Bimodal Discrimination Tasks Involving Different Hierarchical Processing Levels

Reveal a Late Stage Interaction Between Multisensory Integration and Attention

Sarah Byce

Thesis submitted in partial fulfillment of the requirements for the Masters degree in

Psychological Science

Department of Psychology

Lakehead University

Supervisor: Dr. Michael Wesner

Second Member: Dr. Gordon Hayman

Additional Member: Dr. Jane Lawrence-Dewar

Table of Contents

Abstract

Multisensory integration and attention have been studied independently and a vast amount of literature exists for both phenomena. Researchers have recently raised questions, however, as to how these two processes interact. For example, does multisensory integration occur automatically without the need for attention; or does integrative processing require attention for it to occur? If the latter, do attention and integration act in parallel throughout all information processing levels, or does one operation need to exist to advance the other? The present study sought to answer these questions through a series of within-subject tasks spanning multiple layers of the processing hierarchy. Forty-five participants completed three tasks involving audiovisual, integrated stimuli in which they discriminated the location of a visual target stimulus from nontarget distractors while being simultaneously presented with congruent auditory tones. The first task involved the discrimination of shapes and was shown to be pre-attentive in nature, with no facilitatory effect being observed in response to simultaneously presented visual and auditory stimuli. The second task involved the discrimination of die-point stimuli, which required high attentional demand. A trend towards intersensory facilitation was observed in this task, but was not significant. The third and final task involved the discrimination of integrated shapes and die-points, which also required a high attentional load. Findings from this conjunction search revealed a surprising reversal of intersensory facilitation. These results suggest that attention has a limited capacity in terms of multisensory processing, and that specific intersensory facilitation requires a unique amount of attentional involvement. Results are discussed in terms of feature integration theory, the perceptual load hypothesis, and attentional

inhibition. This study also highlights the need for multisensory research to pay close

attention to the influence of methodology, task sensitivity, and cognitive hierarchy when

interpreting results.

Introduction

It is a well-established concept in cognitive psychology that stimuli presented to more than one modality, if congruent in time, space and/or in meaning, will result in faster and/or more accurate stimulus detection and discrimination than either of the same stimuli presented in only one modality. This phenomenon has been studied at cellular, cortical, and behavioral levels and has been referred to as multisensory integration, intersensory facilitation, or the redundant targets effect. If two stimuli occur at essentially the same time and place, or originate from the same object, these stimuli are bound together by the process of multisensory integration, and the above-mentioned facilitation of behavioral responses are usually observed (Spence, 2010). It is important to note, however, that an absence of behavioral facilitation is not necessarily indicative of an absence of multisensory integration (Holmes, 2007). If the two stimuli are separated slightly in time, however, the first stimulus can serve as a valid or invalid cue to the location or presence of the stimulus that follows it (the target). In this instance, the term *crossmodal attention* is said to facilitate the above-mentioned behavioral enhancements (Spence, 2010).

Multisensory integration and crossmodal attention have both been studied extensively, but also rather independently. It is only within the last twenty-or-so years that researchers have started to question how these two processes interact. If one encounters a bimodal stimulus, does multisensory integration occur first, capturing attention; or do you need to evoke attentional processes before multisensory integration can occur? Do these processes occur in parallel; or is multisensory integration an automatic process that can occur outside the scope of attention? Can attentional capture

itself be an automatic process? Different studies draw different conclusions as to the mechanisms and neural substrates underlying multisensory integration and crossmodal attention, their interactions, and the levels of processing at which they occur. At present, there appear to be no studies that draw all of these concepts together to provide conclusive answers as to the relationship between multisensory integration and crossmodal attention (Koelewijn, Bronkhorst, & Theeuwes, 2010; Spence, 2010). My study clarifies this relationship using behavioral visual search methods. Before discussing the methodological specifics, however, I will review the current literature on multisensory integration, feature integration, and crossmodal attention. Once I complete this review, I will conclude by examining current ideas as to how multisensory integration and crossmodal attention interact, as well as the part that attention has to play in feature integration.

Definition of Concepts

Delving into the literature on multisensory integration and crossmodal attention can be confusing, mainly because terminology is applied rather loosely to specific phenomena. The terms multisensory and crossmodal are often used interchangeably and crossmodal relationships between stimuli have been given synonyms ranging from synesthetic associations to crossmodal equivalencies (Spence, 2011). Studies investigating multisensory integration often describe stimuli as crossmodal and vice versa (Stein & Meredith, 1993). In this section I put forward basic sensory and perceptual definitions of multisensory integration and crossmodal attention that will serve to maintain a distinction between these two concepts throughout this paper. The following definitions have been derived from the literature as a whole and are an attempt to make

sense of multiple viewpoints, paradigms, and terminologies. These definitions will be re-iterated and expanded upon in their appropriate sections.

Multisensory integration is a process by which information from different senses is combined. From a sensory perspective, this means that stimuli from different modalities originate from the same place at the same time. Perceptually, these stimuli are bound together into a whole representing one external event or object. This binding of information allows the system to better orient their sensory receptors and/or attention towards said event, either reflexively or consciously. A multisensory stimulus, then, is one in which stimuli from more than one modality occur simultaneously in time and often originate from the same place. Intersensory facilitation is a behavioral consequence of multisensory integration whereby response times and/or accuracy are enhanced by the presence of multisensory stimuli. The majority of research has observed behavioral facilitation in response to congruent, multisensory stimulus presentations (i.e., the intersensory facilitation or redundant targets effects), however multisensory integration can also occur without any behavioral consequences or can even inhibit responses, as in the case of incongruent multisensory stimuli.

In this paper, crossmodal attention is defined as the process by which stimuli can capture attention across modalities. A crossmodal stimulus consists of a cue in one modality that directs attention towards a target in another modality. The cue and target must be separated in time, with the cue preceding the target. They can be coincident in space, but a cue can also be misdirecting in indicating the location of an upcoming target. Perceptually, a cue indicates that a target is forthcoming and usually indicates the upcoming location of the target, making the detection and/or discrimination of the target

easier. Of course, as just mentioned, a cue can also be misleading as to the upcoming location of a target. The correctness of a cue in predicting the location of a target is referred to as cue validity. The directing of attention across modalities can involve shifts of sensory organs and/or cognitive resources, and can occur reflexively or deliberately. Thus, for the purpose of this study, multisensory integration involves the binding of information from multiple modalities. A multisensory stimulus is one in which information from more than one modality occurs at the same time and in the same place. Crossmodal attention, on the other hand, is the process by which attention can be captured by one modality in order to be directed within another. A crossmodal stimulus, then, requires a cue in one modality and a target in another. The cue and target must be separated in time and may or may not be coincident in space.

The Integration of Multisensory Information

Multisensory integration describes a process by which information from the different senses is combined to modulate perception, decisions, and potentially overt behavior. In general, it enhances speed of detection, localization, and reactions to significant multimodal stimuli in the environment (see Stein, Stanford, & Rowland, 2009 for a review). Events occurring in the environment typically produce multiple types of physical energy simultaneously. These energies are often independent and cannot influence one another and must be transduced by the different sensory organs of the nervous system before they can interact. A multisensory stimulus, therefore, is an external event that generates several independent physical energies, each of which is concomitantly detected by different sensory receptors. Once these external energies have been transduced into neural signals, there is potential for interactions between them.

Therefore, this makes multisensory integration an emergent property of the brain (e.g., Meredith, 2002).

Some intracellular research has revealed that some neurons exhibit multisensory integrative processing, and these cells have been localized throughout the neural hierarchy, from the superior colliculus (SC) of the midbrain to heteromodal areas of the cortex (Calvert & Thesen, 2004; Fairhall & Macaluso, 2009). The cellular properties underlying multisensory integration have been studied extensively in the cat SC and have yielded specific spatial and temporal parameters to which stimuli must conform in order to enhance or depress integration. Weak unimodal stimuli can also produce heightened responses when combined (i.e., inverse effectiveness; Stein & Meredith, 1993).

Behavioral effects of multisensory integration have also been observed in phenomenon such as the ventriloquism effect (Bertelson, Vroomen, de Gelder, & Driver, 2000), the intersensory facilitation effect (Hershenson, 1962), and the redundant targets effect (Miller, 1991). There is also an increasing amount of evidence that multisensory integration is not solely a bottom-up, automatic process; rather it is also dependent on feedback projections from the cortex (Jiang, Wallace, Jiang, Vaughan, & Stein, 2001) and on top-down attentional modulation (Koelewijn et al., 2010). Below I will review the phenomenon of multisensory integration from the single cell in the SC to the cortex and its influence on overt behavior.

The role of the CNS in multisensory integration.

The most intensely studied structure to date in the literature on multisensory integration is the superior colliculus, specifically in the cat model. The SC is involved in the overt orienting (physical movement) of sensory receptors (the eyes, ears, and body)

towards external stimuli. As most environmental events produce more than one type of stimuli (e.g., visual, auditory, somatosensory, proprioceptive, olfactory, gustatory), it makes evolutionary sense that overt orienting should be modulated by multimodal stimuli (Stein & Stanford, 2008). The SC receives input from visual, auditory, and somatosensory areas of the cortex and projects to the spinal cord, frontal eye fields, motor cortex, basal ganglia, thalamus, cerebellum, locus coeruleus, reticular formation, and hypothalamus (for a review see Stein & Meredith, 1993).

The SC is located in the quadrigeminal plate of the midbrain and is composed of seven alternating cellular and fibrous layers oriented horizontally in the structure. The top three layers of the SC, referred to as the superficial layers, receive inputs from the retina and primary visual cortex and contain a retinotopic map of visual space. It is in the deep layers of the SC that information from the visual, auditory, and tactile modalities converge and are integrated (Angelaki, Gu, & DeAngelis, 2009; Meredith, 2002; Meredith, Nemitz, & Stein, 1987; Meredith & Stein, 1983, 1986; Rowland, Quessy, Stanford, & Stein, 2007; Stein & Meredith, 1993; Stein & Stanford, 2008; Wallace, Meredith, & Stein, 1998). The deep layers receive input from extraprimary sensory areas (a large portion of which are devoted to visual signals) and contains coarse maps of visual and auditory space as well as the body surface. This convergence means that over half of SC neurons are multisensory and respond to two or three modalities. These multimodal neurons integrate information in such a way that their output no longer resembles the output of individual sensory inputs alone (Meredith, 2002). Thus, at the cellular level, multisensory integration is defined as a statistically significant response difference between the number of impulses evoked in a cell (or group of cells) by a

combination of stimuli from multiple modalities versus the impulse number evoked by the most effective of any one of the stimuli on its own (Stein & Stanford, 2008). Unlike the deep layers, the superficial SC appears to play a less direct role in integration processes and is more involved with individual modal mappings (Stein & Meredith, 1993).

Multisensory integration is an emergent property of the SC, with neural output governed by the organization of multisensory neuron receptive fields. A receptive field is an area of sensory space that characterizes the physiological properties of a particular neuron in response to properly energy- and quality-tuned stimuli. Each multisensory neuron receives inputs from other neurons, each with their own respective receptive fields. Thus, multisensory neuron receptive fields are representative of multiple, presynaptic, overlapping receptive fields, with receptive fields from each modality to which it responds (Kadunce, Vaughan, Wallace, & Stein, 2001). These sensory representations are aligned, although not with the greatest spatial accuracy. This broad spatial bandwidth ensures that slight reception shifts within the smaller clustered receptive fields that converge on multisensory neurons will not disrupt the overall multimodal register of spatial position. Again, the general localization of a stimulus depends on the activation of a sizable enough pool of neurons to produce an effect larger than the surrounding localized neural activity (Stein & Meredith, 1993). Therefore, responses to stimuli from multiple modalities can be defined by neural integration as originating from one location as long as the multimodal stimulation is registered by overlapping convergent receptive fields (Stein & Stanford, 2008).

Multisensory neurons exhibit either enhancement or depression of response to multisensory stimuli. Meredith and Stein (1983) conducted seminal studies investigating multisensory cell responses. Response enhancement is said to occur when multisensory stimuli induce a vigorous neuronal response, observed as greater response reliability, number of action potentials, and longer duration of the discharge train, than a unimodal stimulus. For example, a stimulus presented just above threshold in one modality (i.e., a very low intensity stimulus) can in a multisensory capacity, elicit an enhanced multisensory responses that is greater than even the most highly salient stimulus presented in the same modality. This is referred to as the *inverse effectiveness principle* (Stanford, Quessy, & Stein, 2005), which states that SC cells can amplify the combined effects of very subtle environmental stimulus changes. Inverse effectiveness is also observed in behavior, with multimodal combinations of less salient stimuli producing faster reaction times than what is observed with the same weak stimuli presented unimodally. This magnitude of behavioral change is not observed when comparing strong unimodal stimuli to strong bimodal stimuli (Diederich & Colonius, 2004). Response depression occurs when a multisensory stimulus evokes fewer action potentials, lower peak frequencies, and lower response reliability than one of the same component stimuli presented alone, an example of how multisensory integration can produce inhibitory responses. These responses are dependent on the presence or absence of meaningful relationships between stimuli in space and time and whether stimuli fall on excitatory or inhibitory receptive field zones (Meredith & Stein, 1983).

Stimuli that occur at the same place and time are likely to be related via common causality and are likely to require the coordinated orienting of more than one sensory

organ (Meredith & Stein, 1986, Meredith et al., 1987; Stein & Stanford, 2008). Spatially disparate stimuli produce either response depression or no interaction at all, as stimuli falling outside of the receptive field or on an inhibitory zone of a given multisensory neuron will fail to enhance its response (Meredith & Stein, 1986).

In terms of temporal coincidence, maximal multisensory interactions are not dependent on the simultaneous onset of multiple stimuli or on their latencies, but rather on how the activity trains resulting from the stimuli overlap (Stein & Meredith, 1993). Different sensory stimuli move at different speeds in the environment and are also processed at different speeds in the nervous system. Thus, even if two physical stimuli occur at the same time, it does not mean that the resultant transduced neural signals will reach the SC at the same time (Meredith et al., 1987). Environmental stimuli in close temporal proximity however, do produce maximal levels of response enhancement because their transduced peak neuronal discharge periods often overlap, possibly due to pathway adjustments. Stimuli separated by longer intervals are less likely to achieve temporal overlap and may in fact, as in the case of spatial summation, produce response depression. The temporally interleaved stimuli in this case would then be processed as separate events (Nemitz, Meredith, & Stein, 1984; Stein & Meredith, 1993). The optimal temporal window for multisensory integration is to have stimuli presented within 100 msec of each other, although interaction periods can extend for up to 1500 msec and response enhancement can still occur at a cellular level even when different sensory stimuli are separated by up to 600 msec (Meredith et al., 1987).

Along with the predictive integration principles of time, space, and stimulus intensity (the inverse effectiveness principle), levels of spontaneous neuronal activity and

cell responsiveness also influence integration. Neurons with low spontaneous activity and weak sensory responses have a greater capacity for response enhancement than those with higher spontaneous activity and strong sensory responses (Perrault, Vaughan, Stein, & Wallace, 2003). These rules appear to be applicable solely with the integration of multisensory stimuli. The integration of stimuli from more than one modality consistently yields responses that are significantly greater than those evoked by any component stimulus if they follow the rules of space and time. This is not observed with the presentation of multiple stimuli within the same modality (Alvarado, Vaughan, Stanford, & Stein, 2007).

As for the SC efferents, they follow four output pathways. One pathway ascends to the thalamus, one projects to the opposite contralateral SC, and two descend to the brainstem and spinal cord (one crossed and the other uncrossed). These efferents serve to alert higher centers, coordinate the activity of both colliculi, and initiate behavioral responses, respectively (Stein & Meredith, 1993). Not surprisingly, the majority of efferent cells in the deep layers of the SC are also multisensory and follow the same rules of integration as other multisensory cells. This allows for different sensory modalities to affect the orientation of different receptor organs via the same SC-mediated circuits and allows for greater response flexibility so that any stimulus can evoke an orienting response, whether there is multimodal integration or not (Meredith & Stein, 1985).

Multisensory integration does not stem exclusively from the SC. Studies using the cat model have identified two cortical areas, the anterior ectosylvian sulcus (AES) and rostrolateral suprasylvian sulcus (rLS), whose outputs to the SC are crucial for the integration of multisensory stimuli (Alvarado, Stanford, Rowland, Vaughan, & Stein,

2003; Alvarado, Stanford, Vaughan, & Stein, 2007; Cuppini, Ursino, Magosso, Rowland, & Stein, 2010; Jiang, Jiang, & Stein, 2002; Jiang & Stein, 2003; Jiang, Wallace, Jiang, Vaughan, & Stein, 2001; Stein & Stanford, 2008; Wallace, Meredith, & Stein, 1992; Wilkinson, Meredith, & Stein, 1996). Of the two cortical integration areas, the AES is the better studied and is described as a polysensory region in which inputs from several modalities converge. The AES is composed of modality-specific regions for visual, auditory, and tactile inputs. The borders of these regions contain multisensory neurons that function on the same principles that govern those in the SC (i.e., aligned receptive fields with spatial and temporal summation rules; Stein & Stanford, 2008; Wallace et al., 1992).

It is becoming increasingly evident that multisensory integration in the SC is highly dependent on both the AES and rLS. In the cat model, deactivation of these areas effectively eliminates multisensory response enhancement of nearly all SC neurons, while leaving modality-specific responses intact. The majority of SC neurons exhibit dual dependence on both the AES and rLS, mediated either by synergistic activities from both areas together or by mutually exclusive activities from either area having redundant influences (Alvarado, et al., 2007; Jiang et al., 2001). The capacity of SC neurons to produce response depression (Jiang & Stein, 2003) and orienting responses (Jiang et al., 2002) is also compromised by deactivation of the AES and rLS areas, suggesting that the ability of SC neurons to synthesize multisensory inputs is heavily dependent on cortical inputs (Alvarado et al., 2006; Jiang et al., 2001). Cuppini et al (2010) proposed that cortical inputs from the different sensory modalities facilitate each other through interactions within the SC. Although this evidence is compelling in the cat model, there is

no known primate homologue for the AES or rLS. The majority of primate studies focus on the posterior parietal cortex (PPC), where sensory convergence has been identified (e.g., Cohen, 2009; Guipponi, Wardak, Ibarrola, Comte, Sappey-Marinier, Pinede, & Hamed, 2013; Schroeder, Molhom, Lakatos, Ritter, & Foxe, 2004; Sepulcre, Sabuncu, Yeo, Liu, & Johnson, 2012; Stein & Stanford, 2008; Xing & Andersen, 2000).

Studies mapping areas subserving multisensory integration in humans and other primates are less abundant than those conducted with lower mammal models, however some heteromodal areas in the primate brain similar to lower mammal models have been identified. Subcortical primate heteromodal areas include the SC (Fort, Delpuech, Pernier & Giard, 2002a; Fries, 1984; Wallace & Stein, 2001), claustrum (Pearson, Brodal, Gatter, & Powell, 1982), suprageniculate and medial pulvinar nuclei of the thalamus, the basal ganglia (Cappe, Rouiller, & Barone, 2009), and the amygdaloid complex (Turner, Mishkin, & Knapp, 1980). Cortical primate heteromodal areas include the superior temporal sulcus (STS) (Fairhall & Macaluso, 2009), the ventral and lateral intraparietal areas of the parietal cortex (Bremmer et al., 2001), the premotor and prefrontal cortices (Macaluso & Driver, 2005), the insula (Calvert, 2001), and the anterior cingulate gyrus (Laurienti, Wallace, Maldjian, Susi, Stein, & Burdettet, 2003). The STS is thought to be involved in the integration of audio-visual speech information. The intraparietal areas are involved in the detection and integration of multisensory cues based on spatial location (see Calvert & Thesen, 2004 for a review). Multisensory integration effects have also been observed in traditional unimodal sensory areas such as V1 (Fairhall & Macaluso, 2009). There is no doubt that multisensory areas exist; the question to be asked now is

how do these areas interact with primary sensory areas and subcortical structures to allow for multisensory integration?

The integration of information from multiple modalities has classically been viewed as a hierarchical converging system in which different sensory channels merge in polysensory association areas of the frontal, parietal, and temporal cortices. Recent research, however, has shown that multisensory interactions involve several levels of neuronal processing (Driver & Noesselt, 2008) including higher-level cortico-cortical connections, thalamo-cortical projections, as well as both feedforward and feedback projections between primary sensory areas and subcortical multisensory areas (Cappe et al., 2009). There is also evidence for direct connections between primary sensory areas such as V1 and A1 that may underlie observed early event related potential (ERP) effects (Macaluso & Driver, 2005).

Cappe et al. (2009) offer an extensive review of cortico-cortical and thalamo-cortical connections in cat, rat, and ferret models. In terms of cortico-cortical connections, heteromodal areas in frontal, parietal, and temporal lobes all have connection patterns that link them to one or more sensory modalities. Projections have also been found between primary auditory and visual cortices. Feedback connections also exist from areas such as the superior temporal gyrus to primary visual cortex. These connections all appear to be low density, and may represent projections with low functional impact. Projections do show topological organizations within the localized receiving zones. Although these projections have not yet been traced in humans, they exist in cats, rats, ferrets and monkeys, suggesting that cortical areas are most likely

linked hierarchically by feedback and feedforward connections in our brains as well (Cappe et al., 2009).

The thalamus has also been considered as a major multisensory integration structure due to its connections with multiple sensory areas. The vast majority of incoming stimuli pass through the thalamus on its way to the cortex however cortical areas also send reciprocating projections directly to the thalamus that exert some control over thalamic nuclei. There are also feedforward connections from one cortical area to another via the thalamus that allow for rapid transmission. These cortico-thalamo-cortical connections may partially underlie multisensory exchanges by allowing for the convergence of different modal information in one nucleus that is then integrated and sent to various cortical regions. Some have suggested that various pulvinar nuclei may subserve this role, as they are connected to various cortical areas while also receiving direct input from the mesencephalic superior colliculus (Cappe, Morel, Barone, & Rouiller, 2009; Morel, Liu, Wannier, Jeanmonod, & Rouiller, 2005; Shipp, 2003). The thalamus could thus represent an alternative pathway to direct cortico-cortical interactions. Indeed, radioactive tracer studies in monkeys have shown that some thalamic nuclei do project to cortices that overlap across modalities, with the medial pulvinar nuclei displaying the most significant overlap across modalities with projections to the frontal, parietal, occipital, temporal, and insular cortices (Cappe et al., 2009).

Both neuroimaging and ERP studies show that multisensory stimuli can modify unimodal operators to become more sensitive to specific stimuli in their modality. This raises the issue of how information pertaining to one modality can influence brain regions devoted to a different modality. There are two hypotheses that attempt to address this

issue. The first states that there are direct connections between sensory specific areas that cause these modulations. The second hypothesis suggests that there are top-down influences from multisensory regions on unimodal areas via back-projections. Direct connections between primary auditory and visual cortices have been observed using retrograde tracing in monkeys, however these have yet to be found in the human brain. Human ERP studies such as those conducted by Giard and Peronnet (1999) and Foxe et al. (2005) show ERP responses to bimodal auditory and visual stimuli as early as 50 ms post-stimulus which rules out the second hypothesis, as modulatory feedback via longer multimodal pathways would not be able to act that quickly. There is evidence in the human brain, however, for the existence of back projections from parietal and frontal areas to unimodal areas that are implicated in unimodal spatial attention. Functionally, the presence of preparatory activations in frontal-parietal ERP measurements following endogenous cues also supports a top-down modulatory influence. It may be that unimodal areas interact via a combination of direct links and feedback projections (see Macaluso & Driver, 2005 for a review).

Looking now at human neuroimaging, various studies have examined regions associated with multisensory information and attention (e.g., Calvert, 2001). Functional magnetic resonance imaging (fMRI) responses to concurrent auditory and visual presentations tend to activate the right insula, posterior parietal lobe and prefrontal regions, with increased cortical activity in these areas correlating with increased task demands. Functional interactions are also observed between the insula, posterior thalamus, and superior colliculus, suggesting that human multimodal processing is

mediated via subcortical audiovisual temporal correspondence (Bushara, Weeks, Ishii, Catalan, Tian, Rauschecker & Hallett, 1999).

Calvert, Hansen, Iversen and Brammer (2001) conducted an fMRI study where participants completed an audiovisual temporal correspondence task with synchronous and asynchronous bimodal stimuli and found activation in the superior colliculus, insula/claustrum, left superior temporal sulcus, right intraparietal sulcus, and superior and ventromedial frontal gyri, suggesting that these areas are involved in multisensory processing. Functional MRI studies have revealed active brain areas including the lateral parietal cortex, lateral frontal cortex and superior frontal gyrus, anterior cingulate cortex, and anterior insula revealing coactivation during auditory and visual discrimination tasks. (Lewis, Beauchamp, & DeYoe, 2000). Anterior cingulate gyrus and adjacent medial prefrontal cortices have also been implicated when participants were presented with contextually congruent visual and auditory stimuli as opposed to incongruent stimuli. Examples of stimuli include everyday objects such as alarm clocks, animals, and vehicles (Laurienti et al., 2003). The angular gyrus of the right parietal cortex has also been shown to be active during the orienting of spatial attention (Chambers, Payne, Stokes, & Mattigley, 2004). It is easy to see a general pattern of areas involved in multisensory processing. Specifically, multisensory activation areas can differ slightly depending on the performance task, with prefrontal cortices and the angular gyrus being more active with more difficult discrimination tasks or with tasks involving language representations (Laurienti et al., 2003).

Behavioral studies of multisensory integration.

The same properties found with multisensory integration in neurons can also apply to overt behavioral responses towards multisensory stimuli (Diederich & Colonius, 2004; Stein, Huneycutt, & Meredith, 1988). The facilitatory behavioral effects of multisensory integration are most frequently observed as *intersensory facilitation* and the *redundant targets effects*. The intersensory facilitation effect occurs when response times to a stimulus presented in one modality is shortened by the presence of an unrelated accessory stimulus presented in another modality (Hershenson, 1962). The redundant targets effect posits a similar reduction in reaction time, but in this case reduction is based on bimodal stimuli having a meaningful cognitive congruency or redundancy. Indeed, incongruent or conflicting stimuli result in slower response times, which can counter intersensory facilitation (Miller, 1991; Raab, 1962). The maximal effect of redundant targets occurs when stimuli coincide in space and time (Stein et al., 1988). Interestingly, the presence of trimodal stimulus combinations produces even faster reaction times, providing converging evidence as to the presence of trimodal neurons in the SC (Diederich, 1995).

Congruence of stimuli occurs when each modality contributes redundant information about an event or object in the environment. For example, an aurally presented "A" and a visually presented "A" are congruent, as is a bright light and a loud bang occurring at the same time and location. Congruence across modalities can be quite simple, like the examples given above, or it can occur at a more abstract level and involve cognitive properties such as meaning or valence. Stimulus congruence can thus span the entire cognitive hierarchy, from spatial to semantic (Spence, 2011).

Stimuli can be congruent statistically, structurally, or semantically. Stimulus dimensions that are correlated in nature are said to be statistically congruent, such as the natural relationship between an object's size and its resonant frequency (Coward & Stevens, 2004). Structural congruence is an emergent property of the neural connections that is present at birth. For example, prothetic properties such as size and loudness are perceived as congruent because they appear through neuroimaging to be processed in the same way by the brain (Walsh, 2003). Stimuli can also be semantically congruent, such as the correspondence between the words "high" and "low" with the location of a visual stimulus on a vertical plane or with the pitch of a presented tone (Gallace & Spence, 2006). If stimulus pairs fail to meet one of these criteria for congruence, facilitation is not observed. Examples of this failure include pitch and hue, loudness and lightness, and pitch and visual contrast (Spence, 2011).

The facilitatory effect of intersensory and redundant stimuli on behavior has been displayed in a multitude of studies (e.g., Baier, Kleinschmidt, & Muller, 2006; Bernstein, Chu, Briggs, & Schurman, 1973; Bernstein, Rose, & Ashe, 1970; Dijkstra, Frauenfelder, & Schreuder, 1993; Fort et al., 2002b; Gingras, Rowland, & Stein, 2009; Gondon, Gotze, & Greenlee, 2010; Koene, Arnold, & Johnston, 2007; Miller & Reynolds, 2003). There are, however, some situations in which this facilitatory effect fails. Usually this occurs when tasks become more complex and place higher demands on attentional resources (e.g., Byce & Wesner, 2013; Grice, Canham, & Boroughs, 1984; Grice, Canham & Gwynne, 1984).

There are several models that attempt to explicate the mechanisms behind the behavioral manifestations of multisensory integration. If two stimuli from different

modalities are separated slightly in time, the *preparation enhancement model* assumes that the first stimulus increases the preparedness for responding to the second stimulus. Thus, observed shortened response latencies are attributed to attentional mechanisms (Nickerson, 1973). The *race model* (Raab, 1962) and the *coactivation* (or energy summation) *model* (Miller, 1991) have been proposed as explanations for the facilitation behavioral observed when multisensory stimuli are presented simultaneously.

The race model suggests that each stimulus is processed independently, with a response being generated by the first stimulus to reach the response criterion (i.e., information that "beat the other to the criterion finish line"; Raab, 1962). However, under certain conditions the parameters of the race model have been violated and at times cannot account for very large reductions in reaction times (Miller, 1991; Ulrich & Miller, 1997). The coactivation model was proposed as an alternative to the race model and attributes facilitation to the combination of stimulus energies, which sum to meet the response criterion (Nickerson, 1973). There is still discussion as to whether stimuli produce independent activations that are summated at a single point or whether there is some form of interaction among activations. Evidence exists to support both possibilities, with suggestions that conflicting results may be due to differences between bottom-up stimulus summation and the influence of top-down processes such as attention (Diederich, 1995; Miller, 1991).

Attention

Psychologists have always found the concept of attention difficult to define. This may be due to the fact that attention cannot be defined as a unitary concept. Although we would like to have one simple and overarching definition of what attention is, this is

becoming increasingly impossible as we realize that attention is different depending on context. For example, attention can be not only selective, but also selective along different perceptual scales such as a focus on information from one modality, on one object amongst many, or on features specific that make up an object. Attention can also be hijacked by sudden, distracting stimuli. Attention can involve both spatial and temporal components, and can be influenced differently depending on whether information is relevant or irrelevant (Styles, 2006). Here I will give a broad definition of attention; but I will also define attention in the context of feature integration and crossmodal space, as these are relevant to the present study.

Attention, in general, is a mechanism for selective awareness and response to stimuli (Kolb & Whishaw, 2003). Selective attention is the ability to facilitate awareness and responses in one modality, location, object, or feature over other competing stimuli or modalities (Posner & Bois, 1971). It has also been described at a more basic level as that mechanism which recruits neurophysiological and cognitive resources to one stimulation channel over others (Foxe, Simpson, Ahlfors, & Saron, 2005). The most highly simplified metaphor for selective attention is that of a mental flashlight that illuminates specific objects and locations to enhance their processing (Posner, Snyder, & Davidson, 1980).

Posner and Rothbart (2007) described attention as being regulated by three fundamental components: maintaining an alert state, orientation, and executive function. Alerting is involved in achieving and maintaining a state of high sensitivity to stimuli and is regulated by the locus coeruleus, right prefrontal cortex, and parietal cortex. Orienting is the selection of specific stimuli from the environment and usually involves the

movement of sensory receptors towards the selected stimuli. Brain areas involved in orienting include the superior parietal sulcus, temporo-parietal junction, frontal eye fields, and superior colliculus. Executive attention (also referred to as selective attention) is defined as a higher-order function that regulates conflict between stimuli and responses and is localized to the anterior cingulate cortex, lateral ventral prefrontal cortex, and basal ganglia.

It is becoming increasingly difficult to disentangle the effects of alerting and executive attention on the processing of information. The greatest distinction between alerting and executive attention is that attention allows for the ability to select information from one source over another whereas alerting does not. Alerting functions are non-selective in that being alert influences all incoming stimuli equally. Thus, alertness aids in developing and maintaining optimal sensory sensitivity, which is a requirement for selective attention, but does not enhance perceptual sensitivity (Posner & Boies, 1971; Posner & Pertersen, 1990; Spence, 2010). Being alert does tend to quicken response times in the detection and discrimination of stimuli, but it also leads to more response errors (Spence, 2010). Executive, or selective, attention, on the other hand, has been shown to improve both reaction times and accuracy (Fernandez-Duque & Posner, 1997). An alerting cue will inform a participant that a target will occur, but provides no information about the target, such as where, when, or in what modality it will occur.

Selective attention is evoked when a participant is given a cue or instruction set about a target's information that allows for quicker and more accurate detection and categorization of the target. Being alert is a pre-requisite for selective attention and the majority of stimuli tend to require both processes (Fernandez-Duque & Posner, 1997).

Alerting allows for the system to achieve and maintain a state of high sensitivity to incoming stimuli, while attention monitors the stimuli and resolves any conflict between them. Both mechanisms influence overt and covert orienting towards stimuli (Posner & Rothbart, 2007). Throughout this paper, the term "attention" will be used to refer to selective or executive attention. Although attention is a fairly diffuse concept, it can be studied in various contexts. As noted earlier, attention can operate within different contextual scales. It can be directed to features within one stimulus, to stimuli within one modality, to stimuli in multiple modalities, and all across space and time. This review will focus on literature dealing with attention as it applies to feature integration and crossmodal attention and its direction within space.

Feature Integration Theory

Feature integration theory (FIT) is a model of visual object recognition (Treisman & Gelade, 1980) and was an attempt to solve the binding problem – that is, how separate sensory inputs are combined so that we perceive unified objects. The model is premised on evidence that visual perception occurs in two stages. Processing at the first stage occurs automatically, pre-attentively, and in parallel. At this stage specialized "feature detectors" code for sensory features such as color orientation and size. Visual features are defined as stimulus properties that can activate specific detectors in parallel across the optic array. It is important to clarify that these "feature detectors" defined in this context are actually perceptual in nature and involve visual areas beyond the cell-based "feature detectors" located at, and prior to, the primary visual cortex. These perceptual features fall into a hierarchy, with surface-defining features characterized by color, luminance,

and relative motion and object features defined by the integrative spatial arrangement of one or more surface-defining features (Styles, 2006).

At the first stage of the FIT model, surface-defining features are coded on specific maps. Each feature has its own independent-coding feature map that registers the feature across the entire optic array. Again, information regarding each feature is processed preattentively and in parallel, and there is no coordination of information across the different feature maps (Treisman, 1982). Thus, the location of different features across the feature maps cannot be coded for at the first stage. The establishment of multiple feature location requires higher-level attention (Treisman, 1985).

The second stage of the FIT model deals with the integration of multiple features to form what is referred to as a *conjunction of features* that results in the perception of whole, integrated objects. The conjunction of separable features can occur in three ways (for a review see Treisman & Schmidt, 1992). First, an integrated object can be predicted from how well its coded features fall into 'predicted object frames'. This implies that the features fit into a current context or expectation. Second, features can also be conjoined through the focusing of attention on a master map of locations (Treisman, 1985; 1986). The master map represents the locations of all possible features. When attention is focused on a particular area of the map, it allows for the processing of all active features within that area and creates a temporary representation of the integrated object. The recognition of a conjunction object depends on the matching of the object representation to object descriptions stored in visual long-term memory (see Styles, 2006 and Quinlan, 2003 for reviews). Third and finally, features can become integrated at random, resulting in illusory conjunctions. Illusory conjunctions occur because features registered at the

first, pre-attentive stage are free-floating and not organized within a coordinate system (Treisman & Schmidt, 1992).

Evidence for the FIT model originally came from measures of reaction times in visual searches of targets with either unique or conjunction features (Treisman & Gelade, 1980). In these experiments, participants had to search for targets amongst colored letters. In the unique feature task, the target was either a blue letter with any shape, or the letter S in any color. Non-target distractors consisted of green Xs and brown Ts. In the conjoined feature task, the target was a green T, which integrated the features of color and shape, both of which independently made up the distractors. The set-size varied from a single target to a target amongst 5, 15, or 30 distractors. In the unique feature task, reaction times in identifying the presence of the target were unaffected by the number of distractors. This supports the first stage of the FIT model and the idea that unique features can be processed in parallel, without attention. This pop-out effect has been replicated in multiple studies (Treisman, 1985, 1986, 1988; Treisman & Sato, 1990; Treisman & Schmidt, 1992). In the conjunction target task, however, reaction times with integrated target presentations increased linearly with the number of distractors present suggesting that attention must be focused serially on each object in the visual array, with the search terminating once the target is located and properly discriminated from nontargets. This lends support to the idea that the detection and discrimination of integrated features requires attention (Treisman, 1985; 1986; 1988; Treisman & Sato, 1990; Treisman & Schmidt, 1992). Treisman (1986) also looked at the effect of pre-cuing the target location for both unique feature targets and conjunction targets. She suggested that a cue should eliminate the need for serial search with conjunction targets and indeed valid cues

resulted in improved (i.e., faster) reaction times for conjunction targets. No cue advantages were noted for parallel 'pop-out' unique feature targets, however. This lends support to the idea that attention is needed for the perception of integrated objects but not for the perception of unique, preattentive features.

There is evidence, however, of perceptually difficult unique feature targets requiring attention (Joseph, Chun, & Nakayama, 1997; Prinzmetal, Presti, & Posner, 1989) as well as conjunction targets being processed preattentively, or at least more rapidly than expected from a serial search (Houck & Hoffman, 1986; Nakayama & Silverman, 1986). There is still ongoing debate over the involvement of attention in feature detection, with attentional demand appearing to be dependent mostly on the cognitive difficulty of the task and the discriminability of the target from the distractors (Quinlan, 2004). In 1991, Treisman did integrate findings of very fast conjunction searches into her model. She suggested that rapid, seemingly preattentive conjunction search performance might be due to feature inhibition at the level of the feature maps, with active feature detectors mutually inhibiting non-target feature activity. This may result in the inhibition of processing certain areas of the master location map, thereby minimizing the area that needs to be searched and resulting in faster processing speeds. It may also be that features such as color and form are not processed as independently at early visual centers as previously thought (Cohen, 1997). In light of these rapid conjunction searches, Treisman (1991) suggested a modification of the FIT model in which visual input could be initially processed with respect to the master location map and then broken down into specific features. In this way, features can be either integrated by attention, or objects prone to early integration objects can be deconstructed into their

compositional features. Features and/or integrated objects would then form an object file that is compared against objects stored in visual long-term memory for recognition.

Visual attention, according to the FIT model, is object based, with object perception and visual attention dependent on the interaction between feature maps, the master map of locations, and individual object representations (Styles, 2006). Attentional selection can be early or late depending on the demands of the task and the load it places on perceptual resources (Treisman, 1993). This conclusion is also based on the work of Lavie (1995, 2005) whereby the amount of interference from irrelevant distractors on the processing of the target is proportional to the load or difficulty of processing the target. This is referred to as the perceptual load hypothesis and will be discussed in more detail later. Treisman (1993) allows for four levels of attentional selection in the FIT model. Selection can be based on location, individual features, locations defined by objects, and late selection where attention dictates which object representation should control responses. Thus, according to FIT, attention can operate on a number of different levels depending on the demands of the task.

The influence of distractors.

The central tenant of FIT is the idea that there are differences in the underlying processes involved in feature detection and feature binding. Differences of opinion on this tenant led to a prolonged debate between Duncan and Humphrey's (1989, 1992) and Treisman (1991, 1992). Duncan and Humphreys (1989) argued that search performance was not necessarily dependent on underlying processes like attention, but on the similarity of the target to distractor attributes and on the similarity of distractor attributes to each other. Duncan and Humphreys' Attentional Engagement Theory (AET) is

comprised of three stages. First, input is evaluated across the visual array based on structural and feature similarities. All input at this stage is processed in parallel. At the second stage, input that matches an attentional template is allowed to proceed to higher levels of processing. Finally, information relevant to the detection and/or discrimination of the target is allowed into visual short-term memory for further analysis. They suggested that the difficulty of conjunction search is not due to feature binding or attentional demands but rather to the similarity of target features with the non-targets or the dissimilarity of non-targets attributes to each other. This similarity between targets and non-targets, and the possible variability of non-targets similarities, is what makes conjunction search more time consuming than feature searches as - opposed to demands on cognitive processes such as attention.

Treisman (1991) responded to these arguments with a series of experiments in which targets shared the same degree of similarity to non-targets in both a feature search and a conjunction search. The conjunction search was still much slower than the feature search, prompting Treisman to conclude that performance differences could not be due solely to target/non-target differences. Duncan and Humphreys (1992) countered that Treisman's stimuli were not sufficiently similar across the two types of search. They conducted their own experiments using similar targets/non-targets and found equivalent response time performance for feature and conjunction search. Treisman (1992) got the final word in the argument, however, stating that Duncan and Humphreys (1992) feature search was simply too difficult, and thus required attention. In a final experiment, Treisman (1992) added a unique feature to the non-targets in a conjunction search task, predicting that if AET were correct, this unique feature should actively suppress the non-

targets and improve performance by speeding reaction times. No such improvement was found. In conclusion, although the central tenant of FIT remains robust, Duncan and Humphreys (1989, 1992) demonstrated that some consideration must be taken with regards to the similarity of targets and non-targets in the designing of visual search tasks.

Crossmodal Attention

Crossmodal attention is defined as the extent to which a system can selectively direct attention to information within a particular sensory modality, sometimes at the expense of processing stimuli in other modalities. As we know from the literature, however, that this is at odds with intersensory facilitation, where the presence of stimuli from multiple modalities can actually facilitate processing, even if attention is focused on stimuli presented within only one modality. Crossmodal attention also encompasses the process by which stimuli can capture attention across modalities (this is referred to as crossmodal capture). When crossmodal attention is placed in the context of the spatial distribution of stimuli, the orienting of attention in one sensory modality usually results in a concomitant shift of attention in other modalities to the same location and/or object, usually with an observed facilitation of performance in detecting and/or discriminating stimuli (Spence, 2010).

There are currently four major models that attempt to describe the underlying mechanisms of spatial crossmodal attention. The modality-specific model emphasizes that attentional resources are directed independently within each modality with no crossmodal links in spatial attention (Wickens, 2008). As there is overwhelming evidence supporting the idea of some sort of connection between sensory modalities, this model is generally not supported (see Spence, 2010). In direct counterpoint to the modality-

specific model, the supramodal model proposes that there is an overarching attentional system in the brain that allocates resources to all sensory systems. This model assumes that people cannot split attention between locations or modalities simultaneously (Farah, Wong, Monheit, & Morrow, 1989). In a related vein, the separate-but-linked hypothesis suggests that there are modality-specific attentional systems at the earliest levels of processing but that these independent systems are subsequently linked crossmodally, although sometimes in an asymmetric fashion (Spence & Driver, 1996, 1997). Finally, the hybrid model posits that both modality-specific and supramodal attentional systems exist and interact (Eimer & Van Velzen, 2002).

The parameters of crossmodal attention, as well as the existence of the above-mentioned attention models, are usually investigated using spatial cuing paradigms with stimuli in more than one modality. It is important to define certain attention terms used in this research, such as *overt* and *covert* orienting and *exogenous* and *endogenous* mechanisms and *cues* (see Table 1). Overt orienting involves shifting sensory receptors (i.e., the eyes, body - and if you are a species capable of this - ears) towards an object or spatial location of interest. Overt orienting is usually reflexive. It is triggered by sudden stimuli and is generally considered to be a bottom-up process. In contrast to this, covert orienting is generally considered to be a top-down process that involves internal changes in attention with no shifting of sensory receptors. Covert attention is driven by instruction sets, symbolic cues, context, and/or motivational factors (see Spence, 2010 for a review of this terminology).

Attentional orienting can also be classified as operating using endogenous or exogenous mechanisms. Endogenous (top-down) orienting mechanisms are voluntary

shifts of attention driven by expectations. Endogenous attention is usually manipulated

using symbolic cues such as arrows or words that predict the likely location or modality

of a target. Endogenous cues are thus likely to be valid predictors of the location of an

upcoming target. For example, a cue may be 80% valid and 20% invalid in predicting a

given stimulus target location, and it is up to an individual's correct cognitive

interpretation of these cue probabilities that determines response likelihood. Exogenous

(bottom-up) orienting mechanisms involve reflexive shifts in attention driven by salient,

often unexpected, peripheral stimuli. Exogenous cues are usually sudden auditory tones

or visual cues that, although they occur at potential target locations, are generally not

predictive of the upcoming target's location. Thus, exogenous cues have a 50% chance of

being valid or invalid, allowing for no cognitive interpretation of cue probabilities.

Exogenous cues, however, can be used to manipulate covert orientation towards stimuli,

meaning that they do not always provoke purely reflexive exogenous processes. They can

also be manipulated in order predict an upcoming target's location in the same fashion as

endogenous cues (Santangelo & Spence, 2008; Spence & Driver, 1998; Yantis & Jonides,

1990). This means that there are four combinations of attention-based responses that can

be studied: endogenously cued overt attention, exogenously cued overt attention,

endogenously cued covert attention, and exogenously cued covert attention (see

Appendix A for a summary of the definitions of overt and covert orienting, endogenous

and exogenous mechanisms, and endogenous and exogenous cues).

The most commonly used paradigm for studying crossmodal spatial attention is

the orthogonal cuing paradigm developed by Spence and Driver (1996; 1997) in their

seminal behavioral studies on the links between auditory and visual endogenous and

exogenous overt and covert attention. In this paradigm, participants fixate on a center point on a screen and their attention is drawn either to the left or right with central arrow cues (endogenous) or sudden peripheral cues (exogenous) that can be either visual or auditory. Targets then are visually presented in one of the four corners of the screen or aurally presented from speakers placed at the four corners of the screen. Participants then make a speeded discrimination of the target elevation position (up vs. down) independent of horizontal position (left vs. right) it appears on. In studying covert attention, participants must maintain central fixation through the entire task. When saccades are allowed towards the stimuli, the task becomes overt. The orthogonal cueing task replaced Posner's cueing task, whereby cues directed participant's attention to the left or right and participants made simple left verses right target location discriminations (e.g., Posner, 1980). Although simpler left verses right location paradigms have yielded important results, this method has often been found to produce priming effects that confound true crossmodal attentional effects (Spence & Driver 1998). In light of this, the majority of current crossmodal attention studies now make use of the orthogonal cueing paradigm.

Using the Spence and Driver paradigm, multiple studies have observed the facilitation of reaction times and/or accuracy in detecting and/or discriminating target stimuli in space in the visual, auditory, and tactile modalities if their locations were validly cued by either endogenous or exogenous cues in the same or different modalities (Kida, Inui, Tanaka, & Kakigi, 2011; Koelewijn et al., 2010; Macaluso & Driver, 2005; Spence & Driver, 1996; 1997; 1998a; 1998b; Talsma, Senkowski, Soto-Franco, & Woldorff, 2010). What makes crossmodal attention unique from multisensory integration is the temporal relationship between stimuli. As stated earlier, stimuli presented

simultaneously or up to 100 msec apart in time result in cellular and behavioral multisensory integration. Stimuli separated by 200-300 msec are classified as cues and targets and tap into crossmodal attentional properties (McDonald, Teder-Salejarvi, & Ward, 2001). If cues and targets are separated by more than 400 msec, a cognitive memory phenomenon known as inhibition of return (IOR) is observed. Within a specific latency range between cue and target/non-target presentations, there is a slowed response to validly cued targets to the point of no response at all to a given location. The idea of IOR is that there is an active high-end suppression of attentional resources returning to a specific cued location that has earlier been associated with a no target presentation (Klein, 2000).

Seminal behavioral studies of crossmodal attention.

As mentioned above, Spence and Driver (1996; 1997) conducted seminal studies on audiovisual links in endogenous and exogenous covert and overt attention using the orthogonal cuing paradigm. Appendix B summarizes the major findings of their experiments. Generally speaking, a valid endogenous or exogenous cue will facilitate the discrimination of a target stimulus regardless of the modality of the cue or target, whereas invalid cues inhibit performance. Asymmetrical audiovisual links do exist in exogenous covert orienting, whereby valid exogenous visual cues fail to facilitate the processing of auditory targets. This is referred to as a null vision-on-audition finding. These findings will now be explained in more detail.

In terms of endogenous overt attention (see Appendix B, entry 1), Spence and Driver (1996) found facilitation of performance in elevation discriminations of visual and auditory targets that were validly cued by visual arrow cues. Invalid cues resulted in

inhibited performance. This study looked only at the influence of visual endogenous cues on the processing of visual and auditory targets. The ability of correct cues to enhance performance and of incorrect cues to hinder performance as defined by changes in reaction times is referred to as the cue validity effect. They also found that attentional shifts in one modality tended to be accompanied by complimentary shifts in the other modality, even if the event probabilities in the secondary modality were biased against such a shift. Results also showed that participants could split visual and auditory attention between locations in blocked conditions but not in trials that were constantly re-cued. This finding suggested that participants could split visual and auditory attention only if given a long enough time period to make the necessary associations. Covert endogenous attention (see Appendix B, entry 2) yielded the same results. The results of these experiments suggested that audiovisual endogenous attention is neither completely supramodal nor modality-specific, but rather operates on a separate-but-linked or hybrid system (Spence & Driver, 1996).

In another series of experiments Spence and Driver (1997) found similar results for audiovisual exogenous covert and overt attention. They observed facilitated performance in elevation judgments in both visual and auditory modalities when targets were preceded by sudden, uninformative auditory cues either to the left or right of the display. The most interesting finding was that visual exogenous cues failed to influence performance for auditory target discrimination when saccades were not allowed, but visual cues still influenced performance for visual targets. This suggests an asymmetry in exogenous covert attention in which audition influences vision, but not vice versa, and is referred to as the null vision-on-audition asymmetry (see Appendix B, entry 4). Again,

these results contradict a purely supramodal crossmodal system. The authors presented a few potential contributors to this asymmetry (Spence & Driver, 1997). In the environment, auditory events tend to be transient whereas visual events are more stable (Neumann, Van der Heijen, & Allport, 1986). It may thus be more practical to shift vision to transient auditory events as they may be accompanied by more sustained visual events. We are also visual dominant beings, so we may garner more information by focusing our attention on a visual stimulus than to focus auditory attention on an event that we have already seen (Spence & Driver, 1997).

The observed asymmetry in exogenous attention of visual cues on auditory targets may also be due to the organization of auditory and visual information in the SC, which (as mentioned previously) is integral to overt orienting. Covert orienting mechanisms may be closely related to, or may be dependent on, the same cellular mechanisms that underlie overt orienting (Desimone, Wessiger, Thomas, & Schneider, 1992). There are multimodal (visual, auditory, and somatosensory) spatial maps in the lower layers of the SC, but there are no pure auditory maps of space. There are pure visual and spatiotopic maps in upper layers, however (Meredith & Stein, 1983), giving a relative dominance of vision at the cellular level in the SC. Assuming overlapping overt and covert operations in the SC, this may account for the asymmetry observed in exogenous covert attention.

It is also possible that the observed asymmetry in exogenous covert audiovisual attention is simply the product of the orthogonal cuing paradigm. For example, using a Posner-like cuing paradigm, Ward (1994) found an effect opposite that of the null vision-on-audition finding of Spence and Driver (1997). Ward found an asymmetry in which visual cues influenced reaction times to visual and auditory targets, but auditory cues

influenced only auditory targets. Spence and Driver (1997) suggested that Ward's results might have been due to priming effects. Spence & Driver (1997) concluded that visual cues influenced the detection of auditory targets in the Posner-like cuing paradigm but not in the orthogonal cuing paradigm. However later, Ward, McDonald and Lin (2000) controlled for priming effect and replicated Ward's earlier 1994 findings. The visual cues may have failed to capture auditory attention in the orthogonal cuing paradigm because the auditory targets fell outside of the focus of attention created by the visual cues, which were not in spatial alignment with the auditory targets (Prime, McDonald, Green, & Ward, 2008).

The existence of this null vision-on-audition asymmetry in exogenous covert attention with the orthogonal cuing paradigm has been questioned by McDonald, Teder-Salejarvi, Heraldez, & Hillyard (2001). In their study, participants discriminated peripheral low and high frequency auditory targets that were preceded by non-predictive left vs. right visual light-flash cues. No-go trials consisted of auditory targets presented at fixation. The authors observed facilitation of auditory frequency discrimination with the presence of visual cues. They suggested that the null vision-on-audition effect observed by Spence & Driver (1997) might be a limitation of the orthogonal cuing paradigm or be due to criterion shifts made by the participant about how salient a stimulus must be before responding to it.

The present study made use of the orthogonal cuing paradigm in which centrally presented visual endogenous arrow cues directed participants to covertly attend to the left or right side of a computer monitor where they made an elevation judgment as to the location of an indicated multisensory audiovisual stimulus. All stimuli presented in the

experiments were above suprathreshold in nature and thus eliminated the possibility

results being confounded by criterion shifts based on stimulus saliency.

Interactions between endogenous and exogenous cues.

Using the orthogonal discrimination paradigm, Chen, Chen, Gao and Yue (2012)

recently explored the interaction between auditory and visual exogenous and endogenous

cues on covert attention. In this study, both exogenous and endogenous cues were

presented in the same modality before the presentation of a target in another modality.

The endogenous cue was presented first, followed by the exogenous cue and then the

target/nontarget array. The researchers wished to examine the interaction between the two

orienting mechanisms, as previous research suggests that endogenous and exogenous

mechanisms are linked in some fashion (Santangelo & Spence, 2008). Functional MRI

evidence also suggests that the two mechanisms interact, with endogenous cues activating

the temporal-parietal junction and inferior frontal gyrus and exogenous cues activating

the superior parietal gyrus (Santangelo, Olivetti Belardinelli, Spence, & Macaluso, 2009).

Chen et al. (2012) looked at possible interactions behaviorally, varying the

validity of both types of cues. When auditory cues preceded visual targets, they found

that both endogenous and exogenous cues facilitated covert orienting towards the visual

targets. The authors also found that the cue validity effect of both the endogenous and

exogenous cues on their own was significantly larger when the accompanying subsequent

exogenous or endogenous cues were invalid. This suggests that in auditory attention,

when one type of cue is invalid, the other cue type has to compensate in order to direct

spatial attention to the location of a visual target. With visual cues acting on auditory

targets, facilitation of reaction time and accuracy was only observed when both cues were

valid suggesting that both types of visual cues co-directed spatial attention towards the auditory targets. In addition, the authors found that invalid visual exogenous cues were inhibited by valid visual endogenous cues and that invalid endogenous cues allowed for valid exogenous cues to dominate the orienting process. Overall, the study by Chen et al. (2012) suggests that endogenous and exogenous covert orienting mechanisms are separate mechanisms that interact and compete for shared resources.

Influence of multisensory cues on spatial orienting.

Behavioral studies have also looked at the influence of multisensory bimodal cues (concurrent auditory and visual) on spatial orienting. Generally, bimodal cues do not appear to have an influence on the effectiveness of spatial orienting over-and-above what is observed with unimodal cues (Spence & Santangelo, 2009). ERP studies, however, have shown superadditive neuronal responses following bimodal cues as opposed to unimodal cues, suggesting that bimodal cues can cause multisensory integration at the neuronal level, but that this may not be enough to magnify behavioral responses (Santangelo, Van de Lubbe, Olivetti Belardinelli, & Postma, 2008). In review, multisensory cues do not capture attention any better than unimodal cues, suggesting that spatial attention and multisensory integration may be relatively independent mechanisms (Spence & Santangelo, 2009). It has been shown, however, that unimodal cuing effects fail under conditions of high perceptual load (i.e., dual-task performance) whereas multisensory cues are more effective as attentional directors in the face of increased cognitive load (Santangelo & Spence, 2007). Multisensory cues may be more robust attentional directors because they are unaffected by perceptual load increases in any one modality (Spence & Santangelo, 2009).

The detection of visual targets is facilitated by exogenous auditory cues when cues and targets are separated by 100 ms or more, meaning that the facilitation is not due to passive multisensory integration. To borrow from signal detection theory (SDT), an increase in d', or sensitivity, is observed when a cue significantly precedes a target in time, suggesting that attention has an effect on early perceptual processing by enhancing perceptual salience. Decision criteria, β, tend to improve on valid but not on invalid trials (McDonald, Teder-Salejarvi, & Hillyard, 2000). When cues and targets are placed closer together in time (i.e., less than 100 msec SOA) there are observed improvements in d', but not in β (Schirillo, 2011). This suggests that improvements in target discrimination are correlated with enhancements in sensory processing, but not with systematic changes in decisional processing and that the crossmodal covert orienting of attention within audition and vision occurs at lower perceptual levels as well as higher decision-making levels (Ciaramitaro, Cameron, & Glimcher, 2001; Marks, Ben-Artzi, & Lakatos, 2003). Thus, it appears as though tasks using multisensory stimuli influence lower, sensory levels of processing, whereas crossmodal stimuli influence both sensory sensitivity as well as perceptual decision-making processes.

Event-related potential studies of crossmodal attention.

Event-related potential (ERP) studies also offer valuable insight into how people process crossmodal information. ERP studies look at specific waveforms representing different temporally defined processes. Crossmodal ERP studies typically focus on specific waveforms such as the P1/N1 waveform complex. This waveform reflects early sensory processing and it is generally argued that this is the first moment in brain activity when attention can influence sensory perception (Luck, 2005). Other waveforms used to

study crossmodal attention include the late anterior directing attention negativity

(ADAN) and late directing attention positivity (LDAP) waveforms which occur after cue

presentations and are believed to reflect supramodal orienting of attention. The

measurement of these waveforms allow for good temporal resolution of crossmodal

attention processing in the brain as well as some generalized localization of function.

ERP studies provide some insight into the levels of processing at which crossmodal

attention operates and also helps to distinguish between the separate-but-linked,

supramodal and hybrid attention systems. They also help to fill in the gaps left by

limitations inherent in behavioral studies, mainly a lack of insight into neural processes

and levels of processing (Eimer & Van Velzen, 2002). I will now examine these

waveforms more closely.

In general, ERPs elicited by visual and auditory targets are more negative when a

stimulus appears after a valid location cue than after an invalid location cue (Eimer &

Shcroger, 1998; Green & McDonald, 2006; McDonald et al., 2001; Teder-Salejarvi,

Munte, Sperlich, & Hillyard, 1999). This pattern is referred to as a negative difference

(Nd) and occurs in two stages. The Nd1 component starts at approximately 140 msec

post-target onset and is centered over the parietal scalp. This component is believed to

subserve the shifting of attention in space. The Nd2 component begins approximately 200

msec post-target onset and is centered over the fronto-central scalp. This electrical

activity is believed to reflect the processing facilitation caused by a validly cued target

necessary for behaviorally-defined faster response times (Green & McDonald, 2006).

The Nd effects are similar for both auditory and visual targets, which suggests

that attention is not entirely modality-specific and involves partially common

mechanisms or crossmodal links. The Nd effects are not exactly identical across modalities, however. Generally speaking, attending to visual targets results in a larger negative difference effect than when attending to auditory targets. These coarse potential asymmetries are also observed when visual cues precede auditory targets and vice versa, supporting the behavioral studies of Spence and Driver (1997). Although asymmetrical links are observed in sustained endogenous attention, *symmetrical* links are observed with transient exogenous attention, again in alignment with the Spence and Driver (1997) behavioral findings. These asymmetries point to a separate-but-linked system (Eimer & Schroger, 1998). Other authors argue, however, that the differences observed between visually and aurally induced Nds are not large enough to completely negate the presence of a supramodal system (McDonald et al., 2001). It has also been proposed that visual and auditory attention mechanisms are linked together, but with a central amodal pool of resources, as in the hybrid model of crossmodal attention. Evidence for this idea has been derived from changes in steady-state evoked potentials within and between modalities, with attentional capacity being larger between modalities than within any one modality (Talsma, Doty, Strowd, & Woldorff, 2006).

Eimer and Van Velzen (2002) attempted to resolve the supramodal vs. separate-but-linked argument by suggesting a way of investigating a hybrid system using ERPs. They suggested that observed ERP modulations during shifts of attention (Nds) represent a phasic selection of relevant locations that operates in a supramodal manner. The influence of this supramodal spatial selection on stimulus processing within the target modality may also depend on the tonic state of activity in that modality, which varies with task relevance. According to the authors, the modality-specific differences in

activation level can in turn modulate effects of spatial attention (i.e., attentional effects will be larger when tonic baseline activity in the target modality is high and attenuated when activity levels are lower). Eimer and Van Velzen recommended that this effective baseline be taken into account when designing ERP studies and interpreting their results.

The early latencies of the negative difference effect (~200 msec) suggests that attentional processes involve modulations of early sensory processing in the brain, in line with the signal detection studies mentioned previously (Focker, Hotting, Gondan, & Roder, 2010; McDonald et al., 2001; Talsma & Kok, 2001; Teder-Salejarvi et al., 1999). These modulations are also observed in waveforms that represent the earliest processing of information in the brain. Stimuli are first registered in primary sensory cortices as the C1 waveform, which is not influenced by attention or any top-down mechanism. Following the C1 waveform is a positive waveform referred to as P1, occurring 80-130 msec post-stimulus and a negative waveform referred to as N1 occurring 150-190 msec post-stimulus. These waveforms reflect the spreading of activation to secondary sensory cortices and it is here that one sees attentional influences (Luck, 2005). Stimuli appearing at attended locations generally elicit larger P1 and N1 waves over occipital cortex (Eimer & Schroger, 1998; Green & McDonald, 2006; McDonald et al., 2001).

While P1/N1 and Nd waveforms reveal processing at the beginning of target onset, ADAN and LDAP waveforms are associated with cue onset, reflecting attentional control processes. The ADAN is a relative negativity over anterior scalp sites contralateral to the to-be-attended target location occurring approximately 300-500 msec post-cue onset. The LDAP is a relative positivity over posterior scalp sites contralateral to the to-be-attended location beginning approximately 500 msec post-cue onset and lasting

until target onset. Both ADAN and LDAP have been observed in crossmodal tasks with both auditory and visual cues. Generally, these delayed waveforms in the ERP literature have been tied to supramodal processes involving the orienting of attention and arising from multisensory areas in the frontal and parietal lobes (see Green & McDonald, 2006 for a review). Besides the asymmetry observed in Nd waves to visual and auditory targets, another asymmetry exists in these cue-induced waveforms, in which visual cues directing attention to visual targets elicit sustained biasing effects approximately 100 msec earlier than the same cues directing attention to auditory stimuli, again suggesting a separate-but-linked system (Foxe et al., 2005; Foxe & Simpson, 2005). However, as discussed above with respect to neural and behavioral properties, this asymmetry can also be explained by the fact that vision is our most refined and relied-upon sense (Talsma, Kok, Slagter, & Cipriani, 2008).

The presence of cue induced anticipatory biasing in the brain suggests that crossmodal processing cannot solely occur at early sensory processing levels under all conditions. Activation of specific sensory systems just prior to target presentations also suggests the maintenance of sensory-specific biased attentional states by a frontal-parietal attention system. Crossmodal attention, then, involves interactions between all levels of stimulus processing (Giard & Peronnet, 1999; Foxe et al., 2005; Foxe & Simpson; 2005; Talsma et al., 2008).

Automaticity of Multisensory Integration and Crossmodal Attention

In order for a process to be considered automatic, it must meet two criteria: intentionality and load insensitivity. The intentionality criterion states that an automatic process is one that is not subject to voluntary control. A truly automatic process is not

facilitated by the focusing of attention on a certain stimulus or inhibited by the focusing of attention elsewhere. The load insensitivity criterion states that an automatic process is not hindered when concurrent information load or task difficulty is increased (Jonides, 1981; Koelewijn et al., 2010; Santangelo & Spence, 2008; Yantis & Jonides, 1990). Studies have shown that multisensory integration can be influenced by attention (Fairhall & Macaluso, 2009; Talsma & Woldorff, 2005), although there is also evidence that early multisensory integration does not require or involve attention (Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008; Vroomen & de Gelder, 2000). It may be better, then, to distinguish between early automatic and late attention-mediated, multisensory integration effects (i.e., Koelewijn et al., 2010). The involvement of top-down mechanisms may be at the heart of what distinguishes endogenous covert crossmodal attentional processes from other, more automatic forms of multisensory integration.

Similar to multisensory integration, the question also arises as to whether attentional capture is an automatic process. As endogenous attention is largely considered to be voluntary, the question of automaticity has more relevance for exogenous attention, which is generally considered to be more of a bottom-up operation (Santangelo & Spence, 2008). Unimodal studies of the intentionality of exogenous attention have revealed mixed findings, with some studies showing abrupt-onset stimulus presentations failing to interrupt directed attention (e.g., Yantis & Jonides, 1990) while others show a disruption of attention with comparable abrupt stimulus onsets (e.g., Van der Lubbe & Postma, 2005). Studies involving multiple modalities, however, lean towards the conclusion that exogenous orienting does not meet the intentionality criterion and is not a fully automated process, suggesting that bottom-up orienting may be automatic in

unimodal tasks, but not in crossmodal ones (see Santangelo & Spence, 2008 for a review). For example, when participants monitor a rapid sequence of visual and auditory targets and distractors that engage attention while performing a concurrent exogenous orthogonal cuing task, their exogenous spatial orienting is suppressed at high load conditions in the monitoring task (Santangelo, Olivetti Belardinelli, & Spence, 2007).

If exogenous attention is a truly automatic process it should also conform to the load-insensitivity criterion, whereby target stimuli will be processed and attended to and distractors, no matter what their number, will be ignored. The perceptual load hypothesis (Lavie, 1995) of selective attention states that perception has a limited capacity and that all stimuli are processed in an automatic fashion until available resources are exhausted. In unimodal settings, research shows that increasing perceptual load reduces distractor effects (Lavie, 2005), however in crossmodal settings separate capacities for different sensory modalities are observed (Talsma, Doty, Strowd, & Woldorff, 2006).

As mentioned previously, research by Santangelo et al. (2007; 2008) showed that crossmodal exogenous cuing effects with unimodal cues were eliminated with a concurrent monitoring task. The use of multisensory cues, on the other hand, captured spatial attention regardless of the perceptual load of the concurrent task (Santangelo & Spence, 2007). Engaging cognitive resources in perceptually demanding tasks appear to make it more difficult for peripheral stimuli to capture attention. Multisensory stimuli, on the other hand, are able to capture attention regardless of the difficulty of a competing task. Classic, crossmodal exogenous attention, then, fails to meet the load-insensitivity criterion, although multisensory cueing does. Exogenous orienting, therefore, does not

truly hold to the criteria defining automaticity, although it can under specific circumstances (Santangelo & Spence, 2008).

Current Ideas on Interactions

Current research is looking into how multisensory integration and crossmodal attention interact. Some researchers argue that multisensory integration and crossmodal attention are different processes with separate neural mechanisms. This assumption rests largely upon the observed temporal constraints of stimulus presentation for multisensory integration and crossmodal attention, with multisensory integration being optimal when stimuli co-occur in time and crossmodal attention being evident when temporal separation exits between cues and targets (McDonald, Teder-Salejarvi, & Ward, 2001). Others argue that crossmodal attention is simply a form of multisensory integration, and that a temporal distinction cannot be made, as multisensory cells in the cat cerebral cortex can still show integration effects with asynchronies of up to 600 msec (Calvert & Thesen, 2004; Meredith et al., 1987; Wallace, Meredith, & Stein, 1992). This latency would allow enough time for crossmodal attention to engage and would suggest that multisensory integration and crossmodal attention are based on similar underlying processes (Macaluso, Frith, & Driver, 2001).

For the purposes of this study, multisensory integration and crossmodal attention were treated as two separate processes. The two processes may share similarities at the neural level, however we chose to focus on the delineation of multisensory integration and crossmodal attention as exemplified by behavioral research (McDonald et al., 2001). This controversy also rests partly on the uncertainty of knowing the processing stages associated with multisensory integration. Multisensory integration may occur during

early preattentive stages or during later stages of processing, or it may function in parallel with attention across the processing hierarchy (Koelewijn et al., 2010).

The early integration framework model suggests that multisensory integration occurs at an early, preattentive stage, independent of attention and that integration essentially drives attentional capture. Evidence for this model comes from studies of the "pip and pop" effect (whereby spatially non-informative auditory stimuli can enhance visual search; Van der Burg et al., 2008), the McGurk effect (McGurk & McDonald, 1976), and the ventriloquism effect (Bertelson et al., 2000), all of which occur at pre-attentive processing stages. Multisensory cues can also lead to attentional capture in situations where unimodal stimuli cannot (Santangelo & Spence, 2007).

The late integration framework model posits that attention is required for integration to occur. This model suggests that auditory and visual events are individually enhanced by means of unimodal attention before being integrated at higher heteromodal areas. Talsma & Woldorff (2005) observed multisensory integration effects in enhanced frontal positivity 100 msec after bimodal stimulation, suggesting that integration cannot occur without attention (see also Talsma, Doty, & Woldorff, 2007).

The parallel integration network was proposed by Calvert and Thesen (2004) and suggests that multisensory integration takes place at multiple stages with dynamic modulation by attention occurring between stages. Multisensory integration can occur at early or late stages depending on the task and the resources available. In this way, it is possible that similar resources are used for both multisensory integration and attention (Meredith et al., 1987), which reinforces the idea of parallel interactions between them (Calvert & Thesen, 2004; Koelewijn et al., 2010). For example, near-threshold events

might need attentional resources for integration to occur, and can only take place later in time because top-down control takes a longer time to have an effect. Salient, suprathreshold stimuli, on the other hand, may occur automatically at earlier levels (Calvert & Thesen, 2004). This is consistent with fMRI data showing that spatial attention can affect multisensory integration in both cortical and subcortical areas (Fairhill & Macaluso, 2009). A similar pattern can be derived from considering the influence of task complexity. Multisensory integration will occur more-or-less passively in an environmental scene where there is a low amount of competition between stimuli, whereas integration in complex scenes requires the guidance of top-down attention (Talsma, Senkowski, Soto-Franco, & Woldorff, 2010).

Present Study

Where along the cognitive hierarchy, then, do multisensory integration and attention interact? Is attention a prerequisite for integration, or can multisensory integration occur in simple tasks without it? Are integration and attention parallel systems that interact at all levels of the cognitive hierarchy, from simple detection to complex discrimination? Or is there an exact point in the processing system when attention enters the picture? The present study sought to answer the question of where these two processes merge using several behavioral tasks that varied in their attentional demands on observers, from simple single- to complex integrated feature discriminations. The tasks presented a select set of multisensory auditory-visual stimuli ranging in complexity from simple shape and magnitude detection, to the discrimination of integrated features. MSI will be sampled within a pre-attentive task, a task demanding moderate levels of attention, and a difficult task with a very high attentional load.

Based upon the previous literature, I hypothesized that the behavioral manifestations of multisensory integration would be variable across the three tasks due to differences in the attentional load of each task, and that these changes would manifest in differences in behavioral facilitation to multisensory stimuli as well as differing effectiveness of attentional directing endogenous arrow cues. I anticipated no multisensory integration or arrow cue effect to be observed in a preattentive task due to the pop-out nature of the target stimuli. The task demanding the most attention could result in a reversal of intersensory facilitation, with performance being hindered by the presence of multisensory stimuli due to limited attentional capacity. I also hypothesized that intersensory facilitation would most likely to be observed in a task demanding moderate levels of attention. As the attentional demand of the task increased, I expected to see more reliance on endogenous directing cues as well as less facilitation.

Method

Participants

Forty-five participants were recruited from psychology classes at Lakehead University and the community. Participants had to be between the ages of 18 and 30 to participate. Previous experience with this type of research showed that older adults tend to have difficulty discriminating auditory stimuli as they age, thus the age cut-off at 30. Both males and females were allowed to participate. A total of 8 males and 37 females participated, with an average age of 20.31 years ($SD = 2.79$ years). Participants were screened to ensure normal hearing and normal or corrected-to-normal vision, as well as normal color vision. Visual acuity was measured using the Freiburg Visual Acuity and Contrast Test (FrACT) and auditory functioning was tested using a MAICO MA40

audiometer, which ensured that participants could hear 30 dB (A) hearing level (HL) (or 50 dB (A) sound pressure level (SPL)) with 500-, 2000-, and 4000-Hz tones in each ear. This is a standard measurement for normal hearing. Normal color vision was measured using the Ishihara Color Vision Test. Participants were also asked to report ailments such as head colds, eye infections, or ear infections. These can impair visual and auditory functioning. Participants reporting these ailments were excluded from the study until they had recovered.

Apparatus & Stimuli

Participants viewed all trials on a 40 cm-diagonal CRT computer monitor at 59.00 cd/m^2 in a quiet, darkened room in the Sensory Neuroscience and Perception Laboratory at Lakehead University. They were seated comfortably on a height-adjustable chair and positioned on a chin rest so there was 75 centimeters between their entrance pupils to the computer screen. Visual stimuli were created in PowerPoint^{TM} and auditory stimuli were created in Audacity^{TM}. All stimuli was imported into, and presented with, Superlab version 5.0^{TM} software.

Task 1: Shape discrimination

Task 1 required that participants discriminate between black-and-white outlined circles and squares. Instruction sets directed participants to discriminate the location of a square from amongst three distractor circles or the location of a circle from amongst three distractor squares as shown in Figure 1. The target circle or square was presented in one of four spatial quadrants with three non-target squares or circles presented in the remaining quadrants, all 5° diagonally equidistant from a central "red" crosshair. All visual stimuli subtended 1° visual angle in height. The targets and non-targets were

chosen so as to have prominent shape discriminability. Similarity between targets and nontargets has been shown to make visual searches more difficult (Duncan & Humphreys, 1989, 1992), and this task was designed to be relatively easy as confirmed in my pilot study. Participants viewed the central crosshair for 300 msec, which was then replaced by a left or right, directing endogenous arrow cue or a bidirectional "null" arrow cue (serving as a noninformative neutral cue) for 200 msec. Following the rationale for the orthogonal cuing paradigm, the arrow pointed to the left or right side of the monitor while the participants had to discriminate elements in the upper or lower corner locations. Twenty percent of trials included a bidirectional "null" arrow cue, with the remaining trials having either a 60% or 80% valid left or right directing cue (e.g., Vossel, Thiel, & Fink, 2006). Arrow cues were equiluminant (approximately 8.1 cd/m^2) and either "blue" (CIE 1931, x = 0.2830, y = 0.5565) or "green" (CIE 1931, x = 0.1613, y = 0.0640) in color. The color of the arrow cue indicated its validity, for example 80% valid cues being "blue" and 60% valid cues being "green". Bidirectional "null" cues were presented in both colors. The validity attached to the specific color was counterbalanced among participants so that half the participants were presented with "blue" arrows being 80% valid and visa versa. The shape stimuli were presented 200 msec after the offset of the cue, and lasted for 300 msec to ensure that participants did not have time to overtly orient towards the stimuli. The four shapes were presented simultaneously, one in each quadrant. The target stimulus consisted of either a square or circle (participants were informed of the target stimulus in an instruction set) that was placed amongst either three circle or three square nontragets, respectively.

The participants were made to discriminate the location of the target stimulus using a 5-key response pad with four keys matched to the corner locations of the stimuli. The target stimulus could occupy any of the four quadrants and all permutations of the target/nontarget locations were presented. Half of the trials contained only visual stimuli (unimodal trials). The other half contained both visual and auditory information (bimodal trials). When the target stimulus was a square, the visual array was accompanied by a simultaneous 1000 Hz, 55 dB (A) SPL, 300 msec (high) tone presented to both ears via headphones. When the target stimulus was a circle, the visual array was accompanied by a simultaneous 500 Hz, 55 dB (A) SPL, 300 msec (low) tone. Research shows that statistical crossmodal congruence exists between angular shapes and high frequency tones, with less angular shapes being congruent with low frequency tones (Marks, 1987; Spence, 2011). Based on previous research, we anticipated that this stimulus combination would produce multisensory integration as observed through shorter reaction times than would be observed if either visual or auditory stimulus were presented alone, although the results of the present study yielded different findings as a result of differences in attentional load (see Results and Discussion).
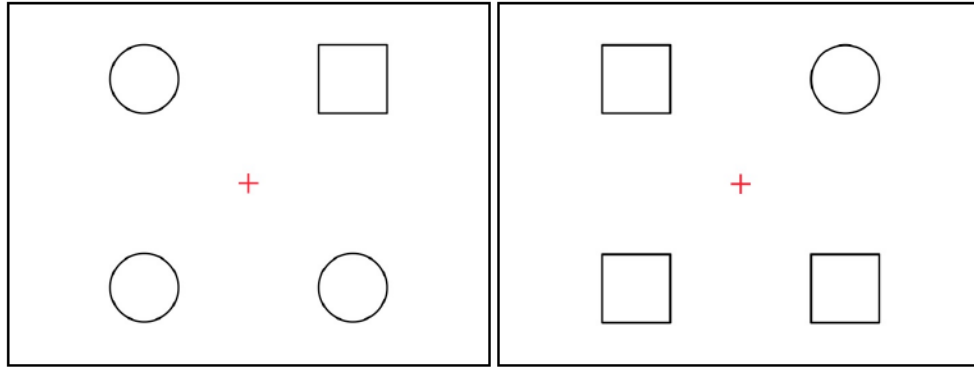
Figure 1. Example of stimuli used for the Shapes task. Left (right) panels show a square (circle) target amongst circle (square) nontargets. Stimuli are to scale. Coincident to these visually-presented stimuli were auditory tone presentations of 500 Hz (circle targets) and 1000 Hz (square targets).

Task 2: Die point discrimination

Task 2 followed the same temporal and spatial constraints as Task 1, but the shape stimuli were replaced by die-point stimuli (dots arranged in formations exactly as they are on dice). As shown in Figure 2, targets were either a 4-point or 5-point die face among three 5-point and 4-point die face non-targets, respectively. Trials consisted of unimodal visual stimuli and bimodal stimuli in which 4-point targets were accompanied by a simultaneous biaural low tone (i.e., 500 Hz) and 5-point targets were accompanied by a simultaneous biaural high tone (i.e., 1000 Hz) to create structural and semantic congruence. Task 2 was designed to be more difficult than Task 1 in that targets and non-targets were more similar to each other both structurally (the physical arrangement of the die points) and semantically (the numerical value of the die-points). Indeed, this increase in difficulty was revealed in the pilot study. Color-coded endogenous arrow cues directed attention to the left and/or right side of the monitor, or were bidirectional "null" cues as a control, as in Task 1.
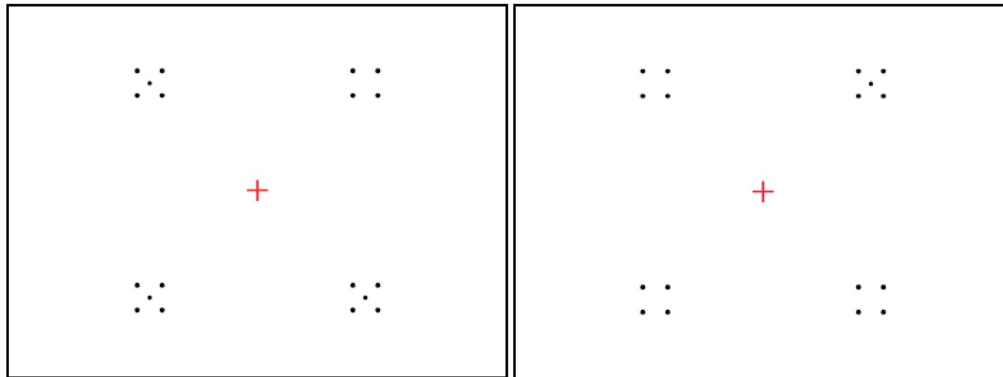
Figure 2. Example of stimuli used for the Die Points task. Stimuli are to scale. Left

(right) panels show a 4-point (5-point) target amongst 5-point (4-point) nontargets.

Coincident to these visually-presented stimuli were auditory tone presentations of 500 Hz

(4-point targets) and 1000 Hz (5-point targets).

Task 3: Discrimination of integrated features

Task 3 followed the same temporal and spatial constraints as mentioned in Tasks

1 and 2. Originally, this task was designed to be the most difficult of the three in that the

visual shape and die-point features were integrated, as shown in Figure 3. The targets

consisted of squares or circles with four or five-die points inside of them. Distractors

consisted of all other possible location permutations. Participants were informed, via

instruction sets, whether the target was a square with four die-points inside of it, a circle

with four die-points inside of it, a square with five die-points inside of it, or a circle with

five-die points inside of it. Trials consisted of unimodal visual stimulus presentations and

bimodal presentations. During the bimodal presentations unique tone combinations

accompanied the targets. The tone combinations consisted of two 150 msec tones

presented sequentially such that high, 1000 Hz and low, 500 Hz frequencies represent

target shapes and magnitude points in that order. Thus, a square target with four die-

points was accompanied by a high-low tone combination (i.e., a 1000 Hz tone followed

by a 500 Hz tone), target circles with four die-point were accompanied by a low-high

tone combination (i.e., a 500 Hz tone followed by a 1000 Hz tone), target squares with

five-die points were accompanied by a high-high tone combination (i.e., a 1000 Hz tone

followed by another 1000 Hz tone), and target circles with five die-points were

accompanied by a low-low tone combination (i.e., a 500 Hz tone followed by another 500

Hz tone). These tone combinations were chosen to maintain the statistical and structural

crossmodal congruence between the visual and auditory stimuli established in Tasks 1

and 2. As in Tasks 1 and 2, color-coded endogenous arrow cues directed attention to the

left and/or right of the visual array, or were bidirectional.
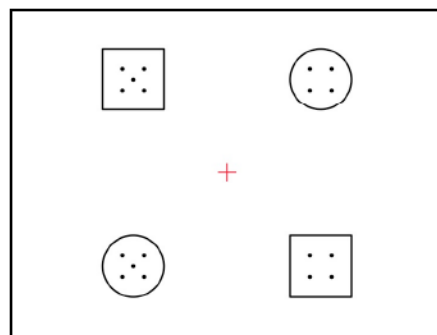
Figure 3. Example of stimuli used for the Integrated Features task. Stimuli are to scale.

Shape and die point stimuli were combined to create integrated targets. Targets were

indicated as either a circle with 4-points, a circle with 5-points, a square with 4-points, or

a square with 5-points. Nontargets consisted of all other permutations. Coincident to

these visually-presented stimuli were auditory presentations of two sequential tones.

Target circles with 4-points were accompanied by a 500-500 Hz tone sequence, target

circles with 5-points were accompanied by a 500-1000 Hz tone sequence, target squares

with 4-points were accompanied by a 1000-500 Hz tone sequence, and target squares

with 5-points were accompanied by a 1000-1000 Hz tone sequence.

Procedure

Before completing any of the tasks, participants signed consent forms detailing the requirements of the study and reviewing their rights to confidentiality and anonymity of participation and that they could withdraw from the study at any time without explanation or penalty. Participants then filled out a short questionnaire asking for their age, gender, and handedness. They then completed the visual and auditory screening tests mentioned above. All participants passed the screening tests and carried on with the three experimental tasks. Tasks 1, 2, and 3 were completed in successive order for each participant in order to maintain an increasing level of difficulty across the session. Task 3 also incorporated elements of Tasks 1 and 2 and was thus completed last. Task instructions were given verbally after which all participants were given a quick practice run to familiarize themselves with the stimuli and the response pad. Participants were allowed breaks between the tasks in order to prevent fatigue. The dependent variables for each task were participant response times (RT) to the target and accuracy (i.e., percent correct of total trials in which the target was properly discriminated from the nontarget distractors). After the tasks were completed, participants were debriefed and were encouraged to ask questions about the details and expected outcomes of the study. The entire laboratory session took approximately two hours for each participant.

Task 1: Shape discrimination

Participants first completed Task 1 trials in which a square was the target and circles were the distractors. Participants were first presented with an instruction set telling them to identify the location of the target square using a four-key response pad whose keys were spatially congruent to the four quadrants of the visual array. Task 1 consisted

of a total of 800 trials and took approximately thirty minutes to complete. Half of the

trials consisted of unimodal visual stimulus presentations and the other half consisted of

visual and aural bimodal presentations. Twenty percent of the total trials contained "null"

arrow cues. The other 80% of trials contained a left or right directing arrow cue, of which

half of the cues were 60% valid and the other half 80% valid (i.e., percentage of trials in

which the left or right visual endogenous arrow cue correctly predicted the side of the

screen that contained the target). The arrow cues were either "blue" or "green", with the

participants being informed of a 60- or 80-percent cue validity designation to each color.

Thus, half of the participants were told that the "blue" arrow cues indicated 80% validity

and "green" arrow cues indicated 60% validity; the other half were instructed that "blue"

arrow cues indicated 60% validity and "green" arrow cues indicated 80% validity.

Unimodal and bimodal stimuli, as well as different cues, were randomized within blocks.

Blocks in which the circle was a designated target and the squares nontargets followed

the same trial organization. Blocks containing squares and circles as targets were

counterbalanced.

Task 2: Die point discrimination

Task 2 followed a procedure identical to that in Task 2, except that the visual

shape stimuli were replaced by die-point stimuli with participants responding to both 5-

point arrangements and 4-point arrangements as targets within randomized, interleaved

blocks.

Task 3: Discrimination of integrated features

Task 3 followed a procedure identical to that in Tasks 1 and 2, except that the

visual stimuli consist of all possible permutations of integrated shape and die-points (i.e.,

4 targets) and the auditory stimuli consisted of two sequential tones instead of one. Task 3 required a longer time commitment (approximately forty-five minutes to an hour), as the total number of trials was doubled to accommodate the necessity of four targets instead of two.

Pilot Study: Multisensory integration

A pilot study was conducted prior to the experimental sessions in order to ensure that multisensory integration was occurring during bimodal audiovisual stimulus presentations. Behaviorally, multisensory integration usually causes responses intersensory facilitation in which responses to bimodal stimuli are significantly faster and more accurate than responses to unimodal stimuli. In order to test this, the pilot compared RT and accuracy for unimodal and bimodal stimuli from all three experimental tasks. The tasks followed the same spatial and temporal parameters as mentioned above except that the pilot study did not include the use of attention directing arrow cues. The tasks were presented sequentially (i.e., shape discrimination, followed by die-point discrimination, followed by integrated feature discrimination), with 15 repeats of each stimulus permutation for both unimodal and bimodal presentations.

A total of four participants, with a mean age of 22.5 years (SD = 1.7 years), completed the pilot study. Although the results of the pilot study did not show significant reaction time ($F(2,3) = 0.043, p = 0.849$) or accuracy ($F(2,3) = 0.080, p = 0.796$) differences between unimodal and bimodal stimulus presentations, both dependent variables showed trends indicating increasing difficulty across tasks. RT ($F(3,3) = 11.853, p = 0.078$) increased in length across the three tasks whereas accuracy ($F(3,3) = 7.706, p = 0.115$) decreased across the three tasks. The results of the pilot study suggested

that the three tasks did differ enough in their level of difficulty to measure the variables of interest across the processing hierarchy. The absence of any sort of trend in terms of intersensory facilitation may have been the product of a limited sample size, however the absence of facilitation did not exclude the possibility that some form of multisensory integration occurred. Later, I assessed further the difficulty levels of each task by conducting a post-hoc search experiment that is detailed later in the Results section (see Methods: Post-Hoc Experiment).

Results

The results of the experiment were analyzed across tasks using two ANOVAs, one for accuracy and one for RT. RT data was further separated into correct and erroneous responses. Individual ANOVAs were also conducted for each task for both dependent variables.

Overall Results for Percent Correct Accuracy

Percent correct accuracy was analyzed using a 3 (Task: Shape vs. Die-Point vs. Integrated Feature) x 2 (Stimulus Presentation: Unimodal vs. Bimodal) x 3 (Cue Validity: Null vs. 60% vs. 80%) within-subjects design. Post-hoc analyses were also conducted within each task.

The three factor, within-subjects ANOVA used to analyze the differences in percent accuracy across all three tasks yielded a significant effect of Task ($F(2, 44) = 49.286, p < 0.001$) and no other significant main effects or interactions for Stimulus Presentation or Cue. Results are displayed in Figure 4, with participants displayed the greatest accuracy in the Shapes task (M = 95.75%, SE = 0.49%), followed by the Die Points task (M = 88.65%, SE = 1.06%). Not surprisingly, the lowest percent correct was

found with the Integrated Features task (M = 84.26%, SE = 1.48%). Individual two factor

ANOVAs conducted separately within each task yielded no significant effects of

Stimulus Presentation or Cue within the Shapes or Die Point tasks. Figure 5 displays a

significant main effect of Cue within the Integrated Features task ($F(2, 44) = 3.625$, $p =$

0.035), with accuracy being significantly lower for the 60% valid cue (M = 83.91%, SE =

1.46%) than for the Null (M = 84.58%, SE = 1.50%) and 80% valid (M = 84.60%, SE =

1.42%) cues (See Table 1).

Table 1

ANOVA Table for Analysis of Accuracy

Variable	df	F	p
Overall Analysis			
Task	2	49.286	0.000*
Stimulus Presentation	1	0.284	0.597
Cue	2	2.565	0.089
Shape			
Stimulus Presentation	1	2.353	0.132
Cue	2	0.132	0.876
Die Point			
Facilitation	1	0.054	0.817
Cue	2	1.324	0.227
Integrated Features			
Stimulus Presentation	1	0.077	0.783
Cue	2	3.625	0.035*

* $p < 0.05$ (significant results)
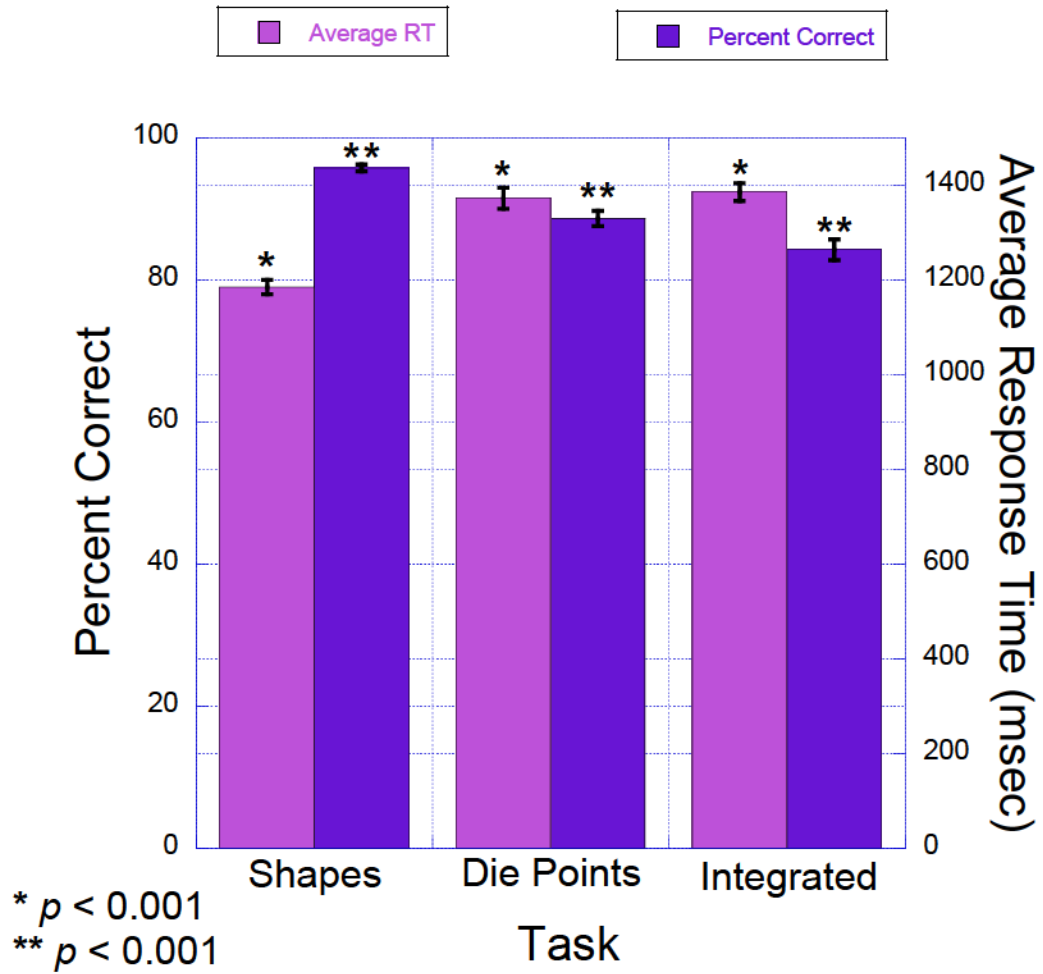
Figure 4. Percent correct (left axis) and average response time (right axis) as a function

of task. Dark purple bars reference the left axis and light purple the right axis.

Participants displayed significant decreases in accuracy and significant increases in RT

across the three tasks. Error bars denote ± 1 SEM.

Figure 5. Percent correct for cue validity as a function of task. There was no significant

effect of cue validity influencing accuracy for the Shapes or Die Points tasks. Cue

validity had a significant effect within the Integrated Features task, with Null (green) and

80% valid cues (beige) yielding greater accuracy than 60% valid cues (orange). Error

bars denote ± 1 SEM.

Overall Results for Response Time

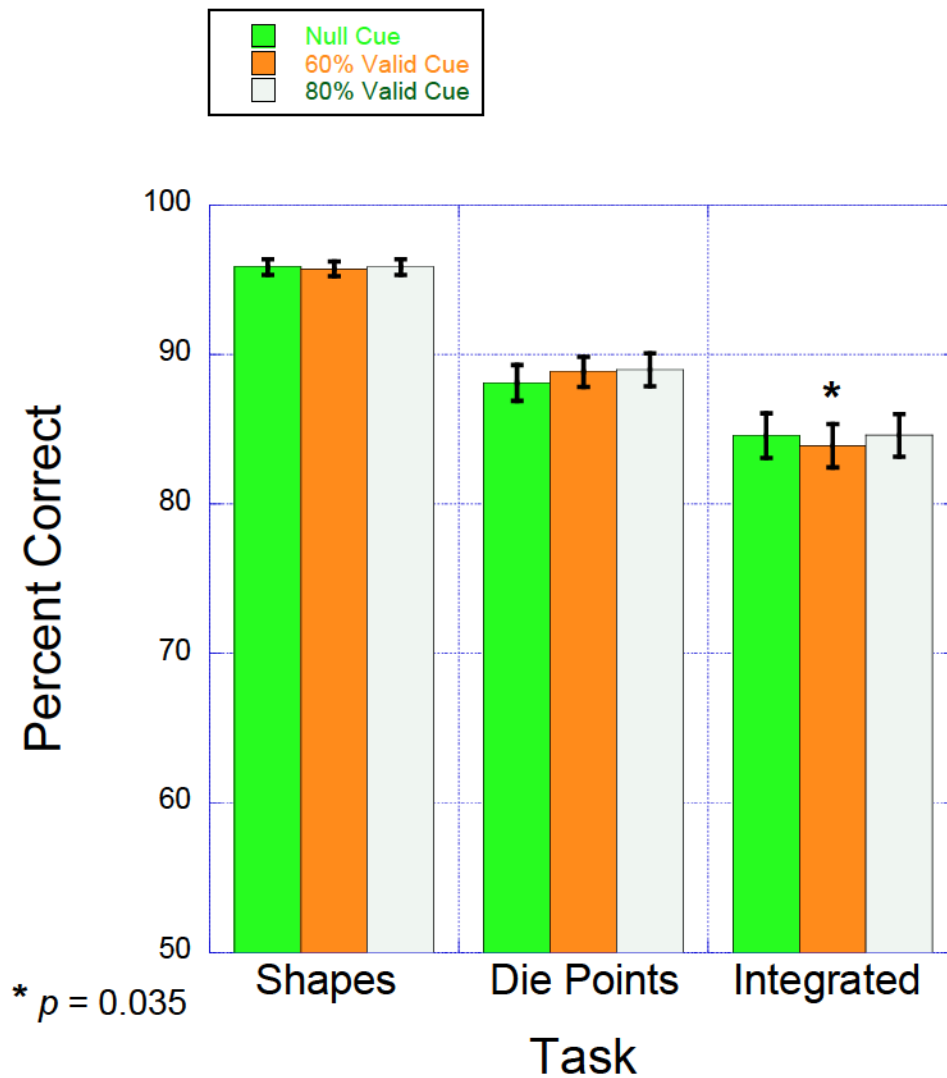Response times were analyzed using a 3 (Task: Shape vs. Die-Point vs. Integrated Feature) x 2 (Stimulus Presentation: Unimodal vs. Bimodal) x 3 (Cue Validity: Null vs. 20% vs. 80%) x 2 (Response: Correct vs. Error Respones) within-subjects design. A significant main effect of Task ($F(2, 44) = 44.971, p < 0.001$) and Response ($F(1, 44) = 11.807, p = 0.002$) was found. Participants displayed the shortest RTs in the Shapes task (M = 1184.72 msec, SE = 15.39 msec) and the longest RTs for the Integrated Features task (1385.21 msec, SE = 18.70 msec), with the RTs for the Die Points task falling inbetween the two (M = 1372.41 msec, SE = 22.37 msec), as shown in Figure 4. A separate set of analyses for RT correct and RT errors revealed significantly shorter correct responses (M = 1293.31 msec, SE = 10.98 msec) than erroneous responses (M = 1334.93 msec, SE = 20.12 msec).

Results for response time by task.

Individual three factor ANOVAs for combined correct and erroneous RTs were conducted separately within the Shapes, Die Points and Integrated Features tasks to look at main effects and interactions of Stimulus Presentation, Cue, and Response. This yielded some task-dependent significant main effects for Stimulus Presentation and Response. As shown in Figure 6, the Shapes task yielded no significant main or interaction effects for stimulus presentation. Within the Die Points task participants had slightly shorter RTs to bimodal stimuli (M = 1391.95 msec, SE = 19.69 msec) than unimodal stimuli (M = 1405.08 msec, SE = 24.65), however this Stimulus Presentation effect was not significant ($F(1, 44) = 0.675, p = 0.416$). Within the Integrated Features task, however, there were significant main effects of Stimulus Presentation ($F(1, 44) =$

15.203, $p < 0.001$) and Response ($F(1, 44) = 23.326$, $p < 0.001$). Participants displayed

significantly longer RTs in response to bimodal stimulus presentations (M = 1416.35

msec, SE = 16.23 msec) in comparison to unimodal stimulus presentations (M = 1394.43

msec, SE = 16.02 msec). Figure 7 shows a significant main effect of Response within the

Die Points task ($F(1, 44) = 29.884$, $p < 0.001$), with correct responses (M = 1333.48

msec, SE = 11.04 msec) being significantly shorter than error responses (M = 1463.48

msec, SE = 32.06 msec). Also, correct responses (M = 1375.83 msec, SE = 11.08 msec)

were significantly shorter than error responses (M = 1435.94 msec, SE = 21.48) (See

Table 2).
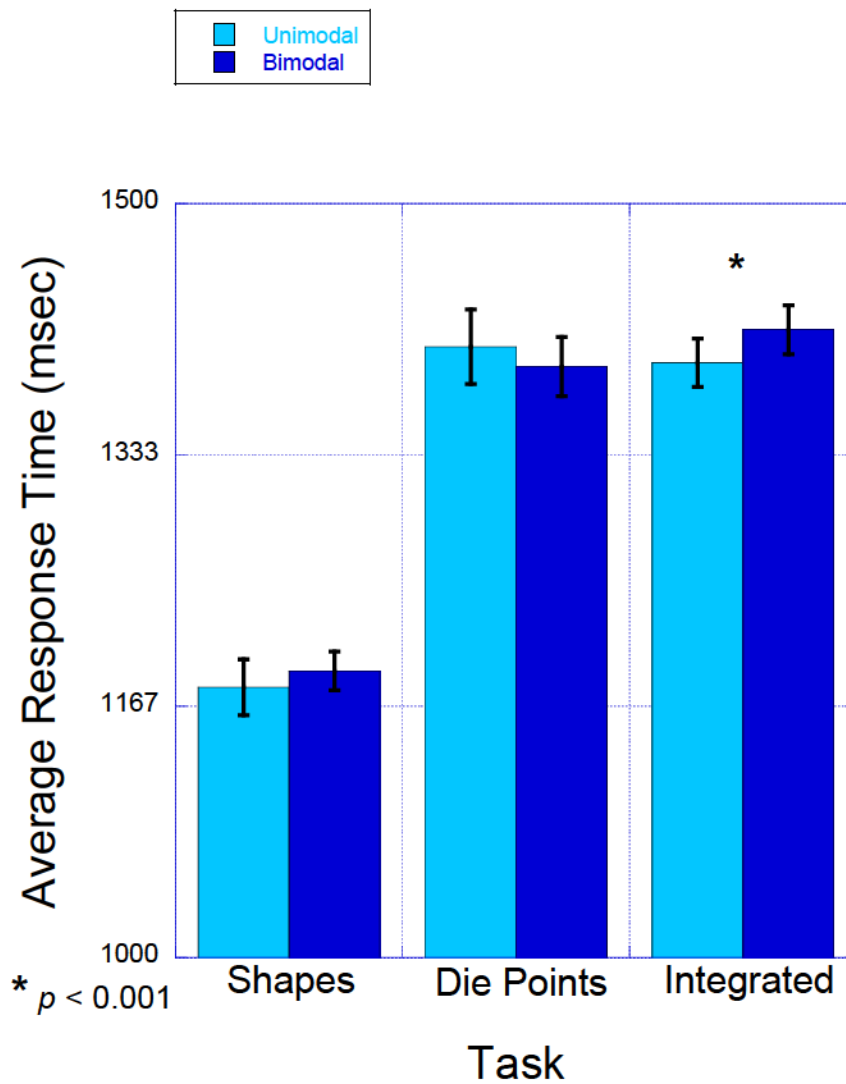
Figure 6. Average response time (msec) to unimodal (light blue) and bimodal (dark blue) stimuli (combined correct and error response times) as a function of task. Participants displayed no significant difference in RT to unimodal and bimodal stimuli in the Shapes task. RT show a trend towards multisensory facilitation (i.e., shorter RT to bimodal over unimodal stimuli) in the Die Points task. A significant reversal of intersensory facilitation was observed in the Integrated Features task. Error bars denote ± 1 SEM.
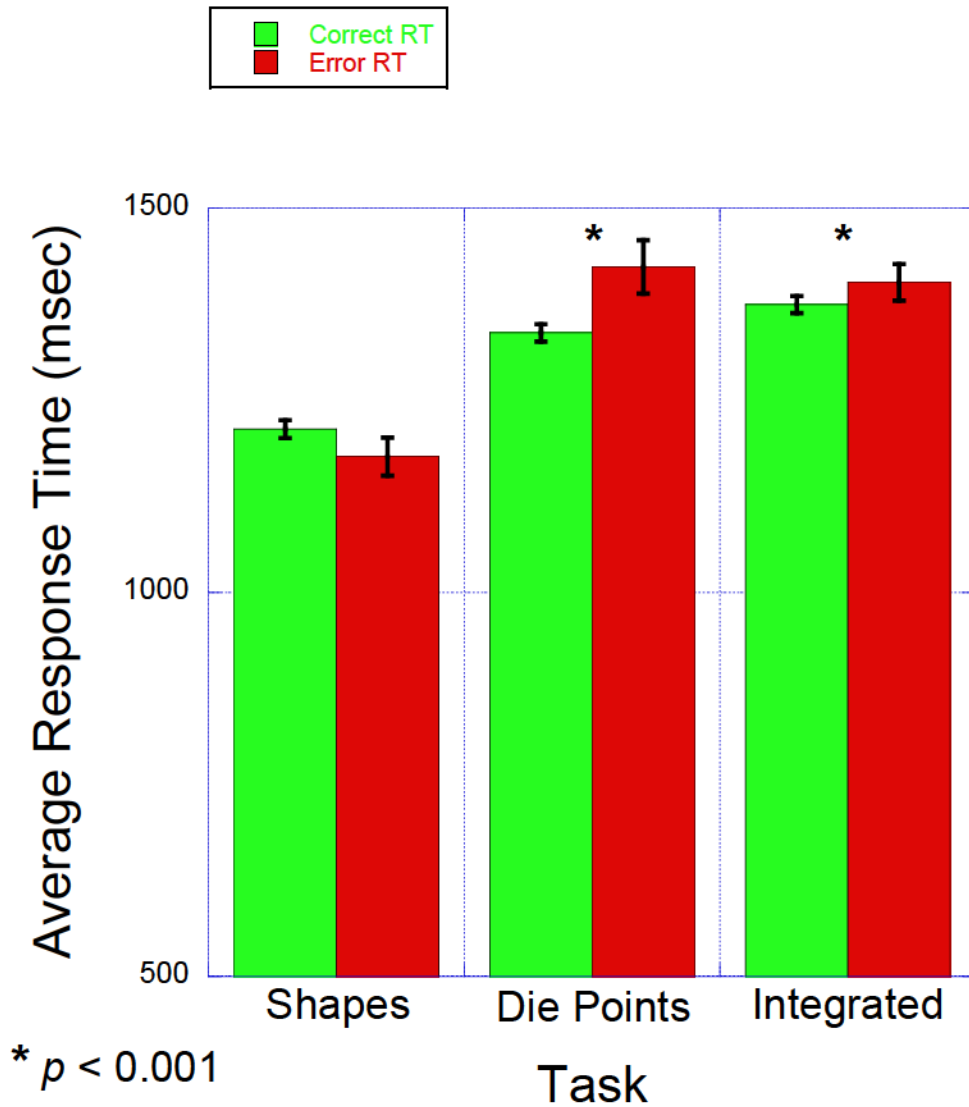
Figure 7. Average response time (msec) for correct (green) and error (red) responses as a function of task. Participants displayed no significant RT differences between correct and error responses in the Shapes task. Both the Die Points and Integrated Features tasks produced significantly longer RT for error responses. Error bars denote ± 1 SEM.

Table 2

ANOVA Table for Analysis of Response Time for Combined Correct and Error Responses

| Variable	df	F	p
Overall Analysis			
Task	2	44.971	0.000*
Stimulus Presentation	1	0.056	0.814
Cue	2	0.357	0.703
Accuracy	1	11.807	0.002*
Shape			
Stimulus Presentation	1	0.984	0.329
Cue	2	0.580	0.566
Accuracy	1	0.689	0.413
Die Point			
Stimulus Presentation	1	0.675	0.416
Cue	2	0.224	0.800
Accuracy	1	29.884	0.000*
Integrated Features			
Stimulus Presentation	1	15.203	0.000*
Cue	2	1.149	0.327
Accuracy	1	23.326	0.000*

* $p < 0.05$ (significant results)

Results separated by RT correct and RT errors.

Differences in response times were also analyzed separately for correct and error responses. In terms of correct RTs, a two factor ANOVA across all three tasks yielded significant main effects of Task ($F(2, 44) = 196.401$, $p < 0.001$) and Stimulus Presentation ($F(1, 44) = 4.793$, $p = 0.034$), as shown in Figure 8. The differences in correct RT between tasks mirrored those observed in the analysis that included both correct and error RTs. Analyses were also conducted within each task looking specifically at correct RT differences. There were no significant effects observed within the Shapes or Die Points tasks, however there was a significant main effect of Stimulus

Presentation in the Integrated Features task ($F(1, 44) = 7.223$, $p = 0.010$). Again, as in the

overall analysis of RT, participants had significantly longer correct RTs in response to

bimodal stimuli (M = 1378.71 msec, SE = 10.77 msec) in comparison to unimodal

stimuli (M = 1370.96 msec, SE = 11.57 msec) (See Table 3).

Table 3

ANOVA Table for Analysis of Response Time for Correct Responses

Variable	df	F	p
Overall Analysis			
Task	2	196.401	0.000*
Stimulus Presentation	1	4.793	0.034*
Cue	2	2.096	0.135
Shape			
Stimulus Presentation	1	1.878	0.178
Cue	2	0.953	0.394
Die Point			
Stimulus Presentation	1	0.848	0.362
Cue	2	1.077	0.350
Integrated Features			
Stimulus Presentation	1	7.223	0.010*
Cue	2	1.039	0.363

* $p < 0.05$ (significant results)

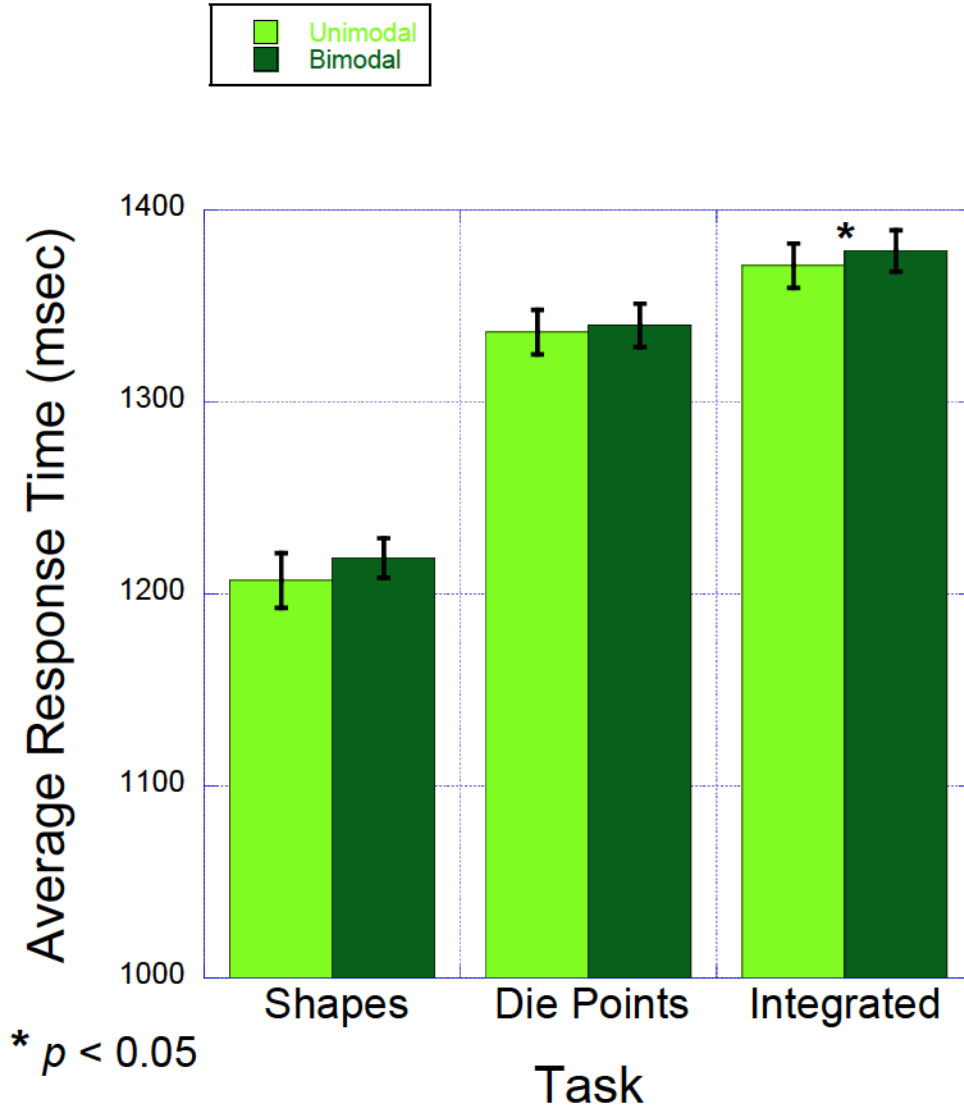*Figure 8.* Average response time (msec) for correct responses to unimodal (light green) and bimodal (dark green) presentations as a function of task. Participants displayed no significant difference in RT correct to unimodal and bimodal stimuli with the Shapes or Die Point tasks. They did, however, show significantly longer RT correct to bimodal over unimodal stimuli with the Integrated Features task. Error bars denote ± 1 SEM.

A three factor ANOVA was also conducted for error RTs across all three tasks, yielding a significant effect of Task ($F(2, 44) = 22.669$, $p < 0.001$). As displayed in Figure 9, error RTs in the Shapes task were significantly shorter (M = 1176.99 msec, SE = 24.77 msec) than those observed in the Die Point (M = 1423.83 msec, SE = 34.82 msec) or Integrated Features (M = 1403.95, SE = 23.79 msec) tasks. Individual analyses conducted within each task for error RTs yielded no significant results in the Shape or Die Point tasks, but did yield a significant main effect of Stimulus Presentation in the Integrated Features task ($F(1, 44) = 10.444$, $p = 0.002$). Errors made in response to bimodal stimuli (M = 1453.98 msec, SE = 23.31 msec) took significantly longer than those made in response to unimodal stimuli (M = 1417.90 msec, SE = 21.02 msec) (Table 4).

Table 4

ANOVA Table for Analysis of Response Time for Error Responses

Variable	df	F	p
Overall Analysis			
Task	2	22.669	0.000*
Stimulus Presentation	1	0.065	0.800
Cue	2	0.721	0.494
Shape			
Stimulus Presentation	1	0.144	0.707
Cue	2	0.479	0.624
Die Point			
Stimulus Presentation	1	1.009	0.321
Cue	2	0.423	0.658
Integrated Features			
Stimulus Presentation	1	10.444	0.002*
Cue	2	0.906	0.412

* $p < 0.05$ (significant results)
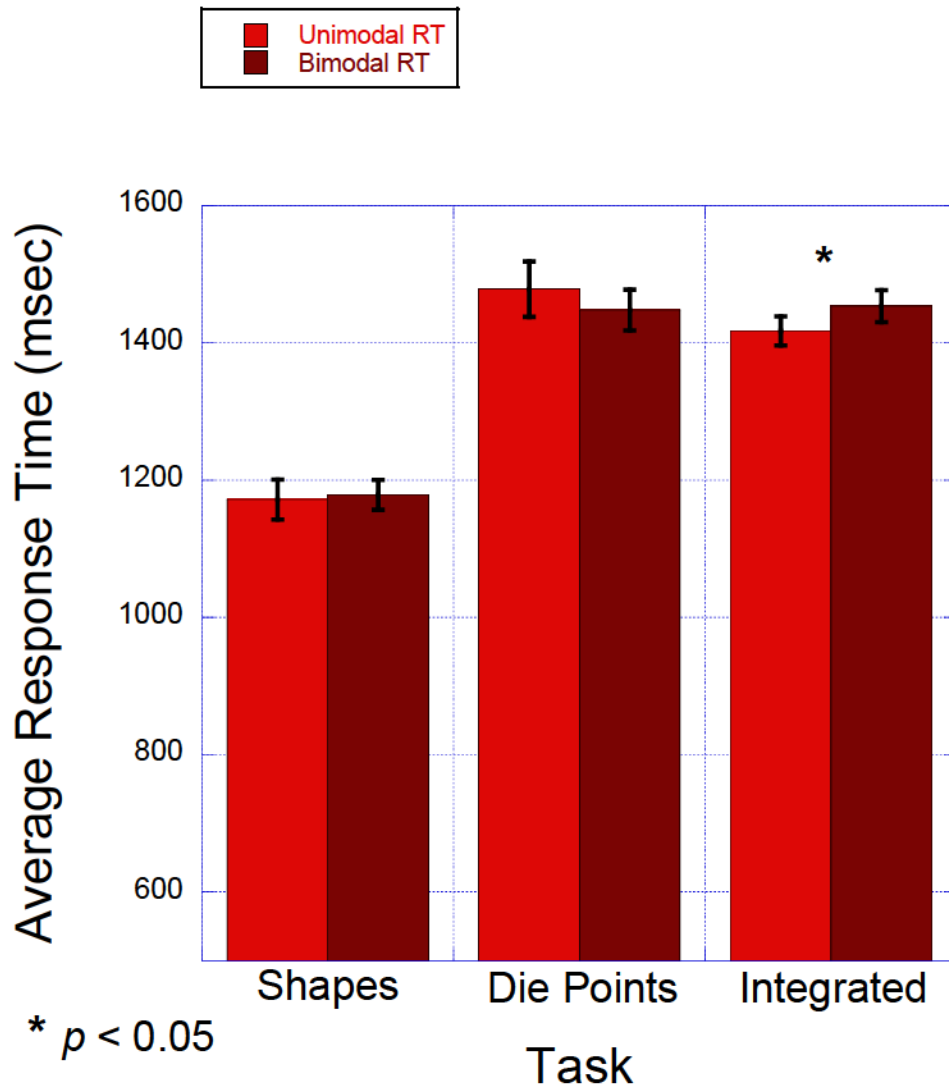
Figure 9. Average response times (msec) for error responses as a function of task for

unimodal (light red) and bimodal (dark red) presentations. Participants displayed no

significant difference between unimodal and bimodal RT for the Shapes task. The Die

Points task displayed slightly shorter error responses for bimodal stimuli. RT was

significantly shorter for unimodal over bimodal stimuli in the Integrated Features task.

Error bars denote ± 1 SEM.

Although the results of the experiment suggest that there were differences in attentional load across the three tasks, with the Shapes task having the lowest attentional demand, the Integrated Features task having the highest demand, and the Die Points task falling somewhere inbetween them, the differences in RT and percent correct were not definitive measures of attention. In order to place the attentional load of each task on firmer quantitative ground, a post-hoc experiment was conducted using a classic FIT paradigm. This experiment made use of the same stimuli as the three tasks mentioned previously (i.e., shapes, die points, and integrated shapes and die points) and varied the number of distractors present with an indicated target. I hypothesized that the shapes task would show no changes in reaction time with increasing numbers of distractors (i.e., be preattentive) but that the die-points and integrated features tasks would show increases in RT in concordance with additional distractors (i.e., be attention-demanding). To confirm my initial expectations, I anticipated that the integrated features search task would display a steeper increase in reaction time across set sizes, and thus be more attentionaly demanding, than the die-points task.

Method: Post-hoc experiment

Participants

Nine participants were recruited from Lakehead University and the community. A total of 2 males and 7 females participated, with an average age of 25.2 years ($SD = 1.92$ years). Participants were screened to ensure normal hearing and normal or corrected-to-normal vision, as well as normal color vision in the same fashion as mentioned in the previous experiment.

Apparatus & Stimuli

The apparatus and stimuli used for the post-hoc experiment were identical to those used in the previous experiment save for a few minor changes. In this experiment, shape, die point, and integrated feature stimuli were arranged in a circle formation 5° from a central fixation point. Stimuli were presented as set sizes with three, five, or eight total elements in the display for all three stimulus groups. Therefore, a set size of 3 denotes those trials with a target accompanied by two distractors or those without a target with three distractor elements, and so on for set sizes of 5 and 8 elements.

Procedure

After obtaining consent and being screened for normal auditory and visual functioning, participants completed a brief practice session in order to familiarize themselves with the demands of the task. For the actual experiment, participants started with shape detection. In each trial, a 500 msec duration fixation cross was proceeded by the stimulus, which remained on the screen until a response was made. An instruction set told the participants to press an indicated key on a 4-key response pad if they detected the presence of a target circle amongst nontarget squares and to press a different key if there was no circle present (i.e., only squares were present). The three set sizes were randomized, with 50 presentations per set size with a target and 30 presentations per set size with no target. A 500 Hz, 55 dB SPL, 500 msec tone was presented to both ears via headphones if an error was made. This procedure was repeated for the die-points stimuli, with a 4-point die face as a target and for the integrated features stimuli, with circle containing 4-points as a target. Accuracy (%C) and RTs for correct responses made to

trials containing targets were analyzed using 3 (Task: Shapes vs. Dies-Points vs.

Integrated Features) x 3 (Set Size: 3 vs. 5 vs. 8) within-subjects ANOVA.

Results: Post-hoc experiment

The two-factor ANOVA used to analyze the differences in percent accuracy

across the Shapes, Die-Points, and Integrated features tasks yielded a significant effect of

Set Size ($F(2, 8) = 9.546$, $p < 0.05$), with no main effect of Task ($F(2, 8) = 1.979$, $p =$

0.208) or interaction effect between Task x Set Size ($F(2, 8) = 2.178$, $p = 0.208$) (see

Table 6). In terms of the main effect of Set Size, accuracy decreased significantly across

tasks, with the Shapes task having the greatest accuracy (M = 96.52 %, SD = 0.78%),

followed by the Die-Points task (M = 95.56%, SD = 1.51%), with the Integrated Features

task showing the least amount of response accuracy (M = 92.00 %, SD = 1.87%).

The two-factor ANOVA used to analyze the differences in RT for correct

responses yielded significant main effects of Task ($F(2, 8) = 201.494$, $p < 0.001$) and Set

Size ($F(2, 8) = 76.503$, $p < 0.001$), and a significant interaction effect of Task x Set Size

($F(2, 8) = 27.156$, $p < 0.05$) (see Table 7). In terms of Task differences, RTs increased

significantly with task difficulty, with the Shapes task eliciting the shortest response

times (M = 476.72 msec, SD = 16.29 msec), the Integrated Features task providing the

longest RTs (M = 859.25 msec, SD = 28.90 msec), and the Die-Points task response

times falling between them (M = 805.25 msec, SD = 37.19 msec). Set Size differences

showed lengthened RTs with increases in set size, with the smallest set size (3) eliciting

the fastest responses (M = 646.23 msec, SD = 22.70 msec) and the largest set size (8)

eliciting the slowest responses (M = 787.99 msec, SD = 26.77 msec). The five-element

set size elicited RTs falling between the two extremes (M = 706.99 msec, SD = 23.89 msec).

The significant interaction effect between Task and Set Size was explored further by analyzing the effect of Set Size within each task. The Shapes task displayed no significant effect of Set Size ($p = 0.180$), however it did produce a significant effect with both the Die-Points ($p < 0.001$) and Integrated Features tasks ($p < 0.001$). For both these tasks, RTs increased within increasing set size (i.e., distractors) (see Table 8). Figure 10 plots the RTs for each Task as a function of Set Size, which allows for the linear fit calculation of the change in RT across set sizes (i.e., fitted slope). Measuring the slopes from the fitted function yields a measure of attentional demand, as put forth in Treisman's Feature Integration Theory (Treisman & Gelade, 1980). The slopes for each function produced the following formulas for Shapes:

$$y = 462.4 + \mathbf{2.6841}x$$

Die Points:

$$y = 565.54 + \mathbf{44.946}x$$

and Integrated Features:

$$y = 661.35 + \mathbf{37.105}x$$

Looking at the slopes for each task function (in bold type), one can see that the Shapes task was preattentive in nature (i.e., the slope is close to zero). The Die Points and Integrated Features tasks, however, displayed similar, steeper slopes, indicating that they were attention-demanding.

Table 6

ANOVA Table for Analysis of Accuracy

Variable	df	F	p
Task	2	1.979	0.208
Set Size	2	9.546	0.010*
Task x Set Size	2	2.178	0.208

*$p < 0.05$ (significant results)

Table 7

ANOVA Table for Analysis of Response Time

Variable	df	F	p
Task	2	201.494	0.000*
Set Size	2	76.503	0.000*
Task x Set Size	2	27.156	0.001*

*$p < 0.05$ (significant results)

Table 8

ANOVA Table for Analysis of Response Time (Task x Set Size Interaction)

Variable	Mean (SD)	F	p
Shapes		2.210	0.180
Set Size 3	465.510 (17.914)		
Set Size 5	484.060 (16.129)		
Set Size 8	480.578 (18.616)		
Die-Points		41.239	0.000*
Set Size 3	700.577 (27.739)		
Set Size 5	789.933 (39.474)		
Set Size 8	925.239 (47.391)		
Integrated Features		66.520	0.000*
Set Size 3	772.603 (30.891)		
Set Size 5	846.994 (31.724)		
Set Size 8	958.152 (27.748)		

*$p < 0.001$ (significant result)
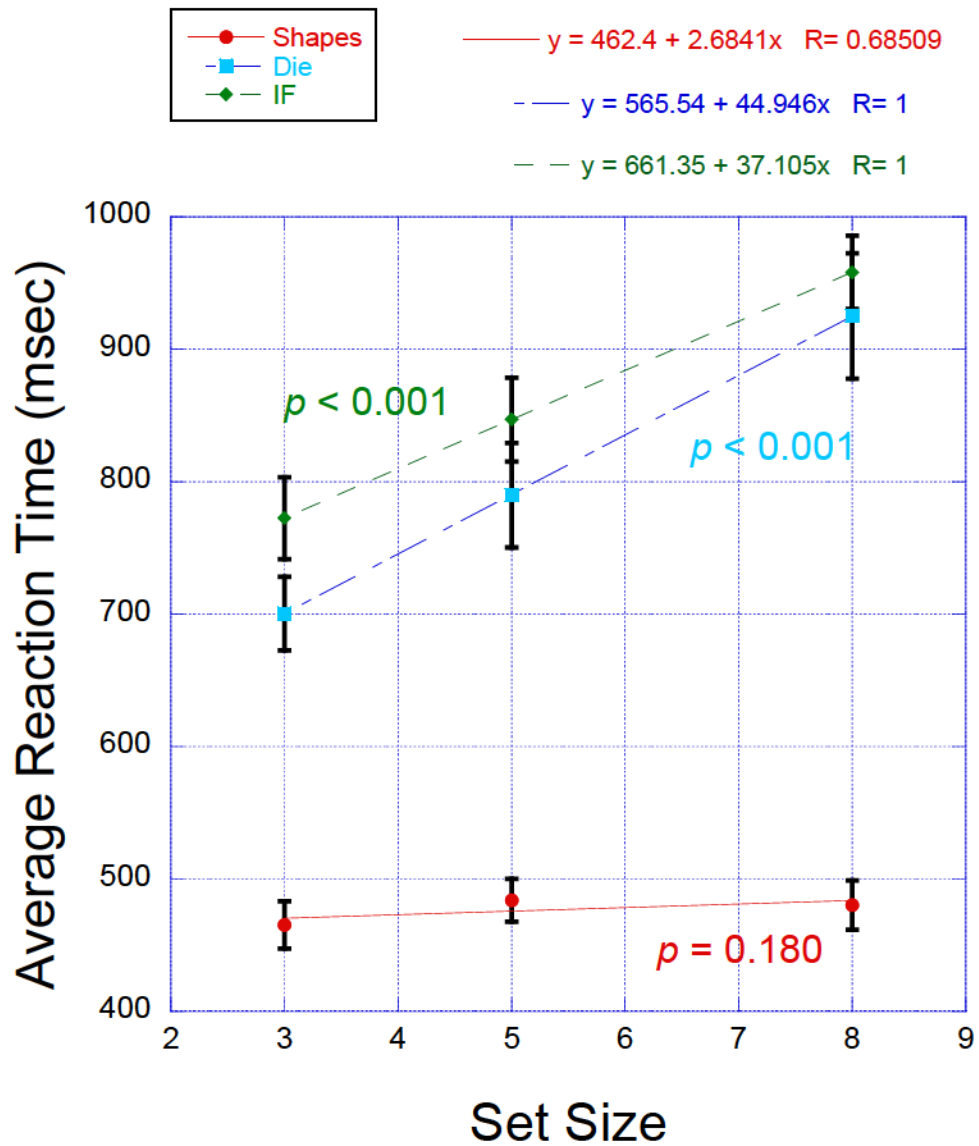
Figure 10. Average response time (msec) as a function of set size for Shapes (in red), Die

Points (in blue) and Integrated Features (in green). Participants displayed no significant

changes in RT as a function of set size for the Shapes task, but displayed significant

increases in RT across set sizes for the Die Points and Integrated Features tasks. Note: the

similar fitted slopes for the Die Point and Integrated Features tasks indicate comparable

seriality of search or attentional load. Error bars denote ± 1 SEM.

Discussion

The field of cognitive psychology has recently begun to ask questions concerning the interactions between two thoroughly investigated processes: multisensory integration and attention. Recent review papers have begun to piece together existing evidence on whether or not integration requires attention in order to occur (Koelewijn et al., 2010; Talsma et al., 2010). After reviewing dozens of independent studies on multisensory integration, the current consensus as to the involvement of attention in this phenomenon has been boiled down to three potential models. The early integration framework model suggests that there is the potential for integration to occur preattentively, without any top-down attentional involvement. The late integration model suggests the opposite: that multisensory integration cannot occur without attention. The third and final model, the parallel integration network model (Calvert & Thesen, 2004) suggests that integration occurs at multiple stages of processing all of which can be dynamically modulated by attention. Individual studies provide evidence both for and against all of these models, however to our knowledge, the present study is the only investigation of these models across the processing hierarchy using a covert, endogenous cuing paradigm. This study is unique in that it attempts to tackle the question of multisensory and attentional interactions using tasks that involve multiple hierarchical levels of cognitive processing. It is also unique in that it makes use of a discrimination paradigm and not the more common detection paradigms used in the majority of multisensory research.

The three tasks, Shapes, Die Points, and Integrated Features, were designed in an attempt to have an even sampling of multisensory integration across the processing hierarchy, from preattentive processing in an easy shape discrimination task to the high-

level processing attentional demands of a feature integration task. The results of the main

experiment yielded significant RT and accuracy differences between the tasks suggesting

an increase in attentional demand across tasks, with the Shapes task being the "easiest"

with the greatest accuracy and shortest RTs and the Integrated Features being the

"hardest" with the least accuracy and slowest RTs (see Figure 4). The Die Points task

fell nicely between these two extremes. Although the RT and accuracy differences were

suggestive of attentional load differences, they were not considered conclusive evidence.

A post-hoc study was conducted to calculate a more exact measurement of attentional

demand between the tasks using a classic FIT paradigm. The results partially supported

the results of the main experiment. The Shapes task displayed no significant change in

response times to targets across increasing set sizes, with a linear fit function displaying a

slope close to zero ($y = 462.4 + \mathbf{2.6841}x$). According to FIT, this indicates that the

Shapes task was preattentive in nature, with the indicated target displaying a "pop-out"

effect regardless of the number of distractors present.

The Die Points ($y = 565.54 + \mathbf{44.946}x$) and Integrated Features ($y = 661.35 +$

$\mathbf{37.105}x$) tasks displayed significant response time changes across set size identical to

patterns found in classic conjunction search tasks, with their linear fit functions

displaying steeper slopes. It was interesting to note that the Die Points task actually

displayed a steeper slope than the Integrated Features tasks. This difference was not

significant, however, suggesting that the Die Points and Integrated Features tasks were

both similarly attention-demanding (see Figure 10). Considering the RT and accuracy

results from the main experiment in combination with the results of the post-hoc

experiment, we conservatively concluded that the Shapes task tapped into potential

integration occurring preattentively while the Die Points and Integrated Features tasks allowed us investigate multisensory integration under high attentional load.

There was no significant effect of the covert, endogenous attention directing cues on RT. The only significant influence of the cues was during the Integrated Features task, with participants displaying greater accuracy when presented with 80% validity cues, and surprisingly with the null validity cues (double arrowheads) in comparison to the near-chance 60% validity cues (see Figure 6). Accuracy within the Shapes and Die Points tasks were unaffected by cue. The absence of any RT or accuracy cue effects within the Shapes task is logical due to the preattentive nature of the task, as one cannot direct attention within a task where there is no attentional demand in the first place. The high attentional demand of the Integrated Features task may have forced participants to rely more on the cues to maintain accuracy. However, there was a lack of any cue effect within the Die Points task, which as we know from the post-hoc experiment had a similar attentional load as the Integrated Features task. It may be that participants relied on the cues for the Integrated Features task due to the complexity of having to discriminate more than one feature. The Die Points task, although having a similar level of attentional load to the Integrated Features task, still only involved the discrimination of a single feature and thus may not have required dependence on the cues. The results observed within each task with regards to cueing effects combined with the attentional load quantifications from the post-hoc experiment allow for two perspectives on how multisensory stimulus processing is influenced by attention.

The effect of intersensory facilitation (i.e., the shortening of RT with multimodal versus unimodal presentations) was analyzed within each task. An analysis of the

difference in RT between bimodal and unimodal stimuli was conducted first for combined correct and error responses (see Figure 6). The Shapes task yielded no significant unimodal versus bimodal RT differences. On the opposite end of the spectrum, the Integrated Features task showed a reversal of intersensory facilitation, with RT to bimodal presentations actually being slower than those observed for unimodal presentations. This finding of facilitation reversal for an attention-demanding discrimination task is consistent with previous findings in our lab (Byce & Wesner, 2013). The Die Points task also did not yield a significant effect of facilitation; however a trend was observed with responses to bimodal stimuli being slightly shorter than to unimodal stimuli. This result is interesting given that the Die Points task had a similar attentional load to the Integrated Features task. The results for the Shapes and Integrated Features tasks were replicated when looking only at RT for correct responses, however the Die Points task no longer showed a trend towards facilitation with this RT correct analysis (see Figure 8). An analysis of accuracy yielded no significant facilitation effect for any of the tasks.

I thought it important to include an analysis of error RT, as this is often overlooked in other studies and looking at error responses can also provide information as to what level participants are processing the incoming stimuli. The analysis yielded no significant effect of facilitation for the Shapes or Die Points tasks, however RT errors to bimodal presentations in the Integrated Features task were significantly longer than to unimodal presentations (see Figure 9). Error RTs were also significantly longer than correct RTs, overall, for both the Die Points and Integrated Features tasks, with no significant difference observed with the Shapes task (see Figure 7). The longer RT error

rates found with the Die Points and Integrated Features tasks indirectly indicates that participants were behaving more thoughtfully prior to making their key-pad responses. On the other hand, the slightly shorter RT error rates with the Shapes tasks could possibly indicate less intentional processing of the stimuli or a general lack of attentional arousal.

Taken together, all of these results help to further differentiate the cognitive demands between the tasks. The Shapes task displayed no significant difference in RT or accuracy across all analyses, which provides almost overwhelming evidence that the Shapes task was preattentive, with the addition of cues and multimodal information having no facilitatory effect whatsoever. This is to be expected given that it is difficult to make an already automatically processed task any easier. The analysis for the Integrated Features task also points to its difficulty and to the possibility that the addition of extra information (i.e., redundant auditory stimuli) to this task actually appears to overwhelm participants' attentional capacity.

By far the most interesting finding was the contradicting data gathered from the Die Points task. This task appears to be similar in nature to the Integrated Features task, displaying the same characteristics in terms of attentional load based on the post-hoc experiment and differences between correct and error response times. The Die Points task, however, unlike the Integrated Features task, did not yield significant reversals in intersensory facilitation. Initially, the Die Points task was designed to have an attentional demand intermediate to that of the Shapes and Integrated Features tasks, and although the post-hoc visual search experiment provide evidence that it is more similar in attention load to the Integrated Features task, my multisensory integration findings suggest that

these tasks are not identical in terms of levels of cognitive processing or that they may not have been identical in terms of their top-down modulatory susceptibilities.

These results can be interpreted in the context of multiple converging lines of evidence. First, these results can be interpreted within the frameworks of Treisman's FIT (Treisman & Gelade, 1980) and Lavie's (1995, 2005) perceptual load hypothesis. Alternatively they can be interpreted in light of research investigating the inhibitory influence of distractors on attention (Duncan & Humphreys, 1989, 1992). One must also consider the logic behind models such as Maximum Likelihood Estimation when it comes to measuring behavioral facilitation (Ernst & Bulthoff, 2004; Ma & Pouget, 2008). A discussion is therefore warranted regarding the influence of methodology on multisensory integration findings.

The vast majority of studies reporting intersensory facilitatory effects make use of overt, exogenous detection tasks. The present study used a more complex, covert, endogenous discrimination task. There is currently still debate over the mechanisms underlying overt and covert orientation and exogenous and endogenous attention. These processes are not identical and there are definitive differences between the detection of a stimulus and making a discriminatory judgment about it. Differences in intersensory facilitation findings have also been reported for different tasks (Grice, Canham, & Gwynne, 1984; Grice & Gwynne, 1987; Grice & Reed, 1992) suggesting that tasks can be differentially responsive to the measurement of multisensory effects (Barrett & Krumbholz, 2012). Finally, the new information from this study must also be interpreted to answer the original question of this paper, which is to explicate the role of attention in multisensory integration and help contribute to a unified theoretical model.

The Use of Feature Integration Theory

FIT (Treisman & Gelade, 1980) suggests that visual perception occurs in two stages. The first stage is automatic and preattentive, and involves the processing of individual features in parallel. The second stage involves the integration of multiple features (i.e., a conjunction of features) and requires attentional involvement. The most solid evidence for the existence of these two stages is the difference observed between feature and conjunction searches. In a feature search, the participant must discriminate a target that differs in only one feature from distractors (i.e., a blue circle amongst red circles). In a feature search task, response times to the target remain constant no matter how many distractors are added. This is referred to as "pop-out" and is a major defining factor in preattentive tasks. In conjunction searches, the target shares one or more features with the distractors (i.e., a blue circle amongst blue squares and red circles) and search for the target occurs serially. Thus, response times to locate the target increase with increasing numbers of distractors, suggesting that the task requires attention. These classic trends were replicated in my post-hoc experiment, which supported the Shapes task as a preattentive, pop-out task, and Die Points and Integrated Features tasks as attention demanding.

This information can be used to interpret the influence of attentional load on multisensory integration. The Shapes task presents itself as a classical, preattentive, pop-out feature search. The lack of any influence of accessory auditory stimuli on the discrimination of the visual shapes target suggests that multisensory integration cannot occur in the absence of attention. This makes intuitive sense, as it is difficult to facilitate an already simple and automatically processed task. In other words, if a task is already at

the floor of what is considered "easy" it cannot be made easier by the addition of extra information. The idea that multisensory integration cannot occur without at least some attentional presence is also supported by previous research (Fairhall & Macaluso, 2009; Talsma et al., 2007; Talsma & Woldorff, 2005).

Of course, evidence also exists to support the idea that multisensory integration occurs automatically. A series of experiments conducted by Van der Burg et al. (2008) are some of the most-cited behavioral studies supporting this claim. Van der Burg and colleagues conducted a series of feature detection tasks in which participants had to indicate the presence of a vertically or horizontally oriented bar amongst relatively large set sizes of distractors at other orientations. During the visual stimulus presentations, the target and/or the distractors underwent color changes. These color changes were occasionally accompanied by an auditory tone or "pip". The most important result of the experiments was that the identification of targets was significantly facilitated when a target color change was accompanied by a simultaneous pip. It is important to note that the pip was non-informative, giving no information about the location of the target or about what the color change would be.

Overall, the visual search task used by Van der Burg and colleagues (2008) was attentionally demanding, with RT increasing with larger set sizes for visual-only stimulus presentations. This effect was eliminated, however, with the presence of the auditory pip stimuli, suggesting that the addition of the auditory stimulus induced a pop-out effect. This "pip and pop" effect was also observed even when the pip was rarely synchronized with the target color change (i.e., when the pip did not necessarily have great validity). Thus, the fact that the auditory pip provided no real information about the location of the

target and was still effective even when it was not highly valid suggested that attentional guidance by audiovisual events is largely automatic. This research is supported by other behavioral (Vroomen & de Gelder, 2000) and neurological (Giard & Peronnet, 1999) research.

Although this paper is cited as evidence for the automaticity of multisensory integration, the authors themselves admit that they could not completely eliminate the possibility that top-down influences played a part in the "pip and pop" effect. First, RT search slopes for targets accompanied by auditory stimuli never quite reached values typical for parallel search. Second, pips that were validly paired with targets yielded greater facilitation than when they were less valid. Thirdly and finally, the fact that the task itself was attentionally demanding and that the presence of audiovisual integration "guided" attention seems counterintuitive to claiming that the processes involved occurred in the absence of attention. Van der Burg and colleagues (2008) concluded that the integration of audiovisual events occurred rapidly and relatively early in the perceptual continuum, however they did not exclude the possibility that this initial integration was not acted on by attention in someway. Thus, the study did not completely exclude the possibility that multisensory integration requires at least some attentional involvement (see also Alsius, Navarra, Campbell, & Soto-Faraco, 2005; Fujisaki, Koene, Arnold, Johnston, & Nishida, 2005; Talsma et al., 2007).

The Integrated Features task lies opposite that of the Shapes task when it comes to attention demands. This task was designed as a conjunction search task and our post-hoc search experiment revealed that it was attentionally demanding. In this task, we observed a reversal of intersensory facilitation in that RTs to bimodal stimuli were actually longer

than RTs measured with unimodal presentations. This reversal of facilitation may be due to an exhaustion of attentional capacity, which may have caused the auditory accessory stimuli to be processed as a distractor, or not at all. Recent research has observed similar results with tasks that place high demands on attention (Alsius et al., 2005; Van der Burg, Awh, & Olivers, 2013).

Alsius and colleagues (2005) measured the effectiveness of the McGurk illusion (a classic example of multisensory integration) under conditions of divided attention. They measured the number of audiovisual fusions reported within the McGurk illusion by participants, who simultaneously engaged in either concurrent visual or auditory tasks. They found no effect of concurrent task engagement on participant's processing of the individual auditory and visual components of the McGurk illusion, but they did find that engagement in a concurrent visual task was detrimental to the number of fusions that occurred with the combined auditory and visual components. In other words, dividing attention prevented audiovisual integration. The authors concluded that exhausting attentional resources could compromise multisensory integration.

The capacity of multisensory integration also appears to be limited temporally. Van der Burg and colleagues (2013) provided evidence that audiovisual processing can only handle one auditory and one visual event at a time. Participants were instructed to remember the location of target disks among distractors. The number of targets was modulated and targets were occasionally accompanied by a non-informative auditory signal. The results showed that participants could only detect a single visual event with an accompanying auditory signal, and thus only one visual event could be linked to one auditory event at a time. This one-to-one ratio most likely serves an adaptive purpose to

help focus attention on one event rather than many. The authors concluded that intersensory capacity limitation serves to focus attention on only one audiovisual event at a time. This study serves to illustrate the contextual limitations of intersensory facilitation. If multisensory facilitation can be constrained by time and attentional capacity, it may be more accurate to classify facilitation as special case of intersensory integration rather than as the rule.

The most puzzling results of the present study are the behavioral outcomes of the Die Points task. The visual search post-hoc experiment revealed that the Die Points task had a similar attentional demand to the Integrated Features Task, yet this task did not yield the same reversal of intersensory facilitation. In fact the Die Points task was interesting because although it was defined as a classical feature search task (i.e., the target and distractors differ only in one feature), it acted like a conjunction search task (i.e., RTs lengthened with increasing set size) in that it appeared to require quite a lot of attention. Feature search tasks that show a set size influence do not exactly follow the original tenants of FIT, however they have been documented previously (see Quinlan, 2003).

The Die Points task was designed taking into account the attentional engagement theory (AET) put forward by Duncan and Humphreys (1989, 1992), which was a direct challenge to FIT. AET suggests that attention can be engaged, even in a feature search task, if targets and distractors share some degree of similarity. In the case of the Die Points task, the 4- and 5-point targets were similar structurally (i.e., shared similar orientations in space) and semantically (i.e., the numbers 4 and 5 are close to each other). Thus, the differences in facilitation observed between the Die Points and Integrated

Features tasks, even though they shared a similar level of attentonal demand, may have been due to differences inherent in the processing of single similar features versus the processing of two conjoined features. It would be interesting to remove the structural aspect of the die points and have participants discriminate random clusters of dots to determine if this produces less attentional demand than the interpretation of stimuli within a semantically-recognizable structural framework (i.e., die-face configurations). Further potential explanation for the differences between the Die Points and Integrated Features tasks are discussed below.

Use of the Perceptual Load Hypothesis

The perceptual load hypothesis (Lavie, 1995) was an attempt to resolve debate about whether or not attentional selection occurs early or late in the processing of targets amongst distractors. The theory proposes that there are two mechanisms of selective attention. The *perceptual selection mechanism* is relatively passive and allows for the exclusion of distractors from perception under conditions of high perceptual load. High perceptual load is defined by an increased set size in a stimulus array. With this mechanism, interference from distractors is prevented because attentional capacity is already exhausted. The *active selection mechanism* operates under conditions of low perceptual load, and allows for the processing and cognizant rejection of irrelevant distractors, as attentional capacity is not completely exhausted. This mechanism is dependent on higher cognitive functions such as working memory and goal processing (Lavie, 1995; Lavie, Hirst, Fockert, & Viding, 2004; Lavie & Tsal, 1994). The original experiments for the perceptual load hypothesis made use of the Eriksen paradigm (Ericksen & Eriksen, 1974). The Eriksen paradigm consists of a target letter surrounded

by distractors. Lavie (1995) had participants both detect and discriminate individual features and conjunctions under low and high perceptual load to derive the principles of her theory. Participants had to indicate the presence of a target letter flanked by increasing numbers of distractor letters. A critical distractor letter that was compatible, neutral, or incompatible with the target letter also accompanied this stimulus. The results showed that incompatible critical distractors induced an interference effect only under conditions of low attentional load (i.e., small flanking letter set size), where more attentional resources were available for distractor processing. The distractors produced no effect under conditions of high attentional load (i.e., large flanking letter set size).

Based on the perceptual load hypothesis, if there is a level of perceptual load that is sufficiently high to exhaust perceptual capacity, distractors will be excluded from perception. Essentially, there will be no attentional capacity left to process the distractors. Under conditions of low perceptual load, attentional capacity left over from the processing of the targets will be available to include the processing of irrelevant distractors (see Lavie et al., 2004 for a review). The perceptual load hypothesis can be used to help explain the results obtained for the Die Points task. Based upon the results of the post-hoc study, we can conclude that the Die Points task placed a sufficiently demanding level of perceptual load on the attentional system. Within the perceptual load hypothesis framework, any distractors should have been excluded from perception due to attentional capacity already being filled by the high demands of the task. If we consider the auditory stimuli as a distractor stimulus instead of a facilitating one, then it could account for the non-significant effect of facilitation observed within the Die Points.

Although it is possible that the Die Points task relied upon the perceptual selection mechanism put forward by the perceptual load hypothesis, with the auditory stimuli not being further processed due to a limited attentional capacity, the perceptual load hypothesis cannot explain why the Integrated Features task did not display similar results. This task appeared to allow for the processing of the auditory stimuli despite having a level of attentional load comparable to the Die Points task. We should also have observed facilitation in response to the Shapes task based on the rules of the load hypothesis, as extra attentional resources should have spilled over to process the auditory accessory stimuli. However given that the Shapes task was completely preattentive, perhaps the perceptual load hypothesis does not apply given that attention is not involved in the first place.

Influence of Attentional Inhibition

Another potential explanation for the reversal of intersensory facilitation in the Integrated Features task may have been attentional inhibition of the stimulus distractors and/or auditory accessory stimuli. The top-down (i.e., attentive) processing of visual features can occur as early as 100 msec following stimulus presentation, as shown in ERP studies (Zhang & Luck, 2009). At present there is debate over whether this early selectivity is due to the activation of target features or the inhibition of distractor features. The activation of target features is often conceptualized as the augmentation of a signal over noise and is based neurologically in the gain of neurons that are tuned to specific target features (Saenz, Buracas, & Boynton, 2002; Wolfe, 1994). Attentive inhibition of distractor features appears to be more dependent on recent experience and goals, with

distractors being actively deprioritized (Braithwaite & Humphreys, 2003; Lleras, Kawahara, Wan, & Ariga, 2008).

It may be that activation and inhibition interact or occur at different times during feature detection/discrimination, and some previous electrophysiological research suggests that feature-based distractor inhibition usually occurs at later stages of processing, approximately 200-300 msec post-stimulus onset (Andersen & Muller, 2010; Shin, Wan, Fabiani, Gratton, & Lleras, 2008). However, a recent evoked potential study conducted by Moher and colleagues (2014) observed attentional inhibition of distractors as early as 100 msec post-stimulus. This study looked at the response of the P1 in the visual cortex in response to colored dot probes. The P1 component represents early visual processing and occurs approximately 100 msec post-stimulus onset. The amplitude of the P1 waveform can also be influenced by neuronal activity reflecting top-down attentional influence. The study showed a reduced P1 response to distractor color probes in early visual processing, but no increase in the P1 response to target color probes. These results suggest that feature-based attention can at least partially modulate sensory processing by inhibiting signals from distractor features (Moher, Lakshmanan, Egeth, & Ewen, 2014). The neuronal mechanism behind this active inhibition may be the suppression of neuron responses to non-preferred features. This neural inhibition has been observed in monkeys (Khayat, Niebergall, & Martinez-Trujilo, 2010; Martinez-Trujilo & Treue, 2004) and humans (Andersen & Muller, Shin et al., 2008; Snyder & Foxe, 2010). It is unlikely that activation plays no role in feature-based attention, however it appears as though early visual processing occurs primarily through inhibition under conditions of strong target-distractor competition (Moher et al., 2014).

Based on these results and previous research, Moher and colleagues (2014) put forward two potential mechanisms for how feature-based attention operates in visual search. The first mechanism is a rapidly initiated activation of target features that can be adjusted in response to a frequently changing environment. The second mechanism modulates visual input relatively early via the inhibition of distractors and is more useful in static environments. Of course, inhibition can also occur at later stages as well (Andersen & Muller, 2010; Shin et al., 2008). This may be the reason why multiple studies have observed intersensory facilitation with detection paradigms and why we have failed to observe the same effect in discrimination tasks. Detection tasks do not place large demands on processing when it comes to interpreting distractors, whereas discrimination tasks must evoke some sort of mechanism to either activate targets above other, irrelevant noise or inhibit that irrelevant noise.

The reversal of facilitation observed in the Integrated Features task, then, may have been due to this attentional inhibition of distractor stimuli, regardless of whether it occurs early or late in processing. The high attentional load of the task may have required a great deal of distractor inhibition, to the extent that the auditory accessory stimuli may have been treated as distractors as well and suppressed. The visual distractors in the Integrated Features task may have also been more demanding on attentional inhibition due to their sharing features with the visual targets. This might also account for the differences observed between the Die Points and Integrated Features tasks, with the Die Points task displaying no effect of facilitation and the Integrated Features task showing a reversal of facilitation, despite their having a similar level of attentional load. It is possible that the distractors in the Integrated Features task were simply more effective

than those in the Die Points task and induced a greater level of distractor inhibition that spread to encompass the auditory stimuli as well.

The Issue of Methodological Influences in the Study of Multisensory Integration

The present study revealed how important task-defined processing is in measuring the influences of multimodal stimuli on behavior. If task level can have such a dramatic effect on behavioral outputs, then certainly a discussion must be had concerning the influences of methodology. Multisensory integration has been studied behaviorally using a huge number of different experimental paradigms. Paradigms include visual search, spatial orienting, language stimuli, working memory, and all tap into different hierarchical processing levels with corresponding differences in susceptibility for attentional modulation. The possibility that these different paradigms can result in task-specific results or that they are actually all addressing different processes altogether is rarely addressed. There have been some instances, however, where the idea of task sensitivity (i.e., the ability of the task to measure the process in question) in multisensory integration has been addressed. I will discuss two examples in an attempt to illustrate the importance of methodological influences: the issue of two-choice verses go/no-go responding brought up by Grice and colleagues in the 1980's and early 90's (Grice et al., 1984a; Grice & Canham, 1990; Grice, Canham, & Gwynne, 1984b; Grice & Gwynne, 1987; Grice & Reed, 1992) and, more recently, the importance of task sensitivity in detecting intersensory attentional facilitation (Barrett & Krumbholz, 2012).

In a series of experiments, Grice and colleagues (Grice et al., 1984a, 1984b; Grice & Canham, 1990; Grice & Gwynne, 1987) investigated the influence of the redundant targets effect (i.e., the usual speeding of RT to redundant targets, a form of sensory

integration) for visually presented letter stimuli under conditions of focused attention. The tasks made use of visual letter stimuli. The first study found no significant facilitatory effect of having redundant targets present (i.e., the same two letters presented on either side of a fixation cross) over and above what was observed for the presentation of only one letter (Grice et al., 1984b). Another set of experiments found that increasing the number of redundant letter targets facilitated responses in the presence of a greater number of distractors, but in the absence of a significant amount of "noise" from the distractors, no facilitation was found (Grice & Gwynne, 1987). The important thing to note is that these experiments made use of choice reaction time measurements, whereby the participants pressed one response key when a target was present and another when an irrelevant distractor was present.

The results of these choice reaction time (CRT) experiments (Grice et al., 1984b, Grice & Gwynne, 1987), which showed no redundancy effects, eventually came into conflict with the results of a similar paradigm that required go/no-go responses (Grice & Canham, 1990). Go/no-go (GNG) paradigms involve making a response to a target and withholding a response when there is no target present. Grice & Canham (1990) did observe redundancy gains in a similar letter task, suggesting that the distracting effect of the redundant non-targets in the CRT paradigm may have been due to response competition. A direct comparison of CRT verses GNG responses for the same task yielded a redundancy effect for the GNG task, but no facilitation for the CRT task, suggesting that GNG tasks may be more sensitive in capturing redundancy effects (Grice & Reed, 1992).

Generally speaking, GNG responses usually result is faster RT and greater accuracy than CRT responses (Gomez, Ratcliff, & Perea, 2007). The differences observed between results derived from CRT and GNG responses has of course raised the question about whether one paradigm is a better choice than the other for investigating cognitive processes. Two theories have been proposed to explain the discrepancy between the two paradigms. First, the two paradigms may cause participants to change how they make decisions about a task (i.e., lead to criterion shifts; Gordon, 1983; Hino & Lupker, 1998). The second possibility is that the two procedures actually change the core mechanisms involved in the task (i.e., attention, memory, redundancy effects, etc) (Grice & Reed, 1992; Perea, Rosa, & Gomez, 2002).

A study using theoretical diffusion models conducted by Gomez and colleagues (2007) showed that the GNG paradigm is in fact simply a type of two-CRT task whereby the go and no-go responses are both associated with an implicit choice. Diffusion models assume that information accumulates towards a decision over time from a starting point to one of two response criteria or boundaries. A response is given when the accumulation of information reaches one of the boundaries. This theoretical model shows that GNG paradigms are associated with two potential criteria, which is the same as what is postulated for two-CRT tasks (Gomez et al., 2007). This theoretical data is supported by neurological evidence. For example, lesions to the orbitofrontal area, an area associated with executive decision-making, are correlated with impaired performance in GNG tasks (Drewe, 1975). ERP studies also observe a negative N2 waveform over frontocentral scalp distributions during GNG tasks, with the N2 component being associated with executive control and conflict detection (e.g., Nieuwenhuis, Yeung, & Cohen, 2004).

These findings suggest that both the go and no-go aspects of GNG tasks require executive control. Similar results have been obtained in fMRI studies, which show activation consistent with inhibitory control (e.g., Menon, Adleman, White, Glover, & Reiss, 2001). These findings converge to suggest that GNG paradigms do involve two executively controlled responses to stimuli, just like 2-CRT tasks.

If the GNG and CRT paradigms are based on the same mechanisms, then why do studies consistently find faster and more accurate responses with GNG tasks? And why did Grice and Reed (1992) find different redundancy effects with the two tasks? Although the Grice studies looked at visual letter redundancy, similar parallels can be drawn to multisensory integration, which is simply redundancy across modalities. Based on this information, multisensory effects may have been masked in the present study by the use of a CRT paradigm. However, the majority of multisensory integration studies make use of detection CRT tasks and find significant facilitatory effects of integration. Perhaps CRT tasks are sensitive enough to detect facilitation in simple detection tasks, but are not sensitive enough to detect the same phenomenon in more complex, discriminatory tasks that contain distractors. These are questions that still require answers.

The issue of task sensitivity has also come up more recently for multisensory attentional capture (Barrett & Krumbholz, 2012). Multisensory cues do not generally capture, or facilitate, attention over and above that is observed with unimodal cues, except under conditions of high attentional load (see Santangelo et al., 2008). Where multisensory perceptual integration appears to reflect a true combination of signals from multiple modalities, multisensory attentional capture appears to be based on more of a

winner-takes-all competition between unimodal cue components for attentional control (Spence, 2010). The recent study by Barrett and Krumbholz (2012) suggests that the absence of behavioral facilitation observed with multisensory cues might simply be the result of a lack of task sensitivity inherent in response time measurements caused by post-perceptual factors. Barrett and Krumbholz (2012) used a temporal order judgment (TOJ) task to look at the influence of bimodal cues on attentional facilitation. TOJ tasks require that participants report their perceived order of occurrence for two asynchronously presented target stimuli and are thought to have increased perceptual sensitivity to cued locations. The results of the study showed that bimodal cues facilitate attention in a way that suggests the combination of intramodal and crossmodal cue components.

The studies mentioned above, as well as others not mentioned here, highlight the influence of methodology on behavioral results. They also highlight the need for researchers to be careful in how they draw conclusions from behavioral data. Data from behavioral studies is, of course, useful in explaining cognitive phenomenon, but must be interpreted in light of the tasks used. More research must be conducted into the sensitivity of behavioral tasks, not only those used to measure multisensory integration, but for other cognitive processes as well. Researchers, when interpreting their results, must give heed to the type of stimuli and responses they use and to the processing demands of their designed tasks. It is not a question of which type of task is "better" but rather a question of task conditions being adequately defined and understood with deference to the cognitive processing systems they activate. In this way, researchers will be able to interpret their findings in a context that advances all behavioral research in a meaningful, productive way.

Limitations of the Present Study and Future Directions

This was the first study, to my knowledge, to directly investigate the interactions between multisensory integration and attention across the cognitive processing hierarchy. We made use of complex, discriminatory behavioral tasks designed to tap into covert, endogenous attentional mechanisms. Although behavioral tasks allow for a direct measurement of how cognitive mechanisms translate into physical action, they do, like all forms of measurement, have their limitations, one of which is an inability to directly measure neuronal activity. Behavioral results also make it difficult to determine if a given manipulation impacted sensory, perceptual, or motor aspects of a response. Of course, every methodology presents itself with inherent and often unavoidable limitations, and great care was taken in the design of the present study to be aware of these limitations and to minimize them as much as possible. Here I will discuss the limitations of the present study and of behavioral research in general, including possible discrepancies between neuronal activity and behavioral outcomes, the issue of task sensitivity, and generalizability. I will also outline future directions for this area of study, such as the use of signal detection theory and maximum likelihood estimation to clarify behavioral results and the use of electroencephalogram (EEG) technology.

A potential limitation of the present study is that no actual behavioral facilitation was observed in response to the congruent audiovisual stimuli presented in any task. There was a trend towards facilitation observed within the Die Points task, however this trend was not significant and disappeared when only correct RTs were analyzed. Although the absence and reversal of facilitation in the Shapes and Integrated Features tasks, respectively, can be explained theoretically, it is also possible that the auditory

accessory stimuli used was simply not effective in inducing multisensory integration. It is possible that the participants did not perceive the auditory tones as congruent to the visual targets, even though it was explained to participants that specific tones accompanied specific targets. However it is also very probable that the absence of multisensory facilitation was due to the complex, discriminatory nature of the tasks.

It may be possible in the future to use more obviously congruent stimuli, such as visual and aural letter stimuli or visual stimuli varying in "brightness" accompanied by tones of high or low pitch. It might also be wise in the future to give lengthy practice sessions for the participants to fully engrain audiovisual congruence, although this may lead to over-learning of the task such that there are no error rates to investigate and also requires extensive, impractical time commitments. During the development of the tasks I did consider presenting tones spatially so that they were congruent with the targets' visual location, however I rejected this method because I wanted to tap into processes more advanced than overt, exogenous spatial orienting and did not want to risk the possibility that participants would simply ignore the visual stimuli in favor simply responding to the aural location.

Because no intersensory facilitation was observed, however, does not mean that the results of the present study cannot still be used to contribute to the question of multisensory integration and attentional interactions. The results of the present study help to narrow down an optimal attentional window for intersensory facilitation. The results of the Shapes task suggest that at least some form of attentional engagement is required for facilitation to occur. The results of the Integrated Features task also indicate that too much attentional demand can actually hinder and reverse behavioral facilitation. This

study also provides evidence that multisensory integration is not necessarily the rule of thumb for the processing of stimuli from more than one modality. The results suggest that the behavioral outcomes of multisensory integration may be more task-specific than once thought and may be confined to more simple situations, such as overt orienting towards a sudden exogenous stimulus.

The lack of cuing effects in the Shapes and Die Points tasks could also be construed as a limitation of the present study. However, the lack of cueing effects in the Shapes task conformed to the preattentive nature of the task and the significant effect of cueing in the Integrated Features task suggested reliance on the cues for complex conjunction searches. It could be argued that some form of cueing effect should have been observed in the Die Points task, however the Die Points task only required the discrimination of one target feature, like the Shapes task, and participants may not have had to rely on the cues due to the semantic nature of the stimuli. It is possible that the endogenous cues may not has been as effective as possible simply because they were largely ignored by the participants in the Shapes and Die Points tasks, however the significant cue effect observed in the Integrated Features task is evidence that cues were attended to. When it comes to the cues themselves, participants may have become confused with the predefined color coding of the cue's validity, however this seems unlikely as participants were repeatedly reminded of the color-coding by the researchers and in the instruction sets. The color-coding of cue validity has also been used successfully in previous research (Vossel et al., 2006). The combination of the cuing effects tasks and attentional load measurements from all three tasks, however, did allow

for a more comprehensive view of the attentional demands of each task and allowed for stronger conclusions.

Studies of multisensory integration have run into problems resolving discrepant findings between neuronal and behavioral responses to multisensory stimuli. For example, the three general "rules" of multisensory integration are the spatial rule, the temporal rule, and the rule of inverse effectiveness (see Introduction). These rules appear to solidly predict the responses of neurons to multisensory stimuli (Meredith & Stein, 1983; 1985). The spatial and temporal rules are also consistent in predicting behavioral responses (e.g., Bolognini, Frassinetti, Serino, & Ladavas, 2005; Holmes & Spence, 2005 although see Zampini, Torresan, Spence, & Murray, 2007), however the link between neuronal responses and behavioral responses for the law of inverse effectiveness is relatively weak (Holmes, 2007; Laurienti, Perrault, Stanford, Wallace, & Stein, 2005).

For example, Holmes (2007) took the results of a study claiming to have found evidence for the law of inverse effectiveness in behavior (Serino, Farne, Rinaldesi, Haggard, & Ladavas, 2007) and showed that some the findings were simply the result of the method of statistical analysis. Holmes (2007) also pointed out that the majority of studies displaying findings of spatial, temporal, inverse effectiveness rules often involve overt orienting and detection of simple sensory targets. Thus, multisensory findings can at times simply be the result of task requirements or statistical analysis and the rules of multisensory integration that exist for substrate do not always apply to behavior. In regards to the results of the present study, it is possible that multisensory integration was occurring at the neuronal level, but was not expressed in participant's behavior. Although this is included in the limitations of the present study, conflicting evidence from different

measurement techniques is quite common in behavioral neuroscience and is an enduring

limitation of the field in general.

The idea that multisensory integration can occur without the presence of an

obvious facilitation effect is accommodated by the maximum likelihood estimation model

of multisensory integration, which allows for multisensory integration to occur in the

absence of obvious response time facilitation to multisensory stimuli (Barrett &

Krumbholz, 2012; Ernst & Bulthoff, 2004; Ma & Pouget, 2008). The maximum

likelihood estimation (MLE) model attempts to find the most efficient way in which to

integrate different sources of information (i.e., different modalities). Individual sensory

signals are considered to be Gaussian-defined estimates, with variable amounts of noise.

The integrated estimate is the sum of these individual sensory estimates, weighted by

their respective variance. The goal is to come up with an estimate with the least amount

of variance (i.e., the most reliability in identifying the stimulus).

The bottom-up aspect of the MLE model is based on the tuning of populations of

neurons to specific sensory signals (i.e., orientation, color, pitch), but top-down

influences can include attention, working memory, and criterion shifts (Ernst & Bulthoff,

2004). In terms of the behavioral outcomes of multisensory integration, the MLE model

suggests that response times to multisensory stimuli do not necessarily have to be

significantly shorter than response times to unimodal stimuli. In fact, 'facilitated'

response times can approximate the average of those derived from each unimodal signal

(Barrett & Krumbholz, 2012; Ma & Pouget, 2008). Conducting an experiment similar to

the present study, with the addition of responses collected to the visual and auditory

components individually, would allow for an alternative measurement of multisensory integration based on the MLE model.

The present study could also be replicated and analyzed using signal detection theory (SDT; Tanner & Swets, 1954). Although signal detection paradigms require large numbers of trials and are thus logistically time limited, they do allow for quantitative measurements of participant's sensitivity (d') and response biases (β). The use of SDT would also allow for the measurement of any criterion shifts across the three tasks and would help to place the level of processing at which discrimination occurs (i.e., low-end sensory processing vs. later perceptual processing). Replication of the above presented study with analyses either conforming to the MLE model or SDT would allow for anther angle from which to interpret results. Of course, this study could also be replicated with the addition of ERP analysis. This would allow for coarse localization of cortical activity (with at least 120 channels), but more importantly this would allow for the investigation of ERP's high-resolution temporal components such as the P1 component, which reflects the first influence of top-down control on sensory processing areas.

The present study is the first, to my knowledge, that directly tackles the question of how multisensory integration and covert, focused attention interact across the processing hierarchy. Future directions include replication with the addition of different methodological paradigms (i.e., go/no-go responses, MLE, SDT, EEG, fMRI) in order to determine any task-specific effects and also to elucidate potential converting neuronal and behavioral evidence. There is also a great need for a systematic investigation into the differences observed between overt, exogenous attention paradigms and those that tap into covert, endogenous attention. The same can be said for differences between detection

and discrimination tasks or language vs. non-language stimuli. In the future it would be highly beneficial to conduct a large-scale behavioral study comparing all of these stimulus environments, ideally within-subjects, to fully understand both the consistencies of multisensory integration and the variation that can occur under different task conditions.

Conclusions

It is difficult to draw solid conclusions regarding the interactions of multisensory integration and attention given that no behavioral intersensory facilitation was actually observed in the present study. However, sometimes the lack of, or a reversal of, an effect can be just as informative as observing that effect, especially as facilitation is simply a unique case of multisensory integration. The fact that I consistently observed no effect of cue or bimodal stimulus presentations within the pre-attentive Shapes task provides solid evidence that multisensory integration does not occur in the absence of attention. This effectively rules out the early integration framework model of multisensory and attentional interaction. The results of the Die Points and Integrated Features tasks were much less clear, with high attentional load in the Die Point feature search task yielding no intersensory facilitation effect, with a similar level of load in the conjunction Integrated Features search task yielding a reversal of facilitation. These results suggest that there is an optimal amount of attentional demand required for intersensory facilitation. Even based on the fact that we did not observe facilitation in the presence of attention, I find it difficult to conclude that integration does not require at least some attentional engagement based on the results of the preattentive Shapes task and the general body of integration research. Thus, the results of this study at least rule out the early integration

model, and lend support to the late integration or parallel integration models of multisensory integration and attentional integration.

It is important to acknowledge that these conclusions are based on the results of a set of complex discrimination tasks that tapped into covert, endogenous attentional mechanisms. Our results probably would have been different if we had made use of a simple detection task, a task involving language stimuli, or exogenous cues directing overt attention. Previous literature exists that has found multisensory integration in each of these conditions, all with differing, and sometimes undefined levels of attentional load or manipulation. This is the first study, to my knowledge, to tackle the large-scale question of multisensory integration and attentional interaction across a large range of the processing levels within one group of participants. This study is also one of the few multisensory integration studies requiring participants to make complex discriminations of targets amongst distractors and not just detect stimuli over-and-above background noise. Thus, there is a caveat to our conclusion that multisensory integration requires attention in order to occur behaviorally, such that we can only apply this to complex, discriminatory situations.

The logical extension of this idea is that a great deal of thought must be put into drawing absolute, global conclusions concerning the nature of attentional involvement in multisensory integration, at least behaviorally. I took great care in the present study to design tasks that tapped into specific cognitive processes. This was also an attempt to try and resolve some of the discrepancies in definitions use within the multisensory integration literature as a whole. While observations of multisensory facilitation may be robust in tasks involving overt orientation using peripherally presented, semantically-void

stimuli, facilitation appears less obvious as researchers start to push the system to handle more complex situations. What may be true about multisensory integration in a detection task may not be true for a discrimination task. The same can be said for overt verses covert orienting, or for exogenous verses endogenous attention. Sometimes the literature regarding multisensory integration makes it seem as though behavioral facilitation to multisensory stimulus presentations is an absolute phenomenon; however the results of this study, and others discussed above, suggest that the behavioral outcomes of integration are much more variable.

The question regarding the interaction between multisensory integration and attention was only raised a handful of years ago, and although some conclusions can be drawn from research pre-dating that question, there is still much to be done in answering it directly. One of the largest hurdles in answering this question and validating one of the three models of attentional and integrative interactions will be clarifying general results from those that are task-specific or mechanism-dependent. Perhaps the three models are independently valid given the correct situation. More recent papers on multisensory integration reveal that the field is moving towards a more thoughtful awareness of the influence of task and task sensitivity. Thus, in line with this current trend, I tried to refrain from making any sweeping statements regarding the nature of attention and multisensory integration and instead formulated my conclusions with the knowledge that the results of this study were partially the product of research design and the associated mechanisms invoked by such a design. The interaction between the sensory energies in our environment and how our brains perceive and act upon these energies is still a topic

of intense investigation and to make an all-encompassing statement regarding that

phenomenon would trivialize the wonderful complexity that is the human brain.

Appendix A

Summary of terminology used in spatial attention research.

Terminology	Definition
1. Overt Orienting	Sensory receptors move towards a stimulus/spatial location
2. Covert Orienting	Internal shifts of attention with no movement of sensory receptors
3. Endogenous Mechanisms	Voluntary shifts of attention driven by expectations; top-down
4. Exogenous Mechanisms	Reflexive shifts of attention driven by salient peripheral stimuli; bottom-up
5. Endogenous Cues	Centrally presented, symbolic arrow or word cues that predict the likely location of a target
6. Exogenous Cues	Sudden, peripherally presented tones or flashes of light that are not predictive of a target's location

Appendix B

Behavioral results for crossmodal auditory and visual spatial attention.

Spatial Attention Orientation Combinations		Behavioral Findings
1. Endogenous Overt Orienting	Spence & Driver (1996)	Facilitation of performance in discrimination of visual and auditory targets with valid visual cues
2. Endogenous Covert Orienting	Spence & Driver (1996)	Facilitation of performance in discrimination of visual and auditory targets with valid visual cues
3. Exogenous Overt Orienting	Spence & Driver (1997)	Facilitation of performance in discrimination of visual and auditory targets with valid visual and auditory cues
4. Exogenous Covert Orienting	Spence & Driver (1997)	Facilitation of performance in discrimination of visual and auditory targets with valid auditory cues
		Facilitation of performance in discrimination of visual targets with valid visual cues
		No influence of valid visual cues on performance in discrimination of auditory targets (null vision-on-auditory finding)

References

Allman, B. L., Bittencourt-Navarrete, R. E., Keniston, L. P., Medina, A. E., Wang, M. Y., & Meredith, M. A. (2008). Do cross-modal projections always result in multisensory integration? *Cerebral Cortex, 18,* 2066-2076. doi:10.1093/cercor/bhm230.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under attention demands. *Current Biology, 15,* 839-843. doi:10.1016/j.cub.2005.03.046.

Alvarado, J. C., Stanford, T. R., Rowland, B. A., Vaughan, J. W., & Stein, B. E. (2006). Multisensory integration in the superior colliculus requires synergy among corticocollicular inputs. *The Journal of Neuroscience, 29,* 6580-6592. doi:10.1523/jneurosci.0525-09.2009.

Alavarado, J. C., Stanford, T. R., Vaughan, J. W., & Stein, B. E. (2007). Cortex mediates multisensory but not unisensory integration in superior colliculus. *The Journal of Neuroscience, 27,* 12775-12786. doi:10.1523/jneurosci.3524-07.2007.

Alvarado, J. C., Vaughan, J. W., Stanford, T. R., & Stein, B. E. (2007). Multisensory versus unisensory integration: contrasting modes in the superior colliculus. *Journal of Neurophysiology, 97,* 3193-3205. doi:10.1152/jn.00018.2007.

Andersen, S. K., & Muller, M. M. (2010). Behavioral performance follows the time course of neural facilitation and suppression during cued shifts of feature selective attention. *Proceedings of the National Academy of Sciences, USA, 107,* 13878-13882.

Angelaki, D. E., Gu, Y., & DeAngelis, G. C. (2009). Multisensory integration: psychophysics, neurophysiology, and computation. *Current Opinion in Neurobiology, 19,* 452-458. doi:10.1016/j.conb.2009.06.008.

Attention, Mental Images and Consciousness. (2003). In B. Kolb & I. Q. Whishaw (Eds.), *Fundamentals of Human Neuropsychology* (6 ed., pp. 622-636). New York, NY: Worth Publishers.

Baier, B., Kleinschmidt, A., & Muller, N. G. (2006). Cross-modal processing in early visual and auditory cortices depends on expected statistical relationship of multisensory information. *The Journal of Neuroscience, 26,* 12260-12265. doi:10.1523/jneurosci.1457-06.2006.

Barrett, D. J. K., & Krumbholz, K. (2012). Evidence for multisensory integration in the elicitation of prior entry by bimodal cues. *Experimental Brain Research, 222,* 11 20. doi:10.1007/s00221-012-3191-8.

Bernstein, I. H., Chu, P. K., Briggs, P., & Schurman, D. L. (1973). Stimulus intensity and foreperiod effects in intersensory facilitation. *Quarterly Journal of Experimental Psychology, 25,* 171-181.

Bernstein, I. H., Rose, R., & Ashe, V. M. (1970). Energy integration in intersensory facilitation. *Journal of Experimental Psychology, 2,* 192-203.

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics, 62,* 321-332.

Bolognini, N., Frassinetti, F., Serino, A., & Ladavas, E. (2005). "Acoustical vision" of

    below threshold stimuli: interaction among spatially converging audiovisual

    inputs. *Experimental Brain Research, 160,* 273-282.

Braithwaith, J. J., & Humphreys, G. W. (2003). Inhibition and anticipation in visual

    search: evidence from effects of color foreknowledge on preview search.

    Perception & Psychophysics, 23, 213-237.

Bremmer, F., Schlack, A., Shah, N. N. J., Zafiris, O., Kubischik, M., Hoffmann, K.,

    Zilles, K., & Fink, G. R. (2001). Polymodal motion processing in posterior

    parietal and premotor cortex: a human fMRI study strongly implies equivalencies

    between humans and monkeys. *Neuron, 29,* 287-296.

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. *Natural Neuroscience, 2,* 759-766.

Byce, S. M., & Wesner, M. F. (2013, May). *Bimodal processing may improve low-end*

    detection but not high-end discrimination performance. Poster presented at the

    Association for Psychological Sciences 25th Annual Convention at the

    Washington Marriott Wardman Park, Washington, DC.

Calvert, G. A. (2001). Crossmodal processing in the human brain: insight from functional

    neuroimaging studies. *Cerebral Cortex, 11,* 1110-1123.

Calvert, G. A., Hansen, P. C., Iversen, S. D., & Brammer, M. J. (2001). Detection of

    multisensory integration sites by application of electrophysiological criteria to the

    BOLD response. *NeuroImage, 14,* 427-438.

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological

approaches and emerging principles in the human brain. *Journal of Physiology

Paris, 98,* 191-205. doi:10.1016/j.physparis.2004.03.018.

Cappe, C., Morel, A., Barone, P., & Rouiller, E. M. (2009). The thalamocortical

projection systems in primate: an anatomical support for multisensory and

sensorimotor integrations. *Cerebral Cortex, 19,* 2025-2037. doi:10.1093/cercor/

bhn228.

Cappe, C., Rouiller, E. M., & Barone, P. (2009). Multisensory anatomical pathways.

Journal of Hearing Research, 258, 28-36. doi:10.1016/j.heares.2009.04.017.

Chen, X., Chen, Q., Gao, D., & Yue, Z. (2012). Interaction between endogenous and

exogenous orienting in crossmodal attention. *Scandinavian Journal of

Psychology, 53,* 303-308. doi: 10.1111/j.1467-9450.2012.00957.x.

Ciaramitaro, V. M., Cameron, E. L., & Glimcher, P. W. (2001). Stimulus probability

directs spatial attention: an enhancement of sensitivity in humans and monkeys.

Vision Research, 41, 57-75.

Cohen, D. J. (1997). Visual detection and perceptual independence: assessing color and

form. *Perception & Psychophysics, 59,* 623-635.

Cohen, Y. E. (2009). Multimodal activity in the parietal cortex. *Hearing Research, 258,*

100-105. doi:10.1016/j.heares.2009.01.011.

Coward, S. W., & Stevens, C. J. (2004). Extracting meaning from sound: nomic

mappings, everyday listening, and perceiving object size from frequency.

Psychological Record, 54, 349-364.

Cuppini, C., Ursino, M., Magosso, E., Rowland, B. A., & Stein, B. E. (2010). An

emergent model of multisensory integration in superior colliculus neurons.

Frontiers in Integrative Neuroscience, 4, 1-15. doi:10.3389/fnint.2010.00006.

Desimone, R., Wessinger, M., Thomas, L., & Schneider, W. (1992). Attentional control

of visual perception: cortical and subcortical mechanisms. *Cold Harbor Symposia*

of Quantitative Biology, 55, 963-971.

Diederich, A. (1995). Intersensory facilitation of reaction time: evaluation of counter and

diffusion coactivation models. *Journal of Mathematical Psychology, 39,* 197-215.

doi:10.1006/jmps.1995.1020.

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement:

effects of stimulus onset and intensity on reaction time. *Perception &*

Psychophysics, 66, 1388-1404.

Dijkstra, T., Frauenfelder, U. H., & Schreuder, R. (1993). Bidirectional grapheme-

phoneme activation in a bimodal detection task. *Journal of Experimental*

Psychology: Human Perception and Performance, 19, 931-950.

Drewe, E. A. (1975). Go no-go learning after frontal lobe lesions in humans. *Cortex, 11,*

8-16.

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences

on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron,*

57, 11-23.

Driver, J., & Spence, C. (1998a). Attention and the crossmodal construction of space.

Trends in Cognitive Sciences, 2, 254-262.

Driver, J., & Spence, C. (1998b). Cross-modal links in spatial attention. *Philosophical Transactions of the Royal Society of London, 353,* 1319-1331.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433-458.

Duncan, J., & Humphreys, G. W. (1992). Beyond the search surface: visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 578-588.

Eimer, M., & Schroger, E. (1998). ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology, 35,* 313-327.

Eimer, M., & Van Velzen, J. (2002). Crossmodal links in spatial attention are mediated by supramodal control processes: evidence from event-related potentials. *Psychophysiology, 39,* 437-449. doi:10.1017.S0048577201393162.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non search task. *Perception & Psychophysics, 16,* 143-149.

Ernst, M. O., & Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8,* 162-169. doi:10.1016/j.tics.2004.02.002.

Fairhall, S. L., & Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience, 29,* 1247-1257. doi:10.1111/j.1460-9568.2009.06688.x

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

Fernandez-Duque, D., & Posner, M. I. (1997). Relating the mechanisms of orienting

and alerting. *Neuropsychologia, 35,* 477-486.

Focker, J., Hotting, K., Gondan, M., & Roder, B. (2010). Unimodal and crossmodal gradients of spatial attention: evidence form event-related potentials. *Brain Topography, 23,* 1-13. doi:10.1007/s10548-009-0111-8.

Fort, A., Delpuech, C., Pernier, J., & Giard, M-H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex, 12,* 1031-1039.

Fort, A., Delpuech, C., Pernier, J., & Giard, M-H. (2002). Early auditory-visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research, 14,* 20-30.

Foxe, J. J., & Simpson, G. V. (2005). Biasing the brain's attentional set: II. Effects of selective intersensory attentional deployments on subsequent sensory processing. *Experimental Brain Research, 166,* 393-401. doi:10.1007/s00221-005-2379-6.

Foxe, J. J., Simpson, G.V., Ahlfors, S. P., & Saron, C. D. (2005). Biasing the brain's attentional set: I. cue driven deployment of intersensory selective attention. *Experimental Brain Research, 166,* 370-392. doi:10.1007/s00221-005-2378-7.

Fries, W. (1984). Cortical projections to the SC in the macaque monkey: a retrograde study using horseradish peroxidase. *Journal of Comparative Neurology, 230,* 55-76.

Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2005). Visual search for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society B: Biological Sciences, 273,* 865-874.

Gallace, A., & Spence, C. (2006). Multisensory synesthetic interactions in the speeded

classification of visual size. *Perception & Psychophysics, 68,* 1191-1203.

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

Gingras, G., Rowland, B. A., & Stein, B. E. (2009). The differing impact of multisensory and unisensory integration on behavior. *The Journal of Neuroscience, 29,* 4897-4902. doi:10.1523/jneurosci.4120-08.2009.

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General, 136,* 389-413. doi:10.1037/0096 3445.136.3.389.

Gondon, M., Gotze, C., & Greenlee, M. W. (2010). Redundancy gains in simple responses and go/no-go tasks. *Attention, Perception, & Psychophysics, 72,* 1692-1709. doi:10.3788/APP.72.6.1692.

Gordon, B. (1983). Lexical access and lexical decision: mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior, 22,* 24-44.

Green, J. J., & McDonald, J. J. (2006). An event-related potential study of supramodal attentional control and crossmodal attentional effects. *Psychophysiology, 43,* 161 171. doi: 10.1111/j.1469-8986.2006.00394.x.

Grice, G. R., & Canham, L. (1990). Redundancy phenomena are affected by response requirements. *Perception & Psychophysics, 48,* 209-213.

Grice, G. R., Canham, L., & Boroughs, J. M. (1984a). Combination rule for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics, 35,* 451-463.

Grice, G. R., Canham, L., & Gwynne, J. W. (1984b). Absence of a redundant-signals

effect in a reaction time task with divided attention. *Perception & Psychophysics,*

36, 565-570.

Grice, G. R., & Gwynne, J. W. (1987). Dependence of target redundancy effects on noise

conditions and number of targets. *Perception & Psychophysics, 42,* 29-36.

Grice, G. R., & Reed, J. M. (1992). What makes a target redundant? *Perception &*

Psychophysics, 51, 437-442.

Guipponi, O., Wardak, C., Ibarrola, D., Comte, J-C, Sappey-Marinier, D., Pinede, S.,

& Hamed, S. B. (2013). Multimodal convergence within the intraparietal sulcus

of the macaque monkey. *The Journal of Neuroscience, 33,* 4128-4139.

doi:10.1523/jneurosci.1421-12.2013.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal*

of Experimental Psychology, 63, 289-293. doi:10.1037/h0039516.

Hino, Y., & Lupker, S. J. (1998). The effects of word frequency for Japanese kana and

kanji words in naming and lexical decision: can the dual-route model save the

lexical-selection account? *Journal of Experimental Psychology: Human*

Perception and Performance, 24, 1431-1453.

Holmes, N. P. (2007). The law of inverse effectiveness in neurons and behavior:

multisensory integration versus normal variability. *Neuropsychologia, 45,* 3340-

3345. doi:10.1016/j.neuropsychologia.2007.05.025.

Holmes, N. P., & Spence, C. (2005). Multisensory integration: space, time, and

superadditivity. *Current Biology, 15,* R762-764.

Houck, M. R., & Hoffman, J. E. (1986). Conjunction of color and form without attention:

evidence from an orientation-contingent color after-effect. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 186-199.

Jiang, W., Jiang, H., & Stein, B. E. (2002). Two corticotectal areas facilitate multisensory orienting behavior. *Journal of Cognitive Neuroscience, 14,* 1240-1255.

Jiang, W., & Stein, B. E. (2003). Cortex controls multisensory depression in superior colliculus. *Journal of Neurophysiology, 90,* 2123-2135. doi:10.1152/jn.00369.2003.

Jiang, W., Wallace, M. T., Jiang, H., Vaughan, J. W., & Stein, B. E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. *Journal of Neurophysiology, 85,* 506-522.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eyes. In J. Long & A. Baddeley (Eds.), *Attention and Performance IX* (pp.187-203). Hillsdale, NJ: Erlbaum.

Joseph, J. S., Chun, M., & Nakayama, K. (1997). Attentional requirements in a "preattentive" feature search task. *Nature, 387,* 805-807.

Kadunce, D. C., Vaughan, J. W., Wallace, M. T., & Stein, B. E. (2001). The influence of Visual and auditory receptive field organization on multisensory integration in the Superior colliculus. *Experimental Brain Research, 139,* 303-310.

Khayat, P. S., Niebergall, R., & Martinez-Trujilo, J. C. (2010). Attention differentially modulates similar neuronal responses evoked by varying contract and direction stimuli in area MT. *Journal of Neuroscience, 30,* 2188-2197.

Kida, T., Inui, K., Tanaka, E., Kakigi, R. (2010). Dynamics of within-, inter-, and crossmodal attentional modulation. *Journal of Neurophysiology, 105,* 674-

686. doi:10.1152/jn.00807.2009.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences, 4,* 138-147. Doi:10.1016/S1364-6613(00)01452-3.

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: a review of audiovisual studies. *Acta Psychologica, 134,* 372-384. doi:10.1016/j.actpsy.2010.03.010.

Koene, A., Arnold, D., & Johnston, A. (2007). Bimodal sensory discrimination is finer than dual single modality discrimination. *Journal of Vision, 7,* 1-11. doi:10.1167/7.11.14.

Laurienti, P. J., Perrault, T. J., Stanford, T. R., Wallace, M. T., & Stein, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research, 166,* 289-297.

Laurienti, P. J., Wallace, M. T., Maldjian, J. A., Susi, C. M., Stein, B. E., & Burdette, J. H. (2003). Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortex. *Human Brain Mapping, 19,* 213-223. doi:10.1002/hbm.10112.

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 451-468.

Lavie, N. (2005). Distracted and confused? selective attention under load. *Trends in Cognitive Sciences, 9,* 76-82.

Lavie, N., HIrst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133,* 339-354. doi:10/1037/0096-3445.133.3.339.

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of

selection in visual attention. *Perception & Psychophysics, 56,* 183-197.

Lewis, J. W., Beauchamp, M. S., & DeYoe, E. A. (2000). A comparison of visual and

auditory motion processing in human cerebral cortex. *Cerebral Cortex, 10,* 888.

Lleras, A., Kawahara, J. I., Wan, X. I., & Ariga, A. (2008). Intertrial inhibition of focused

attention in pop-out search. *Perception & Psychophysics, 70,* 114-131.

Luck, S. (2005). An introduction to event-related potentials and their neural origins.

In *An introduction to the event-related potential technique* (pp. 1-50). Boston,

MASS: MIT Press.

Ma, W. J., & Pouget, A. (2008). Linking neurons to behavior in multisensory perception:

a computational review. *Brain Research, 1242,* 4-12.

doi:10.1016/j.brainres.2008.04.082.

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: a window onto

functional integration in the human brain. *Trends in Neuroscience, 28,* 264-271.

doi:10.1016/j.tins.2005.03.008.

Macaluso, E., Frith, C., & Driver, J. (2001). Multisensory integration and crossmodal

attention effects in the human brain: response. *Science, 292,* 1791a.

doi:10.1126/science.292.5523.1791a.

Marks, L. E. (1987). On cross-modal similarity: auditory-visual interactions in speeded

discrimination. *Journal of Experimental Psychology: Human Perception and

Performance, 13,* 384-394.

Marks, L. E., Ben-Artzi, E., & Lakatos, S. (2003). Cross-modal interactions in auditory

and visual discrimination. *International Journal of Psychophysiology, 50,* 125-

145. doi:10.1016/S0167-8760(30)00129-6.

Martinez-Trujilo, J. C., & Treue, S. (2004). Feature-based attention increases the

selectivity of population responses in primate visual cortex. *Current Biology, 14,*

744-751.

McDonald, J. J., Teder-Salejarvi, W. A., Heraldez, D., & Hillyard, S. A. (2001).

Electrophysiological evidence for the "missing link" in crossmodal attention.

Canadian Journal of Experimental Psychology, 55, 141-149.

McDonald, J. J., Teder-Salejarvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting

to sound improves visual perception. *Nature, 407,* 906-908.

McDonald, J. J., Teder-Salejarvi, W. A., & Ward, L. M. (2001). Multisensory integration

and crossmodal attention effects in the human brain. *Science, 292,* 1791a.

doi:10.1126/science.292.5523.1791a.

McGurk, H., & McDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,*

746-748.

Menon, V., Adleman, N. E., White, C. D., Glover, G. H., & Reiss, A. L. (2001). Error

related brain activation during a go/no-go response inhibition task. *Human Brain*

Mapping, 12, 131-143.

Meredith, M. A. (2002). On the neural basis for multisensory convergence: a brief

overview. *Cognitive Brain Research, 14,* 31-40.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory

integration in superior colliculus neurons I. temporal factors. *The Journal*

of Neuroscience, 7, 3215-3229.

Meredith, M. A., & Stein, B. E. (1983). Interactions among the converging sensory inputs

in the superior colliculus. *Science, 221,* 389-391.

Meredith, M. A., & Stein, B. E. (1985). Descending efferents from the superior colliculus relay integrated multisensory information. *Science, 227,* 657-659.

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

Miller, J. (1991). Channel interaction and the targets effect in bimodal divided attention. *Journal of Experimental Psychology, 17,* 160-169. doi:10.1037/0096 1523.17.1.160.

Miller, J., & Reynolds, A. (2003). The locus of redundant-targets and nontargets effects: evidence from the psychological refractory period paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 1129-1142. doi:10.1037/0096-1523.29.6.1126.

Moher, J., Lakshmanan, B. M., Egeth, H. E., Ewen, J. B. (2014). Inhibition drives early feature-based attention. *Psychological Science, 25,* 315-324. doi:10.1177/0956797613511257.

Morel, A., Liu, J., Wannier, T., Jeanmonod, D., & Rouiller, E. M. (2005). Divergence and convergence of thalamocortical projections to premotor and supplementary motor cortex: a multiple tracing study in macaque monkey. *European Journal of Neuroscience, 21,* 1007-1029.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320,* 264-265.

Nemitz, J. W., Meredith, M. A., & Stein, B. E. (1984). Temporal determinants of multimodal interactions in superior colliculus cells. *Society for Neuroscience*

Abstract, 10, 298.

Neumann, O., Van der Heijden, A. H. C., & Allport, D. A. (1986). Visual selective attention: introductory remarks. *Psychological Research, 48,* 185-188.

Nickerson, R. S. (1973). Intersensory facilitation of reaction time: energy summation or preparation enhancement? *Psychological Review, 80,* 489-509. doi:10.1037/h0035437.

Nieuwenhuis, S., Yeung, N., & Cohen, J. (2004). Stimulus modality, perceptual overlap, and the go/no-go N2. *Psychophysiology, 41,* 157-160.

Pearson, R., Brodal, P., Gatter, K. C., & Powell, T. P. (1982). The organization of connections of the connections between the cortex and claustrum in the monkey. *Brain Research, 234,* 435-441.

Perea, M., Rosa, E., & Romez, C. (2002). Is the go/no-go lexical decision task an alternative to the yes/no lexical decision task? *Memory & Cognition, 30,* 34-45.

Perrault, T. J., Vaughan, J. W., Stein, B. E., & Wallace, M. T. (2003). Neuron-specific response characteristics predict the magnitude of multisensory integration. *Journal of Neurophysiology, 90,* 4022-4026. doi:10.1152/jn.00494.2003.

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

Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review, 78,* 391-408. doi:10.1037/h0026947.

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology, 58,* 1-23.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual*

Review of Neuroscience, 13, 25-42.

Posner, M. I., Snyder, C. R. R., Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology, 109,* 160-174. doi:10.1037/0096 3445.109.2.106.

Prime, D. J., McDonald, J. J., Green, J., & Ward, L. M. (2008). When cross-modal attention fails. *Canadian Journal of Experimental Psychology, 62,* 192-197. doi:10.1037/1196-1961.62.3.192.

Prinzmetal, W., Presti, D. E., & Posner, M. I. (1986). Does attention affect visual feature integration? *Journal of Experimental Psychology: Human Perception and Performance, 12,* 361-369.

Quinlan, P. T. (2003). Visual feature integration theory: past, present, and future. *Psychological Bulletin, 129,* 643-673. doi:10.1037/0033-2909.129.5.643.

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences, 24,* 574-590.

Rowland, B. A., Quessy, S., Stanford, T. R., & Stein, B. E. (2007). Multisensory integration shortens physiological response latencies. *The Journal of Neuroscience, 27,* 5879-5884. doi:10.1523/jneurosci.4986-06.2007.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience, 5,* 631-632.

Santangelo, V., Olivetti Belardinelli, M., Spence, C. (2007). The suppression of reflexive visual and auditory orienting when attention is otherwise engaged. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 137-148.

Santangelo, V., Olivetti Belardinelli, M., Spence, C., & Macaluso, E. (2009). Interactions between voluntary and stimulus-driven spatial attention across sensory modalities. *Journal of Cognitive Neuroscience, 21,* 2384-2397. doi:10.1162/jocn. 2008.21178.

Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception & Performance, 33,* 1311-1321.

Santangelo, V., & Spence, C. (2008). Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies. *Consciousness and Cognition, 17,* 989-1015. doi:10.1162/jocn.2008.21178.

Santangelo, V., Van der Lubbe, R. H. J., Olivetti Belardinelli, M., & Postma, A. (2008). Multisensory integration affects ERP components elicited by exogenous cues. *Experimental Brain Research, 185,* 269-277.

Schirillo, J. A. (2011). Cross-modal detection using various temporal and spatial configurations. *Attention, Perception, & Psychophysics, 73,* 237-246. doi:10.3758/s13414-010-0012-7.

Schroeder, C. E., & Foxe, J. (2005). Multisensory contributions to low-level, 'unisensory' processing. *Current Opinion in Neurobiology, 15,* 454-458. doi:10.1016/j.conb.2005.06.008.

Schroeder, C. E., Molhom, S., Lakatos, P., Ritter, W., & Foxe, J. J. (2004). Human-simian correspondence in the early cortical processing of multisensory cues. *Cognitive Processing, 5,* 140-151. doi:10/1007/s10339-004-0020-4.

Sepulcre, J., Sabuncu, M. R., Yeo, T. B., Liu, H., & Johnson, K. A. (2012). Stepwise

connectivity of multimodal organization in the human brain. *The Journal of Neuroscience, 32,* 10649-10661. doi:10.1523/jneurosci.4515-12.2012.

Serino, A., Farne, A., Rinaldesi, M. L., Haggard, P., & Ladavas, E. (2007). Can vision of the body ameliorate impaired somatosensory function? *Neuropsychologia, 45,* 1101-1107.

Shin, E., Wan, X. I., Fabiani, M., Gratton, G., & Lleras, A. (2008). Electrophysiological evidence of feature-based inhibition of focused attention across consecutive trials. *Psychophysiology, 45,* 804-811.

Shipp, S. (2003). The functional logic of cortico-pulvinar connections. *Philosophical Transactions of the Royal Society B: Biological Sciences, 358,* 1605-1624.

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. *Journal of Neuroscience, 30,* 4024-4032.

Spence, C. (2010). Crossmodal attention. *Scholarpedia, 5*(5), 6309. doi:10.4249/scholarpedia.6309.

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attention, Perception, & Psychophysics, 73,* 971-995. doi: 10.3758/s13414-010-0073-7.

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1005-1030.

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics, 59,* 1-22.

Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: a review. *Hearing Research, 258*, 134-142. doi:10.1016/j.heares.2009.04.015.

Sperdin, H. F., Cappe, C., & Murray, M. M. (2010). Auditory-somatosensory multisensory interactions in humans: dissociating detection and spatial discrimination. *Neuropsychologia, 48*, 3696-3705. doi:10.1016/j.neuro psychologia.2010.09.001.

Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in cat superior colliculus. *The Journal of Neuroscience, 25*, 6499-6508. doi:10.1523/jneurosci.5095-04.2005.

Stein, B. E., Huneycutt, W. S., & Meredith, M. A. (1988). Neurons and behavior: the same rules of multisensory integration apply. *Brain Research, 448*, 355-358.

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge, MA: The MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews: Neuroscience, 9*, 225-266 doi:10.1038/nrn2331.

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2009). The neural basis of multisensory integration in the midbrain: its organization and maturation. *Hearing Research, 258*, 4-15. doi:10.1016/j.heares.2009.03.012.

Styles, E. A. (2006). *The psychology of attention*. East Sussex, England: Psychology Press.

Talsma, D., Doty, T. J., Strowd, R., & Woldorff, M. G. (2006). Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality. *Psychophysiology, 43,* 541-549. doi:10.1111/j.1469.2006.00452.x.

Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for earl integration? *Cerebral 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.

Talsma, D., Kok, A., Slagter, H. A., & Cipriani, G. (2008). Attentional orienting across the sensory modalities. *Brain and Cognition, 66,* 1-10. doi:10.1016/j.bandc.2007.04.005.

Talsma, D., Senkowski, D., Soto-Franco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14,* 400-410. doi:10.1016/j.tics.2010.06.008.

Taslma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience, 17,* 1098-1114.

Tanner, W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review, 61,* 401-409. doi:10.1037/h0058700.

Teder-Salejarvi, W. A., Munte, T. F., Sperlich, F-J., & Hillyard, S. A. (1999). Intra modal and cross-modal spatial attention to auditory and visual stimuli: an event related brain potential study. *Cognitive Brain Research, 8,* 327-343.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and

    for objects. *Journal of Experimental Psychology: Human Perception and*

    Performance, 8, 194-214.

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and*

    Image Processing, 31, 156-177.

Treisman, A. (1986). Features and objects in visual processing. *Scientific American, 255,*

    106-115.

Treisman, A. (1988). Features and objects: the fourteenth Bartlett memorial lecture.

    Quarterly Journal of Experimental Psychology: Human Experimental

    Psychology, 40A, 201-237.

Treisman, A. (1991). Search, similarity, and integration of features between and within

    dimensions. *Journal of Experimental Psychology: Human Perception and*

    Performance, 17, 652-675.

Treisman, A. (1992). Spreading suppression or feature integration? A reply to Duncan

    and Humphreys (1992). *Journal of Experimental Psychology: Human Perception*

    and Performance, 18, 589-593.

Treisman, A. (1993). The perception of features and objects. In A. Baddeley & L.

    Weiskrantz (Eds.), *Attention: selection, awareness, and control. A tribute to*

    Donald Broadbent (pp. 5-35). Oxford, England: Clarendon Press.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention.

    Cognitive Psychology, 12, 97-136.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental*

    Psychology: Human Perception and Performance, 16, 459-478.

Treisman, A., & Schmidt, G. (1992). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107-141.

Turner, B., Mishkin, M., & Knapp, M. (1980). Organization of the amygdalopetal projections from modality-specific cortical association areas in the monkey. *Journal of Comparative Neurology, 191,* 515-543.

Ulrich, R., & Miller, J. (1997). Tests of race models for reaction time experiments with asynchronous redundant signals. *Journal of Mathematical Psychology, 41,* 367-381.

Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1053-1065.

Van der Lubbe, R. H. J., & Postma, A. (2005). Interruption from irrelevant auditory and visual onsets even when attention is in a focused state. *Experimental Brain Research, 164,* 464-471.

Vossel, S., Thiel, C. M., & Fink, G. R. (2006). Cue validity modulates the neural correlates of covert endogenous orienting of attention in parietal and frontal cortex. *NeuroImage, 32,* 1257-1264. doi:10.1016/j.neuroimage.2006.05.019.

Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1583-1590.

Wallace, M. T., Meredieth, M. A., & Stein, B. E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research, 91,* 484-488.

Wallace, M. T., Meredith, A., & Stein, B. E. (1998). Multisensory integration in the

superior colliculus of the alert cat. *Journal of Neurophysiology, 80,* 1006-1010.

Wallace, M. T., & Stein, B. E. (2001). Sensory and multisensory responses in the newborn monkey superior colliculus. *The Journal of Neuroscience, 21,* 8886 8894.

Walsh, V. (2003). A theory of magnitude: common cortical matrices of time, space and quality. *Trends in Cognitive Sciences, 7,* 483-488.

Ward, L. M. (1994). Supramodal and modality specific mechanisms for stimulus-driven shifts in auditory and visual attention. *Canadian Journal of Experimental Psychology, 48,* 242-259.

Ward, L. M., McDonald, J. J., & Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Perception & Psychophysics, 62,* 1258-1264.

Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors, 50,* 449-454.

Wilkinson, L. K., Meredith, M. A., & Stein, B. E. (1996). The role of anterior ectosylvian Cortex in cross-modality orientation and approach behavior. *Experimental Brain Research, 112,* 1-10.

Wolfe, J. (1994). Guided search 2.0: a revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202-238.

Xing, J., & Andersen, R. A. (2000). Models of the posterior parietal cortex which perform multimodal integration and represent space in several coordinate frames. *Journal of Cognitive Neuroscience, 12,* 601-614. doi:10/1162/089892900562363.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary verses automatic allocation. *Journal of Experimental Psychology: Human*

Perception and Performance, 16, 121-134. doi:10.1037/0096-1523.16.1.121.

Zampini, M., Torresan, D., Spence, C., & Murray, M. M. (2007). Auditory-

somatosensory multisensory interactions in front and rear space.

Neuropychologia, 45, 1869-1877.

Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual

processing. *Nature Neuroscience, 12,* 24-25.