity in visual cortex, such that the afferent signal associated with visual-target presentation is more likely to result in detection.

In **Chapter 1** we test a behavioral prediction based on the specific pattern of projections from auditory cortical regions to primary visual cortex. Falchier et al. (2002)

demonstrated that these corticocortical projections terminate most prominently in the areas of visual cortex that process peripheral space (i.e., the areas of cortex that process the visual field 10 to 20 degrees from central fixation), suggesting that audiovisual interactions that occur via this neural pathway might specifically serve to bolster the processing of peripheral stimuli. We therefore investigate whether auditory facilitation of visual-target detection is greater when a near-threshold visual target is presented in the periphery, relative to when the same visual target is presented near central fixation. If these projections from auditory to visual cortex are indeed a primary anatomic substrate for the auditory facilitation of visual-target detection, then the resulting multisensory enhancement should be greater in response to visual targets presented peripheral to central fixation (see **Chapter 2** for a discussion of additional neural pathways that might be involved in multisensory modulation of early sensory cortices).

Is spatial alignment a prerequisite for multisensory integration during the initial stages of cortical processing? Much of the seminal work regarding multisensory integration was conducted using single-cell recordings in the cat superior colliculus, a midbrain structure involved in the generation of saccades and other orienting movements (Meredith & Stein, 1983; Meredith et al., 1986; Meredith & Stein, 1987). This important work led to the formulation of several rules for multisensory integration, including “the spatial rule,” which suggests that neural responses to multisensory stimuli tend to be enhanced when they occur at approximately the same location (Meredith & Stein 1986). Subsequent work, however, has demonstrated that the degree to which this rule is applicable is contingent on several factors, including the task at hand, as well as when and where in the brain the multisensory interactions occur (e.g., Murray et al., 2005; Stein et al., 1996; Teder-Sälejärvi et al., 2005; Van der Burg et al., 2008a). Murray et al. (2005), for example, demonstrated early auditory-somatosensory interactions (at ~50 ms) in low-level sensory cortices that occurred during a simple detection task, despite wide peripersonal separations between stimuli presented in the two sensory modalities (100 degrees along the horizontal meridian).

In **Chapter 1** we test whether spatial alignment is a prerequisite for the auditory facilitation of visual-target detection, which is hypothesized to result from multisensory interactions during the initial stages of cortical processing (Lakatos et al., 2009; Van der Burg et al., 2008a). Visual targets are paired with sounds that originate from locations that are separated by as much as 104 degrees along the horizontal meridian. If auditory facilitation of visual target-detection occurs despite these wide spatial misalignments, it would suggest an underlying neural mechanism that operates through a spatially diffuse modulation of visual responses. It would also exclude brain areas where the spatial rule of multisensory integration is known to dominate, such as the superior colliculus, as the source of these multisensory effects. Instead, auditory facilitation of visual-target detection might result from early interactions in brain regions that have been traditionally considered to be modality specific (e.g., Lakatos et al. 2007; Lakatos et al. 2009; Murray et al. 2005; Noesselt, Bergmann, Hake, Heinze & Fendrich, 2008; Noesselt et al., 2010).

A potential mechanism for multisensory interactions during the initial stages of cortical processing. Early multisensory interactions likely influence the competition for limited processing resources. By shifting baseline activity in the sensory cortices, multisensory effects—in addition to other factors, such as top-down strategic goals—contribute to whether a given stimulation will exceed the threshold necessary to be selected for further downstream processing. Recent evidence suggests that multisensory modulation during the initial stages of cortical processing does not necessarily occur via a direct boost in cortical activity, but rather through a temporal modulation of ongoing neural dynamics (Lakatos et al., 2007; Lakatos et al., 2008; Lakatos et al., 2009).

It is well-established that ongoing oscillations shape our perception (e.g., Busch, Dubois & VanRullen, 2009; Busch & VanRullen, 2010; Foxe, Simpson & Ahlfors, 1998; Lakatos et al., 2009; Mathewson, Gratton, Fabiani, Beck, & Ro, 2009; Monto, Palva, Voipio & Palva, 2008; Snyder & Foxe, 2010; VanRullen, Reddy & Koch, 2006; VanRullen, Carlson & Cavanagh, 2007; Worden, Foxe, Wang & Simpson, 2000). Two recent studies demonstrated that the likelihood of visual-target detection depends on the prestimulus phase of the brain's intrinsic fluctuations in cortical excitability (Busch et al., 2009; Mathewson et al., 2009). Lakatos et al. (2007; 2008; 2009) further demonstrated that the influence of these cortical oscillations on perception is not always random. Instead the phase of ongoing oscillations can be modulated, such that the arrival of stimulus-driven afferent activity in cortex aligns with either a high- or a low-excitability phase (see Lakatos et al., 2007). What's more, attending to a unisensory stimulus resets the phase of ongoing oscillations across multiple sensory cortices, which means that a sound, for example, might influence the detection and selection of a visual stimulus (see Lakatos et al., 2009).

Phase-modulation of ongoing cortical activity, through which the presentation of a stimulus in one modality influences baseline cortical excitability in another modality, represents a potentially powerful tool for sensory selection and perceptual enhancement. But measurement limitations have prevented the neurophysiologic mechanism from being directly linked with its perceptual outcomes (see **Chapter 2** for a more detailed discussion). In **Chapter 2** we devise a behavioral task to test a hypothesis proposed in the neurophysiologic literature: that auditory facilitation of visual-target detection operates through cross-sensory phase reset of ongoing neural oscillations. To bypass measurement limitations inherent in neurophysiologic recordings, we use the behavioral results themselves to probe for cross-sensory phase reset. If attending to a temporally informative sound consistently resets the phase of ongoing oscillations, visual-target detection at subsequent time points should reflect this underlying, stimulus-locked oscillatory activity. In other words, we hypothesize that the influence of a sound on visual perception can persist for seconds; an idea that runs in opposition to more traditional views of multisensory interactions (in the contexts of localization and feature integration), where the influence of sound on visual processing is thought to be limited to asynchronies on the order of hundreds of milliseconds (e.g., Slutsky & Recanzone, 2001; Stone et al., 2001).

From selection to multisensory feature integration. Co-occurring stimuli in different sensory modalities can interact at the onset of cortical processing to increase the likelihood of detection and subsequent selection for further processing, but the specific rules that govern these multisensory interactions are not universal. We will now move to a later stage of cortical processing, where the various components of an object, including its nonvisual components, are combined into a coherent whole.

How the brain selects features and weaves their anatomically distributed representations into the perception of an integrated object (such as a taxi) is a fundamental question in the neurosciences: the so-called “binding problem.” The most influential models aimed at addressing this question have focused on the role of attention, which can be broadly defined as the mechanisms through which the nervous system prioritizes limited processing resources in a complex environment. Feature Integration Theory (FIT), for example, suggests that spatial selection (i.e., the selection of delimited location for further processing) is required for veridical integration, and thus for the perception of unified objects (Treisman & Gelade, 1980). In support of this assertion, Treisman and Schmidt (1982) demonstrated that binding outside the spotlight of spatial selection sometimes leads to the formation of illusory conjunctions, where participants combine the color of one object with the shape of another. Such findings provide compelling evidence for the importance of spatial selection in determining which features should be combined as components of the same object. Here we extend the discussion of feature integration outside the visual domain, but the importance of spatial factors in feature integration will remain a central issue. This differs from our discussion of multisensory interactions at the onset of cortical processing, where the auditory facilitation of visual-target detection is hypothesized to occur despite wide audiovisual misalignments.

An inherent bias to process objects as wholes. Research within the visual domain clearly shows that there is an inherent bias to process objects as wholes, even when either the object is partially outside the spotlight of spatial selection or particular features of the object are irrelevant to the task at hand (e.g., Blaser, Pylyshyn & Holcombe, 2000; Egly, Driver & Rafal, 1994; Martinez et al., 2006; O'Craven, Downing & Kanwisher, 1999; Schoenfeld et al., 2003; Stroop, 1935). Such research suggests that objects serve as basic units of attention (for reviews, see Driver & Baylis, 1998; Hopf, Schoenfeld & Heinze, 2005; Scholl, 2001). Regardless of whether attention is first deployed based on a location (i.e., spatial selection) or based on a specific feature (i.e., feature-based selection), the competition for limited processing resources often continues based on the properties of an object. That is, experimental evidence suggests both that the spotlight of spatial attention (and the enhanced neural processing associated with it) automatically narrows or expands to match the extent of an object's visual boundaries (e.g., Egly et al., 1994; Martinez et al., 2006; Martinez, Teder-Sälejärvi & Hillyard, 2007), and that attending to a single feature leads to the automatic selection of all of an object's features (e.g., O'Craven et al., 1999; Schoenfeld et al., 2003).

Egly et al. (1994) asked participants to covertly direct their attention to the cued end of one of two bar-shaped objects, where a target was most likely to appear. Participants responded regardless of whether the target appeared at the cued location or at an uncued, invalid location. Average reaction times in response to invalid targets at the uncued end of the attended bar were faster than those in response to invalid targets at the equidistant end of the unattended bar. These findings therefore suggested a same-object advantage with at least two interpretations: (1) either there is a behavioral tendency to first search for an invalid target within the visual boundaries of an object (i.e., priority mapping) (Shomstein & Yantis, 2002; Yantis & Serences, 2003), or (2) attention-related processing spreads within the visual boundaries of an object (Vecera & Farah, 1994; Weber, Kramer & Miller, 1997; Davis, Driver, Pavani & Shepherd, 2000).

To resolve this ambiguity, Martinez et al. (2006) recorded event-related potentials (ERPs) while participants performed nearly the same task, with the exception that invalid targets were irrelevant and therefore required no response. Uncued stimuli that formed part of the attended bar received enhanced neural processing relative to uncued stimuli that formed the equidistant end of the unattended bar during the timeframe of the visual N1 component (160 to 196 ms). Inverse dipole modeling during the latency of the N1 component localized this object-based selection effect to activity in the lateral occipital complex (LOC). Since participants never needed to shift their spatial attention to search for invalid targets, the same-object advantage could not be explained by a behavioral tendency to first search for invalid targets within the visual boundaries of an object (though such prioritization would still require object-based guidance of spatial attention). Rather, these results demonstrated that the same-object advantage stems from the spread of attentional processing within the visual boundaries of an object. The selection of a location within an object leads to the selection of the entire object, expanding the scope of enhanced processing beyond those areas first highlighted by spatial selection (i.e., the target location). In the context of an approaching taxi, attending to the front bumper (the closest part of the approaching taxi) leads to enhanced processing of the entire vehicle, including its hood and windshield.

Objects, however, are defined not only by their form, but also their colors, textures, and motion. A central tenet of the influential biased-competition model of visual attention suggests that the selection of a specific feature results in the preferential processing of an entire object, including features irrelevant to the task at hand (Desimone & Duncan 2006; Duncan, 2006). Indeed, substantial evidence from both ERP and fMRI studies now indicates that enhanced processing spreads to an object's task-irrelevant visual features (e.g., O'Craven et al., 1999; Schoenfeld et al., 2003), seemingly reflecting the formation of an object representation across a distributed cortical network.

Schoenfeld et al. (2003) asked participants to view two superimposed transparent surfaces formed by arrays of dots that moved in opposite directions. This stimulus created the percept of dots painted on transparent surfaces sliding past each other. The right-moving or left-moving dots could be either red or white; however, the color of the dots was irrelevant to the task. During each trial, participants were asked to attend to a specific direction of motion and respond if the dots moving in the attended direction did so at a greater than normal velocity. When the dot array moving in the attended direction also displayed a task-irrelevant color (red), fMRI recordings showed increased neural activity in both the task-relevant movement areas of the cortex and the task-irrelevant color-selective areas. Since the transparent dot arrays were presented at the same location, the researchers concluded that the spread of enhanced processing from the attended direction of movement to the task-irrelevant color could not have resulted from spatial selection alone, but must also have involved object-based selection (i.e., that task-irrelevant color did not receive enhanced processing because it occurred at the attended location, but rather because it was a feature of the attended object). ERP and event-related magnetic field (ERF) recordings further showed that processing within color-selective areas in response to a task-irrelevant color onset at approximately 190 ms (i.e., processing within color-selective areas that occurred regardless of whether the task-irrelevant color was a component of the attended or the unattended dot array), whereas additional processing within color-selective areas that occurred only when the task-irrelevant color was a component of the attended object onset at approximately 240 ms. The authors argued the spread of enhanced processing from the attended feature of an object (i.e., its direction of motion) to its task-irrelevant visual features (i.e., its color) happened rapidly enough to reflect a mechanism for feature integration.

In sum, there is an automatic spread of attentional processing across the cortical regions that process the spatial extent of an object (including areas of the object that are outside the initial focus of spatial selection), and across the cortical regions that process an object's task-irrelevant visual features (e.g., its motion or color). We will henceforth refer to this bottom-up, object-based spread of attentional processing as the stimulus-driven spread of attention. The resulting activation of a distributed cortical network is thought to reflect the formation of a coherent object representation. A taxi is therefore experienced as a yellow moving vehicle, rather than as disjointed features: yellow, moving, and a vehicle.

The whole-object bias also encompasses an object's nonvisual features. Although we may commonly favor visual properties when describing objects, and historically scientists have focused on visual stimuli when studying attention and feature binding, the majority of natural objects in our environment have multisensory features. Indeed, from birth we experience an extremely important multisensory object, the primary caretaker, who is associated with a voice, an image, a smell, a feel, and so on, and these individual features are obviously components of the same object. Returning to our example from the inanimate world, a taxi has several characteristic sounds: the rumbling of its engine, the blast from its horn, and in New York, one might go so far as to say, shouted expletives from its driver. If attentional processing spreads to encompass the task-irrelevant visual features of an object, a logical extension is that an object's nonvisual features will also receive enhanced processing. A series of recent ERP studies have set out to test this hypothesis, yielding extremely compelling results.

In the first of these studies, Busse et al. (2005) showed that a task-irrelevant sound received enhanced processing when it was paired with an attended visual stimulus. Because the attended visual stimulus was presented in the periphery while the task-irrelevant sound was presented at central fixation, the researchers argued that this cross-sensory effect must have resulted from object-based selection rather than the well-known multisensory nature of spatial attention (e.g., Hillyard, Simpson, Woods, Van Voorhis & Munte, 1984; McDonald, Teder-Sälejärvi, Di Russo & Hillyard, 2003). That is, Busse et al. (2005) used the ventriloquism illusion, whereby perceptual localization of a sound is captured by a simultaneously presented but spatially disparate visual stimulus, to attribute a cross-sensory spread of processing to an object-based, late attentional selection process (an ERP modulation starting approximately 220 ms poststimulus). Participants maintained central fixation and performed a covert attention task in response to visual stimuli presented in the periphery. The participants were instructed to respond when a visual stimulus (checkerboards containing 0, 1 or 2 dots) on the attended side of space included a target (2 dots). A centrally presented, task-irrelevant sound was sometimes paired with the visual stimulus that appeared on the covertly attended side of space, and sometimes paired with the visual stimulus that appeared on the unattended side of space.

When paired with the covertly attended visual stimulus, the task-irrelevant sound received enhanced processing (relative to when it was paired with the unattended visual stimulus). The scalp distribution and timing of this negative-going ERP effect resembled the selection negativities typically observed when a sound is specifically attended (Näätänen, Gaillard & Mäntysalo, 1978). Enhanced processing of task-irrelevant sounds paired with attended visual stimuli has thus been described as an automatic spread of attentional processing (a stimulus-driven spread of attention) and taken to indicate that the task-irrelevant sound was grouped as part of the same object as the attended visual stimulus (Busse et al., 2005; Talsma, Doty & Woldorff, 2007).

fMRI, and source modeling localize the frontal negativity associated with the cross-sensory spread of attentional processing to auditory cortex (Busse et al., 2005; Talsma et al., 2007). These observations accord well with studies that described a spread of attentional processing across the visual features of an object, where enhanced processing was detected within feature-specific cortices (O'Craven et al., 1999; Schoenfeld et al., 2003). When an attended visual stimulus included a task-irrelevant color, for example, enhanced activity was reported within color-selective cortical regions (fusiform gyrus). As in the previous section, enhanced activity spreading across feature-specific cortex is thought to reflect the activation of a distributed cortical network that together represents the object. The cortical regions that receive enhanced processing are those regions that process the constituent features of the partially attended object, regardless of whether those features are relevant or irrelevant to the present task.

What stimulus properties determine whether multisensory features are grouped as components of the same object? Given the distributed nature of the cortical representations of stimulus features, an important question is which variables govern the selection of features to be grouped; that is, how does the brain establish what goes with what. Both physical stimulus properties (e.g., a shared location) and semantic stimulus properties (i.e., learned associations) might be hypothesized to play key roles in this process.

Spatial overlap is an obvious indication that two features belong to the same object, but distinct objects frequently overlap in space. It therefore seems to follow that the visuospatial boundaries of an object must be established before multisensory feature integration can proceed. **Chapter 3** tests this assertion, using ERPs and Kanizsa-type illusory contour stimuli to measure the influence of differential boundary formation on multisensory feature integration.

The illusory contour effect is a well-established manipulation of object processing where the positioning of inducers (i.e., the pacman-like elements that constitute Kanizsa-type illusory contour stimuli) determines whether a visual stimulus will be perceived as a unified object or as a collection of elements (e.g., Kanizsa, 1976; Murray et al., 2002; Murray, Imber, Javitt & Foxe, 2006). When the inducers combine to form a more coherent object with less ambiguous boundaries (i.e., illusory contours), there is a signature increase in neural activity within the LOC, a brain area repeatedly associated with object processing and visual feature integration (e.g., Doniger et al., 2000; Martinez et al., 2006; Murray et al., 2002; Sehatpour et al., 2008).

In **Chapter 3** we predict that the differential segmentation of a visual stimulus through the formation of illusory contours will modulate multisensory feature integration in a bottom-up manner, such that the cross-sensory spread of attention will be stronger when the visual elements form a more coherent object with well-defined boundaries. In other words, we expect that the refinement of attended space through the formation of visual boundaries will influence the degree to which a task-irrelevant sound is linked to an attended visual stimulus as a multisensory object.

Up to this point, the described studies used simple multisensory objects (e.g., checkerboards paired with tones), rather than well-known multisensory objects (e.g., barking dogs) to investigate the cross-sensory spread of attention. Frequently encountered objects have stored cortical representations that might influence multisensory feature integration. A car, for example, is a well-known multisensory object. Through repeated exposure, we have learned to recognize the sounds that cars produce. Moreover, we are able to use those sounds to identify the presence of a car in the absence of visual stimulation. The question arises; do such highly learned associations become "hard-wired," such that the stimulus-driven cross-sensory spread of attention is facilitated to a greater extent when the inputs have been repeatedly presented together?

To address this question, **Chapter 4** uses ERPs to compare the spread of attentional processing from an attended visual stimulus to its paired task-irrelevant sound when that sound is either object-congruent (e.g., barks paired with dog photos) or object-incongruent (e.g., barks paired with guitar photos). If highly learned associations affect multisensory feature integration (as measured through the cross-sensory spread of attention), a semantically congruent task-irrelevant sound would be expected to receive enhanced processing, relative to a semantically incongruent task-irrelevant sound. We expect that our findings will accord well with subjective experience, where a barking cat would be weird, but the multisensory features might initially be perceived as a coherent object. Operating in this manner would seem to be highly adaptive in that it allows for the rapid binding of inputs, while also leaving room for the learning of new associations.

A second mechanism for the cross-sensory spread of attention. To city-dwellers, taxis are frequently encountered, well-known multisensory objects. Although vision seems to dominate bottom-up object processing, a lifetime of experiences with an object leads to highly learned associations that increase the influence of an object's nonvisual properties. For example, we can recognize the sound of a car screeching to a halt without seeing the vehicle. In addition to the stimulus-driven spread of attention, which may or may not be modulated by highly learned associations, a second, top-down mechanism for the cross-sensory spread of attention has been established that relies entirely on highly learned associations.

Molholm, Martinez, Shpaner and Foxe (2007) recorded ERPs during a one-back visual task where participants attended for consecutive presentations of a target object: a line-drawing of a dog, guitar, or hammer. These well-known multisensory objects were sometimes paired with their congruent sound (e.g., the dog was sometimes paired with a bark). The results demonstrated a target-specific spread of attentional processing from the attended visual object to a congruent, but completely task-irrelevant sound. Unlike the stimulus-driven spread of attention (described in previous sections), the sound associated with the target object received enhanced processing even when it was presented alone, without its corresponding line-drawing. This observation indicates that enhanced processing of a well-known target object's multisensory features is not dependent on the presence of its attended visual features. Since, however, the scalp topography and latency of the target-specific effect are similar to those for the stimulus-driven spread of attention, it is also considered to reflect a cross-sensory spread of attention: the representation-driven spread of attention.

Chapter 4 will use congruent task-irrelevant sounds and consistently paired incongruent task-irrelevant sounds to test whether this target-specific spread of attentional processing results from pre-existing associations among the multisensory features of the well-known objects, or alternatively, from associations formed during the experimental session.

A model for cortical audiovisual integration. Data from the following four studies (**Chapters 1–4**) will be used to inform a model for cortical audiovisual integration, where the rules for integration, as well as its outcomes, depend on neurophysiologic and neuroanatomic underpinnings, strategic goals, and highly learned associations. This model will be used to summarize our findings regarding multisensory interactions that contribute to detection and selection, and to multisensory feature integration (see the **General Discussion**).

CHAPTER 1

Auditory Facilitation of Visual-Target Detection Persists Regardless of Retinal Eccentricity and Despite Wide Audiovisual Misalignments

Several studies have now demonstrated that sounds facilitate the detection of simultaneously presented visual targets (Lippert et al., 2007; Noesselt et al., 2008; Noesselt et al., 2010; Odgaard et al.,2003; Stein et al., 1996; Van der Burg et al., 2008a; Van der Burg et al., 2008b), but the neural mechanisms and pathways underlying this well-established behavioral effect are still being investigated. Here we manipulated the spatial properties of co-occurring sounds and visual targets to test specific hypotheses based on previously reported neurophysiologic and neuroanatomic findings.

Two primary anatomic models of cortical multisensory integration have been proposed. In one model, unisensory information is first processed extensively in unisensory cortex, with these well-elaborated unisensory representations then converging in so-called higher-order multisensory association areas such as the superior temporal sulcus (Calvert et al., 2001). In the other model, unisensory information begins to interact in early sensory cortices near the onset of stimulus processing (Foxe et al., 2000; Foxe & Schroeder, 2005). Whereas it was once thought that only the first of these models was anatomically plausible, neuroanatomic data now also exist to support the second model (Cappe & Barone, 2005; Falchier et al., 2002; Falchier et al., 2010; Rockland & Ojima, 2003). Falchier et al. (2002), for example, described direct projections from auditory cortical regions to primary visual cortex. These corticocortical projections terminated most prominently in the areas of visual cortex that process peripheral visual space (i.e., the areas of cortex that process the visual field 10 to 20 degrees from central fixation), suggesting that audiovisual interactions that occur via this neural pathway might serve to specifically bolster the processing of peripheral stimuli. This is an attractive finding in that it suggests the possibility of an early warning system, whereby potentially threatening stimuli that arise outside central vision are more likely to be detected. We took advantage of this important anatomic detail to test whether these previously described corticocortical projections are involved in auditory facilitation of visual-target detection (**Figure 1**). Specifically, we tested whether improvements in detection are greater when a visual target paired with a co-occurring sound is presented peripheral to central fixation, relative to when the same bisensory pairing is presented at central fixation.

To further probe the neural substrates underlying the auditory facilitation of visual-target detection, we manipulated audiovisual spatial alignment. Previous research has revealed that early auditory-somatosensory interactions in low-level sensory cortices occur despite wide peripersonal separations (100 degrees along the horizontal meridian; Murray et al., 2005). Here we tested whether a co-occurring sound enhances performance in a visual-target detection task, even when the spatial representations of the sound and the visual target are widely misaligned (by up to 104 degrees along the horizontal meridian; **Figure 1**). It is well-established that multisensory neurons typically have overlapping unisensory receptive fields (i.e., the spatial rule of multisensory integration), such that these neurons respond most strongly to stimuli that arise from the same location (e.g., Meredith & Stein, 1986). If auditory facilitation of visual target-detection occurs despite wide spatial misalignments, it would suggest an underlying neural mechanism that operates through a spatially diffuse modulation of visual responses. It would also exclude brain areas where the spatial rule of multisensory integration is known to dominate, such as the superior colliculus, as the source of these multisensory effects.

METHODS

Participants

Twelve neurologically normal volunteers participated in the experiment (mean age 31 ± 6.2 years; 4 female; 2 left-handed). The Institutional Review Board of the City College of CUNY approved the experimental procedures. Written informed consent was obtained from all subjects prior to data collection, in line with the Declaration of Helsinki.

Stimuli and task

Participants, who were positioned 70 cm in front of a 21-inch cathode-ray tube (CRT) monitor, were asked to maintain central fixation and report the occurrence of visual stimuli: sine-wave gratings with 16-ms duration, subtending 2.5 degrees of the visual angle in both the vertical and horizontal planes. These visual targets were equally likely to occur at any of 3 possible target locations: centrally and 2 degrees below central fixation, 14 degrees to the right and 2 degrees below central fixation, and 14 degrees to the left and 2 degrees below central fixation (**Figure 1**). Potential target locations were marked with circles to increase the ease with which participants were able to deploy their attention. When a visual stimulus appeared, participants pressed a button corresponding to its location (i.e., the up arrow to indicate a central target, the right arrow for a peripheral target to the right of central fixation, or the left arrow for a peripheral target to the left of central fixation). Participants were instructed to respond to all visual stimuli, regardless of whether the visual stimuli occurred alone or with a sound. Task-irrelevant sounds (1000 Hz tone, 16 ms) were presented at a comfortable listening level; either from a BOSE (Companion 2) speaker positioned on top of the monitor or from a BOSE (Companion 2) speaker positioned 90 degrees to the left of central fixation, at a distance of 70 cm from the participant (**Figure 1**). The auditory and visual components of the audiovisual stimuli were therefore maximally misaligned by 104 degrees. There were four equiprobable stimulus types: visual-alone (three locations), auditory-alone (two locations), audiovisual (six combinations of visual and auditory locations), and "no-stimulus trials" (i.e., points in time randomized like other stimulus events in which no stimulus actually occurs). Because auditory-alone and audiovisual trials were equiprobable, sounds did not predict the occurrence of a visual stimulus (i.e., a target). To guard against anticipatory or rhythmic responding, the stimulus-onset asynchrony (SOA), which varied randomly throughout the experiment with a rectangular distribution, was 700 to 2500 ms. Participants completed twelve blocks. Each block included 252 trials, and was broken into 6 mini-blocks (i.e., points at which participants were encouraged to take brief breaks). At the conclusion of each mini-block, hit rates and false alarm rates were calculated on-line, and participants were given feedback on their performance. Between mini-blocks, the contrast of the visual stimuli was automatically adjusted as necessary—and separately for central and peripheral stimuli—such that each participant's hit rate in response to visual-alone stimuli was maintained at approximately 65 percent. That is, contrast was increased if participants detected less than 55 percent of the visual stimuli, and contrast was decreased if participants detected more than 75 percent of the visual stimuli. The contrast levels of visual-alone and audiovisual stimuli were always kept identical.

Statistical analysis

Location-specific false alarm rates in response to auditory-alone trials and no-stimulus trials were combined with location-specific hit rates in response to audiovisual trials and visual-alone trials, respectively, to calculate d-prime values (Green & Swets, 1966). A button press was considered a hit if it occurred 150 to 1000 ms after a visual stimulus. D-prime provided a measure of detectability independent of the potential response bias associated with the inclusion of paired sounds. That is, using d-prime to measure detectability eliminates the alternative hypothesis that improved behavioral performance results from participants simply responding to the occurrence of sounds (see Odgaard et al., 2003). To determine whether it was necessary to use d-prime instead of simply analyzing hit rates, we examined criterion, which is a well-established measure of response bias (Macmillan & Creelman, 1991).

To test whether there was significant auditory facilitation of visual-target detection, and whether this enhancement varied with the visual target's retinal eccentricity or its misalignment with task-irrelevant sounds, a two-way repeated-measures ANOVA was conducted with factors of stimulus type (three levels: visual-alone, audiovisual with centrally presented sound, or audiovisual with peripherally presented sound) and target location (three levels: center, right, or left). Following a significant ANOVA, pairwise comparisons were made using planned protected t-tests (Fisher's least significant difference). The same statistical analyses were then repeated with criterion and reaction times as the dependent measure. The alpha level for all statistical analyses was set at 0.05.

RESULTS

Table 1 shows hit rates and false alarm rates for each combination of visual-target location (i.e., left, center, or right) and stimulus type (i.e., visual-alone, audiovisual with centrally presented sound, or audiovisual with peripherally presented sound). A quick inspection of these data suggests that pairing a visual target with a sound improved visual-target detection regardless of retinal eccentricity and despite wide audiovisual misalignments; however, a repeated-measures ANOVA with criterion as the dependent measure revealed that response bias (as measured through criterion) was significantly different across the three target locations ($F_{2,22} = 22.7$, $P < 0.001$, $\eta_p^2 = 0.67$). Pairwise comparisons showed that this difference was attributable to an overall lower criterion for the left-target location, with participants more likely to report having seen a visual target at the leftmost location relative to the central- ($P = 0.001$) and right-target ($P = 0.001$) locations. Participants were also more likely to report seeing a visual target at the rightmost location relative to the central-target location ($P = 0.03$). A statistically significant interaction between visual-target location and sounds ($F_{2,22} = 4.3$, $P < 0.005$, $\eta_p^2 = 0.28$) further revealed that response bias when there was no sound or when there was a sound that occurred at either of the two speaker locations was not uniform across the three visual-target locations (**Figure 2**). Because the analysis of criterion revealed significant differences in response bias across our conditions, it was necessary to use a measure of detectability that is independent of response bias to determine whether co-occurring sounds improved the sensitivity of visual-target detection: d-prime.

A repeated-measures ANOVA with d-prime as the dependent measure revealed that co-occurring sounds did indeed significantly improve the detection of visual targets ($F_{2,22} = 99.6$, $P < 0.01$, $\eta_p^2 = 0.90$). Pairwise comparisons further revealed that this multisensory enhancement of visual-target detection was independent of whether the co-occurring sound was presented from the central speaker or from the peripheral speaker ($P = 0.59$). That is, sounds improved the detection of visual targets even when these unisensory components were misaligned by up to 104 degrees along the horizontal meridian (**Figure 3a**).

The main effect of target location approached significance ($F_{2,22} = 3.3$, $P = 0.06$, $\eta_p^2 = 0.23$), suggesting that the overall detectability of visual targets at central fixation might be greater than that for visual targets peripheral to central fixation. More importantly for our central hypotheses, the interaction between visual-target location and stimulus type failed to reach significance ($F_{4,44} = 1.7$, $P = 0.19$, $\eta_p^2 = 0.13$), suggesting that auditory facilitation of visual-target detection does not significantly differ with regard to retinal eccentricity (**Figure 3a**). Numerically, facilitation was comparable across the three target locations, with d-prime values ranging from 1.9–2.2 in response to visual-alone targets and 2.6–3.0 for audiovisual targets (**Figure 3a**). Although it is unconventional to examine pairwise comparisons when the F-value fails to reach statistical significance, we decided based on a reviewer's comments that it was important to further investigate what was driving the interaction's relatively low P-value (i.e., 0.19). A closer examination of the data confirmed that auditory facilitation of visual-target detection was significant at each target location, regardless of whether the visual target was paired with a sound from the central or the peripheral speaker (all *P-values* < 0.003). There was also a potential trend toward lower detectability of visual-alone stimuli at the leftmost target location relative to the central- ($P = 0.08$) and right-target locations ($P = 0.08$), as well as a smaller increase in detectability when a target at the leftmost location was paired with a sound from the peripheral speaker relative to when it was paired with a sound from the central speaker ($P = 0.06$). None of these potential trends, however, suggest that auditory facilitation of visual-target detection is greater when a visual target is presented peripheral to central fixation rather than at central fixation. These results are therefore inconsistent with what would be predicted based on previously described corticocortical projections from auditory to visual cortex (Falchier et al., 2002). Such projections have been shown to predominately terminate in regions of visual cortex that process the peripheral visual space (Falchier et al., 2002).

A repeated-measures ANOVA with reaction times as the dependent measure mirrored the results from our d-prime analysis, revealing that co-occurring sounds significantly speeded reaction times ($F_{2,22} = 54.8$, $P < 0.01$, $\eta_p^2 = 0.83$). Pairwise comparisons further demonstrated that the influence of co-occurring sounds on reaction times was independent of whether the co-occurring sound was presented from the central speaker or from the peripheral speaker ($P = 0.36$; **Figure 3b**). Combined with the results of our d-prime analysis, these findings indicate that the neural mechanism underlying auditory facilitation of visual-target detection can operate in a spatially diffuse manner, such that the location of a sound does not define the area of the visual field that receives enhancement.

With regard to whether retinal eccentricity influences the speed of visual-target detection, there was a significant main effect of target location ($F_{2,22} = 3.5$, $P < 0.05$, $\eta_p^2 = 0.23$), where reaction times in response to peripheral targets were faster than those in response to central targets, regardless of whether the visual target was presented with a co-occurring sound. These findings make sense in the context of the magnocellular and parvocellular visual pathways, where magnocellular projections, which consist of large, rapidly conducting neurons, dominate in the visual periphery (Kaplan, 1991; Merigan & Maunsell, 1993). In other words, differences between the cell types that populate the central and peripheral retina most likely account for faster reaction times when visual targets are presented in the periphery, relative to when visual targets are presented near central fixation.

DISCUSSION

Although numerous studies have demonstrated auditory facilitation of visual-target detection (Lippert et al., 2007; Noesselt et al., 2008; Noesselt et al., 2010; Odgaard et al., 2003; Stein et al., 1996; Van der Burg et al., 2008a; Van der Burg et al., 2008b), little is known about the underlying neural substrates. One hypothesis put forth in the literature is that these audiovisual interactions are mediated via direct corticocortical projections from auditory cortices to visual cortex (see Van der Burg et al., 2008a). Here, however, we find no evidence that the spatial characteristics of the behavioral effect fit with what one might predict based on the anatomic data. Falchier et al. (2002) reported that projections from belt and parabelt regions of auditory cortex predominately project to regions of visual cortex that process the periphery (10 to 20 degrees from central fixation), but our findings show that auditory facilitation of visual-target detection is similar regardless of the retinal eccentricity at which visual targets are presented. Based on these similarities (see **Figure 3**), it appears that there is a common mechanism operating across the visual field, and that this mechanism therefore does not stem from the previously described direct corticocortical projections (Falchier et al., 2002). It should be noted, however, that future studies could reveal a separate neural pathway that specifically mediates auditory facilitation of visual-target detection near central fixation. The existence of two pathways, one projecting to regions of visual cortex that process stimuli in the periphery and one projecting to regions of visual cortex that process stimuli near central fixation, might also explain these results.

Our findings further demonstrate that a sound's location in no way limits the area of the visual field where auditory facilitation of visual-target detection is observed. Co-occurring sounds therefore seem to improve the detection of near-threshold visual stimuli through a diffuse modulation of visual processing. Several previous studies have similarly described spatially non-specific multisensory interactions, but within different behavioral contexts (e.g., Gondan et al., 2005; Keetels & Vroomen, 2005; Murray et al., 2005; Stein et al., 1996; Teder-Sälejärvi et al., 2005). Vroomen and Keetels (2006), for example, reported that sounds can influence the sensitivity of visual temporal order judgments, even when presented at locations that are separated from the visual stimuli by 90 degrees along the horizontal meridian. These widely misaligned sounds, which occurred both 100 ms before the first visual stimulus and 100 ms after the second visual stimulus, were hypothesized to improve sensitivity through temporal ventriloquism, an illusion whereby a sound presented in close temporal proximity to a visual stimulus alters the perceived temporal occurrence of that visual stimulus. Sounds are thus thought to improve the sensitivity of temporal order judgments by increasing the apparent temporal separation of consecutive visual stimuli. The present experiment, in comparison, shows that spatially misaligned but co-occurring sounds improve the detectability of near-threshold visual stimuli, perhaps through increasing the apparent contrast of those visual stimuli (see Carrasco et al., 2004).

Auditory facilitation of visual-target detection persists even when a co-occurring sound is presented at a location that is separated from the visual target by 104 degrees along the horizontal meridian. These multisensory effects are therefore unlikely to arise from brain regions where spatial overlap of the unisensory components is known to be a prerequisite for multisensory integration (e.g., the superior colliculus). Instead, auditory facilitation might result from interactions in brain regions that have traditionally been considered to be modality specific (e.g., Lakatos et al., 2007; Lakatos et al., 2009; Murray et al., 2005; Noesselt et al., 2007; Noesselt et al., 2010). Murray et al. (2005), for example, used EEG and behavioral measures to demonstrate that auditory-somatosensory interactions occur despite wide peripersonal separations, where the unisensory components of the stimulus were presented at locations that were approximately 100 degrees apart (along the horizontal meridian). These multisensory effects, which were evident in the ERPs at just 50 ms poststimulus, were localized to auditory cortical association areas (see also Gonzalez Andino et al., 2005). Such evidence suggests that multisensory interactions can occur at approximately the same time that initial afferent activity arrives in the primary sensory cortices (see also Foxe et al., 2000; Giard & Peronnet, 1999; Molholm et al., 2002; Schroeder et al., 2004; Wang et al., 2008). Given that auditory facilitation of visual-target detection similarly persists despite wide audiovisual misalignments, it is plausible that its underlying multisensory interactions also occur in the early sensory cortices near the onset of cortical processing.

A Possible Neural Mechanism

Auditory facilitation of visual-target detection despite spatially uninformative (e.g., Van der Burg et al., 2008a) or widely misaligned sounds (see **Figure 3**) suggests that it might be temporal information from the non-target modality, rather than spatial information, that promotes enhanced detection. Several studies have shown that a temporal cue can be used to guide attentional deployment in preparation for a subsequent target, improving reaction times (e.g., Coull & Nobre, 1998; Hackley et al., 2007; Miniussi et al., 1999). Such an anticipatory focusing of attentional resources, however, seemingly cannot explain the multisensory effects observed in our data. Here a short-duration sound (16 ms) was presented simultaneously with a short-duration (16 ms) visual stimulus, not prior to the visual stimulus. It is therefore unlikely that auditory facilitation of visual-target detection resulted from an anticipatory focusing of attentional resources, because the boost in visual processing would need to be more immediate. An alternative hypothesis that better fits with our data has recently been proposed in the neurophysiologic literature: that the auditory facilitation of visual-target detection operates through a modulation of the brain's intrinsic oscillations (Lakatos et al., 2009).

Neurophysiologic recordings in non-human primates have demonstrated that a stimulus in one sensory modality can synchronize neural activity across multiple functionally interconnected cortical regions, including the primary sensory cortices

(Lakatos et al., 2007; Lakatos et al., 2009). This synchronization results from the phase reset of ongoing neural oscillations, which are intrinsic fluctuations in the brain's excitability that have been shown to influence perception (e.g., Busch et al., 2009; Busch et al., 2010; Foxe et al., 1998; Lakatos et al., 2009; Mathewson et al., 2009; Monto et al., 2008; Snyder & Foxe, 2010; VanRullen et al., 2006; VanRullen et al., 2007; Worden et al., 2000). In the context of the present experiment, sounds would be hypothesized to reset the phase of ongoing oscillations in visual cortex, such that co-occurring visual targets align with high-excitability phases. By shifting the baseline of neurons in visual cortex closer to their firing threshold (Kayser et al., 2009; Montemurro et al., 2008; Whittingstall & Logothetis, 2009), such cross-sensory phase reset would be expected to enhance visual-target detection.

Findings from a recent study in our lab provide compelling support for this notion that auditory-facilitation of visual-target detection is attributable to cross-sensory phase reset of ongoing neural oscillations (see **Chapter 2**). Our data demonstrate stimulus-locked periodicity in visual-target detection as a function of when a near-threshold visual target was presented relative to a sound that occurred at the beginning of each trial. Such periodicity in visual-target detection is a predicted outcome of cross-sensory phase reset in visual cortex. That is, if the phase of ongoing oscillations is consistently reset—time-locked to the occurrence of a sound at the beginning of each trial—then high- and low-excitability phases following this consistent reset should align across trials. Whether or not a subsequent visual target is detected should therefore be influenced by where its presentation falls relative to the underlying oscillatory activity (i.e., when it is presented relative to the sound that occurs at the beginning of each trial).

Spatial Constraints and Multisensory Processing

One might ask why it is that multisensory facilitation is observed here for inputs that are so obviously misaligned. Spatial alignment is known to play an important role in other forms of multisensory behavioral facilitation. For example, auditory-driven speeding of saccadic reaction times to a visual target decreases as the spatial disparity between the bisensory stimuli increases (Arndt & Colonius, 2003; Frens et al., 1995). This is probably not surprising given the central role of the superior colliculus in mediating saccades, a midbrain structure in which multisensory integration is well-known to be dependent upon overlapping receptive fields (see Stein et al., 1998). We suggest that the specific constraints on multisensory integration change depending on the neural mechanisms invoked by the demands of the task at hand. Thus when the task requires an eye movement to an easily detectable visual target, key multisensory processes will occur in the superior colliculus and be subject to the constraints therein. Alternatively, in the case of auditory facilitation of near-threshold visual targets, the best way to guarantee that the signal exceeds the noise—and will therefore be detected—might be to amplify initial afferent activity in visual cortex, for example in the manner proposed above (i.e., cross-sensory phase reset). A spatially diffuse model of signal enhancement might well be beneficial here, perhaps increasing the likelihood of detection in cases where the location of the target is not predetermined. Once detection has occurred, a saccade might then be initiated to the location of the near-threshold visual target. Of course various mechanisms of multisensory enhancement need not be mutually exclusive.

Concluding Remarks

Auditory facilitation of visual-target detection occurs regardless of retinal eccentricity and despite wide audiovisual misalignments. The data presented here thus provide important clues about the neuroanatomic substrates underlying auditory facilitation of visual-target detection. Our results suggest that these multisensory effects might arise from interactions in brain regions that have traditionally been considered to be unisensory (in light of Murray et al., 2005), but do not operate via previously described corticocortical pathways between these early sensory cortices (in light of Falchier et al., 2002).

Table 1.

Hit rates and false alarm rates by visual-target location and stimulus type (visual-alone, audiovisual with a centrally presented sound, or audiovisual with a peripherally presented sound)

	LEFT TARGET			**CENTER TARGET**			**RIGHT TARGET**		
	V	**AVC**	**AVP**	**V**	**AVC**	**AVP**	**V**	**AVC**	**AVP**
HR%	70%	80%	82%	67%	75%	74%	67%	80%	76%
FAR%	10%	3%	6%	5%	1%	1%	6%	4%	2%

Notes: V = visual-alone, AVC = visual paired with sound from the central speaker, AVP = visual paired with sound from the peripheral speaker.

Figure 1.

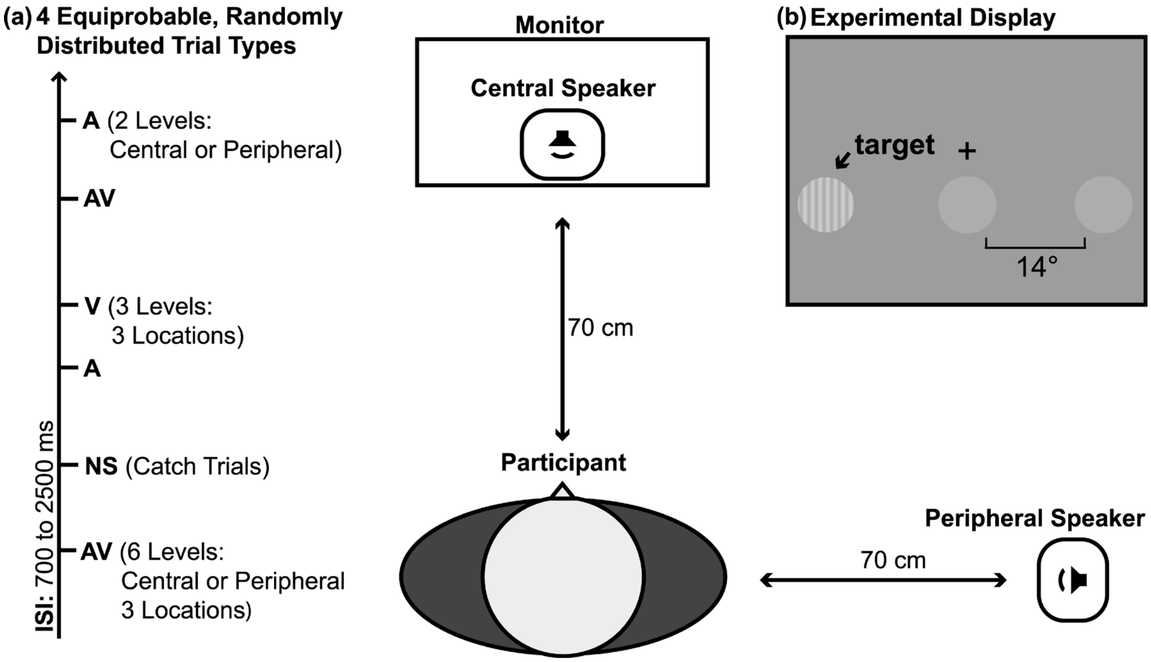

(a) 4 Equiprobable, Randomly Distributed Trial Types
A (2 Levels: Central or Peripheral)
AV
V (3 Levels: 3 Locations)
A
NS (Catch Trials)
AV (6 Levels: Central or Peripheral 3 Locations)
ISI: 700 to 2500 ms
Monitor
Central Speaker
70 cm
Participant
Peripheral Speaker
70 cm
(b) Experimental Display
target
+
14°

Figure 1. A schematic of the experimental design. (a) There were 4 equiprobable trial types: auditory-alone catch trials (A), visual-alone trials (V), no-stimulus catch trials (NS), and audiovisual trials (AV). (b) The experimental display had 3 target locations. Participants were told to respond whenever they saw a visual stimulus (i.e., to press 1 of 3 arrow keys corresponding to the target location).

Figure 2.

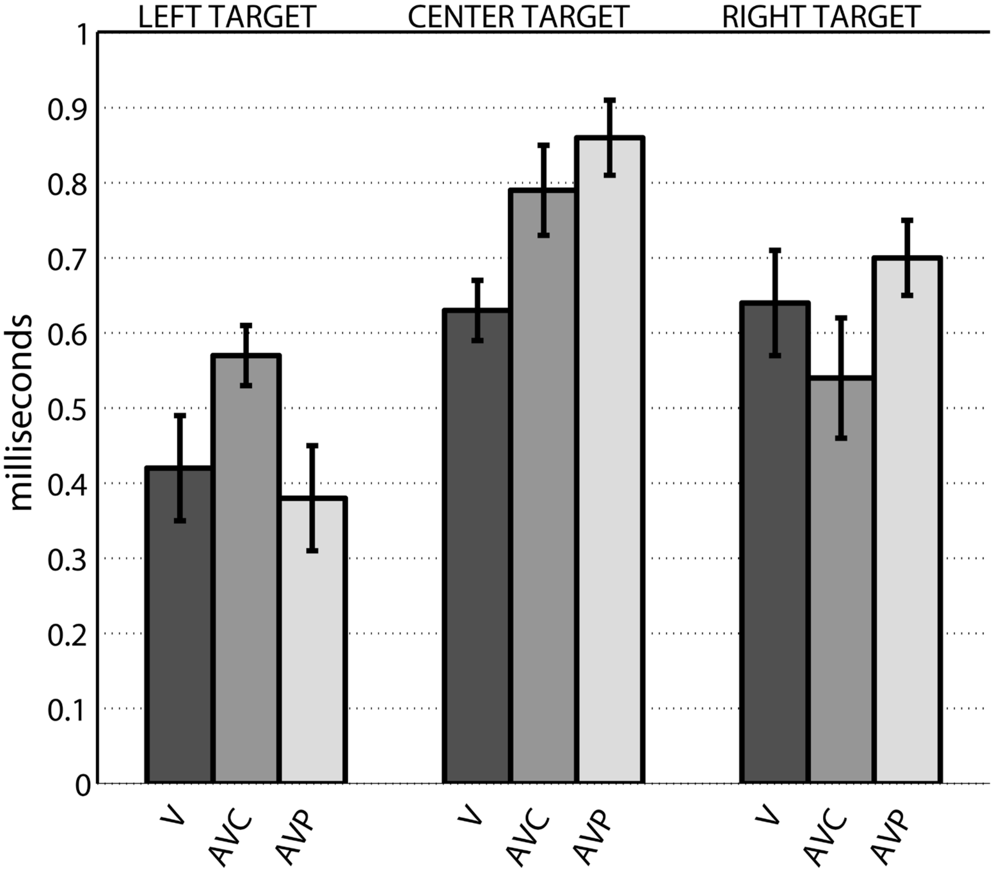

LEFT TARGET
CENTER TARGET
RIGHT TARGET
milliseconds
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
V
AVC
AVP
V
AVC
AVP
V
AVC
AVP

Figure 2. The criterion for reporting occurrences of visual targets at each of the three target locations, sorted by target type: visual-alone (V), visual paired with a sound from the central speaker (AVC), and visual paired with a sound from the peripheral speaker (AVP).

Figure 3.

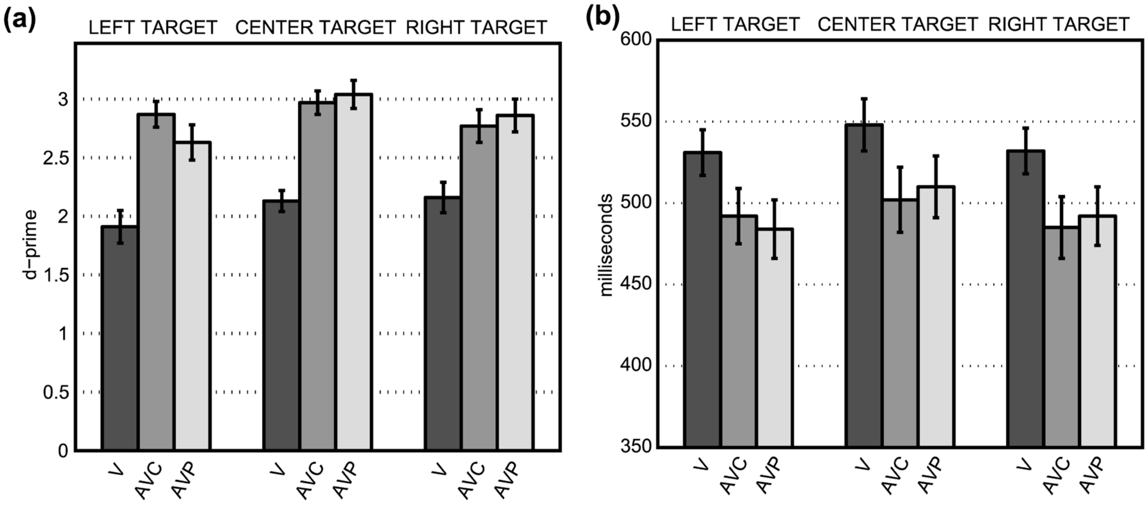

(a)
LEFT TARGET
CENTER TARGET
RIGHT TARGET
d−prime
0
0.5
1
1.5
2
2.5
3
V
AVC
AVP
(b)
LEFT TARGET
CENTER TARGET
RIGHT TARGET
milliseconds
350
400
450
500
550
600
V
AVC
AVP

Figure 3. Auditory facilitation of visual-target detection occurs regardless of retinal eccentricity (right target, center target, left target), and despite wide audiovisual misalignments (as large as 104 degrees). These multisensory effects are demonstrated using both (a) d-prime and (b) reaction times. There were 3 target types: visual-alone (V), visual paired with a sound from the central speaker (AVC), and visual paired with a sound from the peripheral speaker (AVP).

CHAPTER 2

Ready, Set, Reset: Stimulus-Locked Periodicity in Visual-Target Detection Demonstrates the Behavioral Consequences of Cross-sensory Phase Reset

Although sounds have been repeatedly shown to improve the detection of co-occurring visual targets (Frassinetti, Bolognini & Ladavas, 2003; Lippert et al., 2007; McDonald, Teder-Sälejärvi & Hillyard, 2000; Odgaard et al., 2003; Stein et al., 1996; Teder-Sälejärvi et al., 2005; Van der Burg et al., 2008a; Van der Burg et al., 2008b), the mechanisms underlying this multisensory effect are not well-understood. Some researchers have suggested that the multisensory enhancement of visual-target detection operates through a cognitive, attention-related mechanism (e.g., Lippert et al., 2007; Odgaard et al., 2003), whereas others have suggested that these enhancements operate through low-level, multisensory connections between early sensory cortices (e.g., Stein et al., 1996; Van der Burg et al., 2008a). Here we tested the hypothesis that multisensory enhancement of visual-target detection results from an interaction between top-down attention and ongoing neural dynamics (i.e., intrinsic fluctuations in cortical excitability).

There is good reason to suspect the involvement of a top-down attentional mechanism. Consistent with a cognitive account, Lippert et al. (2007) demonstrated auditory facilitation of visual-target detection when sounds were simultaneously presented with visual targets on 50 percent of trials, but not in another condition where sounds were simultaneously presented with visual targets on less than 20 percent of trials. The authors argued that participants in the former condition actively used the sounds to predict the occurrence of visual targets; whereas participants in the latter condition, who were unaware of a relationship between sounds and visual targets, did not. In other words, when sounds reliably predicted the occurrence of visual targets, participants used information from the co-occurring sounds that was unavailable on visual-alone trials to better direct their attentional resources.

These findings fit well with previous observations in the unisensory literature that attention alters the apparent contrast of visual stimuli (Carrasco, Ling & Read, 2004; Carrasco, Fuller & Ling, 2008; Fuller, Rodriguez & Carrasco, 2008; Ling & Carrasco, 2007; Liu, Abrams & Carrasco, 2009; for multisensory evidence see Störmer, McDonald & Hillyard 2009). But studies examining the influence of attention on perceived contrast manipulated the deployment of attention in space, and the multisensory enhancement of visual-target detection occurs even when a paired sound is spatially uninformative (see **Chapter 1**, Van der Burg et al., 2008a; Van der Burg et al., 2008b; Stein et al., 1996; Teder-Sälejärvi et al., 2005). This led us to the notion that it might be temporal information from the non-visual modality, rather than spatial information that promotes enhanced detection. That is, when visual targets are difficult to detect without focused attention, easily perceptible sounds—which sometimes co-occur with visual targets—provide temporal information that can be used to guide attentional deployment (see Coull & Nobre, 1998; Miniussi, Wilding, Coull & Nobre, 1999).

Provided with this temporal information, how does the brain's attentional system then intervene to optimize target detection? Moment-to-moment cortical excitability fluctuates with the phase of ongoing oscillations (e.g., Busch et al., 2009; Busch & VanRullen, 2010; Kayser et al., 2009; Lakatos et al., 2007; Lakatos et al., 2008; Lakatos et al., 2009; Mathewson et al., 2009; Montemurro et al., 2008; Monto et al., 2008;

Whittingstall & Logothetis, 2009). Several studies have now demonstrated that cortical spike timing aligns with the phase of low-frequency oscillations (in the 1–4 Hz frequency range), with action potentials most likely to occur during high-excitability phases (Kayser et al., 2009; Lakatos et al., 2008; Montemurro et al., 2008; Whittingstall & Logothetis, 2009). As a consequence, stimuli that occur during high-excitability phases—when neurons are closer to their firing threshold—are more likely to be detected than stimuli that occur during low-excitability phases. The brain's response to external stimuli is therefore not only a function of stimulus properties (e.g., contrast), but also a function of the brain's internal dynamics. For example, the phase of prestimulus oscillations has been shown to predict the likelihood of subsequent visual-target detection (Busch et al., 2009; Mathewson et al., 2009; Monto et al., 2008).

Recent data from non-human primates reveal that the influence of ongoing neural oscillations on perception in not merely passive, but that selective attention seems to operate in part by modulating ongoing oscillations such that targets arrive during high-excitability phases (Lakatos et al., 2008; Lakatos et al., 2009). In particular, findings from Lakatos et al. (2008) suggest that under conditions where stimulation is rhythmic, attention alters the phase of ongoing oscillations, such that high-excitability phases are aligned with potential targets in the to-be-attended modality (e.g. vision), while low-excitability phases are aligned with distractors in the to-be-ignored modality (e.g., audition). Lakatos et al. (2009) then argued for a second, prerequisite mechanism needed for the entrainment of internal oscillations to an external rhythm. For this second mechanism, referred to as cross-sensory phase reset, attending to a single stimulus synchronizes ongoing oscillation across functionally interconnected cortical regions, including multiple sensory cortices.

We hypothesized that this type of mechanism, through which the presentation of a stimulus in one modality influences baseline cortical excitability in another modality, might underlie the multisensory enhancement of visual-target detection in humans. That is, attending to a temporally informative sound could reset the phase of ongoing oscillations in visual cortex, potentially increasing subthreshold depolarization during stimulus processing and thus the likelihood of visual-target detection. To date, however, the neurophysiologic literature has been unable to link cross-sensory phase reset with its behavioral consequences because significant measurement limitations make it difficult to detect the neurophysiologic signature of phase reset (see Makeig, Debener, Onton & Delorme, 2004; Martinez-Montes et al., 2007; Sauseng et al., 2007; Shah et al., 2004). Simply put, phase alignment across trials might result from either (1) phase reset of ongoing oscillations or (2) a transient sensory response, which is superimposed on ongoing oscillatory activity. In a unisensory context, where, for example, a sound is presented and cross-sensory phase reset is measured in early visual cortices (see Lakatos et al., 2009), there is no measurement limitation because there is no transient sensory response evoked in visual cortex (i.e., there was no visual stimulus). In a multisensory context, on the other hand, where there are transient sensory responses evoked in both auditory and visual cortices (i.e., when a sound and a visual stimulus are simultaneously presented), it becomes difficult to establish whether cross-sensory phase reset occurred in either cortical region. It has thus yet to be proven that cross-sensory phase reset resulting from a stimulus presented in one sensory modality (e.g., a sound) has behavioral consequences for a task in which potential targets are presented in a second sensory modality (e.g., near-threshold visual targets).

If cross-sensory phase reset is indeed a powerful mechanism for sensory selection and perceptual enhancement, then some relatively straight forward behavioral predictions should hold. Here we present the results of an experiment that was designed to bridge the gap between cross-sensory phase reset and behavioral performance. To bypass issues that have complicated the detection of phase reset in neurophysiologic data, we devised a task for which the behavioral results themselves could be used to probe for cross-sensory phase reset (see **Figure 4a**). If attending to a temporally informative sound consistently resets the phase of ongoing oscillations, visual-target detection at subsequent time points should reflect this underlying, stimulus-locked oscillatory activity. We therefore predicted that there would be periodicity in visual-target detection that would persist for seconds following the presentation of a sound. To test the dependence of our results on endogenous factors such as top-down attentional control (Lakatos et al., 2009; Lippert et al., 2007; Odgaard et al., 2003), we also manipulated the probability that a low-contrast visual target would co-occur with the sound (see **Figure 4b**). We predicted that changing the probabilistic context would affect how attention was deployed (whether volitionally or through implicit mechanisms), and therefore influence whether ongoing oscillations were reset such that co-occurring visual targets aligned with high-excitability phases.

RESULTS

Figure 4a provides a schematic of the experimental design. A sound was presented at the beginning of each trial, and a near-threshold visual target was presented either with the sound or at one of twelve timepoints at 500-ms intervals up to 6000 ms after the sound. We thus used hit rates to probe for underlying oscillatory activity, time-locked to a temporally informative sound. We chose a 2-Hz sampling rate to test for the existence of periodicity in behavioral performance because: (1) previous research has suggested that a frequency around approximately 1 Hz (our Nyquist frequency) is typically involved in sensory selection (Lakatos et al., 2007; Lakatos et al., 2008; Schroeder & Lakatos, 2008), and (2) lower-frequency oscillations are known to generally have greater power than higher-frequency oscillations, and their influence on visual-target detection was therefore thought to be more likely to emerge from noise in the behavioral data (Bak, Tang & Wiesenfeld, 1987; Buzsaki & Draguhn, 2004; Van Orden, Holden & Turvey, 2003). While using a 2-Hz sampling rate delimits what one can say about the critical frequencies involved in the multisensory enhancement of visual-target detection, the goal of the present study was simply to establish the existence of periodicity in observed hit rates, not to focus on the specific frequencies that are being reset. If the sound presented at the beginning of each trial consistently reset the phase of ongoing oscillations, it was expected that a periodic pattern of detectability would emerge (see **Figure 4a**).

To test the hypothesis that top-down attentional control is instrumental in the process of phase-resetting (in light of Lippert et al., 2007; Lakatos et al., 2009), the experiment included two conditions (see **Figure 4b**). For the first condition (the "equiprobable" condition), the simultaneous audiovisual stimulus (i.e., visual targets at 0 ms) and visual targets at each of the other twelve time points were equiprobable. For the second condition (the "audiovisual [AV] dominant" condition), the probability of simultaneous audiovisual targets was increased to 20 percent, while the probability of visual targets at each of the other time points was set at 5 percent (i.e., 5 percent multiplied by 12 time points). For both conditions, participants were informed of these probabilities prior to the experiment under the assumption that this knowledge would affect their strategic approach to the task.

Multisensory enhancement of visual-target detection

It is well-established that a simultaneously presented sound can enhance visual-target detection (Frassinetti et al., 2003; Lippert et al., 2007; McDonald et al., 2000; Odgaard et al., 2003; Stein et al., 1996; Teder-Sälejärvi et al., 2005; Van der Burg et al., 2008a; Van der Burg et al., 2008b). Some debate remains, however, regarding the extent to which these multisensory effects are stimulus-driven (e.g., Van der Burg et al., 2008a), or instead operate under top-down attentional control (e.g., Lippert et al., 2007). As a first step in our analysis, we assessed whether (1) our behavioral setup led to auditory facilitation of co-occurring visual targets, and (2) whether auditory facilitation was dependent on the probability of simultaneity (i.e., audiovisual co-occurrence). If the multisensory enhancement of visual-target detection is a bottom-up process that is independent of attentional influences, it should be evident in both experimental conditions. That is, if auditory facilitation is entirely based on the instantaneous properties of the stimulus, it should be observed whenever a visual target is presented with a sound, regardless of the probabilistic context.

To test whether the sound presented at the beginning of each trial enhanced the detection of co-occurring visual targets, we compared hit rates in response to audiovisual targets (i.e., visual targets presented at 0 ms) with those in response to visual-alone targets that were presented 500 ms after the sound. For the equiprobable condition, a two-tailed, paired t-test revealed that hit rates in response to audiovisual targets were actually lower than those in response to visual-alone targets at a nearly significant level (p = 0.06). Rather than facilitating visual-target detection, these data suggest that the sound at the beginning of each trial in the equiprobable condition acted as a distractor. For the AV-dominant condition, on the other hand, a two-tailed, paired t-test revealed a significant multisensory enhancement of visual-target detection (p = 0.03).

To summarize, auditory facilitation of co-occurring visual targets was evident in our second condition, where there was a higher probability of simultaneity (see **Figure 5**), but not in our first condition. These data therefore support earlier findings that a sound's predictive power determines whether or not there is a multisensory enhancement of visual-target detection (Lippert et al., 2007). Because participants were informed of the probabilities prior to the experiment, we propose that their strategic approach to the task, or rather how they deployed their attention, determined how the sound at the beginning of each trial influenced visual-target detection. We discuss this proposal in greater detail in the Discussion.

Although the sound had less predictive power in the equiprobable condition relative to the AV-dominant condition, it still provided temporal information about when participants could begin to expect the occurrence of visual targets. Given that the sound did not provide competing or contradictory information, lower detection of audiovisual targets relative to visual-alone targets (at 500 ms) might seem somewhat surprising. But the present analysis was limited to the first two time points, and therefore does not reveal the whole story. The antipodal results observed here might be better explained in the broader context of phase reset of ongoing oscillations. It is possible, for example, that phase reset occurred in both conditions, but that the exact nature of this phase reset (e.g., the timing of phase reset relative to the sound, or the phase to which oscillations were reset) differed based on the probabilistic context of the task. In the next section we examine visual-target detection across the entire trial duration (i.e., 6 seconds) to test for stimulus-locked periodicity in behavioral performance.

Stimulus-locked periodicity in visual-target detection

Attention is known to increase the apparent contrast of visual stimuli (e.g., Carrasco et al., 2004), and thus it is reasonable to suppose that an attentional mechanism might be involved in the multisensory enhancement of visual-target detection. Further, insofar as periodicity has been hypothesized to be particularly relevant to timekeeping (see Matell & Meck, 2000), if attention is indeed being deployed based on temporal information—as we propose—then cross-sensory phase reset of ongoing oscillations seems an excellent candidate mechanism (Lakatos et al., 2007; Lakatos et al., 2009). Here we probed for evidence of periodicity in visual-target detection at consecutive time points, time-locked to a temporally informative sound.

To test for the presence of periodicity, a bootstrap statistical approach was combined with a two-part sinusoidal model (see Methods for the equations), where the model fit of the observed hit rates for each time point was compared with the model fit of randomly re-sampled data (i.e., hits and misses were randomly re-distributed across the thirteen time points). This procedure was tailored to address the following hypotheses:

(1) that the observed data included a lower-frequency component that resulted from consistent changes in a participant's sustained attention (or vigilance) across the duration of a trial, and (2) that the periodic nature of the observed data was attributable to stimulus-locked oscillatory activity. The model's first sinusoid was constrained to be less than or equal to 0.25 Hz (i.e., 1.5 cycles per 6 seconds). This lower-frequency sinusoid was expected to explain variability associated with consistent changes in sustained attention that occurred within a trial. For example, if a participant's focus consistently waned, such that her performance generally decreased from the first time point (0 ms) to the last time point (6000 ms). The model's second sinusoid was constrained to be greater than 0.25 Hz. This higher-frequency sinusoid was expected to explain variability associated with a physiologic rhythm (i.e., ongoing oscillatory activity), where the phase but not the frequency was influenced by a participant's strategic approach to the task. The amplitudes, phases, and frequencies of these two sinusoids were fit in parallel. To establish whether including the higher-frequency component resulted in a significantly greater proportion of explained variance, we also fit the lower-frequency component of the two-part model by itself (i.e., without the higher-frequency component).

The results of this analysis for the equiprobable condition are shown in **Table 1** (columns 1–3) and **Figure 6**, and for the AV-dominant condition in **Table 2** (columns 1–3) and **Figure 7**. For the equiprobable condition, at the group-level, the full model (i.e., the model with both sinusoids) explained approximately 73 percent of the variability in visual-target detection across the thirteen time points. The lower-frequency component of the model, when fit by itself, explained approximately 33 percent of the variability in visual-target detection, and the addition of the higher-frequency component to the model explained another 40 percent of the variability in visual-target detection. Whereas both the full model ($p < 0.001$) and its lower-frequency component ($p = 0.009$) explained a significantly higher proportion of variance in the observed data than that explained in the randomly re-distributed data, the proportion of variance explained by the higher-frequency component ($p = 0.27$) was not statistically significant.

The pattern of results for the AV-dominant condition, where the probability of simultaneity was higher than the probability of a visual target at each of the other twelve time points, differed from those for the equiprobable condition. At the group-level, the full model explained approximately 75 percent of the variability in visual-target detection across the thirteen time points. The lower-frequency component of the model explained approximately 28 percent of the variability in visual-target detection, and the addition of the higher-frequency component to the model explained another 47 percent of the variability in visual-target detection. Here both the proportion of variance explained by the full model ($p < 0.001$) and its higher-frequency component ($p = 0.02$) were significantly greater for the observed data than for the randomly re-distributed data, while the proportion of variance explained by the lower-frequency component ($p = 0.26$) was not statistically significant.

In addition to the *proportion* of variance explained (i.e., R^2), a second measure of model fit was devised to also consider the *total* amount of variance explained. This second measure was termed the “relevance” of the model. **Figure 8** provides an illustration of why this “relevance” measure provides important information about the model fit. As the amplitude of a sinusoid decreases, it begins to approximate a line. Although such a low-amplitude sinusoid might explain a large proportion of variance, it cannot be taken as evidence of periodicity. To calculate the "relevance" of the model fit, the proportion of variance explained was weighted by the absolute value of the average deviation of the model's predicted values from the predicted mean (see Methods for the equations). As shown in **Tables 1** and **2** (columns 4–6), once the total amount of variance was incorporated into the dependent measure, statistical significance was achieved not only at the group-level for both conditions, but also at the participant-level. On average, the randomly re-distributed data had significantly less variability than the observed data, further supporting the assertion that the observed data included a periodic component.

To further probe for the absence or presence of periodicity, the observed data were also transformed into their frequency-domain representation. The fast Fourier transform (FFT) was used to compare the frequency content of the observed data to that of random-order data. Here the bootstrap statistical approach involved randomly shuffling the observed hit rates across the thirteen time points. Since the frequency spectrum of white noise is flat, any significant spectral peaks in the observed data would provide evidence for a consistent periodic component. **Figure 9** shows the group-level results of this analysis for both experimental conditions. For the equiprobable condition, there were peaks at both 0.17 Hz (1 cycle per 6 seconds) and 1 Hz, but the amplitudes of these peaks were not significantly greater than those of the randomly shuffled data (p = 0.08 and p = 0.14, respectively). In comparison, for the AV-dominant condition, there was a single peak at 1 Hz that was indeed significantly greater than the amplitude of the randomly shuffled data (p = 0.008).

Because of potential aliasing, it is not possible to definitively conclude based on the FFT results that the observed periodic component reflects oscillatory activity at 1 Hz.

Oscillations at 4 Hz and 8 Hz, for example, would also be expected to show peaks (i.e., alias) at 1 Hz with a sampling rate of 2 Hz. It can be concluded, however, that the modeling and FFT results for the AV-dominant condition both provide evidence for stimulus-locked periodicity in the observed hit rates. The modeling results for the equiprobable condition, once the total variance explained is taken into account, also provide evidence—although weaker than the evidence for the AV-dominant condition—for stimulus-locked periodicity. We propose that the emergence of a periodic component across trials reflects the behavioral consequences of cross-sensory phase reset. For the equiprobable condition, there is therefore some evidence of cross-sensory phase reset, but if anything, ongoing neural oscillations are being reset to a low-excitability phase. For the AV-dominant condition, there is strong evidence for cross-sensory phase reset, and ongoing neural oscillations seem to be reset to a high-excitability phase (and auditory facilitation of visual-target detection is therefore observed).

DISCUSSION

The present data strongly support the hypothesis that one of the primary mechanisms driving multisensory enhancement of visual-target detection, which has been observed in a number of behavioral paradigms (Lippert et al., 2007; Odgaard et al., 2003; Stein et al., 1996; Van der Burg et al., 2008a; Van der Burg et al., 2008b), is phase reset of ongoing neural oscillations. Because of measurement limitations (see Makeig et al., 2004; Martinez-Montes et al., 2007; Sauseng et al., 2007; Shah et al., 2004), neurophysiologic research in non-human primates has been limited to describing cross-sensory phase reset in response to unisensory stimulation (e.g., Lakatos et al., 2009).

With the use of a purely behavioral approach we circumvent these issues and provide compelling evidence for cross-sensory phase reset in response to multisensory stimulation. Importantly, these data provide a highly plausible link between this attentional mechanism that has been observed with neurophysiologic recordings (phase reset) and its perceptual consequences. Consistent with the notion that cross-sensory phase reset underlies behavioral facilitation, we found that performance on a visual-target detection task waxed and waned in a periodic fashion, time-locked to a temporally informative sound. These periodic fluctuations in behavior extended beyond five seconds, indicating—at least when attention must be sustained—that cross-sensory phase reset influences perception not only at the moment it occurs but also for seconds thereafter. We attribute this periodicity in behavior to ongoing oscillations continuing to ring at their new phase.

The Role of Endogenous Factors

Our data are consistent with the viewpoint that multisensory enhancement of visual-target detection operates through strategic top-down processes (e.g., Lippert et al., 2007). In the present experiment, auditory facilitation was observed only when participants knew that the visual target was more likely to co-occur with the sound than at any of the other time points (i.e., during the AV-dominant condition) up to six seconds after the sound (see **Figure 5**). Given that participants were informed of the probabilistic context prior to each experimental session, and that the difficult task required highly focused attention, we believe that explicit knowledge may have played a role in how participants deployed their attentional resources. For example, in the AV-dominant condition, the optimal approach to the task was to deploy attentional resources immediately following the button press that initiated each trial. Such anticipatory attention was useful because visual targets were most likely to co-occur with the sound, but the exact timing of the sound following the button press was unpredictable (i.e., the sound occurred 1–3 seconds after the button press). For the equiprobable condition, on the other hand, there was no such optimal approach. As a result, attentional deployment in time relative to the sound was probably less consistent, both across participants and across trials. Because visual targets were just as likely to co-occur with the sound as they were to occur at any of the other 12 timepoints (up to six seconds later), participants in the equiprobable condition had far less impetus to deploy their attention in anticipation of the sound than did participants in the AV-dominant condition. Instead participants in the equiprobable condition might have more often used the sound like a temporal cue to begin attending for the possible occurrence of a visual target, which would make anticipatory attentional deployment weaker in the equiprobable condition than it was in the AV-dominant condition. We propose that it was such differences in attentional deployment across the conditions that led to differences in the observed results. First, a more consistent behavioral approach might explain why evidence of periodicity in visual-target detection was stronger for the AV-dominant condition than for the equiprobable condition. That is, the proportion of variance explained by the higher-frequency component of the two-part sinusoidal model was significant for the AV-dominant condition, even before it was weighted by the total variance explained (see **Figure 6**); whereas the model fit of the higher-frequency component for the equiprobable condition was only significant when both the proportion and the total variance explained were considered together (i.e., as the model relevance; see **Figure 7**). Second, because auditory facilitation of visual-target detection was only observed for the AV-dominant condition, it might be that anticipatory attentional deployment is a prerequisite for cross-sensory phase reset to a high-excitability phase.

Possible support for our proposal that attentional resources were deployed differently across the two conditions arises from the FFT results (see **Figure 9**). The frequency spectrum for the equiprobable condition includes a nearly significant peak at 0.17 Hz ($p = 0.08$) that is entirely absent in the AV-dominant condition. This peak might reflect the ramping up of sustained attention (or vigilance) following the sound that occurs near the beginning of each trial. In other words, the typical strategic approach for the equiprobable condition might best be described by a Gaussian resembling a 0.17-Hz sinusoid, with a crest that occurs either at the beginning, middle, or end of the trial (i.e., a sinusoid with one complete cycle over the six-second period of a trial). In comparison, in the AV-dominant condition, where we suspect that anticipatory attentional deployment was stronger, changes in sustained attention across the duration of a trial might be best described by an even lower-frequency sinusoid, time-locked to the button press that initiated each trial. However, here a frequency less than 0.17 Hz would not be apparent in our FFT results, as 0.17 Hz is the lower limit of our frequency resolution.

If multisensory enhancement of visual-target detection indeed operates through strategic top-down processes, then its underlying neural mechanism should be similarly susceptible to endogenous factors. In support of this prediction, Lakatos et al. (2009) observed phase reset in visual cortex in response to an auditory-alone stimulus during an attend-auditory condition, but no such phase reset was observed in response to the same auditory-alone stimulus during an attend-visual condition. These results demonstrated that in a unisensory context, cross-sensory phase reset only occurs in response to stimuli presented in the to-be-attended modality. In the present experiment, we provide evidence that manipulating the probability of simultaneity—and presumably how participants' deployed their attentional resources—determined whether a co-occurring visual target aligned with a high-excitability phase. As shown in **Figures 6** and **7**, the hit rate in response to co-occurring visual targets (at 0 ms) was at a peak of the modeled sinusoidal cycle for just one participant in the equiprobable condition, but seven (of eight) participants in the AV-dominant condition. Thus the specifics of when and how cross-sensory phase reset occurred may have hinged not only on whether the sound at the beginning of each trial was attended or ignored (Lakatos et al., 2009), but also on a participant's strategic approach to the task.

It is also possible, however, that the probability of audiovisual simultaneity affected how attention was deployed in a purely implicit manner. In this case the participants' awareness of the probabilities was not the governing factor, but rather computations of probability were automatically performed and served to affect the manner in which attention was deployed. Further investigation will be required to test whether explicit knowledge of the weighted probability of co-occurrence—and the subsequent overt strategic modulation—is a prerequisite for auditory facilitation of visual-target detection.

A Possible Anatomic Substrate

The dependence of multisensory enhancement of visual-target detection (and cross-sensory phase reset) on endogenous factors indicates a role for the so-called frontoparietal attentional network, which has been repeatedly associated with goal-oriented behavior and top-down attention (for a review, see Corbetta, Patel & Shulman, 2008). Activation of the frontoparietal network during sustained attention might alter the neurophysiologic properties of neurons in sensory cortices, such that incoming information from the sensory periphery resets the phase of ongoing oscillations. Given the supramodal nature of cross-sensory phase reset, however, the influence of frontoparietal regions on sensory cortices might be mediated indirectly via the thalamus—the brain's sensory hub. Anatomic data suggest that thalamocortical projections are well-situated to synchronize subthreshold oscillatory activity across multiple cortical regions. In addition to the frequently described "core" projections from the sensory nuclei of the thalamus to layer IV of sensory cortices, the thalamus is known to have less-specific, more-diffuse cortical projections (Jones, 1998). These "matrix" neurons from the thalamus project to superficial layers of the sensory cortices, and are thus thought to modulate ongoing processing. The thalamic reticular nucleus receives projections from many cortical regions, including the pre-frontal cortex (Zikopoulos & Barbas, 2007), and is therefore particularly likely to have a role in widespread cortical modulations. The reticular nucleus itself sends inhibitory projections to various thalamic nuclei, including the primary sensory nuclei (e.g., the lateral geniculate nucleus). Perhaps in this way, processing in prefrontal cortex associated with sustained attention (Coull, Frith, Frackowiak & Grasby, 1996; Manly & Robertson, 1997) disinhibits diffuse thalamocortical projections. When a stimulus occurs, these disinhibited thalamocortical projections might then reset the phase of ongoing oscillations across multiple sensory cortices (see also Lakatos et al. 2007; Lakatos et al., 2009).

Phase Reset and Oscillatory Entrainment

Previous evidence has shown that the brain's internal oscillations can entrain to an environmental rhythm (Lakatos et al., 2008), and phase reset has been hypothesized to play a key role in this process. For entrainment to occur there must at least be an initial stimulus-to-phase alignment, as well as some form of continuous updating to maintain this alignment. Cross-sensory phase reset, which synchronizes ongoing neural activity across multiple cortical regions, is the most obvious candidate mechanism for such an alignment (see Lakatos et al., 2009). In the present experiment we demonstrate that cross-sensory phase reset can operate in the absence of an established rhythm, which adds credence to the notion that entrainment is simply continual phase reset in response to each stimulus in the to-be-attended modality (see also Lakatos et al., 2009). This definition of entrainment describes a mechanism with a great degree of flexibility, where the brain’s oscillations can be made to align with stimuli that don’t necessarily follow a metronomic beat (i.e., more complex rhythms).

An important question that requires further investigation is whether the influence of external stimulation on ongoing neural dynamics is limited to phase modulation (i.e., phase reset), or alternatively, whether external stimulation can also modulate the dominant frequencies of the brain's intrinsic oscillations? The former viewpoint implies that phase reset in response to each stimulus in the to-be-attended modality occurs within delimited frequency bands, which are likely determined by membrane and system properties, unaltered by environmental rhythms. If this were the case, then perfect entrainment, where targets in an attended modality align with high-excitability phases and distractors in an unattended modality align with low-excitability phases (e.g., Lakatos et al., 2008), would only be observed if the alternating audio-to-visual stimulation happened to occur at an intrinsic and functionally relevant brain frequency. That is, phase reset seemingly does not occur in an unattended modality (Lakatos et al., 2009), so for potential distractors to align with low-excitability phases the alternating audio-to-visual stimulation must match a naturally occurring brain rhythm (e.g., a low-delta oscillation). Potential distractors align with low-excitability phases because oscillations are reset in response to the preceding stimulus, which occurred in the to-be-attended modality, and are continuing to ring at their new phase.

To test the hypothesis that entrainment is indeed restricted to delimited frequency bands, our lab recently made electrocorticographic (ECoG) recordings over auditory cortex while participants (patients with epilepsy undergoing evaluation) engaged in an auditory-visual intermodal selective-attention task (Gomez-Ramirez et al., 2010). Preliminary analysis of these data indicates that even at a slower stimulation rate of 0.66 Hz, so-called entrainment still occurs at a low-delta oscillation (1.3 Hz), such that an oscillatory cycle intervenes between stimuli. This is in line with the view that the brain's ability to entrain to external rhythms is somewhat limited to the frequencies of its own intrinsic oscillations. What's more, in line with evidence from both the present experiment and previous studies (Lakatos et al., 2007; Lakatos et al., 2008; Schroeder &

Lakatos, 2008), these data suggest that low-delta oscillations are particularly important for sensory selection through phase modulation.

Concluding remarks

Ongoing oscillations shape our perception (e.g., Busch et al., 2009; Busch et al., 2010; Foxe et al., 1998; Lakatos et al., 2009; Mathewson et al., 2009; Monto et al., 2008; Snyder & Foxe, 2010; VanRullen et al., 2006; VanRullen et al., 2007; Worden et al., 2000), and attentional modulation of phase prevents this influence on perception from being random (Lakatos et al., 2009). Here we probed for evidence of cross-sensory phase reset in the context of a visual-target detection task, where phase reset in response to a temporally informative sound was hypothesized to increase the likelihood of detecting low-contrast visual stimuli; however, we suspect that the influence of cross-sensory phase reset on perception extends far beyond this context. For example, by synchronizing subthreshold neural activity across cortical regions, cross-sensory phase reset in response to a single attended stimulus might serve to select an object's co-occurring multisensory features for further processing (see Fiebelkorn, Foxe & Molholm, 2010a; Fiebelkorn, Foxe, Schwartz & Molholm, 2010b). In a complex environment, cross-sensory phase reset might also aid in attentional capture when a sound co-occurs with a sudden change in the visual scene (see Vander Burg et al., 2008a). In both these cases, the stimuli might be easily seen or heard—unlike the present experiment, where we used near-threshold visual stimuli—but cross-sensory phase reset still serves as a tool for attentional selection.

Although auditory facilitation of visual-target detection has been shown to occur even when the sound provides no information about where a visual target occurred (e.g., Van der Burg et al., 2008a), its underlying neural mechanism might be able to operate with some degree of spatial specificity (in a context where space is also a relevant feature). In Lakatos et al. (2007), for example, a somatosensory stimulus administered to the right hand reset the phase of ongoing oscillations in left auditory cortex such that afferent information arrived during a high-excitability phase. In comparison, the same somatosensory stimulus reset the phase of ongoing oscillations in right auditory cortex such that afferent information arrived during a low-excitability phase. Future research will be needed to determine whether such spatially specific phase reset also occurs in visual cortex. If it does, it might be responsible for enhanced visual-target detection in spatial cueing paradigms (e.g., Carrasco et al., 2004; Kelly, Gomez-Ramirez & Foxe, 2008; Störmer et al., 2009); particularly in studies that use a fixed cue-target interval, where the cue provides both temporal and spatial information about subsequent targets.

METHODS

Participants

Eight neurologically normal volunteers participated in each condition of the experiment (mean age 30.3 ± 6.1 years; 5 female; 3 left-handed), and data from all subjects were included in the analyses. Two participants completed both conditions, and the remaining 12 subjects completed one condition each. The Institutional Review Board of the City College of CUNY approved the experimental procedures. Written informed consent was obtained from all subjects prior to data collection, in line with the Declaration of Helsinki.

Stimuli and task

Participants, who were positioned 70 cm in front of a 21-inch cathode-ray tube (CRT) computer monitor, were asked to maintain central fixation and report the occurrence of visual stimuli: sine-wave gratings with 16-ms duration, subtending 2.5 degrees of the visual angle in both the vertical and horizontal planes. The visual stimuli were presented two degrees below central fixation, and sounds (1000 Hz tone, 16 ms) were presented from a BOSE (Companion 2) speaker positioned directly below the computer monitor. After a participant clicked the right mouse button to begin a trial, there was a 1–3 second prestimulus interval, followed by a clearly audible sound (~75 dB SPL). A visual stimulus sometimes co-occurred with the sound (at 0 ms), or at one of twelve evenly spaced time points following the sound: 500 ms, 1000 ms, 1500 ms, 2000 ms, 2500 ms, 3000 ms, 3500 ms, 4000 ms, 4500 ms, 5000 ms, 5500 ms, or 6000 ms (**Figure 4a**). Catch trials (~20 percent), where there was a sound at the beginning of the trial but no visual stimulus throughout the duration of the trial, were included to estimate false alarms. Participants were directed to click the left mouse button whenever they detected a visual stimulus, regardless of when it occurred. For the equiprobable condition (**Figure 4b**), the visual stimulus was equally likely to occur at any of the time points, including simultaneous with the sound (i.e., at 0 ms). For the AV-dominant condition (**Figure 4b**), the probability of audiovisual simultaneity was increased: the visual stimulus was 20 percent likely to co-occur with the sound (i.e., at 0 ms) and 5 percent likely to occur at each of the other time points (up to 6000 ms after the sound). Participants were informed of the probabilities prior to each experimental session. Trials ended either when the participant responded or 7200 ms after the sound was presented (giving participants enough time to respond to a visual stimulus that occurred at 6000 ms). Following each trial, participants were presented with a pause screen and had to click the right mouse button to begin the next trial. An SR Research EyeLink eye tracker was used to discard all trials with blinks or eye movements. After every 20 trials, participants were updated on their behavioral performance.

For six participants from the equiprobable condition, complete datasets were collected for three different contrast levels (3 x 80 trials for each time point); however, this approach proved to be overly intensive. The remaining participants completed a total of 80 trials for each time point with the average visual-target detection across all 13 time points held below 50 percent. For the AV-dominant condition, participants completed 320 multisensory trials and 80 trials for every other time point. To maintain vigilance, participants rarely completed more than 3 blocks of the experiment on a given day (each block took approximately 12 minutes to complete), and all participants were advised to stop the experiment if they were unable to focus on the task at hand. Overall, each participant completed 12 blocks.

Statistical Bootstrapping Procedures

Our aim was to determine if the observed data included a periodic component that would be indicative of cross-sensory phase reset. To do this we looked at the data in two general ways: (1) least-squares fitting of sinusoidal models to the data, and (2) spectral analysis of the data using the fast Fourier transform (FFT). We chose to use a non-parametric bootstrapping procedure to quantify statistical confidence. Non-parametric statistics are robust to violations of assumptions for parametric statistical tests, such as normally-distributed sampling distributions and homogeneity of variance (a more traditional, parametric approach, such as grand averaging would not be appropriate in the context of the present experiment, because such an approach would assume that the frequency and phase of oscillatory activity was the same across participants). For the bootstrapping procedure, each observed *test statistic* of interest is compared to a *reference distribution* of statistics that is derived by iteratively randomizing the original data. We used a one-tailed threshold of $p<0.05$ to define significance. The p-value for a bootstrap test is the proportion of values in the reference distribution that exceed the test statistic (i.e., the observed value from our collected data). Since we wished to evaluate the significance of the temporal pattern of performance, we randomized performance values (i.e., hit rates) across the 13 time points to create our reference distributions. We performed this randomization differently depending on which test statistic was being evaluated. In each case, we chose the more conservative randomization procedure. In the following sections, the applicable test statistics and randomization procedures are described.

Data Modeling

Using an iterative MATLAB algorithm (*lsqcurvefit*), we fit the following model equation to the observed data in a linear least-squares sense:

$$\hat{\boldsymbol{y}}_{full} = A_{lf}e^{(i\omega_{lf}\boldsymbol{x}+\varphi_{lf})} + A_{hf}e^{(i\omega_{hf}\boldsymbol{x}+\varphi_{hf})} + c, \qquad |\omega_{lf}| \leq 0.25, \qquad |\omega_{hf}| > 0.25 \tag{1}$$

Where $\boldsymbol{x}$ is a vector of sample time points in seconds, ω is a frequency in Hertz, φ is a phase angle in radians, A is an amplitude in performance units (i.e., percentage), c is a constant in performance units, and $\hat{\boldsymbol{y}}$ is a vector of modeled values in performance units. The subscripts lf and hf designate lower-frequency and higher-frequency sinusoidal components of the model, respectively. The lower-frequency component was included to model performance variability that could be attributed to consistent within-trial fluctuations in a participant's sustained attention, whereas the higher-frequency component was included to model the aftereffects of phase reset. Note that the sinusoidal nature of equation (1) can be made explicit by substitution by Euler's formula. That is, equation (1) is equivalent to:

$$\begin{aligned}\hat{\boldsymbol{y}}_{full} = {} & a_{lf}\cos(\omega_{lf}\boldsymbol{x}) + b_{lf}\cdot i\sin(\omega_{lf}\boldsymbol{x}) + a_{hf}\cos(\omega_{hf}\boldsymbol{x}) + b_{hf}\cdot i\sin(\omega_{hf}\boldsymbol{x}) \\ & + c\end{aligned} \tag{2}$$

All free parameters of the model were fit in parallel. The iterative curve fitting algorithm is dependent on initial conditions and can potentially return a solution that is locally optimized despite the existence of a better solution far from the local minimum. To avoid such sub-optimal local solutions, we repeated the fitting algorithm 10 times, with randomly-selected initial values for each parameter. We tested several quantities of such repetitions and found that 10 repetitions were sufficient to consistently identify the best-fitting model. This procedure had the additional advantage of removing initial condition bias from the fitting procedure.

After determining the best-fitting model parameters, we calculated the proportion of variance-accounted-for (R^2) for both the full model [$\hat{\boldsymbol{y}}_{full}$, eq. (1)] as well as a model characterized by just the constant term and the low-frequency sinusoidal component:

$$\hat{\boldsymbol{y}}_{lf} = A_{lf} e^{(i\omega_{lf}\boldsymbol{x}+\varphi_{lf})} + c \tag{3}$$

Note that we independently fit equation (3) to the observed data. The proportions of variance-accounted-for were calculated as follows:

$$R^2_{full} = \frac{\Sigma(\hat{\boldsymbol{y}}_{full} - \boldsymbol{y})^2}{\Sigma(\boldsymbol{y} - \bar{y})^2} \tag{4}$$

$$R^2_{lf} = \frac{\Sigma(\hat{\boldsymbol{y}}_{lf} - \boldsymbol{y})^2}{\Sigma(\boldsymbol{y} - \bar{y})^2} \tag{5}$$

Where $\boldsymbol{y}$ is the vector of observed performance values, and $\bar{y}$ is the scalar mean of this vector. We took the difference of these R^2 values to describe the increase in the proportion of variance-accounted-for by including the higher-frequency sinusoidal component (ΔR^2_{hf}):

$$\Delta R_{hf}^2 = R_{full}^2 - R_{lf}^2 \tag{6}$$

Not only were we interested in quantifying the quality of the fit of the higher-frequency sinusoid to the data (i.e., ΔR_{hf}^2), but we moreover wished to know the absolute quantity of variance described by this term of the model. To assess this, we multiplied each model's R^2 by that model's total variance:

$$Rel_{full} = R_{full}^2 \frac{\sum(\hat{\boldsymbol{y}}_{full} - \bar{\hat{y}}_{full})^2}{N} \tag{7}$$

$$Rel_{lf} = R_{lf}^2 \frac{\sum(\hat{\boldsymbol{y}}_{lf} - \bar{\hat{y}}_{lf})^2}{N} \tag{8}$$

Where $\bar{\hat{y}}$ is the mean of model values and N is the number of sample points. We termed this product the *relevance* of the model. We examined the change in *relevance* due to the inclusion of the higher frequency sinusoidal component to the model (ΔRel_{hf}):

$$\Delta Rel_{hf} = Rel_{full} - Rel_{lf} \tag{9}$$

Statistical Bootstrapping of Model Fit Statistics

The test statistics for this analysis were the R^2 values (eqs. 4, 5 & 6) and *relevance* values (eqs. 7, 8 & 9). Reference distributions were created by randomly assigning hits and misses to the 13 time points. The total quantity of hits and misses used for each dataset equaled the average quantity of hits and misses across participants (for the equiprobable condition, hits: N = 357, misses: N=517; for the AV-dominant condition, hits: N=535, misses: N=680). For condition 2, the 0 ms SOA time point contained four times as many targets as each of the other time points, and was therefore assigned the concomitant number of scores for the bootstrapped datasets.

Individual participant statistics

We generated 10^3 bootstrapped datasets by randomly assigning hits and misses to the sampled time points. For each of the bootstrapped datasets, we determined the best-fitting model parameters using the method described in section *Data Modeling*, above. The procedure used to fit models to the bootstrapped data was identical in every way to those used to fit models to the observed data. After fitting the models to the bootstrapped data, we calculated *relevance* and R^2 values for the model terms to create reference distributions.

Group-level statistics

The reference distribution for the group-level analysis was created by drawing 10^3 8-member samples from the individual bootstrap distributions of R^2 and *relevance* values to create a distribution of randomized "group-level" averages. The test statistics were the actual observed group averages for R^2 and *relevance* values.

Statistical Bootstrapping of Frequency Spectra

We tested for frequencies with significantly large amplitudes. The test statistics were thus the amplitudes of each frequency. For the randomization procedure, we shuffled the time points at which performance was measured for each subject while keeping the performance values themselves unchanged. This shuffling procedure preserves the total integrated amplitude of the Fourier-transformed data, which aids in the interpretation of the statistical outcome. That is, we were less interested in testing the total variability of the data than in testing the importance of the temporal order of the data.

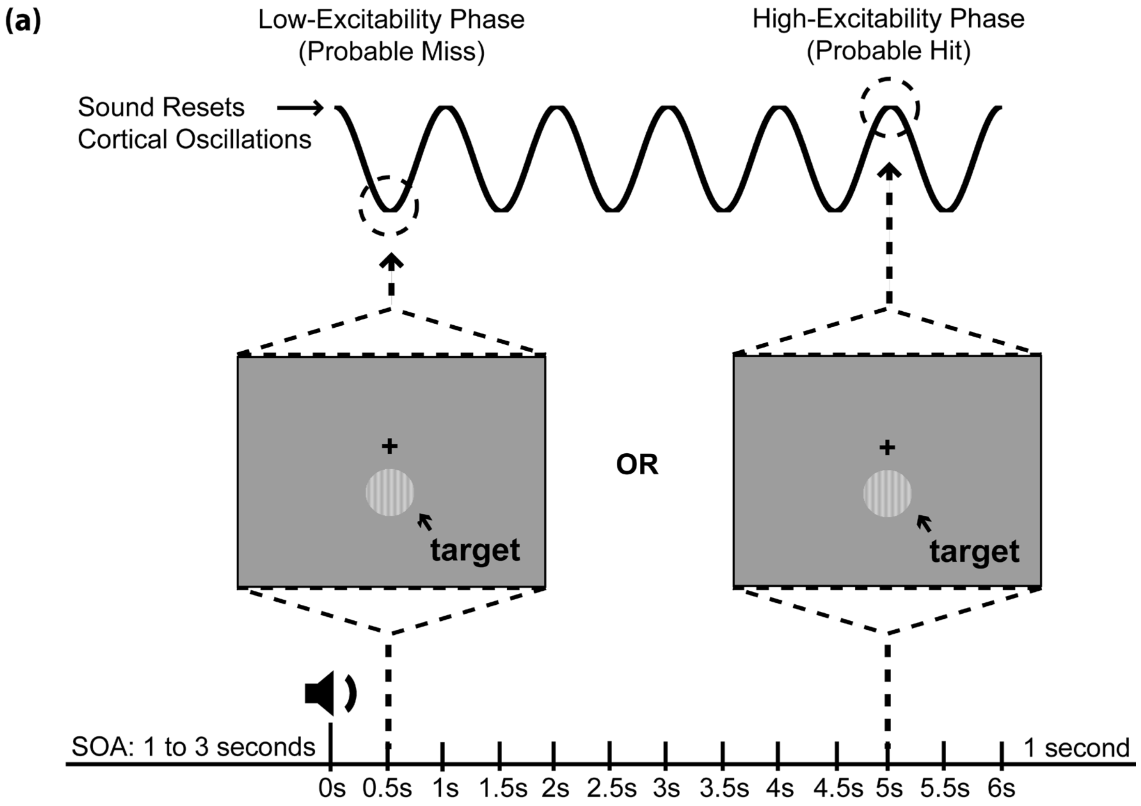
Individual participant statistics

For each participant we created $5x10^3$ bootstrapped datasets by randomly shuffling the temporal order of observed performance values. Each bootstrapped dataset was Fourier-transformed to yield a reference distribution of amplitude spectra. Resultant frequencies ranged from 0-1 Hz with 0.16 Hz resolution. We ignored the 0 Hz component which was always identical to that of the observed data due to the nature of the shuffling procedure.

Group-level statistics

The participant-level bootstrapping procedure identifies significant spectral peaks. We were also interested whether there was group-level commonality regarding peak frequencies. For this analysis, the reference distribution was created by drawing $5x10^3$ 8-member samples from the individual bootstrap distributions to create a distribution of randomized “group-level” averages.

Figure 4.



(b)

Probabilities	0s	0.5s	1s	1.5s	2s	2.5s	3s	3.5s	4s	4.5	5s	5.5s	6s	CT
Equiprobable:	1/17	1/17	1/17	1/17	1/17	1/17	1/17	1/17	1/17	1/17	1/17	1/17	1/17	4/17
AV-dominant:	4/20	1/20	1/20	1/20	1/20	1/20	1/20	1/20	1/20	1/20	1/20	1/20	1/20	4/20

Figure 4. A schematic of the experimental design. (a) On each trial, a 1–3 second SOA was followed by a 1000-Hz tone. A visual stimulus was presented either simultaneously with the tone, or at 1 of 12 other evenly-spaced timepoints after the tone (0.5s, 1s, 1.5s...6s). Catch trials (CT) were also included to estimate false alarm rates. Participants were told to respond whenever they saw a visual stimulus. It was hypothesized that consistent phase reset of underlying neural activity (relative to the tone) would be revealed through periodicity in behavioral performance. (b) For the equiprobable condition, the visual stimulus was equally likely to occur with the tone as it was to occur at each of the other 12 time points. For the audiovisual (AV) dominant condition, the visual stimulus was 4 times more likely to occur with the tone as it was to occur at each of the other 12 time points.

Figure 5.

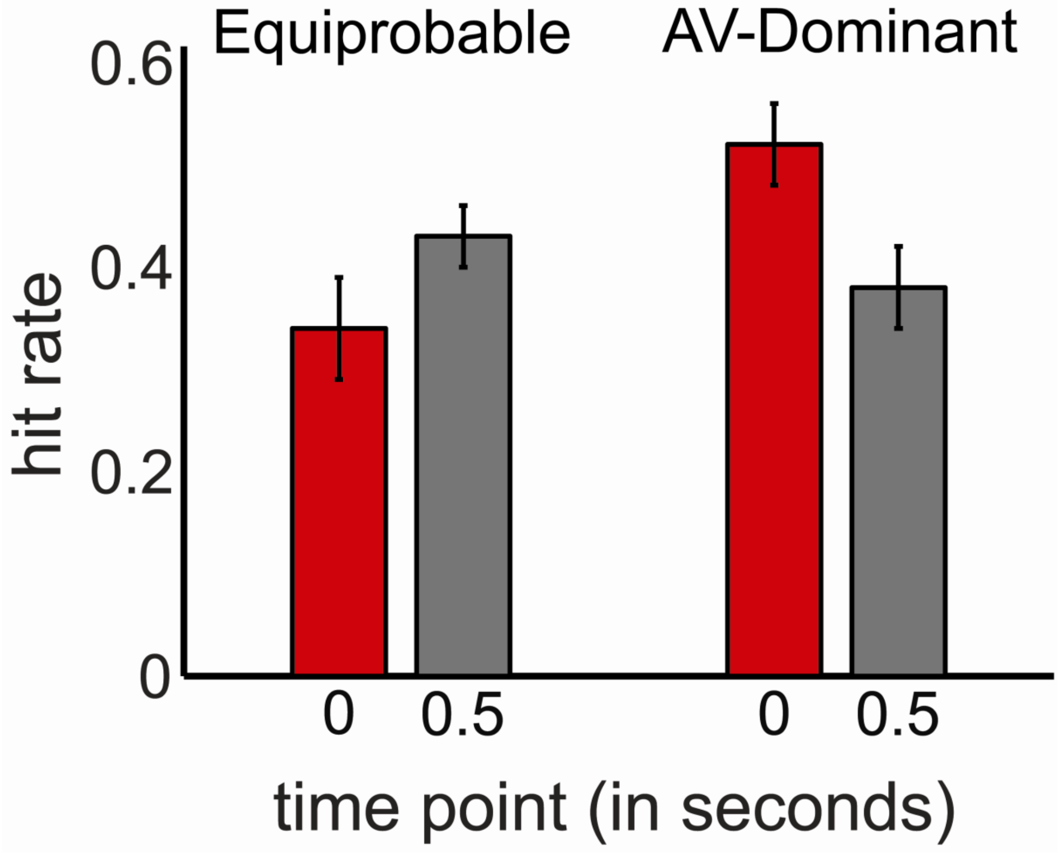

Equiprobable
AV-Dominant
hit rate
0.6
0.4
0.2
0
0
0.5
0
0.5
time point (in seconds)

Figure 5. Auditory facilitation of visual-target detection. Significant auditory facilitation of visual-target detection was observed in the audiovisual (AV) dominant condition (p = 0.03). For the equiprobable condition, on the other hand, there was a nearly significant (p = 0.06) decrease in the detection of visual targets that co-occurred with the sound, relative to visual targets that occurred 0.5 seconds after the sound.

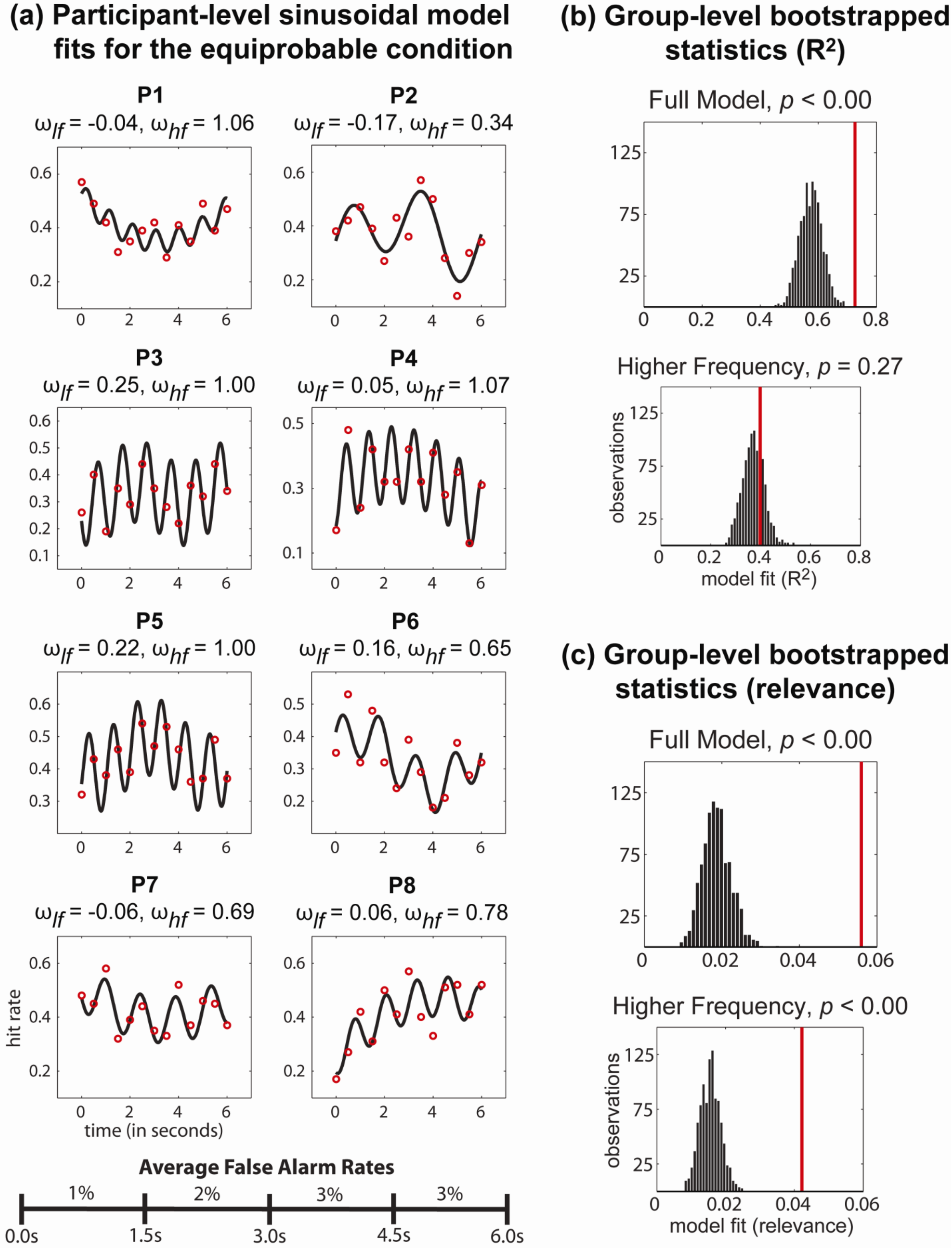
Table 2.

Model fit results for the equiprobable condition.

Participant	**R^2 (*full*) (1)**	**R^2 (*lf*) (2)**	**R^2 (*hf*) (3)**	**Rel (*full*) (4)**	**Rel (*lf*) (5)**	**Rel (*hf*) (6)**
P1	0.72	0.54	0.17	0.038	0.022	0.016
	(0.11)	(0.02)	(0.93)	(0.06)	(0.01)	(0.32)
P2	0.77	0.20	0.58	0.089	0.006	(0.084)
	(0.05)	(0.44)	(0.07)	(0.000)	(0.31)	(0.000)
P3	0.62	0.17	0.45	0.027	0.002	0.025
	(0.34)	(0.54)	(0.25)	(0.17)	(0.42)	(0.13)
P4	0.89	0.20	0.68	0.095	0.005	0.090
	(0.002)	(0.42)	(0.009)	(0.000)	(0.19)	(0.000)
P5	0.69	0.39	0.30	0.027	0.009	0.018
	(0.17)	(0.11)	(0.72)	(0.18)	(0.09)	(0.296)
P6	0.68	0.40	0.28	0.055	0.019	0.036
	(0.19)	(0.09)	(0.77)	(0.007)	(0.01)	(0.03)
P7	0.74	0.24	0.50	0.039	0.004	0.035
	(0.08)	(0.36)	(0.15)	(0.05)	(0.242)	(0.036)
P8	0.71	0.50	0.21	0.079	0.038	0.041
	(0.13)	(0.03)	(0.92)	(0.001)	(0.001)	(0.02)
All	0.73	0.33	0.40	0.055	0.011	0.044
	(0.000)	(0.009)	(0.27)	(0.000)	(0.000)	(0.000)

Notes. *p*-values are in parentheses, *full* = full model (i.e., it includes both sinusoids), *lf* = lower frequency (i.e., $\leq$ 0.25 Hz), *hf* = higher frequency (i.e., > 0.25 Hz).

Figure 6.



(a) Participant-level sinusoidal model fits for the equiprobable condition

(b) Group-level bootstrapped statistics (R^2)

(c) Group-level bootstrapped statistics (relevance)

Figure 6. Modeling results for the equiprobable condition. (a) Each participant's data were fit with a two-part sinusoidal model. The red dots represent the observed data and the black lines represent the model predictions. The frequencies of the two sinusoids that the model fit to the observed data are shown above each participant's plot. Average false alarm rates across 4 time windows show that participants were not biased to respond more at the beginning of each trial. (b) The proportion of variance (R^2) explained by the two-part sinusoidal model for the observed data was compared to that explained for randomly re-distributed data (see also **Table 2**). The red lines represent where the model fit for our observed data falls relative to the bootstrap distributions. (c) An additional measure of model fit that incorporated the total amount of variance explained by the model (i.e., the model relevance) was used to further compare the observed data to randomly re-distributed data (see also **Table 2** and **Figure 8**).

Table 3.

Model fit results for the audiovisual (AV) dominant condition

Participant	**R^2 (*full*)** **(1)**	**R^2 (*lf*)** **(2)**	**R^2 (*hf*)** **(3)**	**Rel (*full*)** **(4)**	**Rel (*lf*)** **(5)**	**Rel (*hf*)** **(6)**
P1	0.68 (0.31)	0.14 (0.70)	0.55 (0.14)	0.035 (0.03)	0.001 (0.52)	0.034 (0.02)
P2	0.89 (0.03)	0.55 (0.06)	0.33 (0.52)	0.071 (0.000)	0.028 (0.001)	(0.043) (0.008)
P3	0.66 (0.35)	0.17 (0.62)	0.49 (0.23)	0.073 (0.000)	0.005 (0.19)	0.068 (0.000)
P4	0.85 (0.04)	0.25 (0.44)	0.60 (0.07)	0.058 (0.003)	0.005 (0.19)	0.053 (0.002)
P5	0.69 (0.29)	0.23 (0.50)	0.46 (0.27)	0.050 (0.006)	0.005 (0.17)	0.045 (0.006)
P6	0.81 (0.09)	0.36 (0.23)	0.45 (0.29)	0.053 (0.005)	0.010 (0.05)	0.042 (0.008)
P7	0.78 (0.13)	0.41 (0.16)	0.37 (0.47)	0.032 (0.03)	0.009 (0.08)	0.023 (0.06)
P8	0.66 (0.36)	0.15 (0.68)	0.51 (0.19)	0.038 (0.02)	0.002 (0.44)	0.036 (0.01)
All	0.75 (0.000)	0.28 (0.26)	0.47 (0.02)	0.051 (0.000)	0.008 (0.005)	0.043 (0.000)

Notes. p-values are in parentheses, *full* = full model (i.e., it includes both sinusoids), *lf* = lower frequency (i.e., $\leq$ 0.25 Hz), *hf* = higher frequency (i.e., > 0.25 Hz).

Figure 7.

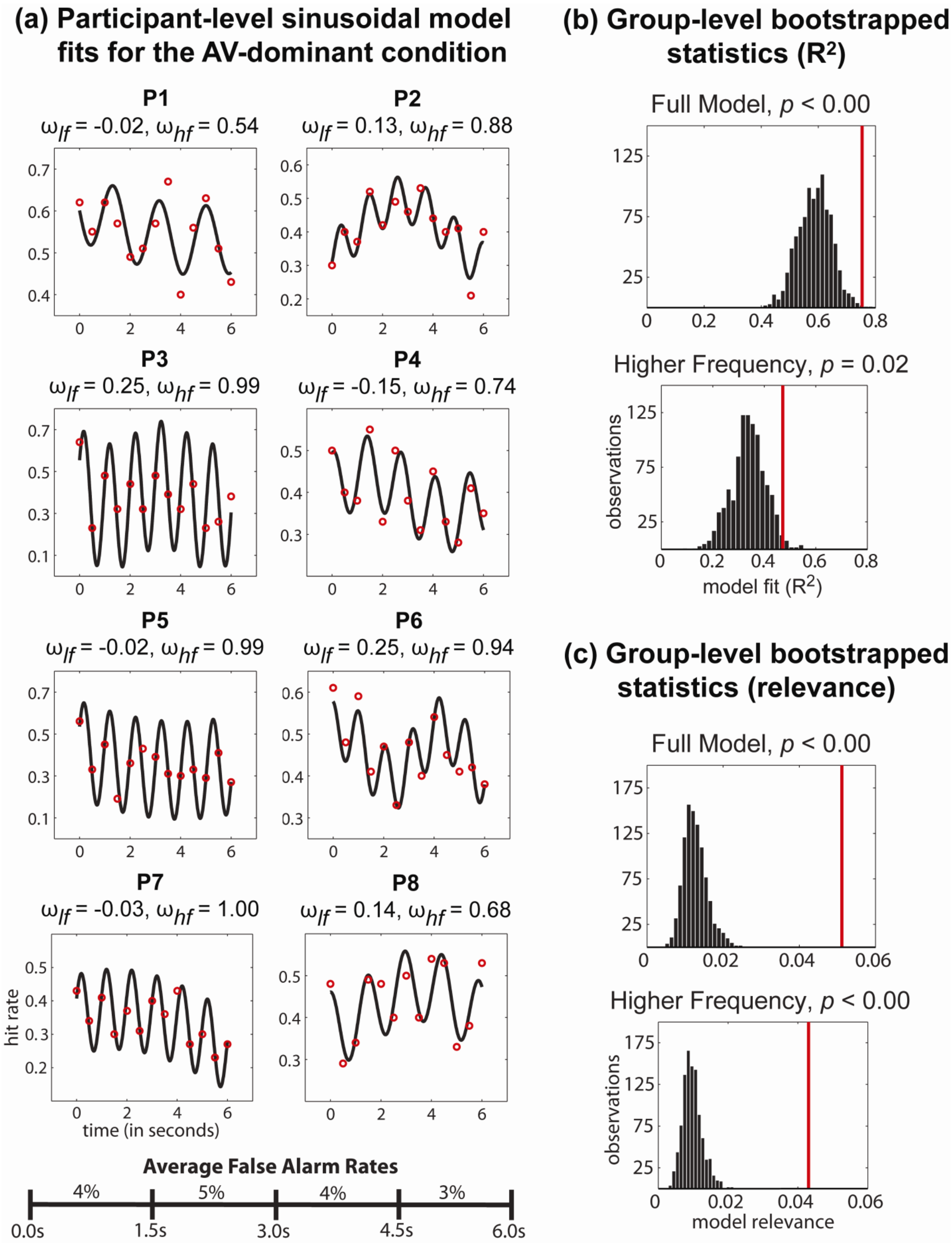
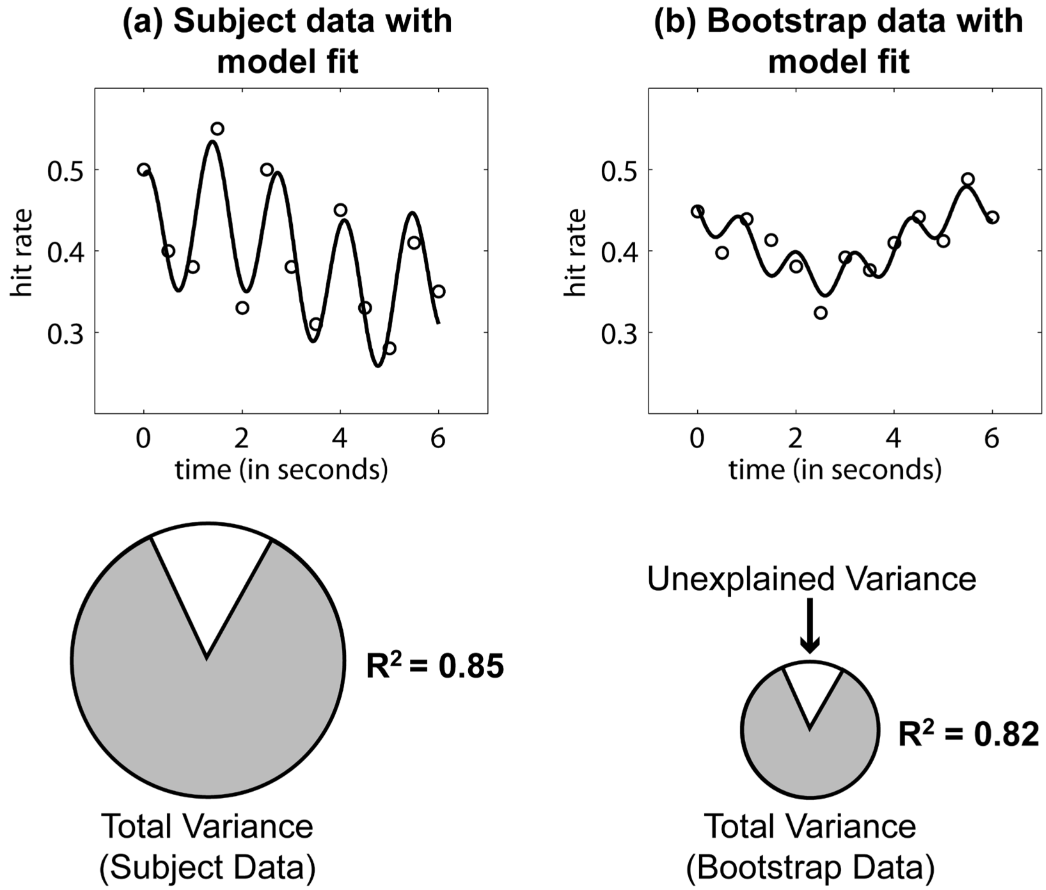

(a) Participant-level sinusoidal model fits for the AV-dominant condition
P1
ωlf = -0.02, ωhf = 0.54
P2
ωlf = 0.13, ωhf = 0.88
P3
ωlf = 0.25, ωhf = 0.99
P4
ωlf = -0.15, ωhf = 0.74
P5
ωlf = -0.02, ωhf = 0.99
P6
ωlf = 0.25, ωhf = 0.94
P7
ωlf = -0.03, ωhf = 1.00
P8
ωlf = 0.14, ωhf = 0.68
hit rate
time (in seconds)
0
2
4
6
Average False Alarm Rates
4%
5%
4%
3%
0.0s
1.5s
3.0s
4.5s
6.0s
(b) Group-level bootstrapped statistics (R2)
Full Model, p < 0.00
Higher Frequency, p = 0.02
observations
model fit (R2)
(c) Group-level bootstrapped statistics (relevance)
Full Model, p < 0.00
Higher Frequency, p < 0.00
observations
model relevance

Figure 7. Modeling results for the audiovisual (AV) dominant condition. (a) Each participant's data were fit with a two-part sinusoidal model. The red dots represent the observed data and the black lines represent the model predictions. The frequencies of the two sinusoids that the model fit to the observed data are shown above each participant's plot. Average false alarm rates across 4 time windows show that participants were not biased to respond more at the beginning of each trial. (b) The proportion of variance (R^2) explained by the two-part sinusoidal model for the observed data was compared to that explained for randomly re-distributed data (see also **Table 3**). The red lines represent where the model fit for our observed data falls relative to the bootstrap distributions. (c) An additional measure of model fit that incorporated the total amount of variance explained by the model (i.e., the model relevance) was used to further compare the observed data to randomly re-distributed data (see also **Table 3** and **Figure 8**).

Figure 8.



(a) Subject data with model fit

(b) Bootstrap data with model fit

Figure 8. An illustration of the additional information that the "relevance" term provides about the model fit. The relevance of the model fit incorporates both the proportion of variance explained and the total amount of variance explained. Here we show (a) observed subject data and (b) randomly re-distributed data where the model explains a similar amount of variance, but the observed subject data has much greater total variance. The total amount of variance is indicated by the size of the pie chart. Given the low-amplitude of the sinusoids that were fit to the randomly re-distributed data, it should not be described as periodic. That is, a low-amplitude sinusoid begins to approximate a line, and therefore a periodic model, although it might provide a good fit in terms of the proportion of variance explained, does not provide a meaningful description of the data beyond what would be provided by a linear model.

Figure 9.

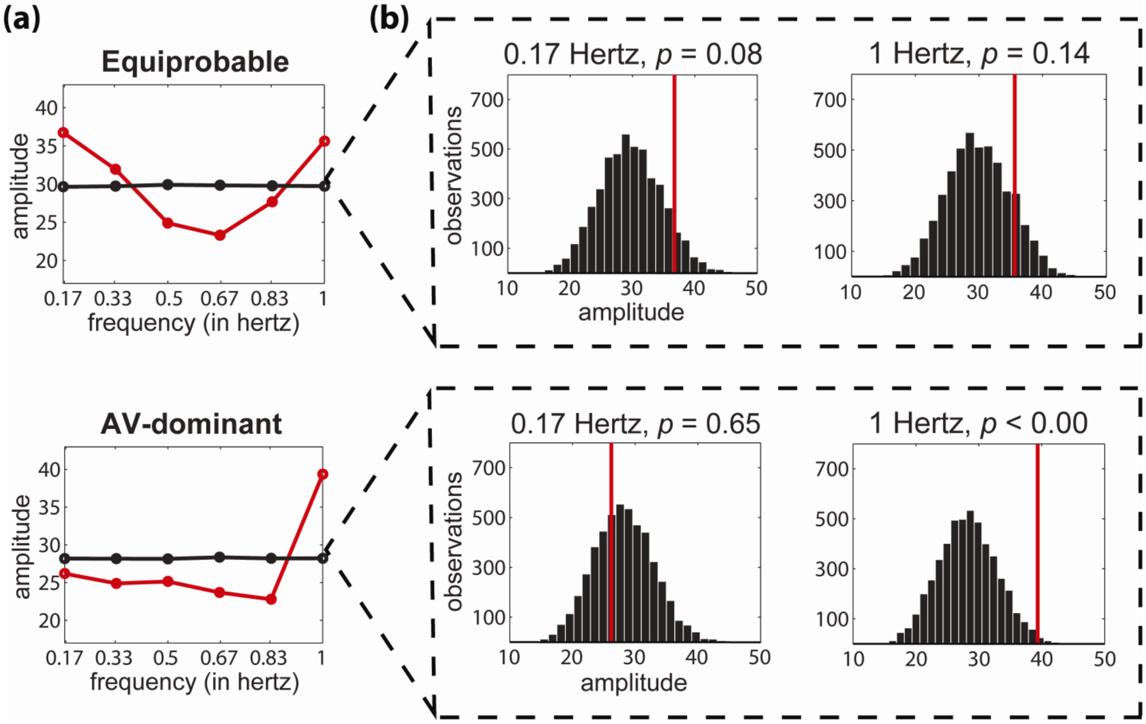

(a)
(b)
Equiprobable
AV-dominant
amplitude
frequency (in hertz)
0.17 0.33 0.5 0.67 0.83 1
20 25 30 35 40
0.17 Hertz, p = 0.08
1 Hertz, p = 0.14
0.17 Hertz, p = 0.65
1 Hertz, p < 0.00
observations
100 300 500 700
10 20 30 40 50
amplitude

Figure 9. FFT results for both experimental conditions. (a) Averaged FFT-amplitude measurements of the observed group-level data (red dots) are compared to distributions of FFT-amplitude measurements taken from randomly re-shuffled data (black dots). Amplitude is measured in performance units (based on hit rates). (b) Bootstrap distributions of FFT-amplitude measurements at 0.17 Hz and 1 Hz. The red lines represent where the FFT-amplitude measurements of our observed data fall relative to the bootstrap distributions.

CHAPTER 3

Staying within the Lines: The Formation of Visuospatial Boundaries Influences Multisensory Feature Integration

In a complex environment, such as the streets of New York, information continually bombards the senses. This information, which is processed in distributed cortical regions, must be recombined to form internal representations of multisensory objects and events (e.g., a taxi screeching to a halt in the crosswalk). An important question is which variables govern the selection of features to be grouped. One obvious variable is spatial alignment; however, the brain does not fully establish an object’s spatial extent until it has established the object’s visual boundaries. We hypothesized that the formation of visuospatial boundaries must play an important role in determining which multisensory features should be bound as components of an object. Here we investigated whether the differential formation of visuospatial boundaries (i.e., the segmentation of a visual stimulus into either a single object or a collection of elements) plays an integral role in multisensory feature integration.

Event-related potential (ERP) studies have demonstrated that task-irrelevant sounds paired with attended visual stimuli receive enhanced processing (Busse et al., 2005; Fiebelkorn et al., 2010a; Molholm et al., 2007; Talsma et al., 2007), which resembles the selective attention effects seen in response to attended sounds (Näätänen et al., 1978). Enhanced processing of task-irrelevant sounds paired with attended visual stimuli has thus been described as an automatic spread of attention (the "cross-sensory spread of attention") and taken to indicate that the task-irrelevant sounds were grouped as part of the same object as the attended visual stimuli (Busse et al., 2005; Fiebelkorn et al., 2010a; Talsma et al., 2007). The spread of attention to unattended locations within an object and to its unattended visual features (e.g., its color) has similarly been taken as an indication of feature integration (Martinez et al., 2006; O'Craven et al., 1999; Schoenfeld et al., 2003). Such spreading of attention, whether within or across sensory modalities, might serve as the glue, so to speak, that binds constituent feature representations into a unified object.

Using the cross-sensory spread of attention as our metric for multisensory feature integration, we collected both scalp and intracranial electrophysiological recordings to test our hypothesis that the formation of an object's visual boundaries is paramount to the linking of its multisensory features. Specifically, we hypothesized that pairing a task-irrelevant tone with a Kanizsa-type illusory contour (IC) stimulus (i.e., a more coherent object with well-defined boundaries; see **Figure 10**) would lead to an earlier, greater-magnitude cross-sensory spread of attention than that occurring when a task-irrelevant tone was paired with a Kanizsa-type non-illusory contour (N-IC) stimulus (i.e., a less coherent object with ambiguous boundaries). In other words, we expected that the refinement of attended space through the formation of visual boundaries would influence the degree to which a task-irrelevant tone was linked to an attended visual stimulus as a multisensory object. Our results provide compelling evidence that this is indeed the case. We therefore propose a model for multisensory feature integration that includes the formation of visual boundaries as a determinative step in establishing which features should be grouped.

METHODS

Subjects. Thirteen neurologically normal, paid volunteers participated in the experiment (mean age 27.4 ± 7.2 years; 7 female; 2 left-handed). Data from 2 additional participants were excluded because fewer than 50 percent of trials were deemed acceptable following artifact rejection procedures. All subjects reported normal hearing and normal or corrected-to-normal vision. In addition to the scalp-recorded data, we collected an additional dataset from a patient with intractable epilepsy who had subdural electrodes implanted to localize seizures (age 35 years; male; right-handed). The Institutional Review Boards of the City College of CUNY, Nathan Kline Institute, and Weill Cornell Presbyterian Hospital approved the experimental procedures. Written informed consent was obtained from all subjects prior to each recording session, in line with the Declaration of Helsinki.

Stimuli and task. Kanizsa-type illusory contour stimuli (IC) and non-illusory contour stimuli (N-IC) were presented on a gray background for 120 ms on a 21-inch cathode-ray tube (CRT) computer monitor, which was positioned 143 cm in front of participants. IC and N-IC stimuli were composed of either 3, 4, or 5 inducers (i.e., the pacman-like elements that constitute the IC and N-IC stimuli) rotated around central fixation at 45°, 90°, or 270° (see **Figure 10** for examples of the visual stimuli). For IC stimuli, the inducers were positioned to form triangles, squares, or pentagons. The visual stimuli subtended 5° of visual angle in the vertical plane and 5° of visual angle in the horizontal plane. A task-irrelevant tone (1000 Hz), which was also 120 ms in duration (with 10-ms rise and fall periods), was presented at a comfortable listening level of approximately 75 dB SPL over two JBL speakers placed at either side of the computer monitor. Because intracranial data were collected in a noisy, potentially distracting hospital environment, the sounds were delivered through headphones (Sennheiser HD 600) rather than speakers. Images and sounds were presented both alone and together, for a total of 5 equiprobable stimulus types: visual-alone IC, visual-alone N-IC, auditory-alone, audiovisual IC, and audiovisual N-IC.	Participants, who were seated in a comfortable chair in a dimly lit and electrically shielded room (the intracranial patient was tested in his hospital room), were asked to make a button press response when they detected a 40 ms flicker in one of the inducers (i.e., a decrease in contrast). Importantly, the target-detection task was independent of the configuration of the inducers: the occurrence of a target was equally likely to occur during the presentation of an IC stimulus or an N-IC stimulus. The specific inducer that flickered varied randomly as did the onset of the flicker (from 20 to 60 ms post stimulus onset). The difficulty of the target-detection task (i.e., the magnitude of the decrease in contrast associated with a flickering inducer) was adjusted automatically throughout the experimental session based on the individual participant’s performance. Each participant's overall hit rate was maintained at approximately 80 percent to ensure that the task remained challenging and did not become trivial with repetition. To minimize distortion from overlapping stimulus responses, the stimulus presentation (both targets and non-targets) was first-order counterbalanced, meaning that each stimulus type was preceded and followed by all stimulus types equally often. Stimulus onset asynchrony (SOA) varied randomly between 400 and 700 ms during scalp recordings, and a total of 252 stimuli were presented within each ~3-minute block. Participants completed a minimum of 8 blocks, and at most 10 blocks. Because it was thought that the target-detection task would be more difficult for epileptic patients than healthy participants, the stimulus SOA varied between 800 and 1100 ms for subdural recordings. **Figure 10** shows a schematic of the experimental paradigm.

Data acquisition and processing. High-density continuous scalp-recorded EEG was obtained from a BioSemi ActiveTwo 168 channel system; continuous subdural-recorded EEG was obtained from a BrainVision amplifier system. Recordings were initially referenced online relative to a common mode active electrode and digitally sampled at 512 Hz (intracranial data were digitally sampled at 1000 Hz). The continuous EEG was divided into epochs (-100 ms pre-stimulus to 500 ms post-stimulus onset), and baseline corrected from -100 ms to stimulus onset. Trials with blinks and eye movements were automatically rejected off-line on the basis of horizontal and vertical electro-oculogram (EOG) recordings. An artifact rejection criterion of ± 80μV was used at all other scalp sites to reject trials with excessive EMG or other noise transients (an artifact rejection criterion of ± 450μV was used for subdural recordings, where the amplitude of the recorded signal is much greater). To prevent contamination from motor responses associated with target detection and false alarms, target trials, trials immediately following target trials, and false-alarm trials were discarded. Across participants, there was an average of 230 accepted sweeps per stimulus type. EEG epochs were averaged for each participant to compute ERPs. The averaged waveforms were algebraically re-referenced to an average reference (i.e., an average of the activity at all electrodes). Separate group-averaged ERPs for each stimulus type were generated for display purposes.

Statistical analysis. To determine whether IC and N-IC stimuli affected target detection differently or whether the co-occurrence of a task-irrelevant tone influenced target detection, reaction times (RTs) and error patterns were examined. Separate two-way, repeated measures ANOVAs were conducted on RTs and hit rates, with factors of object status (2 levels: IC stimuli or N-IC stimuli) and multisensory status (2 levels: visual-alone stimuli or audiovisual stimuli).

Based on findings from previous studies, there were specific hypotheses about the location and timing of the ERP effects under investigation. Amplitude data from 5 frontocentral electrodes were averaged to examine the cross-sensory spread of attention, and amplitude data from 4 right-lateral occipital electrodes and 4 left-lateral occipital electrodes were averaged separately to examine the illusory contour effect. **Figures 11** and **12** highlight (with white circles) the locations of the electrodes that were used for statistical analyses of the electrophysiological data. These electrodes were chosen based on scalp topographies from previous studies that measured either the cross-sensory spread of attention (Busse et al., 2005; Fiebelkorn et al., 2010a; Talsma et al., 2007) or the illusory contour effect (Foxe et al., 2005; Murray et al., 2002; Murray, Imber, Javitt & Foxe, 2006).

For all statistical analyses, amplitude data were averaged across 20-ms windows centered on the latencies of the major ERP components (i.e., the visual and auditory P1 and N1 components). To test for longer latency cross-sensory effects, amplitude data were averaged using consecutive 20-ms windows, starting at 200 ms. An initial latency window from 200 to 220 ms was chosen because previous studies had reported that the cross-sensory spread of attention, which resembles an attention-related processing negativity (PN), onsets anywhere from 180 to 250 ms post stimulus (Busse et al., 2005; Fiebelkorn et al., 2010a; Molholm et al., 2007; Talsma et al., 2007). The use of consecutive 20-ms windows rather than a single window encompassing a longer timeframe allowed for measuring the evolution of differences in the cross-sensory spread of attention based on whether the task-irrelevant tone was paired with an IC or an N-IC stimulus. Since this experiment was not intended to study illusory contours but only to use the IC effect as a metric for pre-attentive visual object processing, a single window from 250 to 300 ms was used to test the significance of longer latency IC effects. The timeframe for this window was chosen based on previous studies that described ERP modulations thought to reflect later stages of visual object processing (Murray et al., 2002; Doniger et al., 2000).

To confirm that there was a significant IC effect and to test whether it was modulated by the inclusion of a task-irrelevant tone, three-way repeated measures ANOVAs were conducted with factors of object status (2 levels: IC stimuli or N-IC stimuli), multisensory status (2 levels: visual-alone or audiovisual), and hemisphere (2 levels: right or left). To test the main hypothesis that the cross-sensory spread of attention from an attended visual stimulus to a simultaneously presented task-irrelevant tone was modulated by pre-attentive visual object processing, one-way repeated measures ANOVAs were conducted with a factor of stimulus type (3 levels: auditory alone, audiovisual IC, or audiovisual N-IC). Prior to all statistical analyses, the unisensory responses to audiovisual stimuli were isolated by subtracting the responses to either visual-alone stimuli (for analyses that examined the cross-sensory spread of attention, AV - V) or auditory-alone stimuli (for analyses that examined the illusory contour effect, AV - A) from the response to audiovisual stimuli. Subtracting unisensory responses from audiovisual responses allows for the isolation of cross-sensory effects at auditory and visual scalp sites (e.g., Busse et al. 2005; Molholm et al., 2007).

Following significant ANOVAs, pairwise comparisons were made using planned protected t-tests (LSD). The alpha level for all statistical analyses was set at 0.05. When appropriate, Greenhouse-Geisser corrections were made. The onset of ERP modulations was determined through running t-tests between amplitude measurements at consecutive time points (e.g., Molholm et al., 2007). Onset was considered to be the first of ten consecutive, statistically significant time points. This criterion was chosen because the likelihood of getting ten false positives in a row is extremely low (Guthrie & Buchwald, 1991).

Intracranial data. Subdural electrodes are highly sensitive to local field potentials (LFP) generated within ~4.0 mm^2 area and much less sensitive to distant activity (Allison et al., 1999; Lachaux et al., 2005), which allows for improved localization of underlying current sources relative to scalp-recorded EEG. An additional dataset was therefore collected from a patient who had undergone surgery for intractable epilepsy. This patient was on anti-seizure medication during the experimental session. The location of the subdural grid electrodes was confirmed using a postoperative MRI. No interictal activity was detected in the structures investigated here. Statistical analyses of intracranial data were identical to those for scalp data except that they were based on the variance measured across single trials rather than the variance measured across participants. This same approach is used to analyze data collected from non-human primates. Because the primary purpose for collecting intracranial data was to confirm that the cross-sensory spread of attention reflected enhanced activity within auditory cortical areas, an electrode adjacent to auditory cortex served as the reference. Choosing an adjacent electrode as the reference serves to isolate local cortical activity.

RESULTS

Behavioral results. As shown in **Table 3**, RTs in response to the 4 target types (visual-alone IC, visual-alone N-IC, audiovisual IC, and audiovisual N-IC) were nearly identical. A repeated measures ANOVA with factors of object status (2 levels: IC stimuli or N-IC stimuli) and multisensory status (2 levels: visual-alone stimuli or audiovisual stimuli) revealed no significant main effects ($F_{1, 12} = 0.0$, $p = 0.92$; $F_{1, 12} = 0.1$, $p = 0.73$) or interactions ($F_{1, 12} = 0.6$, $p = 0.45$). Hit rates in response to the 4 target types, on the other hand, differed significantly based on multisensory status ($F_{1, 12} = 60.2$, $p < 0.001$). This pattern of results held regardless of whether the target (i.e., a change in contrast of one of the pacman-like elements) occurred during the presentation of an IC stimulus ($p < 0.001$) or during the presentation of an N-IC stimulus ($p < 0.001$).

While the co-occurrence of a task-irrelevant tone interfered somewhat with behavioral performance, there was no significant main effect of object status or interactions between object status and multisensory status. This indicates that target detection was equivalent for IC and N-IC stimuli. Based on these findings, it can reasonably be concluded that differences in the electrophysiological data do not result from differential top-down attentional deployment across object status (i.e., whether the visual stimulus was an IC or an N-IC stimulus) or from differential arousal associated with object status. Rather, differences in the electrophysiological data as a function of object status must result from differences in the pre-attentive processing of visual boundaries.

The IC effect. **Figure 11** shows waveforms and scalp topographies associated with the IC effect. The voltage maps of the ERP effects that occurred when inducers were positioned to form illusory contours closely resemble the scalp distribution that previous studies have localized to the LOC (e.g., Foxe, Murray & Javitt, 2005; Murray et al., 2002). Repeated measures ANOVAs with factors of object status (2 levels: IC stimuli or N-IC stimuli), multisensory status (2 levels: visual-alone or audiovisual), and hemisphere (2 levels: right or left) revealed no significant main effects or interactions during the latency of the visual P1 component, but did reveal significant main effects of both object status (i.e., the illusory contour effect, $F_{1, 12} = 14.5$, $p < 0.01$) and hemisphere ($F_{1, 12} = 5.7$, $p = 0.03$) during the latency of the visual N1 component. Multisensory status during the latency of the visual N1 component (~170 ms), however, was insignificant ($F_{1, 12} = 0.2$, $p = 0.64$), as were all interactions (all p's > 0.19). Repeated measures ANOVAs on amplitude data averaged across a window from 250 to 300 ms post stimulus onset revealed a significant illusory contour effect (i.e., the main effect of object status, $F_{1, 12} = 16.5$, $p < 0.01$), but no other significant main effects (hemisphere, $F_{1, 12} = 0.6$, $p = 0.46$; multisensory status, $F_{1, 12} = 0.2$, $p = 0.66$) or interactions (all p's > 0.32). Running t-tests determined that the approximate onset of the illusory contour effect was consistent, regardless of whether the visual stimuli were presented with or without a task-irrelevant tone (~130 to 134 ms).

These findings replicate those from previous studies that describe a pre-attentive modulation of visual object processing associated with the formation of illusory contours

(Grosof, Shapley & Hawken, 1993; Lee & Nguyen, 2001; Murray et al., 2006; Sheth, Sharma Rao & Sur, 1996; von der Heydt, Peterhans & Baumgartner, 1984). The amplitude of the IC effect was not significantly affected by the presence or absence of a task-irrelevant tone. This was true both at the earliest stages of visual object processing (i.e., during the latency of the visual N1 component), and at later stages of visual object processing (from 250 to 300 ms post stimulus onset).

The cross-sensory spread of attention. **Figure 12** shows waveforms and scalp topographies associated with the cross-sensory spread of attention. The voltage maps of the ERP effects resemble the scalp distribution of the late frontal negativity that previous studies have shown to reflect attentional processing in auditory cortex (Busse et al., 2005; Fiebelkorn et al., 2010a; Molholm et al., 2007; Talsma et al., 2007). Repeated measures ANOVAs with a factor of stimulus type (3 levels: auditory-alone, audiovisual IC, or audiovisual N-IC) revealed no significant main effects during the latency of the major auditory components: the auditory P1 ($F_{2, 24} = 1.3$, $p = 0.30$), N1 ($F_{2, 24} = 1.0$, $p = 0.37$), or P2 ($F_{2, 24} = 1.2$, $p = 0.31$). Starting at 200 ms post stimulus, repeated measures ANOVAs on consecutive 20-ms windows revealed a main effect of stimulus type from 220 to 260 ms ($F_{2, 24} = 3.6$–6.9, all p's < 0.04) and from 320 to 500 ms ($F_{2, 24} = 3.4$–16.6, all p's $<$ 0.05). Pairwise comparisons revealed that the response to a task-irrelevant tone paired with an IC stimulus was more negative-going than the response to a task-irrelevant tone presented alone from 220 to 260 ms (all p's < 0.05) and from 320 to 500 ms (all p's $<$ 0.02). The response to a task-irrelevant tone paired with an N-IC stimulus was more negative-going than the response to a task-irrelevant tone presented alone from 340 to

420 ms (all *p*'s < 0.03). Pairwise comparisons further revealed that the response to a task-irrelevant tone paired with an IC stimulus was more negative-going than the response to a task-irrelevant tone paired with an N-IC stimulus from 240 to 280 ms (all *p*'s < 0.02), from 340 to 380 ms (all *p*'s < 0.05), and from 420 to 500 ms (all *p*'s < 0.04). Running t-tests revealed that the cross-sensory spread of attention in response to an audiovisual IC stimulus onset at ~210 ms post stimulus; whereas the cross-sensory spread of attention in response to an audiovisual N-IC stimulus did not onset until ~322 ms post stimulus.

These data provide evidence for a direct link between pre-attentive visual object processing and the cross-sensory spread of attention. Relative to the cross-sensory spread of attention when a task-irrelevant tone was paired with an N-IC stimulus, pairing a task-irrelevant tone with an IC stimulus (i.e., a more coherent object with well-defined boundaries) led to an earlier and greater magnitude cross-sensory spread of attention.

Intracranial results. **Figure 13** shows waveforms from 3 subdural electrodes. Two electrodes (Talairach: x = 49, y = -63, z = 2 mm; Talairach: x = 49, y = -58, z = 15 mm) positioned over visual cortex demonstrated a significant illusory contour effect (i.e., a main effect of object status). A third electrode, which was positioned over auditory cortex (Talairach: x = 57, y = -12, z = 7 mm; on the superior temporal gyrus), demonstrated a significant cross-sensory spread of attention. The position of this third electrode was not only determined based on its Talairach coordinates, but also on the absence of a response when the participant viewed visual-alone stimuli.

Repeated measures ANOVAs on Electrodes 1 and 2 (i.e., the electrodes positioned over visual cortex) with a factor of object status (2 levels: IC stimulus or N-IC stimulus) revealed a significant illusory contour effect (i.e., a main effect of object status) for Electrode 1 from 250 to 300 ms ($F_{1, 248} = 3.81$, $p = 0.05$), and for electrode 2 during the timeframe of the visual N1 component ($F_{1, 248} = 3.76$, $p = 0.05$), then again from 250 to 300 ms ($F_{1, 248} = 5.8$, $p = 0.02$). Repeated measures ANOVAs on electrode 3 (i.e., the electrode positioned over auditory cortex) with a factor of stimulus type (3 levels: auditory-alone, audiovisual IC stimulus, or audiovisual N-IC stimulus) revealed a significant cross-sensory spread of attention from 240 to 280 ms ($F_{2, 494} = 3.21$–3.40, $p =$ 0.04–0.03). Pairwise comparisons revealed that the response to an audiovisual IC stimulus was significantly different from that to either an auditory-alone stimulus ($p =$ 0.05–0.04) or an audiovisual N-IC stimulus ($p =$ 0.02–0.01). These intracranial data, which mimic what was observed on the scalp, provide compelling evidence that the segmentation of a visual stimulus through the formation of illusory contours modulates processing of a task-irrelevant tone at the level of auditory cortex.

Although there was a significant cross-sensory spread of attention from an attended IC stimulus to a task-irrelevant tone, the intracranial ERPs showed no significant cross-sensory spread of attention when the same task-irrelevant tone was paired with an N-IC stimulus. This is somewhat surprising given that previous studies have demonstrated the cross-sensory spread of attention using a variety of visual stimuli, including sinusoidal gratings (Talsma et al., 2007) and photographs of well-known objects (Fiebelkorn et al., 2010a). Because a faster rate of stimulus presentation enables a more selective focusing of attention, these previous studies (and the scalp-recorded portion of the present study) used short interstimulus intervals. In comparison, for the intracranial portion of the present study, a somewhat longer interstimulus interval (680 to

980 ms) was used to decrease the difficulty of the task. Given this slower rate of stimulus presentation, it is possible that the intracranial patient allocated more attention to the task-irrelevant tones, resulting in a smaller difference between attentional processing measured during the auditory-alone and audiovisual trials. Of greatest importance to the main hypothesis, however, is whether there was a difference between the amplitude and latency of the cross-sensory spread of attention when a task-irrelevant tone was paired with an IC stimulus relative to when a task-irrelevant tone was paired with an N-IC stimulus: as reported above, this difference was indeed significant.

DISCUSSION

Here we show that the refinement of spatial selection through the formation of visuospatial boundaries affects the cross-sensory spread of attention to a task-irrelevant sound, our metric for multisensory feature integration. The dominance of vision in audiovisual localization, as demonstrated by the well-known ventriloquism illusion, makes sense given that spatial tuning within the visual system is much more precise that that within the auditory system. A sound will be grouped as part of an object as long as its relatively diffuse spatial representation overlaps the more pinpointed spatial representation of the attended visual stimulus. But the brain does not fully establish an object's visuospatial extent until it has established the object's visual boundaries. It follows that the formation of visuospatial boundaries should influence multisensory feature integration. We predicted that differential segmentation of a visual stimulus through the formation of illusory contours would modulate multisensory feature integration in a bottom-up manner, such that the cross-sensory spread of attention would be stronger when the visual elements formed a more coherent object with well-defined boundaries. The data reported here are consistent with this hypothesis, providing compelling evidence for a direct link between visual object formation and the integration of an object's multisensory features. Since the behavioral data indicate that participants treated IC and N-IC stimuli equivalently, differences in electrophysiological data do not result from differences in arousal or top-down attentional deployment. Instead it seems to have been the differential formation of visual boundaries that influenced multisensory feature integration: pairing a task-irrelevant tone with an IC stimulus (i.e., a more coherent object with well-defined boundaries) led to an earlier, greater-magnitude cross-sensory spread of attention than that occurring when a task-irrelevant tone was paired with an N-IC stimulus (i.e., a less coherent object with ambiguous boundaries).

A Model for Multisensory Feature Integration

The present results, in conjunction with findings from previous studies, suggest a model for multisensory feature integration in which: (1) the visual scene is parsed into a coarse outline of objects (i.e., segmentation, see Driver, Davis, Russell, Turatto & Freeman, 2001; Moore & Egeth, 1997; Razpurker-Apfeld & Hillel, 2008), (2) spatial selection determines the relevant location for further processing (Treisman & Gelade, 1980), (3) attention spreads within the visual boundaries of an object at the selected location (i.e., the attentional spotlight conforms to the shape of the object, see Egly et al., 1994; Martinez et al., 2006; Martinez et al., 2007), and (4) attention spreads to coincident multisensory features (and visual features) that fall within that object's visual boundaries (Busse et al., 2005; Fiebelkorn et al., 2009; Molholm et al., 2007; Talsma et al., 2007).

The first three steps reflect object-based selection, and it is this refinement of attended space that determines whether a task-irrelevant sound will be grouped with the attended visual stimulus (i.e., whether the multisensory features have overlapping spatial representations).

The IC effect, a well-studied manipulation of automatic visual-object processing, serves as a good metric for the object-based selection processes that our model posits as determinative steps preceding multisensory feature integration. The formation of visual boundaries within LOC regions is thought to occur between 90 and 190 ms post stimulus onset with subsequent feedback to lower-tier areas (Murray et al., 2002; Murray et al., 2006). This time course accords with findings that spatial attention spreads within both illusory boundaries and real-object boundaries starting at ~160 ms post stimulus onset, an effect that has also been localized to LOC regions (Martinez et al., 2006; Martinez et al., 2007). According to our model, these object-based selection processes within the ventral visual stream establish the boundaries for subsequent multisensory feature integration.

Human lesion studies have shown that although an intact ventral visual stream is sufficient for boundary completion, parietal areas are needed for non-shape feature integration (Barton, Malcolm & Hafter, 2007; Humphreys & Riddoch, 1994; Robertson, 2003). Patients with large parietal lesions are able to make shape determinations but make mistakes when asked to name the color of a specific object presented among several different-colored objects. That is, patients with bilateral parietal damage sometimes form illusory conjunctions (see Treisman & Schmidt, 1982), combining the color of one object with the shape of another. We posit that multisensory feature integration through the cross-sensory spread of attention likewise requires contributions from both the ventral and dorsal visual streams. As discussed above, ventral stream areas (e.g., LOC regions) are thought to establish the boundaries for feature integration (steps 1–3 of our model), whereas dorsal stream areas (e.g., the posterior parietal cortex), which are known to include multisensory spatial representations (Bushara et al., 1999; Cohen & Andersen, 2004; Linden, Grunewald & Andersen, 1999; Molholm et al., 2006; Stricanne & Andersen, 1996), are thought to co-register the spatial representations of features that fall within an object's visual boundaries. Following such co-registration, there is a spread of attentional processing, as dorsal stream areas provide feedback to feature-specific cortex (step 4 of our model). Localization of enhanced processing associated with the spread of attention to feature-specific cortical areas has been previously confirmed through the use of fMRI and EEG source modeling (Busse et al., 2005; O'Craven et al., 1999; Schoenfeld et al., 2003). The present study also provides confirmatory evidence that the cross-sensory spread of attention localizes to feature-specific cortex (i.e., auditory cortex) through the use of intracranial ERPs, which have excellent spatial resolution relative to scalp-recorded ERPs.

Feedback to feature-specific cortex might limit the response of neurons that have multiple objects within their large receptive fields to the response associated with features that are components of the selected object (i.e., the neuron's response is no longer a weighted sum of all the objects in its receptive field but a response limited to the selected object), or it might enhance the response of neurons specifically tuned to prefer features that are components of the selected object (see Reynolds & Desimone, 1999). Regardless of the exact mechanism or combination of mechanisms, enhanced processing through feedback to feature-specific cortical areas presumably binds anatomically separate feature representations into a coherent object.

Concluding Remarks

The data reported here demonstrate a direct link between pre-attentive visual object processing and the integration of an object's multisensory features. Segmentation of a visual stimulus through the formation of illusory contours increases the degree to which the visual stimulus and its paired task-irrelevant sound are grouped as a multisensory object. We suggest that the visuospatial extent of an object (as determined through the formation of its visual boundaries) must first be parsed before its multisensory features can be grouped.

Figure 10.

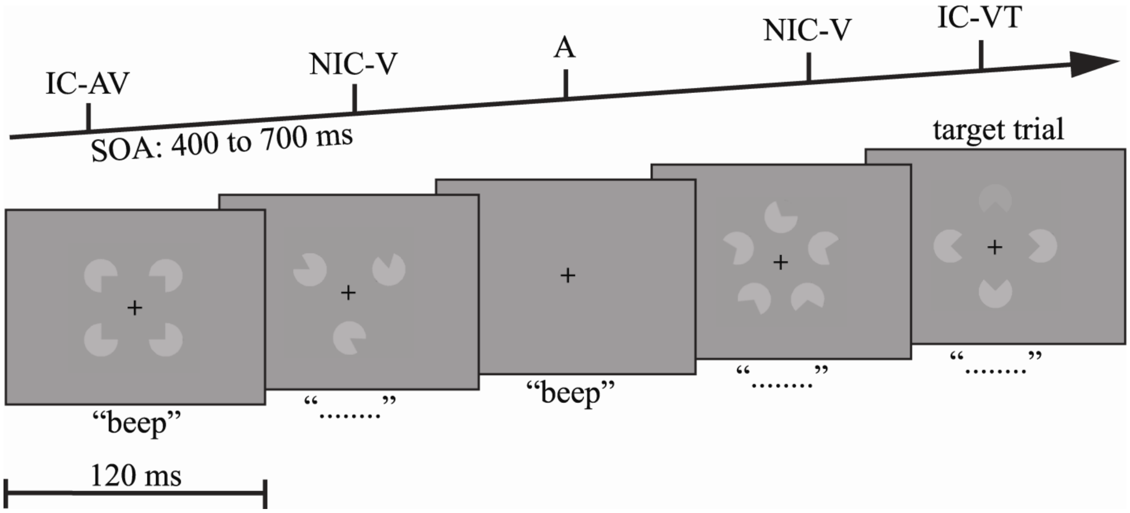

IC-AV
NIC-V
A
NIC-V
IC-VT
SOA: 400 to 700 ms
target trial
"beep"
"........"
"beep"
"........"
"........"
120 ms

Figure 10. A schematic of the experimental design. Audiovisual IC stimulus (IC-AV), visual-alone N-IC stimulus (NIC-V), auditory-alone stimulus (A), and visual-alone IC target (IC-VT).

Table 4

Average reaction times (RTs) and hit rates (HRs) sorted by multisensory status and object status

	Visual-alone		Audiovisual	
	IC	N-IC	IC	N-IC
RTs (ms)	430(25)	426(23)	426(25)	428(24)
HRs (%)	89(5)	90(4)	76(7)	74(7)

Notes. Standard deviations in parentheses.

Figure 11.

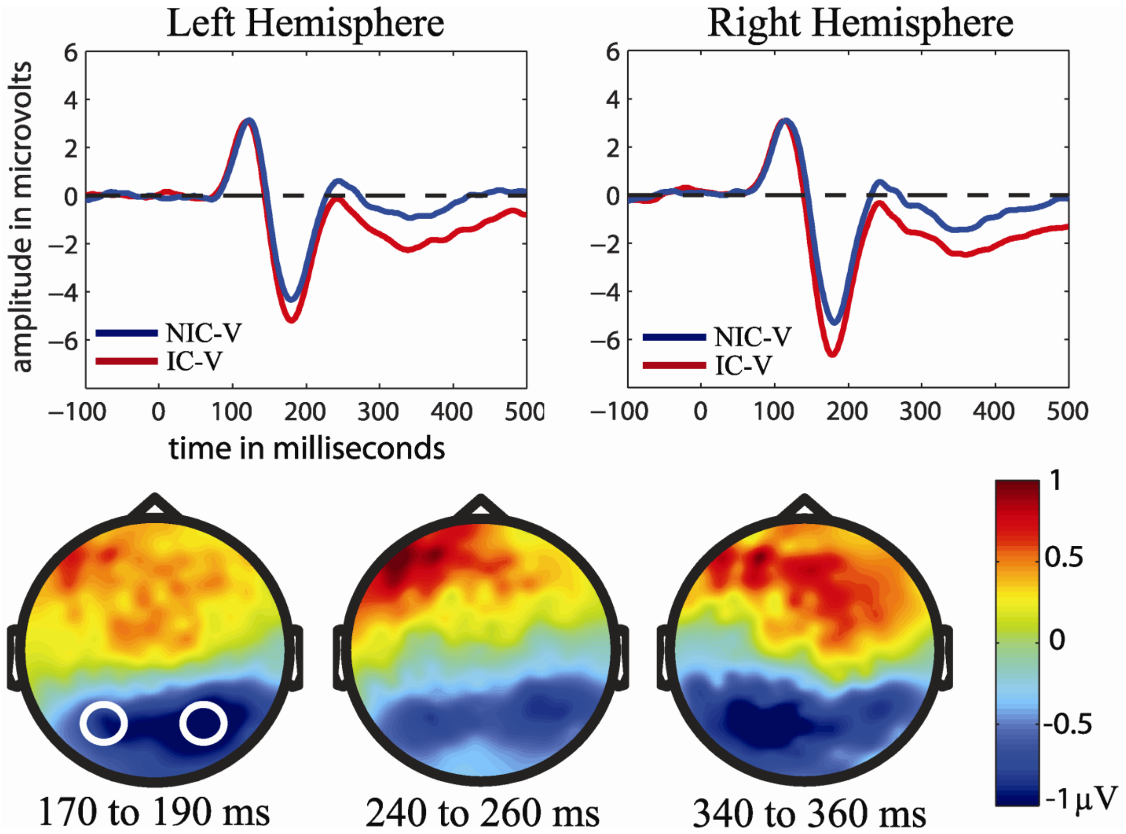

Left Hemisphere
Right Hemisphere
amplitude in microvolts
6
4
2
0
−2
−4
−6
−100
0
100
200
300
400
500
time in milliseconds
NIC-V
IC-V
1
0.5
0
-0.5
-1 µV
170 to 190 ms
240 to 260 ms
340 to 360 ms

Figure 11. The IC effect over the left and right hemispheres. Flattened voltage maps were derived from ERP difference waves. The white circles on the first voltage map represent the location of the electrodes that were averaged to produce the right and left hemisphere waveforms. These electrodes were also used for the statistical analysis. Visual-alone N-IC stimulus (NIC-V), visual-alone IC stimulus (IC-V).

Figure 12.

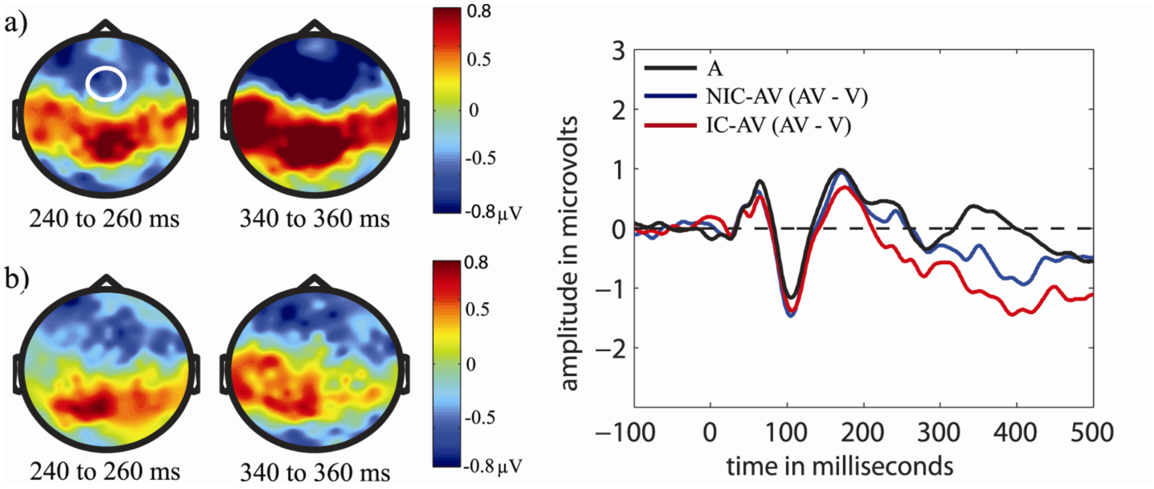

a)
240 to 260 ms
340 to 360 ms
0.8
0.5
0
-0.5
-0.8 µV
b)
240 to 260 ms
340 to 360 ms
0.8
0.5
0
-0.5
-0.8 µV
A
NIC-AV (AV - V)
IC-AV (AV - V)
amplitude in microvolts
3
2
1
0
−1
−2
−100
0
100
200
300
400
500
time in milliseconds

Figure 12. The cross-sensory spread of attention in response to a task-irrelevant tone paired with either an IC or an N-IC stimulus. Flattened voltage maps were derived from ERP difference waves between **(a)** the response to an audiovisual IC stimulus (IC-AV) and the response to an auditory-alone stimulus (A) and **(b)** the response to an IC-AV and the response to an audiovisual N-IC stimulus (NIC-AV). The white circle on the first voltage map represents the location of the electrodes that were averaged to produce the waveforms. These electrodes were also used for the statistical analysis.

Figure 13.

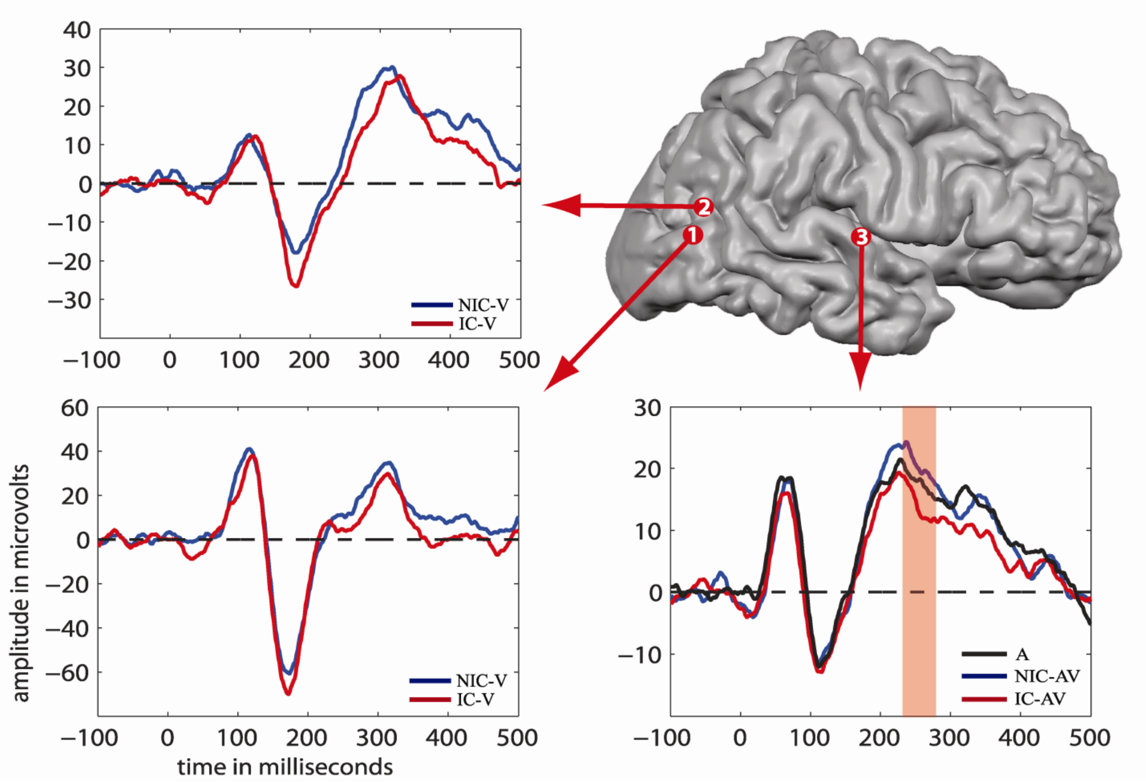

40
30
20
10
0
−10
−20
−30
−100
0
100
200
300
400
500
NIC-V
IC-V
2
1
3
60
40
20
0
−20
−40
−60
amplitude in microvolts
time in milliseconds
NIC-V
IC-V
30
20
10
0
−10
A
NIC-AV
IC-AV

Figure 13. ERP waveforms recorded from 3 subdural-grid electrodes that were implanted on the cortical surface (as part of a larger array) of a single patient undergoing surgery for intractable epilepsy. Electrodes 1 and 2 were found to demonstrate a significant IC effect, whereas electrode 3, which was localized to auditory cortex, was found to demonstrate a significant cross-sensory spread of attention when a task-irrelevant tone was paired with an IC stimulus. The pink bar represents the timeframe during which the response to a task-irrelevant tone paired with an IC stimulus was statistically more negative than the response to either a task-irrelevant tone paired with an N-IC stimulus or a task-irrelevant tone presented alone.

CHAPTER 4

Dual Mechanisms for the Cross-Sensory Spread of Attention: How Much Do Learned Associations Matter?

Imagine a dog standing outside a screen door, barking to be let back into the house. The different features of this object—the barking dog—stimulate different types of specialized sensory receptors, and these various inputs provide complementary or redundant information that is used to identify and respond to the object. The constituent features of this or any object, whether within or across sensory modalities, are represented in anatomically separated cortical regions. A fundamental question is how these features are selected and recombined to form what is ultimately perceived as a coherent object: the so-called "binding problem."

There is substantial evidence that attention-related mechanisms play a crucial role in object-binding processes (Robertson, 2003). In the influential biased-competition model of visual attention, focal spatial selection results in the preferential processing of an entire object, including features that are irrelevant to the task at hand (Desimone & Duncan, 1995; Duncan, 2006). Indeed, experimental evidence has revealed that attention spreads both within the visual boundaries of an object—even when part of that object is outside the region of interest highlighted by spatial selection—and to the task-irrelevant visual features of that object (Egly et al., 1994; Martinez et al., 2006; Martinez et al., 2007; Melcher & Vidnyanszky, 2006; O'Craven et al., 1999; Schoenfeld et al., 2003; Wylie, Javitt & Foxe, 2004). When participants are asked to attend to an object's motion, for example, its other visual features, such as its color, also receive enhanced processing (O'Craven et al., 1999; Schoenfeld et al., 2003; Wylie et al., 2006).

Real-world objects frequently have multisensory features, and hence one would expect an effective binding mechanism to operate beyond the boundaries of visual representations. Consistent with this notion, recent studies have demonstrated that attention also spreads to a task-irrelevant sound presented with an attended visual stimulus (Busse et al., 2005; Molholm et al., 2007; Talsma et al., 2007). Molholm et al. (2007) hypothesized the existence of two separate processes that could lead to this object-based, cross-sensory spread of attention: (1) a *"stimulus-driven"* process for which no preexisting relationship between the features need be present, and (2) a *"representation-driven"* process that involves the activation of preexisting and highly associated feature-representations.[1] These two processes might be separate mechanisms through which attention binds the distributed features of an object. The stimulus-driven spread of attention is hypothesized to occur whenever a task-irrelevant sound is presented at the same time and location as an attended visual stimulus (e.g., as in Talsma et al., 2007, for arbitrarily paired tones and gratings), and is thought to reflect bottom-up feature binding. The representation-driven spread of attention, on the other hand, is hypothesized to occur in response to a task-irrelevant sound that is semantically related to a visual target (e.g., the bark of a dog when the target stimulus is the image of a dog), and is thought to reflect top-down interactions between memory and attention. In this case, the activation of one feature within the stored representation of a well-known object (e.g., a dog) leads to

---

[1] In Molholm et al. (2007), the stimulus-driven and representation-driven processes are referred to as the cross-sensory spread of attention and the cross-sensory priming effect, respectively. Here we adopt new terminology for the sake of clarity.

enhanced processing of all that object's features, including its task-irrelevant auditory features (e.g., a bark).

The representation-driven spread of attention serves as one possible mechanism through which the multisensory features of well-known objects might be treated differently than the multisensory features of novel objects. It remains to be tested, however, if the representation-driven spread of attention results from either long-term, semantically based associations among an object's multisensory features, or from associations established during the experimental session. It also remains to be tested if the more automatic stimulus-driven spread of attention is influenced by highly learned associations among an object's multisensory features. Previous studies have only demonstrated the existence of this bottom-up mechanism for the cross-sensory spread of attention using novel objects (Busse et al., 2005; Talsma et al., 2007).

In the present study we used high-density scalp-recorded ERPs to resolve these questions and to disentangle the respective contributions of the stimulus-driven and representation-driven processes to the object-based, cross-sensory spread of attention. Participants were instructed to respond to the second of consecutively presented images of a well-known object (dogs, cars, or guitars) while ignoring all sounds (i.e., a one-back task). Unlike Molholm et al. (2007), who used a small stimulus set consisting of congruent multisensory objects (e.g., barking dogs) to examine the representation-driven spread of attention, we also included consistently paired, incongruent multisensory objects (e.g., guitars consistently paired with barks), as well as multiple exemplars of each object. The use of this expanded stimulus set allowed us to establish if the representation-driven spread of attention resulted from the activation of a preexisting, canonical representation of the well-known object, rather than experiment-specific associations. The inclusion of both object-congruent and object-incongruent multisensory trials further allowed us to establish if bottom-up feature binding, as measured through the stimulus-driven spread of attention, differs based on whether the task-irrelevant sound and its paired visual stimulus are components of the same well-known object or components of different well-known objects.

Our findings reveal dual mechanisms for the cross-sensory spread of attention with overlapping latencies and scalp topographies: (1) a stimulus-driven process that reflects an inherent bias to bind co-occurring multisensory features, regardless of whether those features are semantically congruent or incongruent, and (2) a representation-driven process that requires the existence of preexisting, highly learned associations among the multisensory features. We also show that the cross-sensory spread of attention through the stimulus-driven and representation-driven processes appear additive when an image of the visual target (e.g., a dog) is paired with a congruent, task-irrelevant sound (e.g., a bark).

METHODS

Subjects

Twelve neurologically normal, paid volunteers participated in the first experiment (mean age 26.7 ± 5.4 years; four female; one left-handed). Data from two additional subjects were excluded either because of unusually poor task performance or because of bridging among the scalp electrodes. An additional ten neurologically normal, paid volunteers participated in a second, control experiment (mean age 26.9 ± 4.3 years; four female; all right-handed). All subjects reported normal hearing and normal or corrected-to-normal vision. The Institutional Review Boards of both the Nathan Kline Institute and the City College of CUNY approved the experimental procedures. Written informed consent was obtained from all subjects prior to each recording session, in line with the Declaration of Helsinki.

Stimuli

Twenty unique black-and-white photographs and twenty unique sounds were used to represent each of three well-known multisensory objects: dogs, cars, and guitars (for a total of 60 images and 60 sounds). These photographs and sounds were collected through Internet searches, and later standardized. The photographs were presented for 400 ms on a 21-inch cathode-ray tube (CRT) computer monitor, which was positioned 143 cm in front of participants. The centrally presented images subtended an average of 4.4° of visual angle in the vertical plane and 5.8° of visual angle in the horizontal plane, and appeared on a gray background. The sounds, which were also 400 ms in duration (with 40-ms rise and fall periods), were presented at a comfortable listening level of approximately 75 dB SPL over two JBL speakers placed at either side of the computer monitor. The photographs and sounds were presented alone, and combined to form congruent pairs (e.g., dogs paired with barks) and incongruent pairs (e.g., dogs paired with the sounds of car engines), for a total of four stimulus types and 240 stimuli. Incongruent photographs and sounds were consistently paired throughout the experimental session: photographs of dogs were presented with sounds of car engines, photographs of cars were presented with strums of guitars, and photographs of guitars were presented with barks of dogs.

Because the inclusion of semantically incongruent audiovisual stimuli created a somewhat ecologically invalid situation where sounds might be actively suppressed, a second dataset was collected after their removal. It was hypothesized that the removal of incongruent audiovisual stimuli would decrease the suppression of task-irrelevant sounds, and thus increase the amplitude of the observed cross-sensory spread of attention through both the stimulus-driven and representation-driven processes. This second experiment included a total of three stimulus types (visual-alone, auditory-alone, and congruent audiovisual) and 180 stimuli.

Procedure

Participants were seated in a comfortable chair in a dimly lit and electrically shielded room and asked to keep head and eye movements to a minimum, while maintaining fixation on a central cross. Eye position was monitored with horizontal and vertical electro-oculogram (EOG) recordings. Visual-alone, auditory-alone, congruent audiovisual, and incongruent audiovisual stimuli were presented equiprobably and in pseudo-random order (incongruent audiovisual stimuli were excluded from the second experiment). Stimulus onset asynchrony varied randomly between 800 and 1100 ms, and a total of 180 stimuli were presented within blocks of approximately three minutes. The order of target-dog, target-car, and target-guitar blocks was pseudo-randomized within a total of 36 blocks (the second experiment included a total of 27 blocks). **Figure 14** shows a schematic of the experimental paradigm during a target-guitar block. During each block, participants were asked to click the left mouse button in response to the second of consecutively presented photographs of the target object (dogs, cars, or guitars), while ignoring all sounds. For example, when dogs were the target object, participants were instructed to make a button press response when the photograph of a dog was followed by the photograph of another (or the same) dog, regardless of intervening auditory-alone stimuli. The probability of target presentations was maintained at six percent.

Data Acquisition and Analysis

High-density continuous EEG recordings were obtained from a BioSemi ActiveTwo 168 channel system. Recordings were initially referenced online relative to a common mode active electrode and digitally sampled at 512 Hz. The continuous EEG was divided into epochs (-100 ms pre-stimulus to 500 ms post-stimulus onset), and baseline corrected from -100 ms to stimulus onset. Trials with blinks and eye movements were automatically rejected off-line on the basis of EOG recordings. An artifact rejection criterion of ± 100μV was used at all other scalp sites to reject trials with excessive EMG or other noise transients. To prevent contamination from motor responses associated with target detection and false alarms, all target trials (i.e., the second of consecutive trials that included an image of the target object) and false-alarm trials were discarded. Despite the removal of target trials, the use of a one-back task allowed for the examination of responses to target objects without contamination from motor responses. EEG epochs, sorted according to stimulus type (visual-alone, auditory-alone, congruent audiovisual, incongruent audiovisual) and target condition (target object, nontarget object), were averaged for each subject to compute ERPs. In the absence of any significant behavioral differences in response to stimuli within each of the object classes (dogs, cars, and guitars), ERPs in response to each stimulus type (4 levels) within each target condition (2 levels) were collapsed across object to increase the signal-to-noise ratio for all electrophysiological analyses. Across participants, there was an average of 355 accepted sweeps per condition. The waveforms were algebraically re-referenced to the nasion prior to data analysis. Separate group-averaged ERPs for each stimulus type within each target condition were generated for display purposes.

Statistical Analysis of Behavior

First experiment. To determine both whether participants successfully ignored the task-irrelevant sounds and whether there were differences in behavioral performance across objects (dogs, cars, or guitars), reaction times (RTs) and error patterns were examined. Separate two-way, repeated measures ANOVAs were conducted on RTs and hit rates (HRs), with factors of target object (3 levels: dogs, cars, or guitars) and target type (3 levels: visual-alone, congruent audiovisual, or incongruent audiovisual). Pairwise comparisons were made using planned protected t-tests (LSD). False alarms were examined to quantify the number of times each participant responded when an auditory representation of the target object was either presented alone or with an incongruent image (i.e., the image of a nontarget object). These errors, if frequent, would suggest that participants were unable to ignore task-irrelevant sounds.

Second Experiment. The analysis of behavioral results for the second experiment was identical to that from the first experiment, with the exception that there were only two target types (2 levels: visual-alone, congruent audiovisual) instead of three. Behavioral results were also compared across the experiments to determine whether the decreased task interference that resulted from the removal of incongruent audiovisual stimuli affected performance. To this end, between-groups ANOVAs with a factor of target type (2 levels: visual-alone, congruent audiovisual) were conducted on RTs and HRs.

Statistical Analysis of Electrophysiology

Based on findings from previous studies, there were specific hypotheses about the timing and location of the ERP effects under investigation. For the first experiment, amplitude data from seven frontocentral electrodes were averaged over a 100-ms latency window (see **Figure 15**). These seven frontocentral electrodes were chosen following an examination of the grand-averaged ERP waveforms in response to task-irrelevant auditory-alone stimuli. The selected electrodes had a strong auditory N1 response, and it was reasoned that the scalp topographies for the N1 component and the attention-related processing negativity (PN) are similar (e.g., Talsma et al., 2007). A latency window from 200 to 300 ms post stimulus was chosen because the cross-sensory spread of attention from an attended visual stimulus to a simultaneously presented task-irrelevant sound, whether through the stimulus-driven or representation-driven processes, had previously been shown to onset anywhere from 180 to 250 ms post stimulus (Busse et al., 2005; Molholm et al., 2007; Talsma et al., 2007).

For the second experiment, amplitude data from the same seven frontocentral electrodes were averaged over a shorter 50-ms latency window that was based on the timing of the cross-sensory spread of attention observed in the first experiment. This same 50-ms latency window was then used to make comparisons between the first and second experiments.

The following approaches were used to isolate the cross-sensory spread of attention-related processing that resulted from the *stimulus-driven* and *representation-driven* processes: (1) to isolate the stimulus-driven spread of attention, ERPs in response to task-irrelevant auditory representations of nontarget objects presented with attended images of nontarget objects (e.g., dogs with barks when guitars were the target object) were compared within blocks to ERPs in response to auditory-alone representations of nontarget objects, and (2) to isolate the representation-driven spread of attention, ERPs in response to task-irrelevant sounds presented with attended images of the target object (e.g., dogs with barks when dogs were the target object) were compared across blocks to ERPs in response to the same audiovisual stimuli when their visual component was a nontarget object (e.g., dogs with barks when guitars were the target object) [2]. **Figure 16** illustrates these experimental comparisons.

First experiment: (a) The Stimulus-Driven Spread of Attention. A one-way, repeated measures ANOVA, which was limited to ERPs in response to nontarget objects, was conducted with a factor of stimulus type (3 levels: auditory-alone, congruent audiovisual, incongruent audiovisual). The average amplitude (from 200 to 300 ms) of the response to an auditory-alone stimulus (A) was compared with the average amplitude of the response to an audiovisual stimulus minus the average amplitude of the response to its corresponding visual-alone stimulus (AV - V). Subtracting the visual-alone response

---

[2] Because the behavioral task required participants to identify images as either the target object or as a nontarget object, the visual stimuli were always attended. In comparison, the task-irrelevant sounds, which also included representations of both target and nontarget objects, were always unattended.

from the audiovisual response allows for the isolation of cross-sensory effects at auditory scalp sites (e.g., Busse et al., 2005; Molholm et al., 2007). That is, the electrical signal elicited in response to the visual stimulus may volume conduct across the scalp to areas where auditory evoked potentials are typically recorded, and thus the visual response associated with the presentation of an audiovisual stimulus could be misinterpreted as an enhanced (or reduced) auditory response.

First experiment: (b) The Representation-Driven Spread of Attention. A two-way, repeated measures ANOVA was conducted with factors of stimulus type (3 levels: auditory-alone, congruent audiovisual, incongruent audiovisual) and target condition (2 levels: target object, non-target object). To test for the representation-driven spread of attention in response to audiovisual stimuli, the auditory response (A^+V^+ - V^+ or A^-V^+ - V^+) to an audiovisual stimulus when its visual component was an image of the target object was compared with the auditory response (A^-V^- - V^-) to the same audiovisual stimulus when its visual component was an image of a nontarget object. For *congruent* audiovisual stimuli, for example, the auditory response (A^+V^+ - V^+) to the bark of a dog presented with the image of a dog when *dogs* were the target object, was compared with the auditory response (A^-V^- - V^-) to the same stimulus when either *cars* or *guitars* were the target object. For *incongruent* audiovisual stimuli, the auditory response (A^-V^+ - V^+) to the sound of a car engine presented with the image of a dog when *dogs* were the target object, was compared with the auditory response (A^-V^- - V^-) to the same stimulus when *guitars* were the target object.

Second Experiment. A nearly identical analysis of the electrophysiological data from the second experiment was undertaken to test for the cross-sensory spread of attention after the removal of incongruent audiovisual stimuli from the stimulus set (e.g., dogs paired with the sounds of car engines). To test for the *stimulus-driven* spread of attention, a one-way repeated measures ANOVA, which was limited to ERPs in response to nontarget objects, was conducted with a factor of stimulus type (2 levels: auditory-alone, congruent audiovisual). To test for the *representation-driven* spread of attention, a two-way, repeated measures ANOVA was conducted with factors of stimulus type (2 levels: auditory-alone, congruent audiovisual) and target condition (2 levels: target object, nontarget object).

Electrophysiological data were also compared across experiments to determine whether the removal of incongruent audiovisual stimuli, which presumably decreased the suppression of task-irrelevant sounds, increased the amplitude of the cross-sensory spread of attention. To this end, between-groups ANOVAs were conducted on the average amplitudes of the stimulus-driven (from 246 to 296 ms) and representation-driven (from 193 to 243 ms) processes in response to congruent audiovisual stimuli. The amplitude of the cross-sensory spread of attention, whether through the stimulus-driven or the representation-driven process, was expected to increase in the absence of interference from task-irrelevant sounds (i.e., with the removal of incongruent audiovisual stimuli).

An examination of the ERP waveforms representing the stimulus-driven spread of attention from the second experiment further revealed a potential window of earlier cross-sensory integration during the timeframe of the auditory N1 component at ~120 ms. A post hoc, one-way repeated measures ANOVA was conducted with a factor of stimulus type (auditory-alone, congruent audiovisual). This analysis, which was limited to ERP responses to nontarget objects, used a 20-ms latency window that was centered on the auditory N1 component.

Both Experiments. As with the behavioral analyses, all pairwise comparisons within the electrophysiological analyses were made using planned protected t-tests (LSD). For all of the statistical tests, both behavioral and electrophysiological, the alpha level was set at 0.05. Where appropriate, Greenhouse-Geisser corrections were made.

The onset of ERP modulations was determined through running t-tests between amplitude measurements at consecutive time points (e.g., Molholm et al., 2002; Wylie et al., 2003; Molholm et al., 2007). Onset was considered to be the first of ten consecutive, statistically significant time points. This criterion was chosen because the likelihood of getting ten false positives in a row is extremely low (Guthrie & Buchwald, 1991). The use of running t-tests with a strict criterion controls for the type I errors that result from multiple comparisons.

Scalp Topographies. To test whether attention-related processing from the stimulus-driven and representation-driven processes was attributable to different intracranial current sources, scalp topographies of the significant attention-related effects were compared. These comparisons were made using topography-normalized voltages (based on the amplitudes of the difference waves) from 20 frontocentral electrodes (**Figure 15**) (McCarthy & Wood, 1985). A two-way, repeated measures ANOVA was conducted with factors of cross-sensory attention effect (4 levels: a stimulus-driven spread of attention in response to congruent audiovisual stimuli, a stimulus-driven spread of attention in response to incongruent audiovisual stimuli, a representation-driven spread of attention in response to congruent audiovisual stimuli, a representation-driven spread of attention in response to auditory-alone stimuli) and anterior-posterior position (4 levels: frontal, frontocentral, central, and posterior). Significant interactions between the attention-related effects and the electrode-position factor would indicate significant differences among the scalp topographies and thus differences among the intracranial current sources (e.g., Talsma & Kok, 2001; Talsma et al., 2007).

RESULTS

Behavioral Results

First experiment. As shown in **Table 4**, reaction times (RTs) when images of guitars were the target were somewhat slower than RTs when images of either dogs or cars were the target (~14 ms slower). RTs were also somewhat slower in response to congruent audiovisual targets (i.e., when the visual target was paired with a semantically congruent, task-irrelevant sound) relative to RTs in response to either incongruent audiovisual targets or visual-alone targets (~ 8 ms slower). A repeated measures ANOVA with factors of target objects (3 levels: dogs, cars, guitars) and target types (3 levels: visual-alone, congruent audiovisual, incongruent audiovisual) revealed a significant main effect of target objects ($F_{2,22} = 9.0$, $p = 0.001$), while the main effect of target types did not reach significance ($F_{2,22} = 2.7$, $p = 0.09$). There were also no significant interactions between target objects and target types ($F_{4,44} = 0.6$, $p = 0.95$). Pairwise comparisons across the target objects revealed that RTs when images of guitars were the target were indeed significantly slower than RTs when images of either dogs ($p = 0.002$) or cars ($p = 0.006$) were the target. This difference might reflect the fact that participants were less familiar with guitars than dogs or cars.

Molholm, Ritter, Javitt, and Foxe (2004), who asked participants to simultaneously detect visual and auditory representations of a specific well-known animal (e.g., a dog), reported that RTs in response to congruent audiovisual targets were faster than those in response to both incongruent audiovisual targets (e.g., a barking cat when *cats* were the target object) and visual-alone targets. The absence of this trend in the present data was interpreted as evidence that participants had successfully ignored the task-irrelevant sounds. This interpretation was further supported by the negligible number of false alarms in response to auditory-alone representations (e.g., a bark) of the visual target (e.g., the image of a dog) or false alarms in response to auditory representations of the visual target paired with incongruent images of nontargets (e.g., the image of a guitar). Across all twelve participants, less than one percent of responses were categorized as false alarms, and only three false alarms fit the above description.

The hit rate (HR) collapsed across target objects and target types was 93 percent, and varied little within or across these factors (Table 1). A repeated measures ANOVA revealed no significant main effects or interactions, indicating that task performance was approximately equivalent across target objects ($F_{2,22} = 0.2$, $p = 0.81$) and target types ($F_{2,22} = 0.1$, $p = 0.89$). The absence of significant differences among HRs, as well as the similarity among RTs, served as justification to collapse the ERPs in response to each stimulus type (auditory-alone, visual-alone, congruent audiovisual, incongruent audiovisual) across the object classes (dogs, cars, guitars), which increased the signal-to-noise ratio for all electrophysiological analyses.

Second Experiment. The behavioral results from the second experiment, which did not include incongruent audiovisual stimuli (e.g., dogs paired with sounds of car engines), revealed no significant main effects or interactions for RTs or HRs. That is, both RTs and HRs were approximately equivalent across target objects ($F_{2,14} = 0.09, p = 0.92$; $F_{2,14} = 1.32, p = 0.30$) and target types ($F_{1,7} = 0.10, p = 0.78$; $F_{1,7} = 0.22, p = 0.66$). Similar to the results from the first experiment, less than one percent of the responses across all ten participants were categorized as false alarms, and only two false alarms occurred in response to an auditory-alone representation of the visual target.

Table 5 compares the behavioral results from the first and second experiments. It was reasoned that the inclusion of incongruent audiovisual stimuli in the first experiment would create a somewhat ecologically invalid situation where task-irrelevant sounds might be actively suppressed to prevent them from interfering with the behavioral task. In support of this notion, performance improved when task interference from task-irrelevant sounds decreased (i.e., when incongruent audiovisual stimuli were removed from the stimulus set). Between-groups ANOVAs revealed that RTs from the second experiment, relative to those from the first experiment, were significantly faster in response to both visual-alone targets ($F_{1,20} = 4.80, p = 0.04$) and congruent audiovisual targets ($F_{1,20} = 7.04, p = 0.02$). In addition to these faster RTs, HRs from the second experiment were also higher in response to both visual-alone targets and congruent audiovisual targets. Between-groups ANOVAs, however, revealed that HR differences across the experiments did not reach statistical significance ($F_{1,20} = 2.81$, $p = 0.11$; $F_{1,20} = 3.24$, $p = 0.09$).

Electrophysiological Results

First experiment: a) The Stimulus-Driven Spread of Attention. The ERP waveforms shown in **Figure 17** illustrate the stimulus-driven spread of attention that occurs whenever a task-irrelevant sound is paired with an attended visual stimulus. A repeated measures ANOVA with a single factor of stimulus type (auditory-alone, congruent audiovisual, incongruent audiovisual), which was limited to ERPs in response to nontarget objects, revealed a significant main effect ($F_{2,22} = 7.1$, $p = 0.01$). Pairwise comparisons across the stimulus types further revealed that ERPs in response (from 200 to 300 ms) to congruent and incongruent audiovisual stimuli (AV - V) were significantly more negative than ERPs in response to auditory-alone stimuli ($p = 0.03$; $p = 0.01$). In comparison, the ERPs in response to congruent and incongruent audiovisual stimuli (AV - V, limited to ERPs in response to nontarget objects) did not differ significantly ($p = 0.48$). Running t-tests revealed that the onset of the stimulus-driven spread of attention in response to congruent and incongruent audiovisual stimuli was also approximately equivalent, starting at 246 ms and 236 ms, respectively. These data indicate that the stimulus-driven spread of attention from an attended visual stimulus to its paired, task-irrelevant sound is independent of highly learned associations among a well-known object's multisensory features.

First experiment: b) The Representation-Driven Spread of Attention. The ERP waveforms shown in **Figure 18** illustrate the representation-driven spread of attention, hypothesized to result from the activation of a cortical representation established through a lifetime of experiences with an object. A repeated measures ANOVA with factors of stimulus type (3 levels: auditory-alone, congruent audiovisual, incongruent audiovisual) and target condition (2 levels: target object, nontarget object) revealed both significant main effects ($F_{2,22} = 4.8, p = 0.02$; $F_{1,11} = 5.6, p = 0.04$) and a significant interaction ($F_{2,22} = 4.2, p = 0.02$). Pairwise comparisons between target conditions (2 levels: target object, nontarget object) within the stimulus types revealed that ERPs in response to auditory-alone stimuli ($p = 0.04$) and ERPs (AV - V) in response to congruent audiovisual ($p = 0.04$) stimuli were significantly more negative when those stimuli were representations of the visual target than when those stimuli were representations of visual nontargets. However, ERPs (AV - V) in response to incongruent audiovisual stimuli did not significantly differ across the target conditions ($p = 0.91$). These data indicate that the representation-driven spread of attention depends on pre-existing, highly learned associations among a well-known object's multisensory features. As in Molholm et al. (2007), the representation-driven spread of attention occurred somewhat earlier when task-irrelevant sounds were paired with a visual target, starting at 193 ms, compared to 236 ms when task-irrelevant sounds were presented alone.

First experiment: Scalp Topographies. The voltage maps associated with the cross-sensory spread of attention that results from both the stimulus-driven and representation-driven processes (see **Figures 17** and **18**) resembled the distribution of the late frontal negativity typically observed in response to an attended sound (a PN or an Nd component [Näätänen et al., 1978, Hansen & Hillyard, 1980]). Previous studies have reported similar scalp topographies corresponding to the stimulus-driven spread of attention from an attended visual stimulus to its paired, task-irrelevant sound (Busse et al., 2005; Talsma et al., 2007), and corresponding to the representation-driven spread of attention from a visual target to a semantically congruent, task-irrelevant sound (Molholm et al., 2007). For the data presented here, a comparison of the scalp topographies was conducted to test whether different underlying current sources generated the stimulus-driven and representation-driven processes. For this analysis, significant interactions between the attention-related effects and electrode positions would indicate significant differences among the scalp topographies; however, a two-way, repeated measures ANOVA conducted with attention effects (4 levels: a stimulus-driven spread of attention in response to congruent audiovisual stimuli, a stimulus-driven spread of attention in response to incongruent audiovisual stimuli, a representation-driven spread of attention in response to congruent audiovisual stimuli, a representation-driven spread of attention in response to auditory-alone stimuli) and anterior-posterior position (4 levels: frontal, frontocentral, central, and posterior) as factors, revealed no significant interactions ($F_{9,99} = 1.9$, $p = 0.12$).

Second Experiment. Repeated measures ANOVAs revealed that both mechanisms leading to the cross-sensory spread of attention remained significant when incongruent audiovisual stimuli were removed from the stimulus set. For the stimulus-driven spread of attention, a within blocks analysis limited to stimuli representing nontarget objects revealed that ERPs in response (from 246 to 296 ms) to congruent audiovisual stimuli (AV - V) were significantly more negative than ERPs in response to auditory-alone stimuli ($F_{1,9} = 11.6$, $p = 0.01$). For the representation-driven spread of attention, a between blocks analysis that included all stimuli (i.e., stimuli representing either target or nontarget objects) revealed that ERPs in response (from 236 to 286 ms) to auditory-alone representations of the visual target and ERPs in response (from 193 to 243 ms) to congruent audiovisual targets (AV - V) were both significantly more negative than ERPs in response to the same stimuli when those stimuli were representations of visual nontargets ($F_{1,9} = 12.9$, $p = 0.01$).

The ERP waveforms shown in **Figure 19** illustrate the combined, cross-sensory spread of attention (i.e., from both attention-related mechanisms) in response to a congruent audiovisual stimulus when its visual component was a target. A comparison between the electrophysiological results from the first and second experiments, which used the shorter 50-ms latency windows, indicated that the presence of incongruent audiovisual stimuli diminished the stimulus-driven spread of attention from an attended visual stimulus to its paired, task-irrelevant sound ($F_{1,20} = 4.3$, $p = 0.05$). Counter to our expectations, the presence of incongruent audiovisual stimuli did not diminish the representation-driven spread of attention ($F_{1,20} = 0.0$, $p = 0.87$).

An examination of the ERP waveforms displayed in **Figure 19b** (recorded when incongruent audiovisual stimuli were removed from the stimulus set) also suggests the possible emergence of an earlier window of cross-sensory integration during the timeframe of the auditory N1 component (at ~120 ms). A post hoc, one-way repeated measures ANOVA (limited to nontarget objects) with a single factor of stimulus type (auditory-alone, congruent audiovisual), however, only approached statistical significance ($F_{1,9} = 4.2$, $p = 0.07$).

DISCUSSION

Research from the visual domain strongly suggests an inherent bias to process objects as wholes, even when attention is focused on constituent features (e.g., Blaser et al., 2000; Egly et al., 1994; Martinez et al., 2006; O'Craven et al., 1999; Schoenfeld et al., 2003; Stroop, 1935). Both brain imaging and electrophysiological studies have demonstrated an object-based spread of attention to feature-representations in anatomically distributed cortical regions (e.g., O'Craven et al., 1999; Schoenfeld et al., 2003). These studies collectively suggest that attention-related processing serves as a mechanism through which visual features are bound to form the perception of a coherent object (e.g., Desimone & Duncan 2006; Duncan, 2006; Robertson, 2003; Serences & Yantis, 2006). Although such models for feature binding are typically discussed in the context of an object's visual features, there is no obvious reason why the binding of an object's multisensory features would not similarly occur through attentional mechanisms. Indeed, recent research suggests that the whole-object bias, which seems to operate through the spread of attention, extends across sensory boundaries (Busse et al., 2005; Molholm et al., 2007; Talsma et al., 2007).

Here we used electrophysiological measures of selective attention to investigate the cross-sensory spread of attention from an attended visual stimulus to a task-irrelevant sound, examining the interplay between bottom-up and top-down attentional processes. More specifically, we investigated whether highly learned associations among the multisensory features of a well-known object, formed through a lifetime of experiences with that object, modulate the spread of attention that contributes to the binding of its multisensory features.

Our findings distinguish dual mechanisms for the object-based cross-sensory spread of attention: (1) a bottom-up, stimulus-driven spread of attention that seems to be unaffected by highly learned associations among the multisensory features of a well-known object, and (2) a top-down, representation-driven spread of attention that is dependent on highly learned associations among the multisensory features of a well-known object. We propose that the stimulus-driven spread of attention reflects an inherent bias to process objects as wholes, with its induction based entirely on the temporal and spatial properties of a multisensory object's features. The representation-driven spread of attention, on the other hand, only occurs in response to sounds that are semantically related to a visual target, reflecting the activation of a preexisting cortical representation of a well-known object that includes the object's multisensory features. We also find that when a semantically congruent, task-irrelevant sound is simultaneously presented with a visual target, the effects of the stimulus-driven and representation-driven processes appear to be additive (see **Figure 19**).

The Representation-Driven Spread of Attention

Molholm et al. (2007) used well-known multisensory objects to demonstrate a cross-sensory spread of attention that occurred only when a task-irrelevant sound was semantically congruent with a visual target. It was hypothesized that this spread of attention occurred because selectively attending for an object in the visual modality led to the activation of a preexisting object representation that included the highly-associated multisensory features of the visual target. The first goal of the present study was to establish whether this “representation-driven” spread of attention in fact results from long-term, semantically based associations among an object’s multisensory feature-representations, or instead from associations formed as a result of repeated pairings during the experimental session, as Molholm et al. (2007) used only a single exemplar for each of the repeatedly presented multisensory objects.

We devised the following approach to limit the formation of associations between specific audiovisual features during the experimental session: (1) multiple exemplars were used within each object class so participants were forced to use a conceptual definition of the visual target rather than rely on the repetition of specific features (lines and curves at specific locations), and (2) sounds were task irrelevant throughout the experimental session to avoid task carry-over effects. Despite these controls, there was significant attention-related processing of task-irrelevant sounds when those sounds were semantically congruent with the visual target, both when the sounds were presented with the visual target and when the sounds were presented alone. The absence of an attentional spread to semantically incongruent task-irrelevant sounds when paired with a visual target, despite the consistent pairing of mismatched sounds and images on incongruent trials (e.g., the sound of a car engine was consistently paired with the image of a dog), further demonstrates that experiment-specific associations do not account for the representation-driven spread of attention.

The above findings point to a process whereby a lifetime of experiences with an object (e.g., a dog) establishes strong associations among its cortically distributed multisensory representations, such that the activation of its visual feature-representation leads to the activation of a distributed network of feature-representations. In other words, the sustained activation of a well-known object's visual feature-representation also creates a processing advantage for incoming sensory signals that match that object's highly associated nonvisual feature-representations. This biasing towards the constituent features of a well-known object, including those features that are irrelevant to the task at hand, could conceivably contribute both to binding multisensory objects and to resolving competitions among the multitude of objects simultaneously present in a real-world environment (Desimone & Duncan, 1995; Duncan, 2006).

It is of note that the representation-driven spread of attention that occurs when a semantically congruent, task-irrelevant sound is simultaneously presented with a visual target does not elicit a *sustained* response as seen for the representation-driven spread of attention in response to an auditory-alone stimulus and the stimulus-driven spread of attention. It might therefore be argued that this ERP modulation instead reflects a non-attention-related cognitive process. Based on the following rationale, however, we conclude that the representation-driven effect is indeed associated with attention-related processing: (1) an apparent reduction in the amplitude of the auditory P2 component is a constituent of the PN and a well-documented marker for attention-related processing in general (e.g., Hansen & Hillyard, 1980; Näätänen et al., 1978; Crowley et al., 2004), (2) attention effects are not necessarily sustained (e.g., Hansen & Hillyard, 1980), and (3) known electrophysiological markers of cognitive processes that are sensitive to the semantic content of the stimulus onset at ~300 ms at the earliest (the class of so-called "N400" effects: e.g., Kutas & Hillyard, 1980; Molholm et al., 2004). To elaborate on this second point, a more negative ERP (at ~300 ms or later) has previously been associated with the processing of semantically *incongruent* stimuli, relative to the processing of semantically *congruent* stimuli. In the present findings, it is instead the semantically congruent stimuli that evoke a more negative response (**Figure 18b**). One possibility that arises from examination of the waveforms in this condition is that the attention effect does not appear to be sustained because there is an additional process superimposed in the later timeframe; perhaps a process related to multisensory integration.

The Stimulus-Driven Spread of Attention

Previous research has also demonstrated a non-target-specific, cross-sensory spread of attention that occurs whenever a task-irrelevant sound is simultaneously presented with an attended visual stimulus (Busse et al., 2005; Talsma et al., 2007). This “stimulus-driven” spread of attention has been hypothesized to reflect bottom-up feature binding. Previous research, however, has not addressed whether the multisensory features of frequently encountered objects become “hard-wired,” such that this bottom-up attentional spread is facilitated for well-known multisensory objects. Another goal of the present study was thus to determine whether the stimulus-driven spread of attention is modulated by highly learned feature-associations.

To this end, we compared the stimulus-driven spread of attention from an attended visual stimulus to its paired task-irrelevant sound when that sound was either object-congruent (e.g., barks paired with dog images) or object-incongruent (e.g., the sounds of car engines paired with dog images). If highly learned associations affect bottom-up feature binding through the stimulus-driven spread of attention, we would expect to observe enhanced processing when an attended visual stimulus is paired with a semantically congruent task-irrelevant sound, relative to when the same visual stimulus is paired with a semantically *incongruent* task-irrelevant sound. On the contrary, our findings indicate that the stimulus-driven spread of attention is independent of highly learned associations among multisensory features: the stimulus-driven spread of attention from the image of a dog to the congruent, task-irrelevant bark of a dog, for example, was equivalent to the stimulus-driven spread of attention from the image of a dog to the incongruent, task-irrelevant sound of a car engine (see **Figure 17**). This leads us to conclude that the brain's default mode is to bind multisensory features based on their temporal and spatial properties, without regard for the learned associations (or lack thereof) among those features. This accords well with subjective experience, where a barking cat would be odd, but the multisensory features would still be perceived as a coherent object.

The Combined Cross-Sensory Spread of Attention

The data presented here demonstrate that the stimulus-driven and representation-driven processes are separate mechanisms through which attention spreads across sensory modalities to encompass an object's task-irrelevant nonvisual features. The first line of evidence is that the stimulus-driven spread of attention occurs in response to both semantically congruent and semantically incongruent multisensory objects, whereas the representation-driven spread of attention occurs only in response to semantically congruent multisensory objects. Further evidence that these are distinct processes is gained from another defining characteristic: unlike the stimulus-driven spread of attention, the representation-driven spread of attention does not require the co-occurrence of an attended visual stimulus.

Despite being separate mechanisms, the stimulus-driven and representation-driven processes seem to have similar outcomes: both processes lead to a cross-sensory spread of attention with overlapping latencies and scalp topographies. Visual examination of the scalp topographies associated with stimulus-driven and representation-driven processes does reveal a somewhat more anterior fronto-central distribution for the stimulus-driven attentional spread (**Figures 17** and **18**), however, suggesting the possibility of different underlying neural generators. But these differences between the scalp distributions were not statistically significant, and another possibility is that the smaller effect size associated with the representation-driven spread of attention simply yielded a noisier distribution. It is also of note that in Molholm et al. (2007), where the representation-driven spread of attention was of greater amplitude, the corresponding scalp distribution appeared to be somewhat more anterior (see Figure 3 in Molholm et al., 2007), in line with the more anterior distribution of the stimulus-driven spread of attention seen here and elsewhere (Busse et al., 2005; Talsma et al., 2007). Nevertheless, such eye-ball comparisons are far from definitive, and neuroimaging studies would be much better suited to detecting subtle differences in the underlying neural generators of these two processes.

Since the inclusion of semantically incongruent multisensory objects in the stimulus set created a somewhat ecologically invalid situation where sounds might be actively suppressed, we collected a second dataset after the removal of incongruents from the stimulus set. This second dataset was meant to test whether the inclusion of semantically incongruent multisensory objects in the present study had led to the smaller effect sizes associated with the representation-driven spread of attention relative to the greater effect sizes observed by Molholm et al. (2007), where only semantically congruent multisensory objects were presented. While the removal of incongruents from the stimulus set increased the amplitude of the stimulus-driven spread of attention, the amplitude of the representation-driven spread of attention was statistically equivalent across the two datasets collected here (see **Figure 19**). Instead, it might have been the inclusion of multiple exemplars of each well-known object in the present study that led to the smaller-amplitude representation-driven spread of attention. The inclusion of multiple exemplars required participants to decode a given stimulus as a size-invariant, position-invariant member of an object class (e.g., “a dog”), rather than as a specific object (e.g., “the dog”). Molholm et al. (2007), on the other hand, used a single exemplar to represent each well-known object. The repetition of a single exemplar, which constrained the representation-set that needed to be matched to determine the presence or absence of a target on each trial, might have facilitated the decoding of the task-irrelevant sound’s identity. This easier and more consistent decoding of the task-irrelevant sound might have led to a larger-amplitude representation-driven spread of attention.

A Model for the Cross-Sensory Spread of Attention and Concluding Remarks

Our findings suggest a model of attentional spread in which object-based selection following spatial selection leads to the *stimulus-driven* spread of attention, independent of semantic processing: (1) spatial selection determines the relevant location for further processing (Treisman & Gelade, 1980), (2) attention spreads within the visual boundaries of the object (object-based selection [Egly et al., 1994; Martinez et al., 2006; Martinez et al., 2007]), and (3) attention spreads to coincident multisensory features that fall within the visual boundaries established through object-based selection. This model for the stimulus-driven spread of attention from an attended visual stimulus to its task-irrelevant multisensory features is based on both the timing of the associated ERP effects, and the apparent dominance of vision in feature integration and object recognition (e.g., Amedi et al., 2005; Greene et al., 2001; James et al., 2002; Molholm et al., 2004; Molholm et al., 2007; Talsma et al., 2007). First, our results, as well as those of previous studies, show that the stimulus-driven spread of attention from an attended visual stimulus to its paired task-irrelevant sound onsets between 220 and 250 ms post stimulus: a timeframe that (1) follows the spread of attention within the visual boundaries of an object, which has been shown to onset at ~160 ms post stimulus (e.g., Martinez et al., 2006), and (2) is similar to the timeframe for the spread of attention to task-irrelevant visual features, which has been shown to also onset between 220 and 250 ms post stimulus (e.g., Schoenfeld et al., 2003). Second, several previous studies have shown that the stimulus-driven spread of attention from an attended sound to its paired task-irrelevant visual stimulus is either significantly reduced in magnitude relative to the visual-to-auditory spread of attention, or altogether non-existent (e.g., Molholm et al., 2007; Talsma et al., 2007; Teder-Sälejärvi et al., 1999). Talsma et al. (2007), for example, showed no stimulus-driven spread of attention from an attended tone to its paired, task-irrelevant horizontal grating (i.e., during an attend-auditory condition), but did show a significant stimulus-driven spread of attention from an attended horizontal grating to its paired, task-irrelevant tone (i.e., during an attend-visual condition).

The model presented here posits that induction of the stimulus-driven spread of attention occurs whenever a task-irrelevant sound overlaps an attended visual stimulus in time and space, but previous findings might seem to challenge this assertion. Van der Burg et al. (2008a), for example, have demonstrated that a spatially non-informative auditory stimulus guides attention towards a synchronized visual event within a cluttered, continuously changing visual environment. The spatially non-informative auditory stimulus increases the salience of the visual target, leading to attentional capture. This shows that some multisensory integration effects might not be constrained by space. Data from our laboratory demonstrated that auditory-somatosensory interactions starting at approximately 50 ms poststimulus were not affected by the spatial alignment of the unisensory signals (Murray et al., 2005). However, in comparison to the later multisensory effects described in the present study (i.e., the spread of attentional processing from an attended visual stimulus to a task-irrelevant sound), these earliest multisensory integration effects seem likely to contribute to signal detection rather than directly contributing to higher-order multisensory object processing (see **Chapters 1** and **2**). Busse et al. (2005), on the other hand, made clever use of the ventriloquism illusion, whereby perceptual localization of a sound is captured by a simultaneously presented but spatially separated visual stimulus, to attribute the stimulus-driven spread of attention to an object-based late selection process (an ERP modulation starting approximately 220 ms poststimulus). This spread of attention-related processing across modalities and space, occurring during the timeframe of the effects we describe in the present study, may initially seem incompatible with our model for the stimulus-driven spread of attention, which states that attention spreads to task-irrelevant multisensory features that fall within the visual boundaries established through object-based selection. That is, Busse et al. (2005) did not present the task-irrelevant sound at the same physical location as the attended visual stimulus. However, induction of the ventriloquism illusion, whereby the localization of the sound was drawn toward the location of the visual stimulus, indicates that subjectively the task-irrelevant sound did fall within the visual boundaries of the object. According to this interpretation, the findings of Busse et al. (2005) fit well within our model for the stimulus-driven spread of attention.

Unlike the stimulus-driven spread of attention, the *representation-driven* spread of attention relies on the semantic properties of the stimulus. The activation of a well-known object's visual representation leads to the activation of its highly associated auditory representations, and a stimulus that partially matches an activated auditory representation will receive enhanced processing even when it is presented alone (i.e., without a semantically congruent visual target). Because the representation-driven spread of attention occurs in response to an auditory-alone stimulus, it cannot be strictly considered a vehicle for binding the simultaneously presented multisensory features of an object into a coherent whole. But when a semantically congruent, task-irrelevant sound (e.g., a bark) is paired with a visual target (e.g., the image of a dog when *dogs* are the visual target), the cross-sensory spread of attention that results from the stimulus-driven and representation-driven processes appears to be additive (**Figure 19**), suggesting that the representation-driven spread of attention can contribute to multisensory feature binding.

Our model implies that object processing (object-based selection) leads to the stimulus-driven spread of attention, following the selection of a relevant location (spatial selection). An alternative model, based entirely on the supramodal properties of spatial attention, would posit that spatial selection leads to the stimulus-driven spread of attention independent of object-based selection (Hillyard et al., 1984; Eimer & Shroger, 1998; Teder-Sälejärvi et al. 1999; McDonald et al., 2000; McDonald et al., 2003). In contrast to the stimulus-driven spread of attention, we hypothesize that the top-down, representation-driven spread of attention operates independently of spatial selection. In this case, a task-irrelevant sound that is semantically congruent with the visual target will receive enhanced processing even when it is presented at an unattended location. (For examples of visual features receiving enhanced processing outside the spotlight of spatial selection, see the following: Chelazzi, Miller, Duncan & Desimone, 1993; Chelazzi, Duncan, Miller & Desimone, 1998; Treue & Martinez-Trujillo, 1996; Saenz, Buracas & Boynton, 2003). Such findings would indicate that induction of the representation-driven spread of attention, antipodal to the stimulus-driven spread of attention, is relatively independent of a multisensory object's temporal and spatial properties but dependent on its semantic properties. The resolution of these remaining issues will further expand our understanding of the interplay between attentional deployment and object processing that leads to multisensory feature binding.

Figure 14.

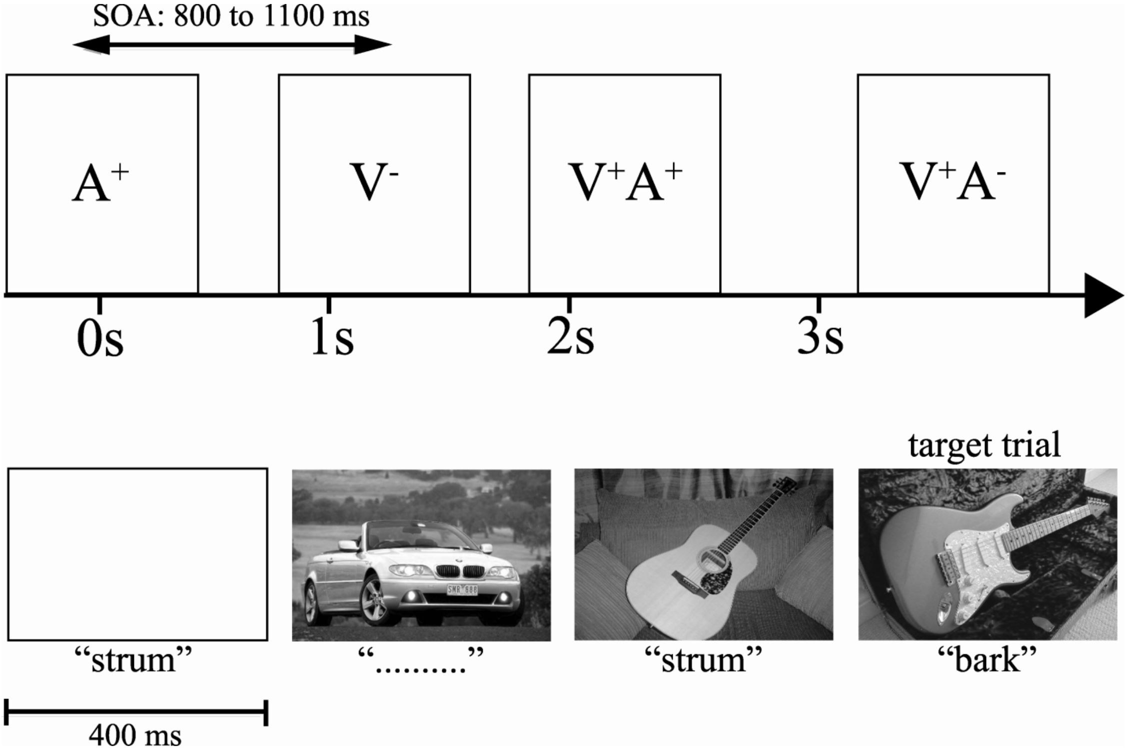

SOA: 800 to 1100 ms
A^+
V^-
V^+A^+
V^+A^-
0s
1s
2s
3s
target trial
"strum"
".........."
"strum"
"bark"
400 ms

Figure 14. A schematic of the experimental design during a target-guitar block. A guitar sound (A^+), the image of a car (V^-), the image of a guitar paired with the sound of a guitar (V^+A^+), the image of a guitar paired with the sound of a dog (V^+A^-).

Figure 15.

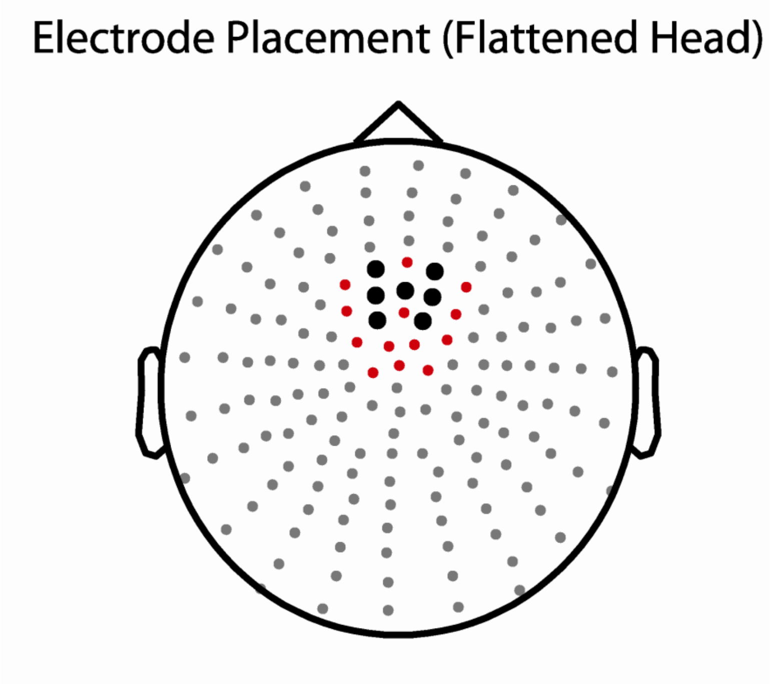


Figure 15. For all statistical analyses of ERPs, amplitude data was averaged across seven fronto-central electrodes (highlighted in black). An additional thirteen electrodes were included in the analysis of scalp topographies (highlighted in red).

Figure 16.

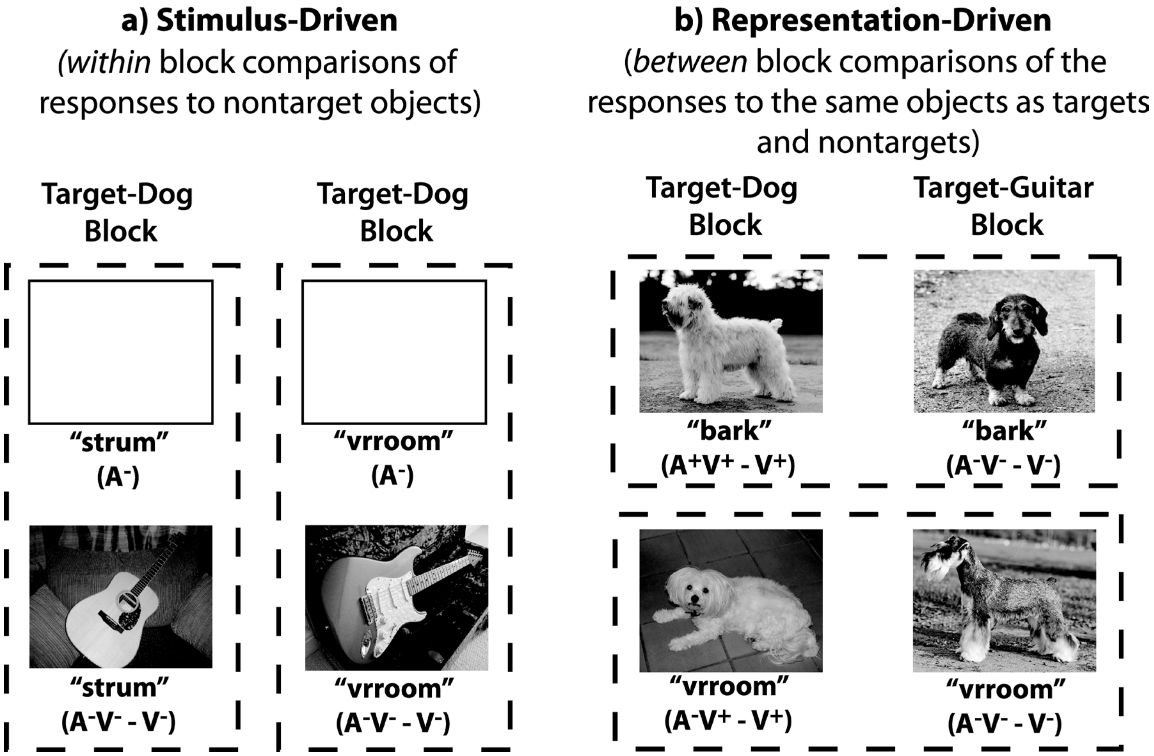

a) Stimulus-Driven
(*within* block comparisons of responses to nontarget objects)
Target-Dog Block
Target-Dog Block
"strum"
(A$^-$)
"vrroom"
(A$^-$)
"strum"
(A$^-$V$^-$ - V$^-$)
"vrroom"
(A$^-$V$^-$ - V$^-$)
b) Representation-Driven
(*between* block comparisons of the responses to the same objects as targets and nontargets)
Target-Dog Block
Target-Guitar Block
"bark"
(A$^+$V$^+$ - V$^+$)
"bark"
(A$^-$V$^-$ - V$^-$)
"vrroom"
(A$^-$V$^+$ - V$^+$)
"vrroom"
(A$^-$V$^-$ - V$^-$)

Figure 16. Examples of the experimental comparisons used to isolate the (**a**) stimulus-driven and (**b**) representation-driven processes. Responses to target and nontarget objects were collapsed across the object classes (dog, cars, and guitars).

Table 5.

Average reaction times and percent hits sorted by target objects and target types.

	Dogs			**Cars**			**Guitars**		
	V^+	V^+A^+	V^+A^-	V^+	V^+A^+	V^+A^-	V^+	V^+A^+	V^+A^-
RTs	452(71)	461(66)	452(69)	454(70)	463(59)	456(71)	468(72)	475(65)	469(78)
%Hit	93(7)	94(8)	94(6)	93(5)	93(8)	94(6)	94(7)	93(7)	92(7)

Notes. Visual-alone (V^+), congruent audiovisual (V^+A^+), incongruent audiovisual (V^+A^-). All times in milliseconds. Standard deviations in parentheses.

Table 6.

Average reaction times and percent hits sorted by experiment and target types.

	First Experiment		Second Experiment	
	V^+	V^+A^+	V^+	V^+A^+
RTs	457(69)	461(65)	398(40)	398(43)
%Hits	93(6)	93(7)	97(2)	97(2)

Notes. Visual-alone (V^+), congruent audiovisual (V^+A^+). All times in milliseconds. Standard deviations in parentheses.

Figure 17.

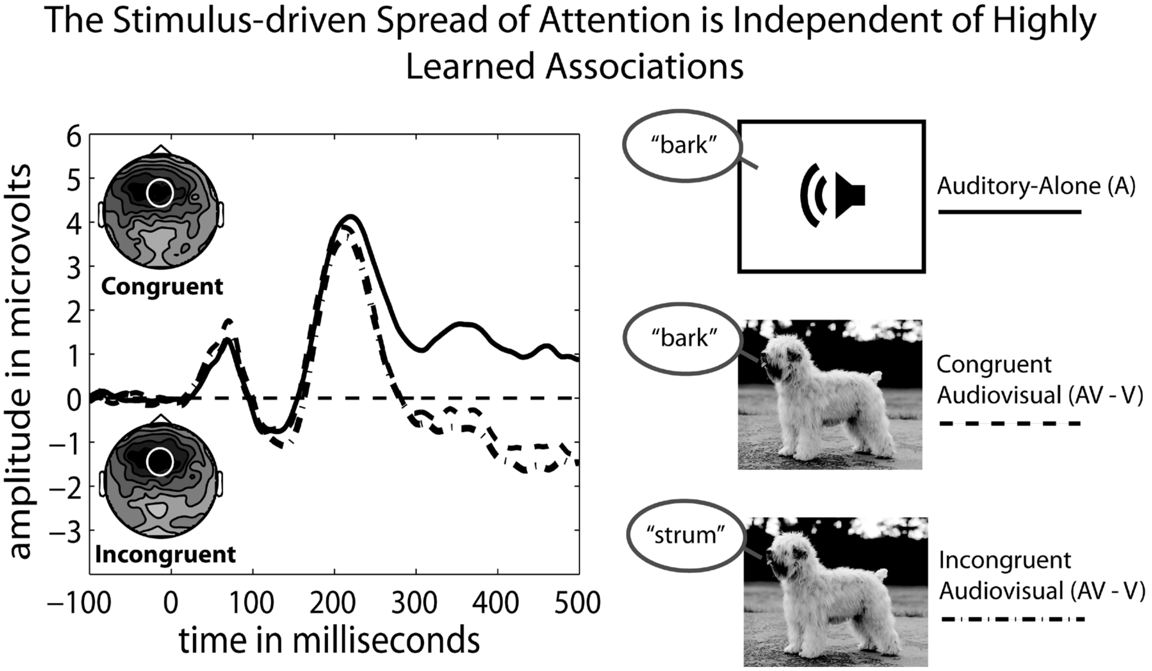

The Stimulus-driven Spread of Attention is Independent of Highly Learned Associations
amplitude in microvolts
6
5
4
3
2
1
0
−1
−2
−3
Congruent
Incongruent
−100
0
100
200
300
400
500
time in milliseconds
"bark"
Auditory-Alone (A)
"bark"
Congruent Audiovisual (AV - V)
"strum"
Incongruent Audiovisual (AV - V)

Figure 17. The stimulus-driven spread of attention is independent of highly learned associations. The stimulus-driven spread of attention from an attended visual stimulus to a paired congruent or incongruent, task-irrelevant sound. In both cases, ERPs were limited to those in response to nontarget objects. Flattened voltage maps were derived from ERP difference waves.

Figure 18.

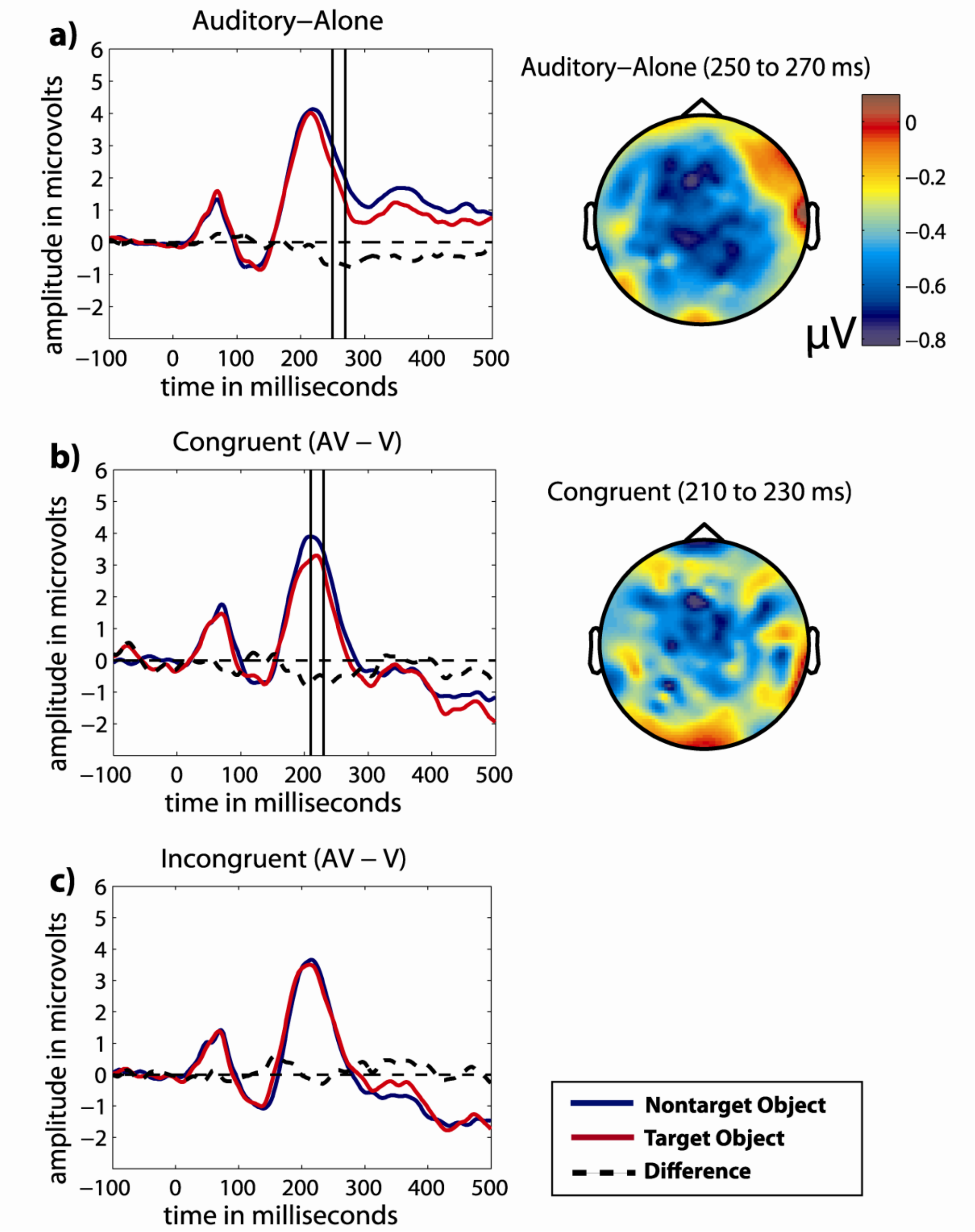
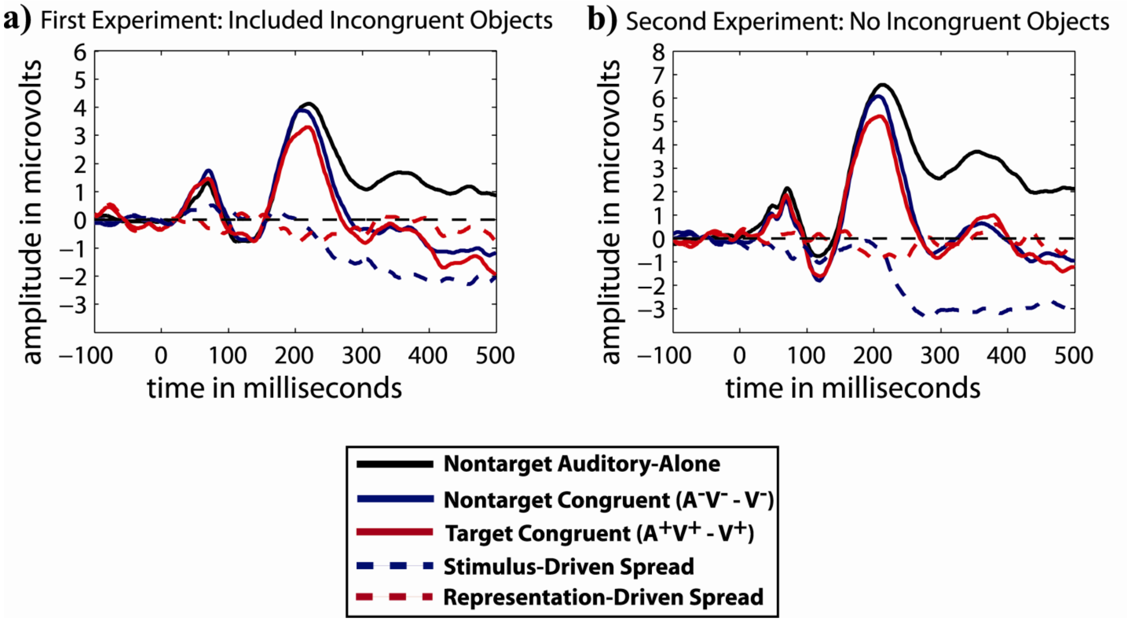

a)
Auditory–Alone
amplitude in microvolts
time in milliseconds
6
5
4
3
2
1
0
−1
−2
−100
0
100
200
300
400
500
Auditory–Alone (250 to 270 ms)
0
−0.2
−0.4
−0.6
−0.8
μV
b)
Congruent (AV – V)
Congruent (210 to 230 ms)
c)
Incongruent (AV – V)
Nontarget Object
Target Object
Difference

Figure 18. The representation-driven spread of attention is dependent on highly learned associations. The representation-driven spread of attention that occurs in response to a task-irrelevant sound that is semantically congruent with a visual target. ERP waveforms are shown for **(a)** auditory-alone, **(b)** congruent audiovisual, and **(c)** incongruent audiovisual stimuli. Flattened voltage maps, derived from the ERP difference waves, are shown for **(a)** auditory-alone and **(b)** congruent audiovisual stimuli.

Figure 19.



Notes. The combined cross-sensory spread of attention from the **(a)** first and **(b)** second experiments.

GENERAL DISCUSSION

The preceding chapters investigated multisensory interactions at two stages of cortical processing. An overview of our results suggests the following conclusions: (1) at the onset of cortical processing, multisensory interactions seem to be largely dependent on temporal relationships among the stimuli, without regard for their spatial alignment (see **Chapters 1** and **2**), (2) during cortical processing that contributes to feature integration, multisensory interactions seem to be dependent on both temporal and spatial relationships among the stimuli (e.g., Bonath et al., 2007; Donohue et al., 2011; see **Chapter 3**),[3] and (3) at both stages of processing, top-down attentional biasing has a powerful influence on multisensory interactions, meaning that strategic goals are able to shape how the various senses interact (see **Chapters 2** and **4**). Because these conclusions are based entirely on the neural mechanisms that we investigated in the preceding chapters, there are likely to be instances of multisensory integration that violate them. Either way, our results emphasize that the rules for multisensory interactions depend on neurophysiologic and neuroanatomic underpinnings, as well as strategic goals.

The rules for multisensory interactions are not set in stone. Traditional views of multisensory integration have focused on the importance of spatial and temporal alignment (Meredith & Stein, 1983; Meredith & Stein, 1986; Meredith et al., 1987), but here our results demonstrate that the senses sometimes interact despite wide spatial and temporal misalignments. That is, auditory facilitation of visual-target detection persists

---

[3] In Chapter 4 we hypothesize that the representation-driven spread of attention, which is dependent on highly learned associations, might occur regardless of spatial alignment. This mechanism for the cross-sensory spread of attention might therefore be an exception to the above statement.

despite spatial separations greater than 100 degrees on the horizontal meridian (**Chapter 1**), and despite temporal separations on the order of seconds (**Chapter 2**). Whereas our results are purely behavioral, previous studies have suggested that such violations of the traditional rules for multisensory integration (as first described by Barry Stein and colleagues) can occur in early sensory cortices near the onset of cortical processing (Lakatos et al., 2007; Lakatos et al., 2009; Murray et al., 2005). The conditions, under which these multisensory interactions occur, regardless of the stage of cortical processing, provide important clues about their specific neuroanatomic and neurophysiologic underpinnings. For example, spatial mapping, which is known to exist in early sensory cortices, does not seem to be preserved in the neuroanatomic connections that support the auditory facilitation of visual-target detection, perhaps suggesting a role for non-specific thalamocortical projections (see Jones, 1998). In comparison, the relative importance of spatial alignment in multisensory feature integration suggests the possible involvement of parietal cortex, where co-registration of multisensory spatial representations is thought to occur (Bushara et al., 1999; Cohen & Andersen, 2004; Linden, Grunewald & Andersen, 1999; Molholm et al., 2006; Stricanne & Andersen, 1996).

Researchers who conduct future examinations of multisensory interactions should remain cognizant of the traditional rules for multisensory integration, but should not be limited by these rules. For example, when defining a temporal window of integration, researchers should consider that the underlying mechanism might be periodic, and thus avoid defining the temporal window of integration based on a local minimum.

Multisensory interactions and attentional selection. Strategic goals determine top-down attentional deployment. And the way in which the brain's limited resources are deployed in anticipation of an upcoming stimulus strongly influences how the senses will interact. Intersensory selective attention, for example, determines whether enhanced processing associated with feature integration will spread among the sensory cortices. A task-irrelevant sound presented at approximately the same time and location as an attended visual stimulus will receive enhanced processing via the cross-sensory spread of attention, but the opposite statement is not true: a task-irrelevant visual stimulus presented at approximately the same time and location as an attended sound will *not* receive enhanced processing via the cross-sensory spread of attention (Molholm et al., 2007; Talsma et al., 2007).

Cross-sensory phase reset has also been shown to depend on intersensory selective attention (Lakatos et al., 2009). But unlike the cross-sensory spread of attention, cross-sensory phase reset occurs in response to both attended visual stimuli and attended sounds. The occurrence of a stimulus in the attended modality resets the phase of ongoing neural oscillations across multiple sensory cortices. What's more, evidence from **Chapter 2** suggests that the exact phase to which oscillations are reset might also depend on an individual's strategic approach to the task. Ongoing oscillations seem to be reset to a high-excitability phase when there is a relatively high likelihood of audiovisual simultaneity (i.e., when the sound is a good predictor of visual-target presentation), but might instead be reset to a low-excitability phase when there is a low likelihood of audiovisual simultaneity (i.e., when the sound serves as a distractor).

In addition to intersensory selective attention, biasing of attention in anticipation of a well-known object (e.g., a dog) also influences multisensory interactions. Results from **Chapter 4**, for example, confirmed that attending for a specific, well-known object enhances the visual-to-auditory spread of attention when a task-irrelevant sound (e.g., a bark) is semantically congruent with the target object (e.g., the image of a dog). Here attentional biasing is thought to activate a pre-existing cortical representation of the well-known object that includes that object's multisensory features (Molholm et al., 2007; see **Chapter 4**). The notion is that the multisensory features of well-known objects are so tightly bound, that attending for a single feature of that object results in the activation of its entire cortical representation.

Up to this point, we have provided examples of how attentional selection (through top-down biasing) influences multisensory interactions, but multisensory interactions also influence attentional selection (or the competition for limited processing resources). For example, the occurrence of a sound improves visual-target detection within a continuously changing, complex visual scene (see Van der Burg et al., 2008a). **Chapter 2** argues that this behavioral effect (i.e., auditory facilitation of visual-target detection) results from cross-sensory phase reset, where the occurrence of a sound resets the phase of ongoing oscillations in visual cortex, such that a co-occurring visual target (or change in the visual scene) now aligns with a high-excitability phase. By shifting neurons in visual cortex closer to their firing threshold, cross-sensory phase reset in response to an attended sound increases the likelihood that a visual target will be selected for further downstream processing. Multisensory interactions are thus both a contributor to the competition for limited processing resources and an outcome of that competition.

A model for cortical audiovisual integration. Here we use results presented in the preceding chapters, as well as the work of others, to track the neural response to a visual target simultaneously presented with a sound. Substantial portions of this model have already been outlined in **Chapters 3** and **4**, but are repeated here to summarize our findings across all four chapters**.**

The breadth of research reviewed thus far suggests a model for cortical audiovisual integration in which: (1) the visual scene is parsed into a coarse outline of objects (i.e., segmentation, see Driver et al., 2001; Moore & Egeth, 1997; Razpurker-Apfeld & Hillel, 2008), (2) spatial selection determines the relevant location for further processing (Treisman & Gelade, 1980), (3) attention spreads within the visual boundaries of an object at the selected location (i.e., the attentional spotlight conforms to the shape of the object, see Egly et al., 1994; Martinez et al., 2006; Martinez et al., 2007; see **Chapter 3**), and (4) attention spreads to coincident multisensory features (and visual features) that fall within that object's visual boundaries (Busse et al., 2005; Molholm et al., 2007; Talsma et al., 2007; see **Chapters 3** and **4**). The first three steps reflect object-based selection, and it is this refinement of attended space that determines whether a task-irrelevant input (e.g., a sound) will be grouped with the attended visual stimulus (i.e., whether the multisensory features have overlapping spatial representations). The forth step results in enhanced processing within a distributed cortical network, and presumably a unified object representation. This enhanced processing can receive an additional boost if the multisensory stimuli match the pre-existing cortical representation of a target object (see **Chapter 4**).

By contributing to the detection of a visual target within a complex, continuously changing environment, cross-sensory phase reset can help guide the selection process (see **Chapter 2)**. At the initial stages of cortical processing, the co-occurrence of a sound aids in detection and selection regardless of spatial alignment (see **Chapter 1**). However, the co-occurring sound will only later be combined with the visual target as part of the same object if the sound's diffuse spatial representation overlaps with the visual target's boundaries (see **Chapter 3**). This means that multisensory feature integration is directly linked with pre-attentive visual object processing, as the boundaries of the visual target need to be parsed before it can be determined whether a co-occurring sound should be included as part of the same object.

Chapter 3 used the illusory contour effect to probe the influence of pre-attentive visual object processing on the cross-sensory spread of attention. The illusory contour effect itself serves as a good metric for the object-based selection processes that our model posits as determinative steps preceding multisensory feature integration. The formation of visual boundaries within LOC regions is thought to occur between 90 and 190 ms post stimulus onset with subsequent feedback to lower-tier areas (Murray et al., 2002; Murray et al., 2006). This time course accords with findings that spatial attention spreads within both illusory boundaries and real-object boundaries starting at ~160 ms post stimulus onset, an effect that has also been localized to LOC regions (Martinez et al., 2006; Martinez et al., 2007). According to our model, these object-based selection processes within the ventral visual stream establish the boundaries for subsequent multisensory feature integration, which occurs at approximately 220 ms (Busse et al., 2005; Talsma et al., 2007; see **Chapters 3** and **4**).

Human lesion studies have shown that although an intact ventral visual stream is sufficient for boundary completion, parietal areas are needed for non-shape feature integration (Barton et al., 2007; Humphreys & Riddoch, 1994; Robertson, 2003). Patients with large parietal lesions are able to make shape determinations but make mistakes when asked to name the color of a specific object presented among several different-colored objects. That is, patients with bilateral parietal damage sometimes form illusory conjunctions (see Treisman & Schmidt, 1982), combining the color of one object with the shape of another. We posit that multisensory feature integration through the cross-sensory spread of attention likewise requires contributions from both the ventral visual stream and parietal areas.

As discussed above, the ventral visual stream (e.g., LOC regions) is thought to establish the boundaries for feature integration (steps 1–3 of our model), whereas parietal areas (e.g., the posterior parietal cortex), which are known to include multisensory spatial representations (Bushara et al., 1999; Cohen et al., 2004; Linden et al., 1999; Molholm et al., 2006; Stricanne et al., 1996), are thought to co-register the spatial representations of features that fall within an object’s visual boundaries. Following such co-registration, there is a spread of attentional processing, as parietal cortical areas provide feedback to sensory-specific cortex (step 4 of our model). Prior to this feedback from parietal cortex, interactions between the early sensory cortices seem to be based on temporal alignment, independent of spatial alignment.

Feedback to sensory-specific cortex might limit the response of neurons that have multiple objects within their large receptive fields to the response associated with features that are components of the selected object (i.e., the neuron’s response is no longer a weighted sum of all the objects in its receptive field but a response limited to the selected object), or it might enhance the response of neurons specifically tuned to prefer features that are components of the selected object (see Reynolds & Desimone, 1999). Regardless of the exact mechanism, or combination of mechanisms, enhanced processing through feedback to sensory-specific cortical areas presumably binds anatomically separate feature representations into a coherent object.

REFERENCES

Allison, T., Puce, A., Spencer, D.D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cereb. Cortex, 9*, 415-430.

Amedi, A., Von, K.K., Van Atteveldt, N.M., Beauchamp, M.S., & Naumer, M.J. (2005). Functional imaging of human crossmodal identification and object recognition. *Exp. Brain Res., 166,* 559-571.

Bak, P., Tang, C., & Wiesenfeld, K. (1987). Self-organized criticality: an explanation of 1/*f* noise. *Physical Review Letters, 59*, 381-384.

Barton, J.J.S., Malcolm, G.L., & Hefter, R.L. (2007). Spatial processing in Balint syndrome and prosopagnosia: a study of three patients. *J. Neuro. Ophthalmol., 27,* 268-274.

Blaser, E., Pylyshyn, W., & Holcombe, A. (2000). Tracking an object through feature space. *Nature, 408,* 196-199.

Brovelli, A., Ding, M., Ledberg, A., Chen, Y., Nakamura, R., & Bressler, S.L. (2004). Beta Oscillations in a Large-Scale Sensorimotor Cortical Network: Directional Influences Revealed by Granger Causality. *Proc. Natl. Acad. Sci. U.S.A., 101(26),* 9849-9854.

Bushara, K.O., Weeks, R.A., Ishii, K., Catalan M.J., Tian, B., Rauschecker, J.P., & Hallett, M. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat. Neurosci., 2(8),* 759-765.

Busse, L., Roberts, K.C., Crist, R.E., Weissman, D.H., & Woldorff, M.G. (2005). The spread of attention across modalities and space in a multisensory object. *P. Natl. Acad. Sci. U. S. A., 102,* 18751-18756.

Busch, N.A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *J. Neurosci., 29,* 7869-7876.

Busch, N.A., & VanRullen, R. (2010). Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proc. Natl. Acad. Sci. U.S.A., 107,* 16048-16053.

Buzsaki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science, 304,*1926-1929.

Calvert, G.A., Hansen, P.C., Iversen, S.D., & Brammer, M.J. (2001). Detection of audio-visual integration in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage, 14,* 427-438.

Cappe, C., & Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *Eur. J. Neurosci., 22,* 2886-2902.

Carrasco, M., Fuller, S., & Ling, S. (2008). Transient attention does increase perceived contrast of suprathreshold stimuli: a reply to Prinzmetal, Long and Leonhard (2008). *Percept. Psychophys., 70,* 1151-1164.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neurosci., 7,*308-313.

Chelazzi, L., Miller, E.K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature, 363,* 345-347.

Chelazzi, L., Duncan, J., Miller, E.K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *J. Neurophysiol., 80,* 2918-2940.

Cohen, Y.E., & Andersen, R.A. (2004). Multimodal spatial representations in the primate parietal lobe. In Spence, C., & Driver, J. (eds) *Crossmodal Space and Crossmodal Attention*. New York, Oxford UP, pp. 99-121.

Corbetta, M., Patel, G., & Shulman, G.L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron, 58,* 306-324.

Coull, J.T., Frith, C.D., Frackowiak, R.S.J., & Grasby, P.M. (1996). A fronto-parietal network for rapid visual information processing: a PET study of sustained attention and working memory. *Neuropsychologia, 34,* 1085-1095.

Coull, J.T., & Nobre, A.C. (1998). Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *J. Neurosci., 18,* 7426-7435.

Crowley, K.E., & Colrain, I.M. (2004). A review of the evidence for P2 being and independent component process: age, sleep and modality. *Clin Neurophysiol., 115,* 732-744.

Davis G., Driver J., Pavani F., & Shepherd A. (2000). Reappraising the apparent costs of attending two separate visual objects. *Vision Res*, *40,* 1323-1332.

Desimone R, & Duncan J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci, 18,* 193–222.

Doniger, G.M., Foxe, J.J., Murray, M.M., Higgins, B.A., Snodgrass, J.G., Schroeder, C.E., & Javitt, D.C. (2000). Activation timecourse of ventral visual stream object-recognition areas: high-density electrial mapping of perceptual closure processes. *J. Cogn. Neurosci., 12,* 612-621.

Donohue, S.E., Roberts, K.C., Grent-‘t-Jong, T., & Woldorff, M.G. (2011). The cross-modal spread of attention reveals differential constraints for the temporal and spatial linking of visual and auditory stimulus events. *Under Review.*

Driver, J., & Baylis G.C. (1998). Attention and visual object segmentation. In R. Parasuraman (Ed.), *The attentive brain*. Cambridge, MA, MIT Press, pp.299-325.

Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). Segmentation, attention and phenomenal visual objects. *Cognition, 80,* 61-95.

Duncan J. (2006). EPS Mid-Career Award 2004: brain mechanisms of attention. *Q. J. Exp. Psychol., 59,* 2-27

Egly, R., Driver, J., & Rafal, R.D. (1994). Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *J. Exp. Psychol. Gen., 123,* 161-177.

Eimer, M, & Shröger, E. (1998). ERP effects of inter-modal attention and cross-modal links in spatial attention. *Psychophysiology. 35,* 313-337.

Engel, A.K., Konig, P., & Singer, W. (1991). Direct physiological evidence for scene segmentation by temporal coding. *Proc. Natl. Acad. Sci. U.S.A., 88,* 9136-9140.

Fiebelkorn, I.C., Foxe, J.J., & Molholm, S. (2010a). Dual mechanisms for the cross-sensory spread of attention: how much do learned associations matter? *Cereb. Cortex, 20,* 109-120.

Fiebelkorn, I.C., Foxe, J.J., Schwartz, T.H., & Molholm, S. (2010b). Staying within the Lines: The Formation of Visuospatial Boundaries Influences Multisensory Feature Integration. *Euro. J. Neurosci., 31,* 1737-1743.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *J. Neurosci., 22,* 5749-5759.

Falchier, A., Schroeder, C.E., Hackett, T.A., Lakatos, P., Nascimento-Silva, S., Ulbert, I., Karmos, G., & Smiley, J.F. (2010). Projections from visual areas V2 and prostriata to caudal auditory cortex in the monkey. *Cereb. Cortex, 20,* 1529-1538.

Foxe, J.J., Morocz, I.A., Murray, M.M., Higgins, B.A., Javitt, D.C., & Schroeder, C.E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Brain Res. Cogn. Brain Res., 10,* 77-83.

Foxe, J.J., Murray, M.M., & Javitt, D.C. (2005). Filling-in in schizophrenia: a high-density electrical mapping and source-analysis investigation of illusory contour processing. *Cereb. Cortex, 15,* 1917-1927.

Foxe, J.J., & Schroeder, C.E. (2005). The case for feedforward multisensory convergence during early cortical processing. *NeuroReport, 16,* 419-423.

Foxe, J.J., Simpson, G.V., & Ahlfors, S.P. (1998). Cued shifts of intermodal attention: Parieto-occipital ~10 Hz activity reflects anticipatory state of visual attention mechanisms. *NeuroReport, 9,* 3929-3933.

Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancment of visual perception by crossmodal visuo-auditory interaction. *Exp. Brain Res., 147,* 332-343.

Fuller, S., Rodriguez, R.Z., & Carrasco, M. (2008). Apparent contrast differs across the vertical meridian: visual and attentional factors. *J. Vision, 8,* 11-16.

Giard, M.H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J. Cogn. Neurosci., 11,* 473-490.

Gomez-Ramirez, M., Kelly, S.P., Molholm, S., Sehatpour, P., Schwartz, T.H., & Foxe, J.J. (2010). Oscillatory sensory mechanisms mediate intersensory attention to rhythmic auditory and visual inputs: a human electro-corticographic (ECoG) investigation. *In Preparation.*

Gondan, M., Niederhaus, B., Rosler, F., & Roder, B. (2005). Multisensory processing in redundant-target effect: a behavioral and event-related potential study. *Percept. Psychophys., 67,* 713-726.

Gonzalez Andino, S.L., Murray, M.M., Foxe, J.J., & Grave de Peralta Menendez, R. (2005). How single-trial electrical neuroimaging contributes to multisensory research. *Exp. Brain Res., 166,* 298-304.

Green, D.M., & Swets, J.A. (1966). Signal detection and psychophysics. Wiley, New York.

Greene, A.J., Easton, R.D., & LaShell, L.S. (2001). Visual-auditory events: cross-modal perceptual priming and recognition memory. *Conscious Cogni., 10,* 425-435.

Grosof, D.H., Shapley, R.M., & Hawken, M.J. (1993). Macaque V1 neurons can signal ‘illusory’ contours. *Nature*, *365,* 550-552.

Guthrie, D., & Buchwald, J.S. (1991). Significance testing of difference potentials. *Psychophysiology*, *28,* 240-244.

Hackley, S.A., Schankin, A., Wohlschlaeger, A., & Wascher, E. (2007). Localization of temporal preparation effects via trisected reaction time. *Psychophysiology 44,* 334-338.

Hansen, J.C., & Hillyard S.A. (1980). Endogenous brain potentials associated with selective auditory attention. *Electroen. Clin. Neuro., 499,* 277-290.

Hillyard, S.A., Simpson, G.V., Woods, D.L., VanVoorhis, S., & Munte, T.F. (1984). Event-related brain potentials and selective attention to different modalities. In: F. Reinoso-Suarez & C. Aimone-Marsan (Eds). *Cortical Integration*. New York, Raven Press. pp 395-413.

Hopf, J.M., Schoenfeld, M.A., & Heinze, H.J. (2005). The Temporal Flexibility of Attentional Selection in the Visual Cortex. *Current Opinion in Neurobiology, 15,* 183-187.

Humphreys, G.W., & Riddoch, M.J. (1994). Attention to within-object and between-object spatial representations: multiple sites for visual selection. *Cogn. Neuropsychol*., *11,* 207-241.

James, T.W., Humphrey, G.K., Gati, J.S., Servos, P., Menon, R.S., & Goodale, M.A. (2002).

Haptic study of three-dimensional objects activates extrastriate visual areas.

Neuropsychologia, 40, 1706-1714.

Jones, E.G. (1998). Viewpoint: the core and matrix of thalamic organization. *Neuroscience 85,* 331- 345.

Kaplan, E. (1991). The receptive field structure of retinal ganglion cells in cat and monkey. In: Leventhal A.G. (ed) Vision and visual dysfunction. CRC, Boston, pp 10-40.

Kanizsa, G. (1976). Subjective contours. *Sci. Am., 234,* 48-52.

Kayser, C., Montemurro, M.A., Logothetis, N.K., & Panzeri, S. (2009) Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron, 61,* 597-608.

Keetels, M., & Vroomen, J. (2005). The role of spatial disparity and hemifield in audio-visual temporal order judgments. *Exp. Brain Res. 167,* 635-640.

Kelly, S.P., Gomez-Ramirez, M., & Foxe, J.J. (2008). Spatial attention modulates initial afferent activity in human primary visual cortex. *Cereb. Cortex, 18,* 2629-2636.

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207,* 203-205.

Lakatos, P., Chen, C.M., O'Connell, M.N., Mills, A., & Schroeder, C.E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53,* 279-292.

Lakatos, P., Karmos, G., Mehta, A., Ulbert, I., & Schroeder, C.E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science, 320,* 110-113.

Lakatos, P., O'Connell, M.N., Barczak, A., Mills, A., Javitt, D.C., Schroeder, C.E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron, 64,* 419-430.

Lachaux, J.P., George, N., Tallon-Baudry, C., Martinerie, J., Hugueville, L., Minotti, L., Kahane, P., & Renault, B. (2005). The many faces of the gamma band response to complex visual stimuli. *Neuroimage*, *25,* 491-501.

Lee, T.S., & Nguyen, M. (2001). Dynamics of subjective contour formation in the early visual cortex. *Proc. Natl. Acad. Sci. U. S. A.*, *98,* 1907-1911.

Linden, J.F., Grunewald, A., & Andersen, R.A. (1999). Responses to auditory stimuli in macaque lateral intraparietal area II: Behavioral modulation. *J. Neurophysiol., 82,* 343-358.

Ling, S., & Carrasco, M. (2007). Transient covert attention does alter appearance: a reply to Schneider (2006). *Percept. Psychophys., 69,* 1051-1068.

Lippert, M., Logothetis, N.K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Res. 1173,* 102-109.

Liu, T., Abrams, J., & Carrasco, M. (2009). Voluntary attention enhances contrast appearance. *Psychol. Sci., 20,* 354-362.

Macmillan, N.A., & Creelman, C.D. (1991). Detection theory: a user's guide. Cambridge, UK: Cambridge University Press.

Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics. *Trends Cogn. Sci., 8,* 204-210.

Manly, T., & Robertson, I.H. (1997). Sustained attention and the frontal lobes**.** In Methodolog**y** of Frontal and Executive Function, P. Rabbitt, ed. (Hove, UK: Psychology Press), pp.135-150.

Martinez, A., Teder-Sälejärvi, W.A., & Hillyard, S.A. (2007). Spatial attention facilitates selection of illusory objects: Evidence from event-related brain potentials. *Brain Res., 1139,* 143-152.

Martinez, A., Teder-Salejarvi, W., Vazquez, M., Molholm, S., Foxe, J.J., Javitt, D.C., Di Russo, F.,Worden, M.S., & Hillyard, S.A. (2006). Objects are highlighted by spatial attention. *J. Cogn. Neurosci., 18,* 298-310.

Martinez-Montes, E., Cuspineda-Bravo, E.R., El-Deredy, W., Sanchez-Bornot, J.M., Lage-Castellanos, A., & Valdes-Sosa, P.A. (2007). Exploring event-related brain dynamics with tests on complex valued time-frequency representations. Statist. Med. DOI: 10.1002/sim.3132.

Mathewson, K.E., Gratton, G., Fabiani, M., Beck, D.M., & Ro, T. (2009). To see or not to see: prestimulus alpha phase predicts visual awareness. *J. Neurosci., 29,* 2725-2732.

Matell, M.S., & Meck, W.H. (2004). Neuropsychological mechanisms of interval timing behavior. *BioEssays, 22,* 94-103.

McCarthy, G., & Wood, C.C. (1985). Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroen. Clin. Neuro., 62,* 203-208.

McDonald, J.J., Teder-Sälejärvi, W.A., Di Russo, F., & Hillyard, S.A. (2003). Neural substrates of perceptual enhancement by cross-modal spatial attention. *J. Cogni.*

Neurosci., 15, 10-19.

McDonald, J.J., Teder-Sälejärvi, W.A., & Hillyard, S.A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407,* 906-908.

Melcher, D., & Vidnyanszky, Z. (2006). Subthreshold features of objects: Unseen but not unbound. *Vision Res., 46,* 1863-1867.

Meredith, M.A., & Stein, B.E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res. 365,* 350-354.

Meredith, M.A., Nemitz, J.W., & Stein, B.E. (1987). Determinants of multisensory integration in superior colliculus neurons: I. Temporal factors. *J. Neurosci., 7,* 3215-3229.

Merigan, W., & Maunsell, J. (1993). How parallel are the primate visual pathways? *Annu. Rev. Neurosci., 5,* 347-352.

Miniussi, C., Wilding, E.L., Coull, J.T., & Nobre, A.C. (1999). Orienting attention in time: modulation of brain potentials. *Brain, 122,* 1507-1518.

Molholm, S., Martinez, A., Shpaner, M., Foxe, J.J. (2007). Object-based attention is multisensory: co-activation of an object’s representations in ignored sensory modalities. *Euro. J. Neurosci., 26,* 499-509.

Molholm, S., Ritter, W., Javitt, D.C., & Foxe, J.J. (2004). Multisensory visual–auditory object recognition in humans: A high-density electrical mapping study. *Cereb Cortex., 14,* 452-465.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., & Foxe J.J. (2002) Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Cogn. Brain Res., 14,* 115-129.

Molholm, S., Sehatpour, P., Mehta, A.D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., Dyke, J.P., Schwartz, T.H., & Foxe, J.J. (2006). Audio-Visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *J. Neurophyisol., 96,* 721-729.

Montemurro, M.A., Rasch, M.J., Murayama, Y., Logothetis, N.K., & Panzeri, S. (2008) Phase-of-firing coding of natural stimuli in primary visual cortex. *Curr. Biol. 18,* 375-380.

Monto, S., Palva, S., Voipio, J., & Palva, J.M. (2008). Very slow EEG fluctuations predict the dynamics of stimulus detection and oscillation amplitudes in humans. *J. Neurosci. 28,* 8268-8272.

Moore, C.M., & Egeth, H. (1997). Perception without attention: evidence of grouping under conditions of inattention. *J. Exp. Psychol. Hum. Percept. Perform., 23,* 229-352.

Murray, M.M., Molholm, S., Michel, C.M., Heslenfeld, D.J., Ritter, W., Javitt, D.C., Schroeder, C.E., & Foxe, J.J. (2005). Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cereb.Cortex,15,* 963-974.

Murray, M.M., Wylie, G.R., Higgins, B.A., Javitt, D.C., Schroeder, C.E., & Foxe, J.J. (2002). The spatiotemporal dynamics of illusory contour processing: Combined high-density electrical mapping, source analysis, and functional magnetic resonance imaging. *J. Neurosci., 22,* 5055-5073.

Murray, M.M., Imber, M.I., Javitt, D.C., & Foxe J.J. (2006). Boundary completion is automatic and dissociable from shape discrimination. *J. Neurosci., 26(46),* 12043-12054.

Näätänen, R., Gaillard, A.W.K., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta. Psychol., 42,* 313-329.

Noesselt, T., Bergmann, D., Hake, M., Heinze, H.J., & Fendrich, R. (2008). Sound increases the saliency of visual events. *Brain Res., 1220,* 157-163.

Noesselt, T., Tyll, S., Boehler, C.N., Budinger, E., Heinze, H.J., & Driver, J. (2010) Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies related to perceptual enhancement of visual detection sensitivity. *J. Neurosci., 30,* 13609-13623.

O'Craven, K.M., Downing, P.E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature, 401,* 584-587.

Odgaard, E.C., Arieh, Y., & Marks, L.E. (2003). Cross-modal enhancement of perceived brightness: sensory interaction versus response bias. *Percept. Psychophys., 65,* 123-132.

Razpurker-Apfeld, I., & Pratt, H. (2008). Perceptual visual grouping under inattention: electrophysiological functional imaging. *Brain Cogn., 67,* 183-196.

Recanzone, G.H. (2000). Spatial processing in the auditory cortex of the macaque monkey. *P. Natl. Acad. Sci. U.S.A., 97,* 11829-11835.

Reynolds, J.H., & Desimone, R. (1999). The role of neural mechanisms in solving the binding problem. *Neuron, 24,* 19-24.

Robertson, L.C. (2003). Binding, spatial attention and perceptual awareness. *Nature Rev., 4,* 9-102.

Rockland, K.S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *Inter. J. Psychophysiol., 50,* 19-26.

Saenz, M., Buracas, G.T., & Boynton, G.M. (2003). Global feature-based attention for motion and color. *Vision Res., 43,* 629-637.

Sauseng, P., Klimesch, W., Gruber, W.R., Hanslmayr, S., Freunberger, R., & Doppelmayr M. (2007). Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion. *Neurosci., 146,* 1435-1444.

Schoenfeld, M.A., Tempelmann, C., Martinez, A., Hopf, J.M., Sattler, C., Heinze, H.J., & Hillyard, S.A. (2003). Dynamics of feature binding during object-selective attention. *P. Natl. Acad. Sci. U.S.A., 100,* 11806-11811.

Scholl, BJ. (2001). Objects and Attention: The State of the Art. *Cognition, 80,* 1-46.

Schroeder, C.E., & Lakatos, P. (2008). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci., 32,* 9-18.

Schroeder, C.E., Molholm, S., Lakatos, P., Ritter, W., & Foxe, J.J. (2004). Human-Simian correspondence in the early cortical processing of multisensory cues. *Cogni. Processing, 5,* 140-151.

Sehatpour, P., Molholm, S., Schwartz, T.H., Mahoney, J.R., Mehta, A.D., Javitt, D.C., Stanton, P.K., & Foxe, J.J. (2008). A human intracranial study of long-range oscillatory coherence across a frontal-occipital-hippocampal brain network during object processing. *Proc. Natl. Acad. Sci. U.S.A., 105,* 4399-4404.

Serences, J.T., & Yantis, S. (2006). Selective visual attention and perceptual coherence. *Trends Cogn. Sci., 10,* 38-45.

Shah, A.S., Bressler, S.L., Knuth, K.H., Ding, M., Mehta, A.D., Ulbert, I., & Schroeder, C.E. (2004). Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cereb. Cortex,* 14, 476-483.

Sheth, B.R., Sharma, J., Rao, S.C., & Sur, M. (1996). Orientation maps of subjective contours in visual cortex. *Science, 274,* 2110-2115.

Shomstein, S., & Yantis, S. (2002). Object-based attention: sensory modulation or priority setting? *Percept Psychopys, 64,* 41-51.

Slutsky D.A.., & Recanzone G.H. (2001). Temporal and spatial dependency of the ventriloquism effect. *NeuroReport*, *12,* 7-10.

Snyder, A.C., & Foxe, J.J. (2010). Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: a high-density electrical mapping study. *J. Neurosci., 30,* 4024-4032.

Stein, B.E., London, N., Wilkonson, L.K., & Price, D.D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *J. Cogn. Neurosci., 8,* 497-506.

Stone, J.V., Hunkin, N.M., Porrill, J., Wood, R., Keeler, V., Beanland, M., Port, M. & Porter, N.R. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society of London B-Biological Sciences, 268,* 31-38.

Störmer, V.S., McDonald, J.J., & Hillyard, S.A. (2009). Cross-modal cueing of attention alters appearance and early cortical processing of visual stimuli. *Proc. Natl. Acad. Sci. U.S.A., 106,* 22456-22461.

Stricanne, B., & Andersen, R.A. (1996). Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *J. Neurophysiol., 76,* 2071-2076.

Stroop J.R. (1935). Studies of interference in serial verbal reactions. *J Exp Psychol., 18,* 643-662.

Talsma, D., Doty, T.J., & Woldorff, M.G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cereb. Cortex, 17,* 679-690.

Talsma, D., & Kok, A. (2001). Nonspatial intermodal selective attention is mediated by sensory brain areas: evidence from event-related potentials. *Psychophysiology, 38,* 736-751.

Teder-Sälejärvi, W.A., Di Russo, F., McDonald, J.J., & Hillyard, S.A. (2005). Effects of spatial congruity on audio-visual multimodal integration. *J. Cogn. Neurosci., 17,* 1396-1409.

Teder-Sälejärvi, W.A., Munte, T.F., Franz-Jurgen, S., Hillyard, S.A. (1999). Intramodal and cross-modal spatial attention to auditory and visual stimuli. An event-related brain potential study. *Cogni. Brain Res., 8,* 327-343.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cogni. Psychol.,12,* 97-136.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cogn. Psychol., 14,* 107-141.

Treue, S., & Martinez-Trujillo, J.C. (1996). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature, 399,* 575-579.

Van der Burg, E., Olivers, C.N.L, Bronkhorst, A.W., & Theeuwes, J. (2008a). Pip and pop: auditory signals improve spatial visual search. *J. Exp. Psychol. Hum. Percept. Perform. 34,* 1053-1065.

Van der Burg, E., Olivers, C.N.L., Bronkhorst, A.W., & Theeuwes, J. (2008b). Audiovisual events capture attention: Evidence from temporal order judgments. *J. Vis., 8,* 1-10.

Van Orden, G.C., Holden, J.G., & Turvey, M.T. (2003). Self-organization of cognitive performance. *J. Exp. Psychol. Gen., 132,* 331-350.

VanRullen, R., Reddy, L., & Koch, C. (2006). The continuous wagon wheel illusion is associated with changes in EEG power around 13 Hz. *J. Neurosci., 26,* 502-507.

VanRullen, R., Carlson, T., & Cavanagh, P. (2007). The blinking spotlight of attention. *Proc. Natl. Acad. Sci. U.S.A., 104,* 19204-19209.

Vecera, S.P., & Farah, J.M. (1994). Does visual attention select objects or locations? *J. Exp. Psychol. Gen., 123,* 146-160.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224,* 1260-1262.

Vroomen, J., & Keetels, M. (2006). The spatial constraint in intersensory pairing: no role in temporal ventriloquism. *J. Exp. Pschol. Hum. Percept. Perform. 32,* 1063-1071.

Wang, Y., Celebrini, S., Trotter, Y., & Barone, P. (2008). Visuo-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. *BMC Neurosci., 9,* 79.

Weber, T.A., Kramer, A.F., & Miller, A.G. (1997). Selective processing of superimposed objects: an electrophysiological analysis of object-based attentional selection. *Biol. Psychol., 45,* 159-182.

Whittingstall, K., & Logothetis, N.K. (2009). Frequency-band coupling in surface EEG reflects spiking activity in monkey visual cortex. *Neuron*, *64,* 281-289.

Worden, M.S., Foxe, J.J., Wang, N., & Simpson, G.V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. *J. Neurosci., 20,* RC63.

Wylie, G.R., Javitt, D.C., Foxe, J.J. (2003). Task switching: a high-density electrical mapping study. *NeuroImage, 20,* 2322-2342.

Wylie, G.R., Javitt, D.C., & Foxe, J.J. (2004). Don't think of a white bear: an investigation of the effects of sequential instructional sets on cortical activity in a task-switching paradigm. *Human Brain Mapping*, *21,* 279-297.

Wylie, G.R., Javitt, D.C., Foxe, J.J. (2006). Jumping the gun: Is effective preparation contingent upon anticipatory activation in task-relevant neural circuitry. *Cereb. Cortex., 16,* 394-404.

Yantis, S., & Serences, T.J. (2003). Cortical mechanisms of space-based and object-based attentional control. *Curr. Opin. Neurobiol., 13,* 187-193.

Zikopoulos, B., and Barbas, H. (2007). Circuits for multisensory integration and attentional modulation through the prefrontal cortex and the thalamic reticular nucleus in primates. *Rev. Neurosci., 18,* 417-438.