



HHS Public Access

Author manuscript

Neurosci Biobehav Rev. Author manuscript; available in PMC 2017 February 01.

Published in final edited form as:

Neurosci Biobehav Rev. 2016 February ; 61: 208–224. doi:10.1016/j.neubiorev.2015.11.002.

The interactions of multisensory integration with endogenous and exogenous attention

Xiaoyu Tang^{a,c}, Jinglong Wu^{b,c,*}, and Yong Shen^{d,e}

^aCollege of Psychology, Liaoning Normal University, 850 Huanghe Road, Shahekou District, Dalian, Liaoning, 116029, China

^bKey Laboratory of Biomimetic Robots and System, Ministry of Education, State Key Laboratory of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology, 5 Nandajie, Zhongguancun, Haidian, Beijing, 100081, China

^cBiomedical Engineering Laboratory, Graduate School of Natural Science and Technology, Okayama University, 3-1-1 Tsushima-naka, Okayama, 700-8530, Japan

^dNeurodegenerative Disease Research Center, School of Life Sciences, University of Science and Technology of China, CAS Key Laboratory of Brain Functions and Disease, Hefei, China

^eCenter for Advanced Therapeutic Strategies for Brain Disorders, Roskamp Institute, Sarasota, FL 34243, USA

Abstract

Stimuli from multiple sensory organs can be integrated into a coherent representation through multiple phases of multisensory processing; this phenomenon is called multisensory integration. Multisensory integration can interact with attention. Here, we propose a framework in which attention modulates multisensory processing in both endogenous (goal-driven) and exogenous (stimulus-driven) ways. Moreover, multisensory integration exerts not only bottom-up but also top-down control over attention. Specifically, we propose the following: (1) endogenous attentional selectivity acts on multiple levels of multisensory processing to determine the extent to which simultaneous stimuli from different modalities can be integrated; (2) integrated multisensory events exert top-down control on attentional capture via multisensory search templates that are stored in the brain; (3) integrated multisensory events can capture attention efficiently, even in quite complex circumstances, due to their increased salience compared to unimodal events and can thus improve search accuracy; and (4) within a multisensory object, endogenous attention can spread from one modality to another in an exogenous manner.

*Corresponding author at Key Laboratory of Biomimetic Robots and System, Ministry of Education, State Key Laboratory of Intelligent Control and Decision of Complex Systems, Beijing Institute of Technology, 5 Nandajie, Zhongguancun, Haidian, Beijing, 100081, China. Tel.: +81 86 251 8052; fax: +81 86 251 8266.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

multisensory integration; multisensory processing; attention; exogenous attention; endogenous attention; attentional selectivity; multisensory search templates; cross-modal spread of attention

1. Introduction

1.1. Multisensory integration

When we look for a friend in a rowdy crowd, it is easier to find our target if that person waves his/her arms and shouts loudly. To help us complete this search task more rapidly, information from different sensory modalities (i.e., visual: the waving arms; and auditory: the shout) not only interacts but also converges into a coherent and meaningful representation. These interactions and convergences between individual sensory systems have been termed multisensory integration (Lewkowicz and Ghazanfar, 2009; Talsma et al., 2010). There are two main types of behavioral outcomes of multisensory integration. The first type includes the multisensory illusion effects that have been demonstrated to illustrate the merging of information across senses, e.g., the *ventriloquism effect*¹ (Hairston et al., 2003), the *McGurk effect* (McGurk and MacDonald, 1976), the *freezing effect* (Vroomen and de Gelder, 2000), and the *double-flash illusion* (Shams et al., 2000). The second type includes multisensory performance improvement effects, such as the redundant signals effect (RSE), in which responses to the simultaneous presentation of stimuli from multiple sensory systems can be faster and more accurate than responses to the same stimuli presented in isolation (Hershenson, 1962; Kinchla, 1974). In this paper, we focus on the multisensory performance improvement effects, such as the RSE, that are used to underscore the combining of information from separate modalities.

The multisensory integration research field has produced enormous gains in interest and popularity since the late 19th century (Stratton, 1897). In the last few decades, many studies have used technological advances in neuroimaging and electrophysiology to address where and when multisensory integration should be expected. Evidence of multisensory processing has been demonstrated in a number of cortical and subcortical human brain areas (see Figure 1a). The superior colliculus (SC) is part of the midbrain and contains a large number of multisensory neurons that play an important role in the integration of information from the somatosensory, visual and auditory modalities (Fairhall and Macaluso, 2009; Meredith and Stein, 1996; Wallace et al., 1998). The superior temporal sulcus (STS), which is an association cortex, mediates multisensory benefits at the level of object recognition (Werner and Noppeney, 2010b), especially for biologically relevant stimuli from different modalities; such stimuli include speech (Senkowski et al., 2008), faces/voices (Ghazanfar et al., 2005), and real-life objects (Beauchamp et al., 2004; Werner and Noppeney, 2010a). Posterior parietal regions such as the superior parietal lobule (SPL) and intraparietal sulcus (IPS) can mediate behavioral multisensory facilitation effects (Molholm et al., 2006; Werner and Noppeney, 2010a) through anticipatory motor control (Krause et al., 2012b). The posterior parietal and premotor cortices act at guiding and controlling action in space and are also

¹Terms with the format of *bold-italics* have been explained in **Glossary**. See supplementary material.

important for the integration of neural signals from different sensory modalities (Bremmer et al., 2001; Driver and Noesselt, 2008). Prefrontal cortex neurons have been found to participate in meaningful cross-modal associations (Fuster et al., 2000). For example, the ventrolateral prefrontal cortex (vlPFC) mediates multisensory facilitation of semantic categorization (Sugihara et al., 2006; Werner and Noppeney, 2010a). Moreover, integration between the senses can influence activity at some of the lowest cortical levels, e.g., the primary visual cortex (Martuzzi et al., 2007; Romei et al., 2007), primary auditory cortex (Calvert et al., 1997; Van den Brink et al., 2014), and primary somatosensory cortex (Cappe and Barone, 2005; Zhou and Fuster, 2000). These presumptive unimodal sensory areas have also been suggested to be multisensory (Ghazanfar and Schroeder, 2006).

In addition, multisensory integration has been attributed to anatomical connections between different brain areas. On the one hand, connections between sensory-related subcortical structures and the corresponding cortical areas play a role in multisensory processing. Such connections include those between the medial geniculate nucleus (MGN) and primary auditory cortex and between the lateral geniculate nucleus (LGN) and primary visual cortex (Noesselt et al., 2010; Van den Brink et al., 2014). Multisensory integration in the SC has also been shown to be mediated by cortical inputs (Bishop et al., 2012; Jiang et al., 2001). On the other hand, connections between cortical areas can mediate multisensory improvements. For example, synchronous auditory stimuli may amplify visual activations by increasing the connectivity between low-level visual and auditory areas and improve visual perception (Beer et al., 2011; Lewis and Noppeney, 2010; Romei et al., 2009).

The neural areas that are correlated with multisensory integration (especially its improvement of behavioral/perceptual outcomes) have been summarized above. Obviously, multisensory integration can occur across multiple neural levels (i.e., at subcortical levels, at the level of association cortices, and at the lowest cortical levels), which indicates that multisensory integration can be modulated by a variety of factors. Previous studies have shown that the *intensity*, *temporal coincidence*, and *spatial coincidence* [at least in some circumstances; see the review by (Spence, 2013)] of multisensory stimuli are determinants of multisensory integration (Meredith et al., 1987; Meredith and Stein, 1986a, b; Stein and Meredith, 1993; Stein et al., 1993). Although multisensory integration is typically considered an automatic process, it can be affected by *top-down* factors, such as attention (Talsma and Woldorff, 2005).

1.2. Endogenous and exogenous attention

Attention plays a key role in selecting relevant and ruling out irrelevant modalities, spatial locations, and task-related objects. Two mechanisms, endogenous and exogenous, are involved in this filtering process. Endogenous attention is also called voluntary or goal-driven attention and involves a more purposeful and effort-intensive *orienting* process (Macaluso, 2010), e.g., orienting to a red table after someone tells you that your friend is at a red table. In contrast, exogenous attention, which is also called involuntary or *stimulus-driven* attention, can be triggered reflexively by a salient sensory event in the external world (Hopfinger and West, 2006), e.g., the colorful clothing of your friend causes him/her to stand out.

The relationship between endogenous and exogenous attention has been extensively explored. In studies of the visual system, endogenous and exogenous attention are generally considered to be two distinct attention systems that have different behavioral effects and partially unique neural substrates (Berger et al., 2005; Chica et al., 2013; Mysore and Knudsen, 2013; Peelen et al., 2004). Unlike endogenous attention, exogenous attention does not demand cognitive resources and is less susceptible to interference (Chica and Lupiáñez, 2009). The effects that are induced by exogenous attention are faster and more transient than those induced by endogenous attention (Busse et al., 2008; Jonides and Irwin, 1981; Shepherd and Müller, 1989). Neuroimaging studies have revealed that the two mechanisms are mediated by a largely common fronto-parietal network (Peelen et al., 2004). However, studies have also suggested that endogenous attention is associated with the dorsal attention network, which is illustrated in blue in Figure 1b, whereas exogenous attention is associated with the ventral attention network (Fox et al., 2006), which is illustrated in red in Figure 1b. The ventral attention network is right lateralized and includes the right temporal-parietal junction (TPJ), the right ventral frontal cortex (VFC), and parts of middle frontal gyrus (MFG) and of the inferior frontal gyrus (IFG). The ventral network is involved in involuntary (stimulus-driven) orienting, which directs attention to salient events (Chica et al., 2013; Fox et al., 2006). The dorsal attention network is bilateral and includes the SPL, IPS, and frontal eye field (FEF) of the prefrontal cortex. The dorsal network is involved in voluntary (top-down) orienting, and its activity increases after the presentation of cues that indicate where, when, or to what subjects should direct their attention (Corbetta and Shulman, 2002). Chica et al. (2013) put forward the hypothesis of “a dorsal frontoparietal network in the orienting of both endogenous and exogenous attention, a ventral frontoparietal counterpart in reorienting to task-relevant events” (Chica et al., 2013; Corbetta et al., 2008). Event-related potential (ERP) studies have shown that endogenous and exogenous attention modulate different stages of stimulus processing. Specifically, endogenous attention exerts its effects on the N1 (Hopfinger and West, 2006) and P300 (Chica and Lupiáñez, 2009) components, whereas exogenous attention modulates the P1 component (Chica and Lupiáñez, 2009; Hopfinger and West, 2006). A previous study, however, showed that the amplitude of P1 can be modulated by endogenous attention, while that of N1 can be modulated by exogenous attention (Natale et al., 2006).

The relationship between endogenous and exogenous attention has also been considered in other models. Studies have suggested that the mechanisms of endogenous and exogenous attention constitute two distinct attention systems but that they also draw on the same capacity-limited system (Busse et al., 2008). Within this capacity-limited system, the mechanisms of endogenous and exogenous attention are not independent; instead, they compete with each other for the control over attention (Godijn and Theeuwes, 2002; Yantis, 1998, 2000; Yantis and Jonides, 1990). The winner of the competition between the exogenous and endogenous mechanism takes control of attention and determines where or what is to be attended.

Regardless of whether endogenous and exogenous attention are two distinct attentional systems or are two modes of a single attention system, the majority of studies in the field have at least shown that the two mechanisms differentially modulate stimulus processing.

This finding suggests that endogenous and exogenous attention may also differentially modulate multisensory integration.

1.3. Interactions of multisensory integration with endogenous and exogenous attention

Attention enables the selection of stimuli from a multitude of sensory information to help the brain integrate useful stimuli from various sensory modalities into coherent cognition (Giard and Peronnet, 1999). Conversely, due to its increased salience, an integrated multisensory stimulus can capture attention more efficiently in complex contexts (Van der Burg et al., 2008b). Recently, investigations of the interplay between multisensory integration and attention have blossomed in a spectacular fashion. To date, however, it is unclear under what circumstances and through what mechanisms multisensory integration and attention interact. Although some studies have demonstrated that multisensory integration can occur independently of attention (Bertelson et al., 2000a; Bertelson et al., 2000b; Spence and Driver, 2000; Vroomen et al., 2001a; Vroomen et al., 2001b), other studies have found that attention can modulate multisensory integration (Alsius et al., 2005; Alsius et al., 2007; Harrar et al., 2014; Talsma et al., 2007; Talsma and Woldorff, 2005).

To explain the relationship between multisensory integration and attention, multiple proposals have been put forth. For example, Talsma et al. (2010) proposed that multisensory integration has a stimulus-driven effect on attention but that top-down directed attention also influences multisensory processing (Talsma et al., 2010). A review by De Meo et al. (2015) proposed that early multisensory integration is independent of top-down attentional control (De Meo et al., 2015). The interaction between multisensory integration and attention has also been proposed to depend on the level of processing at which the integration occurs (Koelewijn et al., 2010). These studies focused primarily on the interaction between top-down attentional control (referred to here as endogenous attention) and multisensory integration. As discussed in section 1.2, however, endogenous and exogenous attention may modulate stimulus processing in distinct manners. These two mechanisms might modulate multisensory processing differently, as discussed in detail below. Therefore, in this review, we describe the interactions between multisensory integration and both endogenous and exogenous attention.

Based on reports in the literature, we propose the framework illustrated in Figure 1c, which shows that attention modulates multisensory processing in both a goal-driven endogenous [illustrated in Figure 1c (1)] and stimulus-driven exogenous [illustrated in Figure 1c (4)] manner. Further, multisensory integration exerts both *bottom-up* [Figure 1c (3)] and top-down [Figure 1c (2)] control over attention. The following sections discuss the contents of the framework in greater detail.

Moreover, we note that the majority of studies of the interaction between multisensory integration and endogenous and exogenous attention that are reviewed here used auditory/visual materials (a few used tactile materials). The framework proposed here may well apply to other combinations of sensory modalities (e.g., tactile, smell, taste, or proprioception), but such an examination is beyond the scope of the present article.

2. Effects of endogenous attentional selectivity on multisensory performance improvements

In this section, we mainly review studies of the modulation of multisensory performance improvements by endogenous, goal-driven selectivity. The term “multisensory performance improvements” has been considered in a broad sense here and is used to describe situations in which a stimulus from one modality can cause faster, more accurate, and/or more precise perception of a stimulus from another different modality. Within the auditory and visual modalities, multisensory performance improvements include reports of faster responses to visual (Cornel et al., 2002) or auditory (Li et al., 2010) targets, an increase in the perceived visual salience (Van der Burg et al., 2008b; Van der Burg et al., 2011), a decrease in the visual contrast thresholds (Lippert et al., 2007; Noesselt et al., 2010), and a change in the response bias (Odgaard et al., 2003).

According to the literature, endogenous attention can modulate multisensory performance improvements through spatial or modality selectivity (see Figure 2). However, studies have also suggested that endogenous and exogenous spatial attention do not influence the ventriloquism effect (Bertelson et al., 2000a; Bertelson et al., 2000b; Spence and Driver, 2000; Vroomen et al., 2001a; Vroomen et al., 2001b). Therefore, the modulation of endogenous attention on multisensory illusions must be considered carefully.

2.1. Modulation of multisensory performance improvements by the spatial selectivity of endogenous attention

Based on task instructions or informative visual central cue stimuli, attention can be focused on a spatial location, such as the left or right side of fixation, which is called selective spatial attention (Attend to a location, Figure 2 & Table 1). Attention can also be allocated to multiple locations, e.g., to both the left and right, which is called divided spatial attention (Attend to multiple locations, Figure 2 & Table 1). Previous studies have reported that this endogenous attentional selectivity can facilitate responses to unimodal (visual-V or auditory-A) signals at the attended (expected) spatial locations compared with the unattended (unexpected) locations (Coull and Nobre, 1998; Li et al., 2012; Posner et al., 1980; Tang et al., 2013). This analogous attention effect (attended vs. unattended) has also been found for stimuli from multiple sensory modalities, e.g., the simultaneous presentation of auditory and visual stimuli (audiovisual-AV) (Table 1). For example, the amplitude of the P1 component elicited by attended audiovisual stimuli is larger than the P1 elicited by unattended audiovisual stimuli (Talsma et al., 2007). When spatially-coincident stimuli from different modalities are presented simultaneously, the multisensory performance improvement that is the outcome of multisensory integration has been demonstrated in situations of both selective (Li et al., 2010; Wu et al., 2009) and divided spatial attention (Li et al., 2015). The multisensory performance improvement is typically accompanied by brain responses to multisensory stimuli that diverge from the summed brain responses to the constituent unimodal stimuli (e.g., AV vs. A+V). This nonlinear response is the hallmark of multisensory integration (De Meo et al., 2015; Giard and Peronnet, 1999).

Selective spatial attention has been shown to modulate multisensory nonlinear responses; specifically, greater ERP responses to stimuli at attended locations than to those at unattended locations have been observed at 280 ms post-stimulus onset over the centro-medial area (Li et al., 2010), at 100 ms post-stimulus onset over the fronto-central area (Talsma and Woldorff, 2005), and even as early as 40 ms in oscillatory gamma-band responses (Senkowski et al., 2005). Selective spatial attention can also modulate higher-level multisensory integration, e.g., the interaction between speaking lips in the visual stream and spoken words in the auditory stream (Fairhall and Macaluso, 2009). In the functional magnetic resonance imaging (fMRI) study that reported this last result, endogenous attention enhanced the activity in cortical and subcortical sites, including the STS, striate visual cortex, extrastriate visual cortex and SC only when the simultaneously spoken words matched the attended speaking lips.

Further, in many circumstances, spatial attention is required to be distributed across different locations (e.g., the left and right sides) and not focused on only one location (see Figure 2: Example 3 & 4). In divided spatial attention conditions, the perceptual sensitivity to a visual target can be enhanced by audiovisual interactions when a simultaneous auditory signal is presented in the spatially congruent/coincident location instead of a spatially incongruent or different location (Frassinetti et al., 2002; Gao et al., 2014). Moreover, audiovisual interaction can be modulated by spatial congruency via the neural activities over the ventral occipito-temporal and superior temporal area starting at 100 ms after the onset of stimuli (Teder-Sälejärvi et al., 2005).

The majority of studies have found that endogenous spatial attention can enhance multisensory integration (Fairhall and Macaluso, 2009; Talsma and Woldorff, 2005), although one study demonstrated the larger benefits of multisensory stimulation on visual target detection in the endogenously unattended half of the stimulus display compared with the endogenously attended half of the display (Zou et al., 2012). Taken together, the results mentioned above provide evidence for the ability of endogenous attentional spatial selectivity to modulate multisensory performance improvement based on both low-level meaningless stimuli and high-level meaningful stimuli. Further, the endogenous attentional modulation effect on multisensory processing is influenced by the spatial or semantic congruency.

2.2. Modulation of multisensory performance improvements by the modality selectivity of endogenous attention

Attention can be allocated to a specific modality in multisensory streams via instructions or goals through the use of the endogenous attentional selectivity mechanism. For example, when reading a book in noisy circumstances, people must concentrate on the task-relevant modalities, i.e., the book in the visual modality as well as the motion of turning the pages, while ignoring task-irrelevant modalities, such as the noise in the auditory modality. Paying attention to a specific modality can speed up information processing in low-level cortical areas; this effect is also referred to as the *prior-entry effect* (Vibell et al., 2007).

Recently, behavioral and ERP responses to audiovisual stimuli have been found to be increased when subjects pay attention to visual (Wu et al., 2009), auditory (Li et al., 2010),

or audiovisual streams (Giard and Peronnet, 1999). However, modality-specific selective attention (attending to a modality) and divided-modality attention (attending to multiple modalities) differentially modulate multisensory processing (see Table 1: Attend to a modality vs. Attend to multiple modalities). The effect of multisensory integration on behavioral performance can be attenuated or even eliminated under conditions of modality-specific selective attention (Mozolic et al., 2008; Wu et al., 2012a). Multisensory facilitation of response times and accuracy was also found to be optimal when both auditory and visual stimuli were targets (Barutchu et al., 2013). Furthermore, relative to conditions in which attention is focused on a single specific modality, when attention is distributed across modalities, sensory gating can be modulated (Anderson and Rees, 2011; Talsma et al., 2007) such that multisensory integration occurs earlier, e.g., within 100 ms after stimulus onset (the P50 component) (Giard and Peronnet, 1999; Talsma et al., 2007). Both early multisensory ERP processing (Magnée et al., 2011; Talsma et al., 2007) and multisensory-related fMRI responses in the superior temporal cortex (Degerman et al., 2007) are enhanced when attention is divided across modalities relative to under conditions of modality-selective attention.

Although multisensory neural processing enhancements have been found when attention is divided across modalities or is focused on a single modality, multisensory neural processing enhancements are associated with null (Talsma et al., 2007, attend auditory stimuli task) or negative multisensory behavioral performance (Degerman et al., 2007). Specifically, multisensory neural processing enhancements were found by Talsma et al. (2007) and Degerman et al. (2007). However, in the former study, no significant differences between behavioral responses to audiovisual stimuli and behavioral responses to auditory stimuli were found when the subjects were attending to the auditory modality. In the latter study, behavioral responses to audiovisual stimuli were more accurate when the subjects were attending to the visual or auditory modality compared to when they were attending to multiple modalities. Moreover, an ERP study showed that multisensory performance improvements are associated with reduced neural processing during divided-modality compared with modality-selective attention (Mishra and Gazzaley, 2012). These inconsistent results are indicative of the differential modulation of multisensory processing by selective- and divided-modality attention. However, it is difficult to determine the direction of the link between behavioral performance and the underlying multisensory neural activity. Differences in the experimental parameters and tasks might be manifested in the different underlying neural changes and behavioral consequences. Although neural responses reflect brain activity more sensitively than behavioral data, all explanations of the changes in neural responses that occur due to experimental differences should be based on behavioral performance (Cappe et al., 2010; Van der Burg et al., 2011).

Endogenous attention influences multisensory performance improvements at multiple stages through selectivity that is based on spatial location or modality. The monitoring of two locations or modalities has been found to cost more than the monitoring of a specific location or modality, and this cost is correlated with more intense activity in the fronto-parietal region or in the superior temporal cortex when attention is divided between locations or modalities (Degerman et al., 2007; Santangelo et al., 2010). In addition, the behavioral costs of monitoring multiple modalities at two locations are smaller than those of

monitoring multiple modalities at a single location, and this difference has been associated with increased activity in the left and right precuneus (Santangelo et al., 2010). These results suggest that the two types of attentional selectivity are not independent; attending to multiple locations or focusing on a single location interplays with attending to multiple modalities or focusing on a single modality. An ERP study suggested that drawing attention to sensory modalities instead of to spatial locations might more readily speed-up information processing (Vibell et al., 2007). Therefore, the differential modulation of multisensory integration by modality or spatial location selectivity is worthy of further investigation in the future.

Additionally, interactions between multisensory integration and endogenous top-down attentional control are affected by the semantic congruence of multisensory stimuli, although the semantic congruence of multisensory stimuli does not directly modulate early multisensory processing (Fort et al., 2002; Molholm et al., 2004; Yuval-Greenberg and Deouell, 2007). Studies have suggested that changing the proportion of congruent trials will result in shifts in attentional control and will thus indirectly modulate the outcome of multisensory integration (Sarmiento et al., 2012). This indirect modulation effect has also been demonstrated when the spread of attention across modalities, space and time was investigated (Busse et al., 2005; Donohue et al., 2011; Fiebelkorn et al., 2010), as discussed in section 5.

In this section, we review studies of interactions between endogenous attentional selectivity and multisensory performance improvements. The majority of the results support the hypothesis that behavioral or neural responses to multisensory stimuli are facilitated when the stimuli are presented in endogenously attended locations or modalities compared with unattended ones. With respect to the interaction between exogenous attention and multisensory integration, however, a reduced multisensory performance improvement was observed when audiovisual targets were exogenously attended (Van der Stoep et al., 2015). Van der Stoep et al. (2015, Experiment 1) adopted a left/right/central sound as an exogenous cue to trigger spatial orienting attention. The cue was then followed by a visual (V), auditory (A), or audiovisual (AV) target after a randomly determined short inter-stimulus interval (ISI: 200–250 ms). Participants were asked to detect the V/A/AV targets that were presented centrally or at the same side (valid condition) or that were presented on the side opposite (invalid condition) to the auditory cue. Significantly larger race model violations were observed under the invalid condition compared with the valid condition, suggesting that multisensory integration is decreased at exogenously attended locations compared with exogenously unattended locations. This study verified our hypothesis that both endogenous and exogenous attention can influence multisensory processing and that the two attention mechanisms might modulate multisensory processing in different ways.

3. Multisensory templates exert top-down control on contingent attentional capture

As illustrated in the framework, not only can endogenous attentional selectivity modulate multisensory processing; integrated bimodal signals can also influence the allocation of attention. As an example, consider a procedure in which a visual target search display is

presented following an uninformative exogenous visual spatial cue that is or is not accompanied by a tone (Matusz and Eimer, 2011). An audiovisual cue of this type has been found to elicit a larger *spatial cueing effect* than corresponding visual cue, and the audiovisual enhancement of attentional capture has been shown to be automatic and independent of top-down search goals, such as those related to the search for specific or non-specific colors (see Table 2).

Furthermore, attentional capture has been suggested to be contingent on the attentional control settings that are induced by the demands of the task (Folk et al., 1992). For example, in a typical visual contingent capture paradigm, subjects are instructed to respond to a red square that has been defined as the target; before a search array containing the target is presented, a spatially uninformative red or blue singleton cue is presented. Under these conditions, responses to red targets that appeared at the same location as the cue are faster than those to targets that appeared at the opposite location but only when the cue was the red singleton and not when the cue was the blue one; this phenomenon is referred to as the task-set contingent attentional capture effect. Additionally, the different waveform of the N2pc component, which is obtained by subtracting ERPs recorded from electrodes located over the posterior area (e.g., PO7/8) ipsilateral to the side of the color singleton cues from the corresponding contralateral ERPs, is used to estimate the spatial selection of potential targets among distractors (Luck and Hillyard, 1994). A recent ERP study used cues only in the visual modality, and the targets were defined as visual color singletons in the visual search task or as visual color singletons accompanied by a sound in the audiovisual task (see Figure 3a). The authors of this study found that the behavioral spatial cueing effect on the visual target was smaller than that observed on the audiovisual target (see Figure 3b). Moreover, the ERP results revealed that the amplitude of the cue-locked N2pc component during the audiovisual search task was smaller than the task-set contingent attentional capture effect during the unimodal visual search task (see Figure 3c). This effect may have occurred because bimodal attentional templates are involved in the guidance of audiovisual search such that the attentional capture ability of the unimodal visual cue is diminished (Matusz and Eimer, 2013), which has also been confirmed in a visual-tactile search task (Mast et al., 2015).

The results of the studies by Matusz & Eimer (2013) and Mast et al. (2015) provide evidence that integrated multisensory signals control attentional capture during multisensory search in a top-down manner, indicating that multisensory integration influences attention in both a stimulus-driven (Talsma et al., 2010) and top-down fashion.

4. Effects of multisensory integration on exogenous attention

It is well known that responses to visual stimuli can be enhanced by simultaneous auditory stimuli (Kinchla, 1974), even when the auditory stimuli are task-irrelevant. This so-called redundant signals effect or bimodal enhancement effect is a direct reflection of the beneficial effects of multisensory integration on behavioral performance. In the visual search paradigm, the direct effect of multisensory integration on exogenous attention has been investigated using behavioral and neural responses to visual stimuli both with and without synchronously presented task-irrelevant tones. Moreover, multisensory integration also acts

on the exogenous attention in an indirect manner; this effect can be observed by comparing the spatial cueing effect elicited by unimodal cues with that elicited by multisensory cues in the exogenous cueing paradigm.

4.1. Multisensory integration acts to generate exogenous orienting of spatial attention

Since 1980, the Posner paradigm has been widely used to study two qualitatively different attentional orienting mechanisms; one is endogenous, and the other is exogenous. In the classic endogenous cueing paradigm (Posner, 1980), a left- or right-pointing arrow is presented centrally as a cue to guide the participant's attention to the location of the subsequently presented target stimulus. When the exogenous orienting mechanism is the topic of study, the exogenous cueing paradigm is most often utilized (Posner and Cohen, 1984; Zhang et al., 2013). In this paradigm, a change in an uninformative singleton that is presented on either the left or right side of a peripheral box (e.g., a brightened outline) serves as a cue that reflexively summons the participant's attention. After an ISI, a target stimulus appears at either the same (cued condition) or opposite location (uncued condition) of the cue. The exogenous spatial cueing effect is estimated as the difference between the mean reaction time to the targets presented at the uncued location and the mean reaction time to the targets presented at the cued location. This spatial cueing effect can be modulated by different cue-target stimulus onset asynchrony (SOA). Responses to targets are facilitated at the cued location relative to the uncued location in short-ISI conditions (e.g., shorter than 150 ms, the facilitation effect), whereas responses to targets are inhibited at the cued location relative to the uncued location in long-ISI conditions (e.g., longer than 300 ms; the inhibition effect or inhibition of return, i.e., IOR). The time course of the cross-modal cueing effect has found to be longer than that of the intermodal case (Tassinari and Campara, 1996). As an example, the time course of the IOR effect refers to the SOA in which the IOR effect is elicited. The time course of the IOR effect in the cross-modal condition (e.g., visual cue but auditory target) is delayed relative to the time course of the IOR effect in the unimodal condition (e.g., visual cue and visual target). Specifically, previous studies found a significant IOR during cross-modal orienting (visual cue with auditory target) at an SOA of 1050–1350 ms (Spence et al., 2000) but not at an SOA of 575 ms (Schmitt et al., 2000) or 650 ms (Yang and Mayer, 2014).

In contrast to the typical visual exogenous cueing procedure, a modified paradigm has been used to investigate whether multisensory integration can modulate exogenous orienting mechanisms (Table 2). In this version of the paradigm, the presentation of auditory, visual, or audiovisual cue stimuli followed by the presentation of the visual target stimuli. If multisensory integration can facilitate attentional orienting, then an audiovisual cue is expected to elicit a larger spatial cueing effect than a visual or auditory cue. Inconsistent with this prediction, the spatial cueing effect that is triggered by multisensory cues has been found to be analogous to the cueing effect that is triggered by auditory or visual cues (Mahoney et al., 2012; Santangelo et al., 2006). Nevertheless, ERP results have revealed evidence of the integration of audiovisual cues: the posterior P1 component that is elicited by bimodal signals is larger than the sum of those elicited by the single auditory and visual cues (Santangelo et al., 2008b). However, different results are observed when the perceptual/attentional load is manipulated. Only the peripheral spatial cueing task is utilized

in the no-load condition (Figure 4a, left panel), whereas the *rapid serial visual presentation (RSVP)* task is added during the high-load condition (Figure 4a, right panel) to otherwise engage the participant's resources at the center of the display. Spatial cueing effects of comparable magnitudes are elicited by single- and multiple-modality cues in the no-load condition (Figure 4b). However, only the multisensory cues trigger significant spatial cueing effects during the high-perceptual-load condition (Santangelo et al., 2008a; Santangelo and Spence, 2007). Similar results demonstrating that audiovisual cues elicit larger spatial cueing effects than visual cues in the high-load condition have also been found when participants are asked to complete a visual search task (Matusz and Eimer, 2011) or a *temporal order judgment (TOJ)* task (Barrett and Krumbholz, 2012). A larger cueing effect during the high-load condition was also demonstrated when audiotactile cues were compared with auditory cues (Ho et al., 2009). The study by Ho et al. (2009, Experiment 2) showed that in the no-load condition, a significant spatial cueing effect was elicited by auditory, tactile, or audiotactile cues. However, only audiotactile cues triggered a significant cueing effect in the high-load condition.

As mentioned previously, multisensory stimuli have been hypothesized to intensify the allocation of attention towards the cued location. Support for this hypothesis has been found when participants are engaged in other demanding tasks, such as a visual search task. During these tasks, audiovisual cues elicit larger cueing effects than unimodal cues do. These results could occur because the magnitude of the spatial cueing effect that is elicited by the audiovisual cues results from the combination of the spatial cueing effects that are elicited by the auditory and visual cue components. Thus, multisensory stimuli, compared to unimodal stimuli, can capture attention more intensively in a stimulus-driven fashion (Krause et al., 2012a). This finding can also explain why spatially congruent multisensory cues can increase the attention effect and are more effective in biasing access to visual spatial working memory compared to unimodal visual cues (Botta et al., 2011). All of the results that have been described above suggest that multisensory integration acts on the process responsible for exogenous orienting of spatial attention.

4.2. Multisensory integration facilitates visual search

A response to a single visual event can be facilitated by a synchronous single auditory event through improved perceptual sensitivity (Stein et al., 1996), altered responses bias (Odgaard et al., 2003), or temporal attentional capture (Spence and Ngo, 2012). However, every moment, our senses are bombarded with huge amounts of sensory information, as described in the opening story about finding a friend in a rowdy crowd. Recently, the effects of a single tone on the competition between multiple, concurrently presented visual objects in a spatial layout have been investigated (Van der Burg et al., 2008b). In this study, participants searched for a horizontal or vertical line segment among distractor line segments of various orientations that were all continuously changing color. The authors of this study found that both the search time and the search slopes were drastically reduced when the target color change was accompanied by a spatially uninformative tone relative to a condition in which no auditory signal was presented. This audition-driven visual search benefit was called the "pip and pop" effect and resulted from multisensory interaction rather than from an increase in alertness or top-down temporal cueing (Van der Burg et al., 2008a; Van der Burg et al.,

2008b). The pip and pop effect was also observed when the target color change was accompanied by an uninformative tactile signal (Van der Burg et al., 2009) or an olfactory stimulus (Chen et al., 2013). These results suggest that multisensory interaction can aid in the resolution of competition between multiple stimuli.

Furthermore, the underlying neural mechanisms of the pip and pop effect, i.e., how sounds can affect competition among visual stimuli, have been investigated with electrophysiological techniques. In an ERP study (Van der Burg et al., 2011), participants were told to search for a horizontal or vertical target line that was presented at a lateral location in the lower visual field among distractors, all of which were continuously changing orientation (see Figure 4c). The authors of this study observed behavioral search benefits for visual targets that were accompanied by synchronous sounds relative to those without sounds (see Figure 4d, up), which is the so-called pip and pop phenomenon. Moreover, an early interaction between the visual target and the accompanying sound was observed. This interaction began at 50 ms post-stimulus onset over the left parieto-occipital cortex and was followed by a strong N2pc component, which is indicative of attentional capture (Luck and Hillyard, 1994), and a subsequently increased *CNSW* component, which has been linked to visual short-term memory (Klaver et al., 1999). More interestingly, the earliest multisensory interaction was correlated with the behavioral pip and pop effect (see Figure 4d, bottom), indicating that participants with strong early multisensory interactions benefited the most from the synchronized auditory signal. These findings are consistent with the notions that the behavioral benefit results from increased sensitivity (Staufenbiel et al., 2011) and that auditory signals can enhance visual processing within early, low-level visual cortex (Romei et al., 2009; Romei et al., 2007), i.e., an auditory signal can enhance the neural response to a synchronous visual event. Further, multisensory integration of a visual target and a simultaneous auditory event enhances orienting to the location of the visual target and suppresses orienting to the locations of visual distractors (Pluta et al., 2011). These results provide evidence that multisensory integration affects the facilitation of search efficiency. This hypothesis is consistent with previous studies in which the audiovisual ventriloquism effect occurred prior to exogenous orienting (Vroomen et al., 2001b) such that attention was attracted toward the illusory location of the ventriloquized sound (Spence and Driver, 2000; Vroomen et al., 2001a). However, an alternative account has been proposed in which, rather than performance improvements being mediated by multisensory interaction processing, the presentation of any synchronous cue can facilitate a participant's visual target identification performance as long as the signal results in the target being perceived as an "oddball" among distractor stimuli (Ngo and Spence, 2010a; Ngo and Spence, 2012). Therefore, studies are needed to further elucidate the underlying mechanisms of the influences of multisensory interaction on visual search.

The pip and pop effect is modulated not only by stimulus-driven properties, e.g., the transience of signals (Van der Burg et al., 2010b), but also by top-down effects, such as the temporal expectations that are triggered by rhythmic events (Kosem and van Wassenhove, 2012) or a spatial hint concerning the location of a visual target that is provided by temporally synchronous auditory and vibrotactile cues (Ngo and Spence, 2010b). Multisensory integration can take place pre-attentively; this type of integration occurs during

early multisensory interactions between visual distractors and synchronized sound (Van der Burg et al., 2011). The pip and pop effect occurs automatically as long as the spatial attention that is modulated by top-down control is divided across the visual field (Theeuwes, 2010; Van der Burg et al., 2012). In addition, attention to at least one modality has been found to be necessary to speed-up the processing of multisensory stimuli (Van der Burg et al., 2010a).

Generally, multisensory interactions can modulate exogenous/involuntary attention in either a direct manner (e.g., a visual search task with or without a tone) or an indirect manner (e.g., a unimodal or multisensory cue followed by target stimuli). As described in the previously mentioned studies, in contrast to unimodal cues, multisensory cues can attract attention involuntarily, even under conditions of high perceptual/attentional loads (Santangelo et al., 2008a; Santangelo and Spence, 2007) as well as in the visual search task conditions (Matusz and Eimer, 2011). In addition, multisensory interactions can act to increase visual search efficiency (Van der Burg et al., 2008b; Van der Burg et al., 2011). Further, one study used a visual search task followed by a visual or audiovisual cue to trigger exogenous spatial attention and found that, relative to the visual cue, the audiovisual cue triggered a greater spatial cueing effect; this effect suggests that multisensory interactions more efficiently modulate attentional capture. However, this finding seems to contradict previous findings that the spatial cueing effect that is elicited by visual cues is analogous to that elicited by audiovisual cues (Santangelo et al., 2006, 2008b). These contradictory results can be accounted for by the experimental context, i.e., whether distractor interference was designed to potentially compete with the target for the participants' attention. According to the theory of perceptual load (Lavie, 2005), compared to no- or low-load conditions, high-attentional/perceptual-load conditions are required to fully engage all attentional resources in the task. Thus, in simple contexts, a similar shift of exogenous spatial attention or an analogous spatial cueing effect can be triggered regardless of cue type, whereas in more demanding contexts, such as those in which the target is presented in high-load conditions or among distractors, only the more salient or effective multisensory signals can trigger visual target processing enhancements relative to unimodal signals.

5. Cross-modal spread of attention within a multisensory object

Within the visual modality, attention has been suggested to spread across different parts of a visual object (Blaser et al., 2000; Egly et al., 1994). Whether attention spreads across modalities has been investigated (Busse et al., 2005). As shown in the illustration in the left of Figure 5a, participants were asked to pay attention to the right visual event while ignoring the left visual event and all of the center auditory events. By subtracting the ERP responses to only the visual event from the ERP responses to the audiovisual events (i.e., AV-V), the auditory component could be extracted from the responses to a multisensory object in visual attended or unattended conditions. Relative to the unattended condition, late sustained (220–700 ms) activity was observed in response to the central sound over frontal areas in the attended visual stimulus condition. Corresponding fMRI data also showed an attention effect, i.e., enhanced activity in the auditory cortex when the visual stimulus was attended relative to when it was unattended. These results indicate that attention can spread from an

attended visual event to an ignored simultaneous sound; this phenomenon is called the cross-modal spread of attention.

Here, the distinction between the cross-modal spread of attention and cross-modal spatial linking should be noted. The latter occurs when an ignored auditory stimulus appears at a visually cued location (Driver and Spence, 1998). However, the ignored auditory stimulus and the attended visual stimulus are presented asynchronously; thus, they can be considered to be two distinct objects. In contrast, the cross-modal spread of attention occurs within a multisensory object. A multisensory object refers to a single unified percept that is formed when stimuli from different sensory modalities are presented simultaneously. In other words, the peripheral visual stimuli and the center auditory stimuli are presented simultaneously or nearly simultaneously [i.e., the interval between the two stimuli is difficult to perceive, such as an interval of 100 ms, see (Donohue et al., 2011)] such that the visual and auditory stimuli can be integrated into a unified percept (Figure 6). Within the formed multisensory object, attention to the visual component can spread across locations and modalities to the auditory component in an automatic or exogenous manner.

Furthermore, the dual mechanisms are contained during the cross-modal spread of attention (Fiebelkorn et al., 2010). On one hand, stimulus-driven spread of attention occurs whenever an ignored tone appears simultaneously or nearly simultaneously with an attended visual stimulus. This spread of attention is correlated with the activity of the visual and auditory cortices (Busse et al., 2005) and constrained by the temporal and spatial link between the multisensory stimuli. Specifically, temporal linking and the spreading of attention occur only when the auditory and visual stimuli are judged to be simultaneous, whereas, more strictly, spatial linking (like the ventriloquism effect) occurs only when the two stimuli are actually presented simultaneously (Donohue et al., 2011) (see Figure 5b). These results are consistent with the notion that both temporal and spatial parameters are critical in the perception of real-world objects (Slutsky and Recanzone, 2001). On the other hand, representation-driven spread of attention occurs when there is an object-related congruency between relevant visual stimuli and irrelevant auditory stimuli, such as visual presentation of the letter “A” or “X” that is accompanied by the sound of a letter “A” being spoken (Zimmer et al., 2010a; Zimmer et al., 2010b) or the presentation of a picture of a dog/guitar that is accompanied by the sound of barking (Fiebelkorn et al., 2010). The representation-driven spread of attention is modulated by higher cognitive processes, e.g., congruency or other learned associations. The anterior cingulate cortex (ACC) is known to be involved in conflict resolution (Van Veen and Carter, 2005; Weissman et al., 2003). Thus, in addition to the visual and auditory cortices, the ACC, which is correlated with congruency processes, participates in representation-driven spread of attention. There are no consistent conclusions regarding whether a congruent sound or an incongruent sound with a task-relevant visual stimulus can trigger a greater spread of attention. For example, Zimmer et al. (2010) proposed that an incongruent sound is a stronger distraction than a congruent one and captures attention more intensively. However, the cross-modal spread of attention has also been demonstrated to occur only when a task-irrelevant sound is semantically congruent with a visual stimulus (Molholm et al., 2007). Thus, the details of the representation-driven spread of attention deserve further study.

To date, in studies of the cross-modal spread of attention, attention has been entirely triggered in a pattern of endogenous selectivity (see the blue box in Figure 2). However, the cross-modal spread of attention is thought to be an exogenously driven, automatic process. The task-relevant visual stimulus and the synchronous task-irrelevant auditory stimulus can be integrated into a coherent multisensory object in a process that is attributable to endogenous attentional selectivity. Within a multisensory object, attention can spread automatically across modalities. Finally, the task-irrelevant auditory stimulus that was unattended becomes attended (see Figure 6). Thus, it is difficult to determine whether multisensory integration interacts with either endogenous or exogenous attention because the two attention mechanisms interact with each other during the cross-modal spread of attention. After acknowledging the limitations associated with this issue, we cannot help but ask whether attention that is triggered in an exogenous manner can spread across modalities and locations. If so, the entire process of the cross-modal spread of attention might be considered a completely exogenous or involuntary process. Moreover, ERP data have revealed that the cross-modal spread of attention occurs during late-stage processing, e.g., 220 ms post-stimulus onset, which is consistent with the modulation of late-stage stimulus processing by sustained endogenous attention (P300 component, Chica and Lupianez, 2009). However, transient exogenous attention has been found to mediate early-stage stimulus processing (P1 component, Chica and Lupianez, 2009; N1 component, Natale et al., 2006; see also the review by Chica et al., 2013). If the cross-modal spreads of endogenous and exogenous attention were both controlled in the same experimental setting and the spread of each was compared, it might be possible to determine whether multisensory processes are differentially modulated by endogenous and exogenous attention. This question needs to be further investigated.

6. Concluding remarks and future directions

In this review, we proposed a comprehensive framework for the interactions of multisensory integration with endogenous and exogenous attention. The effects of exogenous and endogenous attention on multisensory integration as well as the adverse effects have been summarized to illustrate the interactions from multiple perspectives. Specifically, endogenous attentional selectivity acts on multiple levels of multisensory processing and determines the extent to which simultaneous stimuli from different modalities can be integrated. Integrated multisensory events, which have greater salience compared with unimodal signals, capture attention effectively, and improve search accuracy even in quite complex circumstances. Additionally, multisensory templates that are stored in the brain exert top-down control over attentional capture. Endogenous attention can spread from a component in one modality to another modality within a multisensory object in an exogenous manner. In summary, the novel points proposed in this review article are as follows. (I) Attention modulates multisensory processing in both goal-driven endogenous and stimulus-driven exogenous patterns. Endogenous and exogenous attention differentially but mutually modulate multisensory processing. (II) Multisensory integration exerts bottom-up and top-down control over attention.

Frameworks for the interactions between attention and multisensory integration have been proposed; however, a few unresolved questions are worthy of investigation. For example,

what key stimulus properties are necessary to link multisensory processing and attention? Can we use noninvasive techniques, such as transcranial magnetic stimulation (TMS), to determine the circumstances under which multisensory integration interacts with exogenous or endogenous attention? By utilizing the known effects of brain damage and the cognitive symptoms of patients, can we dissociate multisensory integration from attentional effects? Finally, although exogenous and endogenous attention have been reported to result from two independent attentional systems, some common neural substrates are shared between these systems. Thus, it will be worthwhile to control the two types of orienting attention within the same paradigm to clarify whether multisensory integration differentially interacts with exogenous and endogenous attention.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported in part by JAPAN SOCIETY FOR THE PROMOTION OF SCIENCE (JSPS) KAKENHI, grant numbers: 25249026 and 25303013 and a Grant-in-Aid for Strategic Research Promotion from Okayama University; National Natural Science Foundation of China (NSFC61473043); NIHNIAR01025888 (YS), JSPS Award (YS) and University of Science and Technology of China Fund (YS).

References

- Alsius A, Navarra J, Campbell R, Soto-Faraco S. Audiovisual integration of speech falters under high attention demands. *Current Biology*. 2005; 15:839–843. [PubMed: 15886102]
- Alsius A, Navarra J, Soto-Faraco S. Attention to touch weakens audiovisual speech integration. *Experimental Brain Research*. 2007; 183:399–404. [PubMed: 17899043]
- Anderson EJ, Rees G. Neural correlates of spatial orienting in the human superior colliculus. *Journal of Neurophysiology*. 2011; 106:2273–2284. [PubMed: 21753026]
- Barrett DJK, Krumbholz K. Evidence for multisensory integration in the elicitation of prior entry by bimodal cues. *Experimental Brain Research*. 2012; 222:11–20. [PubMed: 22975896]
- Barutcu A, Freestone DR, Innes-Brown H, Crewther DP, Crewther SG. Evidence for enhanced multisensory facilitation with stimulus relevance: an electrophysiological investigation. *PLoS One*. 2013; 8:e52978. [PubMed: 23372652]
- Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience*. 2004; 7:1190–1192. [PubMed: 15475952]
- Beer AL, Plank T, Greenlee MW. Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. *Experimental Brain Research*. 2011; 213:299–308. [PubMed: 21573953]
- Berger A, Henik A, Rafal R. Competition between endogenous and exogenous orienting of visual attention. *Journal of Experimental Psychology: General*. 2005; 134:207–221. [PubMed: 15869346]
- Bertelson P, Pavani F, Ladavas E, Vroomen J, de Gelder B. Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia*. 2000a; 38:1634–1642. [PubMed: 11074086]
- Bertelson P, Vroomen J, de Gelder B, Driver J. The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*. 2000b; 62:321–332. [PubMed: 10723211]
- Bishop CW, London S, Miller LM. Neural time course of visually enhanced echo suppression. *Journal of Neurophysiology*. 2012; 108:1869–1883. [PubMed: 22786953]
- Blaser E, Pylyshyn ZW, Holcombe AO. Tracking an object through feature space. *Nature*. 2000; 408:196–199. [PubMed: 11089972]

- Botta F, Santangelo V, Raffone A, Sanabria D, Lupianez J, Belardinelli MO. Multisensory integration affects visuo-spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*. 2011; 37:1099–1109. [PubMed: 21553989]
- Bremmer F, Schlack A, Shah NJ, Zafiris O, Kubischik M, Hoffmann K-P, Zilles K, Fink GR. Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron*. 2001; 29:287–296. [PubMed: 11182099]
- Busse L, Katzner S, Treue S. Temporal dynamics of neuronal modulation during exogenous and endogenous shifts of visual attention in macaque area MT. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105:16380–16385. [PubMed: 18922778]
- Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG. The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences of the United States of America*. 2005; 102:18751–18756. [PubMed: 16339900]
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, McGuire PK, Woodruff PWR, Iverson SD, David AS. Activation of auditory cortex during silent lipreading. *Science*. 1997; 276:593–596. [PubMed: 9110978]
- Cappe C, Barone P. Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *European Journal of Neuroscience*. 2005; 22:2886–2902. [PubMed: 16324124]
- Cappe C, Thut G, Romei V, Murray MM. Auditory–visual multisensory interactions in humans: timing, topography, directionality, and sources. *The Journal of Neuroscience*. 2010; 30:12572–12580. [PubMed: 20861363]
- Chen K, Zhou B, Chen S, He S, Zhou W. Olfaction spontaneously highlights visual saliency map. *Proceedings of the Royal Society of London B: Biological Sciences*. 2013; 280:20131729.
- Chica AB, Bartolomeo P, Lupiáñez J. Two cognitive and neural systems for endogenous and exogenous spatial attention. *Behavioural Brain Research*. 2013; 237:107–123. [PubMed: 23000534]
- Chica AB, Lupiáñez J. Effects of endogenous and exogenous attention on visual processing: an inhibition of return study. *Brain Research*. 2009; 1278:75–85. [PubMed: 19374885]
- Corbetta M, Patel G, Shulman GL. The reorienting system of the human brain: from environment to theory of mind. *Neuron*. 2008; 58:306–324. [PubMed: 18466742]
- Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*. 2002; 3:215–229.
- Corneil BD, Van Wanrooij M, Munoz DP, Van Opstal AJ. Auditory-visual interactions subserving goal-directed saccades in a complex scene. *Journal of Neurophysiology*. 2002; 88:438–454. [PubMed: 12091566]
- Coull JT, Nobre AC. Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *The Journal of Neuroscience*. 1998; 18:7426–7435. [PubMed: 9736662]
- De Meo R, Murray MM, Clarke S, Matusz PJ. Top-down control and early multisensory processes: chicken vs egg. *Frontiers in Integrative Neuroscience*. 2015; 9:1–6. [PubMed: 25698945]
- Degerman A, Rinne T, Pekkola J, Autti T, Jaaskelainen IP, Sams M, Alho K. Human brain activity associated with audiovisual perception and attention. *Neuroimage*. 2007; 34:1683–1691. [PubMed: 17204433]
- Donohue SE, Roberts KC, Grent-'t-Jong T, Woldorff MG. The cross-modal spread of attention reveals differential constraints for the temporal and spatial linking of visual and auditory stimulus events. *The Journal of Neuroscience*. 2011; 31:7982–7990. [PubMed: 21632920]
- Driver J, Noesselt T. Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron*. 2008; 57:11–23. [PubMed: 18184561]
- Driver J, Spence C. Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*. 1998; 2:254–262. [PubMed: 21244924]
- Egley R, Driver J, Rafal RD. Shifting visual attention between objects and locations: evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*. 1994; 123:161–177. [PubMed: 8014611]

- Fairhall SL, Macaluso E. Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience*. 2009; 29:1247–1257. [PubMed: 19302160]
- Fiebelkorn IC, Foxe JJ, Molholm S. Dual mechanisms for the cross-sensory spread of attention: how much do learned associations matter? *Cerebral Cortex*. 2010; 20:109–120. [PubMed: 19395527]
- Folk CL, Remington RW, Johnston JC. Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*. 1992; 18:1030–1044. [PubMed: 1431742]
- Fort A, Delpuech C, Pemier J, Giard MH. Early auditory-visual interactions in human cortex during nonredundant target identification. *Cognitive Brain Research*. 2002; 14:20–30. [PubMed: 12063127]
- Fox MD, Corbetta M, Snyder AZ, Vincent JL, Raichle ME. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences of the United States of America*. 2006; 103:10046–10051. [PubMed: 16788060]
- Frassinetti F, Bolognini N, Ladavas E. Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*. 2002; 147:332–343. [PubMed: 12428141]
- Fuster JM, Bodner M, Kroger JK. Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature*. 2000; 405:347–351. [PubMed: 10830963]
- Gao Y, Li Q, Yang W, Yang J, Tang X, Wu J. Effects of ipsilateral and bilateral auditory stimuli on audiovisual integration: a behavioral and event-related potential study. *NeuroReport*. 2014; 25:668–675. [PubMed: 24780895]
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *The Journal of Neuroscience*. 2005; 25:5004–5012. [PubMed: 15901781]
- Ghazanfar AA, Schroeder CE. Is neocortex essentially multisensory? *Trends in Cognitive Sciences*. 2006; 10:278–285. [PubMed: 16713325]
- Giard M, Peronnet F. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*. 1999; 11:473–490. [PubMed: 10511637]
- Godijn R, Theeuwes J. Programming of endogenous and exogenous saccades: evidence for a competitive integration model. *Journal of Experimental Psychology: Human Perception and Performance*. 2002; 28:1039–1054. [PubMed: 12421054]
- Hairston WD, Wallace MT, Vaughan JW, Stein BE, Norris JL, Schirillo JA. Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience*. 2003; 15:20–29. [PubMed: 12590840]
- Harrar V, Tamnam J, Perez-Bellido A, Pitt A, Stein J, Spence C. Multisensory integration and attention in developmental dyslexia. *Current Biology*. 2014; 24:531–535. [PubMed: 24530067]
- Hershenson M. Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*. 1962; 63:289–293. [PubMed: 13906889]
- Ho C, Santangelo V, Spence C. Multisensory warning signals: when spatial correspondence matters. *Experimental Brain Research*. 2009; 195:261–272. [PubMed: 19381621]
- Hopfinger JB, West VM. Interactions between endogenous and exogenous attention on cortical visual processing. *Neuroimage*. 2006; 31:774–789. [PubMed: 16490366]
- Jiang W, Wallace MT, Jiang H, Vaughan JW, Stein BE. Two cortical areas mediate multisensory integration in superior colliculus neurons. *Journal of Neurophysiology*. 2001; 85:506–522. [PubMed: 11160489]
- Jonides J, Irwin DE. Capturing attention. *Cognition*. 1981; 10:145–150. [PubMed: 7198529]
- Kinchla R. Detecting target elements in multielement arrays: a confusability model. *Perception & Psychophysics*. 1974; 15:149–158.
- Klaver P, Talsma D, Wijers AA, Heinze HJ, Mulder G. An event-related brain potential correlate of visual short-term memory. *NeuroReport*. 1999; 10:2001–2005. [PubMed: 10424664]
- Koelewijn T, Bronkhorst A, Theeuwes J. Attention and the multiple stages of multisensory integration: a review of audiovisual studies. *Acta Psychologica*. 2010; 134:372–384. [PubMed: 20427031]

- Kosem A, van Wassenhove V. Temporal structure in audiovisual sensory selection. *PLoS One*. 2012; 7:e40936. [PubMed: 22829899]
- Krause H, Schneider TR, Engel AK, Senkowski D. Capture of visual attention interferes with multisensory speech processing. *Frontiers in Integrative Neuroscience*. 2012a; 6:1–8. [PubMed: 22319479]
- Krause V, Bashir S, Pollok B, Caipa A, Schnitzler A, Pascual-Leone A. 1 Hz rTMS of the left posterior parietal cortex (PPC) modifies sensorimotor timing. *Neuropsychologia*. 2012b; 50:3729–3735. [PubMed: 23103789]
- Lavie N. Distracted and confused?: selective attention under load. *Trends in Cognitive Sciences*. 2005; 9:75–82. [PubMed: 15668100]
- Lewis R, Noppeney U. Audiovisual synchrony improves motion discrimination via enhanced connectivity between early visual and auditory areas. *The Journal of Neuroscience*. 2010; 30:12329–12339. [PubMed: 20844129]
- Lewkowicz DJ, Ghazanfar AA. The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*. 2009; 13:470–478. [PubMed: 19748305]
- Li C, Chen K, Han H, Chui D, Wu J. An fMRI study of the neural systems involved in visually cued auditory top-down spatial and temporal attention. *PLoS One*. 2012; 7:e49948. [PubMed: 23166800]
- Li Q, Wu J, Touge T. Audiovisual interaction enhances auditory detection in late stage: an event-related potential study. *NeuroReport*. 2010; 21:173–178. [PubMed: 20065887]
- Li Q, Yang H, Sun F, Wu J. Spatiotemporal relationships among audiovisual stimuli modulate auditory facilitation of visual target discrimination. *Perception*. 2015; 44:232–242. [PubMed: 26562250]
- Lippert M, Logothetis NK, Kayser C. Improvement of visual contrast detection by a simultaneous sound. *Brain Research*. 2007; 1173:102–109. [PubMed: 17765208]
- Luck SJ, Hillyard SA. Spatial filtering during visual search: evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*. 1994; 20:1000–1014. [PubMed: 7964526]
- Macaluso E. Orienting of spatial attention and the interplay between the senses. *Cortex*. 2010; 46:282–297. [PubMed: 19540475]
- Magnée MJCM, de Gelder B, van Engeland H, Kemner C. Multisensory integration and attention in autism spectrum disorder: evidence from event-related potentials. *PLoS One*. 2011; 6:e24196. [PubMed: 21887382]
- Mahoney JR, Verghese J, Dumas K, Wang CL, Holtzer R. The effect of multisensory cues on attention in aging. *Brain Research*. 2012; 1472:63–73. [PubMed: 22820295]
- Martuzzi R, Murray MM, Michel CM, Thiran JP, Maeder PP, Clarke S, Meuli RA. Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cerebral Cortex*. 2007; 17:1672–1679. [PubMed: 16968869]
- Mast F, Frings C, Spence C. Multisensory top-down sets: evidence for contingent crossmodal capture. *Attention, Perception, & Psychophysics*. 2015; 77:1970–1985.
- Matusz PJ, Eimer M. Multisensory enhancement of attentional capture in visual search. *Psychonomic Bulletin & Review*. 2011; 18:904–909. [PubMed: 21748418]
- Matusz PJ, Eimer M. Top-down control of audiovisual search by bimodal search templates. *Psychophysiology*. 2013; 50:996–1009. [PubMed: 23834379]
- McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature*. 1976; 264:746–748. [PubMed: 1012311]
- Meredith MA, Nemitz JW, Stein BE. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience*. 1987; 7:3215–3229. [PubMed: 3668625]
- Meredith MA, Stein BE. Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*. 1986a; 365:350–354. [PubMed: 3947999]
- Meredith MA, Stein BE. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*. 1986b; 56:640–662. [PubMed: 3537225]

- Meredith MA, Stein BE. Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology*. 1996; 75:1843–1857. [PubMed: 8734584]
- Mishra J, Gazzaley A. Attention distributed across sensory modalities enhances perceptual performance. *The Journal of Neuroscience*. 2012; 32:12294–12302. [PubMed: 22933811]
- Molholm S, Martinez A, Shpaner M, Foxe JJ. Object-based attention is multisensory: co-activation of an object's representations in ignored sensory modalities. *European Journal of Neuroscience*. 2007; 26:499–509. [PubMed: 17650120]
- Molholm S, Ritter W, Javitt DC, Foxe JJ. Multisensory visual–auditory object recognition in humans: a high-density electrical mapping study. *Cerebral Cortex*. 2004; 14:452–465. [PubMed: 15028649]
- Molholm S, Sehatpour P, Mehta AD, Shpaner M, Gomez-Ramirez M, Ortigue S, Dyke JP, Schwartz TH, Foxe JJ. Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *Journal of Neurophysiology*. 2006; 96:721–729. [PubMed: 16687619]
- Mozolic JL, Hugenschmidt CE, Peiffer AM, Laurienti PJ. Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research*. 2008; 184:39–52. [PubMed: 17684735]
- Mysore SP, Knudsen EI. A shared inhibitory circuit for both exogenous and endogenous control of stimulus selection. *Nature Neuroscience*. 2013; 16:473–478. [PubMed: 23475112]
- Natale E, Marzi CA, Girelli M, Pavone EF, Pollmann S. ERP and fMRI correlates of endogenous and exogenous focusing of visual-spatial attention. *European Journal of Neuroscience*. 2006; 23:2511–2521. [PubMed: 16706858]
- Ngo M, Spence C. Crossmodal facilitation of masked visual target identification. *Attention, Perception, & Psychophysics*. 2010a; 72:1938–1947.
- Ngo MK, Spence C. Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli. *Attention, Perception, & Psychophysics*. 2010b; 72:1654–1665.
- Ngo MK, Spence C. Facilitating masked visual target identification with auditory oddball stimuli. *Experimental Brain Research*. 2012; 221:129–136. [PubMed: 22760584]
- Noesselt T, Tyll S, Boehler CN, Budinger E, Heinze HJ, Driver J. Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. *The Journal of Neuroscience*. 2010; 30:13609–13623. [PubMed: 20943902]
- Odgaard EC, Arieh Y, Marks LE. Cross-modal enhancement of perceived brightness: sensory interaction versus response bias. *Perception & Psychophysics*. 2003; 65:123–132. [PubMed: 12699315]
- Peelen MV, Heslenfeld DJ, Theeuwes J. Endogenous and exogenous attention shifts are mediated by the same large-scale neural network. *Neuroimage*. 2004; 22:822–830. [PubMed: 15193611]
- Pluta SR, Rowland BA, Stanford TR, Stein BE. Alterations to multisensory and unisensory integration by stimulus competition. *Journal of Neurophysiology*. 2011; 106:3091–3101. [PubMed: 21957224]
- Posner MI. Orienting of attention. *Quarterly Journal of Experimental Psychology*. 1980; 32:3–25. [PubMed: 7367577]
- Posner, MI.; Cohen, Y. Components of visual orienting. In: Bouma, H.; Bouwhuis, DG., editors. *Attention & Performance X: Control of Language Processes*. Hillsdale, NJ: Erlbaum: Psychology Press; 1984. p. 531-556.
- Posner MI, Snyder CR, Davidson BJ. Attention and the detection of signals. *Journal of Experimental Psychology: General*. 1980; 109:160–174.
- Romei V, Murray MM, Cappe C, Thut G. Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Current Biology*. 2009; 19:1799–1805. [PubMed: 19836243]
- Romei V, Murray MM, Merabet LB, Thut G. Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: implications for multisensory interactions. *The Journal of Neuroscience*. 2007; 27:11465–11472. [PubMed: 17959789]
- Santangelo V, Fagioli S, Macaluso E. The costs of monitoring simultaneously two sensory modalities decrease when dividing attention in space. *Neuroimage*. 2010; 49:2717–2727. [PubMed: 19878728]

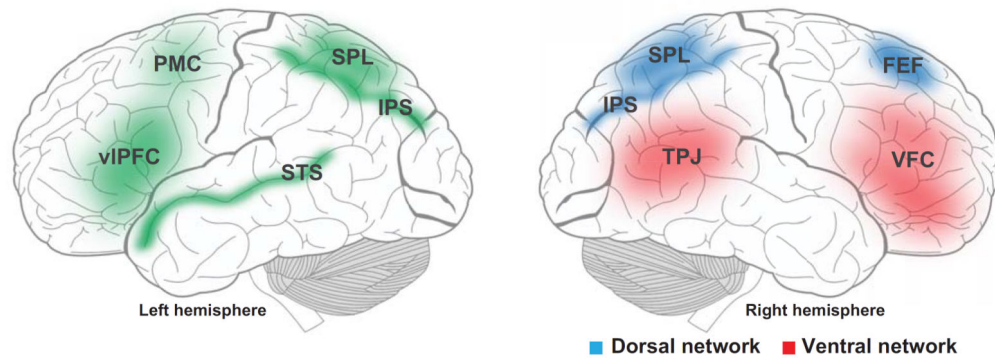
- Santangelo V, Ho C, Spence C. Capturing spatial attention with multisensory cues. *Psychonomic Bulletin & Review*. 2008a; 15:398–403. [PubMed: 18488658]
- Santangelo V, Spence C. Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology: Human Perception and Performance*. 2007; 33:1311–1321. [PubMed: 18085945]
- Santangelo V, Van der Lubbe RH, Belardinelli MO, Postma A. Spatial attention triggered by unimodal, crossmodal, and bimodal exogenous cues: a comparison of reflexive orienting mechanisms. *Experimental Brain Research*. 2006; 173:40–48. [PubMed: 16489435]
- Santangelo V, Van der Lubbe RH, Belardinelli MO, Postma A. Multisensory integration affects ERP components elicited by exogenous cues. *Experimental Brain Research*. 2008b; 185:269–277. [PubMed: 17909764]
- Sarmiento BR, Shore DI, Milliken B, Sanabria D. Audiovisual interactions depend on context of congruency. *Attention, Perception, & Psychophysics*. 2012; 74:563–574.
- Schmitt M, Postma A, Haan ED. Interactions between exogenous auditory and visual spatial attention. *The Quarterly Journal of Experimental Psychology A*. 2000; 53:105–130.
- Senkowski D, Saint-Amour D, Gruber T, Foxe JJ. Look who's talking: the deployment of visuo-spatial attention during multisensory speech processing under noisy environmental conditions. *Neuroimage*. 2008; 43:379–387. [PubMed: 18678262]
- Senkowski D, Talsma D, Herrmann CS, Woldorff MG. Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. *Experimental Brain Research*. 2005; 166:411–426. [PubMed: 16151775]
- Shams L, Kamitani Y, Shimojo S. Illusions - what you see is what you hear. *Nature*. 2000; 408:788–788. [PubMed: 11130706]
- Shepherd M, Müller H. Movement versus focusing of visual attention. *Perception & Psychophysics*. 1989; 46:146–154. [PubMed: 2762102]
- Slutsky DA, Recanzone GH. Temporal and spatial dependency of the ventriloquism effect. *NeuroReport*. 2001; 12:7–10. [PubMed: 11201094]
- Spence C. Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*. 2013; 1296:31–49. [PubMed: 23710729]
- Spence C, Driver J. Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. *NeuroReport*. 2000; 11:2057–2061. [PubMed: 10884070]
- Spence C, Lloyd D, McGlone F, Nicholls MER, Driver J. Inhibition of return is supramodal: a demonstration between all possible pairings of vision, touch, and audition. *Experimental Brain Research*. 2000; 134:42–48. [PubMed: 11026724]
- Spence, C.; Ngo, MK. Does attention or multisensory integration explain the crossmodal facilitation of masked visual target identification. In: Stein, BE., editor. *The new handbook of multisensory processing*. Cambridge MA: MIT Press; 2012. p. 345-358.
- Staufenbiel SM, Van der Lubbe RH, Talsma D. Spatially uninformative sounds increase sensitivity for visual motion change. *Experimental Brain Research*. 2011; 213:457–464. [PubMed: 21805200]
- Stein BE, London N, Wilkinson LK, Price DD. Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. *Journal of Cognitive Neuroscience*. 1996; 8:497–506. [PubMed: 23961981]
- Stein, BE.; Meredith, MA. *The merging of the senses*. Cambridge MA, US: The MIT Press; 1993.
- Stein BE, Meredith MA, Wallace MT. The visually responsive neuron and beyond - multisensory integration in cat and monkey. *Progress in Brain Research*. 1993; 95:79–90. [PubMed: 8493355]
- Stratton GM. Vision without inversion of the retinal image. *Psychological Review*. 1897; 4:341–360.
- Sugihara T, Diltz MD, Averbek BB, Romanski LM. Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *The Journal of Neuroscience*. 2006; 26:11138–11147. [PubMed: 17065454]
- Talsma D, Doty TJ, Woldorff MG. Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*. 2007; 17:679–690. [PubMed: 16707740]

- Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG. The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*. 2010; 14:400–410. [PubMed: 20675182]
- Talsma D, Woldorff MG. Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*. 2005; 17:1098–1114. [PubMed: 16102239]
- Tang X, Li C, Li Q, Gao Y, Yang W, Yang J, Ishikawa S, Wu J. Modulation of auditory stimulus processing by visual spatial or temporal cue: an event-related potentials study. *Neuroscience Letters*. 2013; 553:40–45. [PubMed: 23896527]
- Tassinari G, Campara D. Consequences of covert orienting to non-informative stimuli of different modalities: a unitary mechanism? *Neuropsychologia*. 1996; 34:235–245. [PubMed: 8868280]
- Teder-Sälejärvi WA, Di Russo F, McDonald J, Hillyard SA. Effects of spatial congruity on audio-visual multimodal integration. *Journal of Cognitive Neuroscience*. 2005; 17:1396–1409. [PubMed: 16197693]
- Theeuwes J. Top-down and bottom-up control of visual selection. *Acta Psychologica*. 2010; 135:77–99. [PubMed: 20507828]
- Van den Brink RL, Cohen MX, van der Burg E, Talsma D, Vissers ME, Slagter HA. Subcortical, modality-specific pathways contribute to multisensory processing in humans. *Cerebral Cortex*. 2014; 24:2169–2177. [PubMed: 23529004]
- Van der Burg E, Brederoo SG, Nieuwenstein MR, Theeuwes J, Olivers CNL. Audiovisual semantic interference and attention: evidence from the attentional blink paradigm. *Acta Psychologica*. 2010a; 134:198–205. [PubMed: 20176341]
- Van der Burg E, Cass J, Olivers CNL, Theeuwes J, Alais D. Efficient visual search from synchronized auditory signals requires transient audiovisual events. *PLoS One*. 2010b; 5:e10664. [PubMed: 20498844]
- Van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J. Audiovisual events capture attention: evidence from temporal order judgments. *Journal of Vision*. 2008a; 8:1–10. [PubMed: 18842073]
- Van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J. Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*. 2008b; 34:1053–1065. [PubMed: 18823194]
- Van der Burg E, Olivers CNL, Bronkhorst AW, Theeuwes J. Poke and pop: tactile-visual synchrony increases visual saliency. *Neuroscience Letters*. 2009; 450:60–64. [PubMed: 19013216]
- Van der Burg E, Olivers CNL, Theeuwes J. The attentional window modulates capture by audiovisual events. *PLoS One*. 2012; 7:e39137. [PubMed: 22808027]
- Van der Burg E, Talsma D, Olivers CNL, Hickey C, Theeuwes J. Early multisensory interactions affect the competition among multiple visual objects. *Neuroimage*. 2011; 55:1208–1218. [PubMed: 21195781]
- Van der Stoep N, Van der Stigchel S, Nijboer TCW. Exogenous spatial attention decreases audiovisual integration. *Attention, Perception, & Psychophysics*. 2015; 77:464–482.
- Van Veen V, Carter CS. Separating semantic conflict and response conflict in the Stroop task: a functional MRI study. *Neuroimage*. 2005; 27:497–504. [PubMed: 15964208]
- Vibell J, Klinge C, Zampini M, Spence C, Nobre AC. Temporal order is coded temporally in the brain: early event-related potential latency shifts underlying prior entry in a cross-modal temporal order judgment task. *Journal of Cognitive Neuroscience*. 2007; 19:109–120. [PubMed: 17214568]
- Vroomen J, Bertelson P, de Gelder B. Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychologica*. 2001a; 108:21–33. [PubMed: 11485191]
- Vroomen J, Bertelson P, De Gelder B. The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*. 2001b; 63:651–659. [PubMed: 11436735]
- Vroomen J, de Gelder B. Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*. 2000; 26:1583–1590. [PubMed: 11039486]
- Wallace MT, Meredith MA, Stein BE. Multisensory integration in the superior colliculus of the alert cat. *Journal of Neurophysiology*. 1998; 80:1006–1010. [PubMed: 9705489]

- Weissman DH, Giesbrecht B, Song AW, Mangun GR, Woldorff MG. Conflict monitoring in the human anterior cingulate cortex during selective attention to global and local object features. *Neuroimage*. 2003; 19:1361–1368. [PubMed: 12948694]
- Werner S, Noppeney U. Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *The Journal of Neuroscience*. 2010a; 30:2662–2675. [PubMed: 20164350]
- Werner S, Noppeney U. Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cerebral Cortex*. 2010b; 20:1829–1842. [PubMed: 19923200]
- Wu J, Li Q, Bai O, Touge T. Multisensory interactions elicited by audiovisual stimuli presented peripherally in a visual attention task: a behavioral and event-related potential study in humans. *Journal of Clinical Neurophysiology*. 2009; 26:407–413. [PubMed: 19952565]
- Wu J, Yang J, Yu Y, Li Q, Nakamura N, Shen Y, Ohta Y, Yu S, Abe K. Delayed audiovisual integration of patients with mild cognitive impairment and Alzheimer's Disease compared with normal aged controls. *Journal of Alzheimer's Disease*. 2012a; 32:317–328.
- Wu J, Yang W, Gao Y, Kimura T. Age-related multisensory integration elicited by peripherally presented audiovisual stimuli. *NeuroReport*. 2012b; 23:616–620. [PubMed: 22643234]
- Yang Z, Mayer AR. An event-related fMRI study of exogenous orienting across vision and audition. *Human Brain Mapping*. 2014; 35:964–974. [PubMed: 23288620]
- Yantis, S. Control of visual attention. In: Pashler, H., editor. *Attention*. Hove: Psychology Press; 1998. p. 223-256.
- Yantis, S. Goal-directed and stimulus-driven determinants of attentional control. In: Monsell, S.; Driver, J., editors. *Attention and performance*. Cambridge: MIT Press; 2000. p. 73-103.
- Yantis S, Jonides J. Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance*. 1990; 16:121–134. [PubMed: 2137514]
- Yuval-Greenberg S, Deouell LY. What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *The Journal of Neuroscience*. 2007; 27:1090–1096. [PubMed: 17267563]
- Zhang M, Tang X, Wu J. Blocking the link between stimulus and response at previously attended locations: evidence for inhibitory tagging mechanism. *Neuroscience and Biomedical Engineering*. 2013; 1:13–21.
- Zhou YD, Fuster JM. Visuo-tactile cross-modal associations in cortical somatosensory cells. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97:9777–9782. [PubMed: 10944237]
- Zimmer U, Itthipanyanan S, Grent-T-Jong T, Woldorff M. The electrophysiological time course of the interaction of stimulus conflict and the multisensory spread of attention. *European Journal of Neuroscience*. 2010a; 31:1744–1754. [PubMed: 20584178]
- Zimmer U, Roberts KC, Harshbarger TB, Woldorff MG. Multisensory conflict modulates the spread of visual attention across a multisensory object. *Neuroimage*. 2010b; 52:606–616. [PubMed: 20420924]
- Zou H, Müller HJ, Shi Z. Non-spatial sounds regulate eye movements and enhance visual search. *Journal of Vision*. 2012; 12:1–18.

Endogenous attentional selectivity modulates multisensory performance improvements
Multisensory templates exert top-down control on contingent attentional capture
Multisensory integration acts to generate exogenous orienting of spatial attention
Multisensory integration facilitates search efficiency
Cross-modal spread of attention can occur in a multisensory object

(a) *Multisensory cortical regions involved in multisensory integration* (b) *Brain areas correlated with endogenous and exogenous attention*



(c) *A framework for the interactions of multisensory integration with endogenous and exogenous attention*

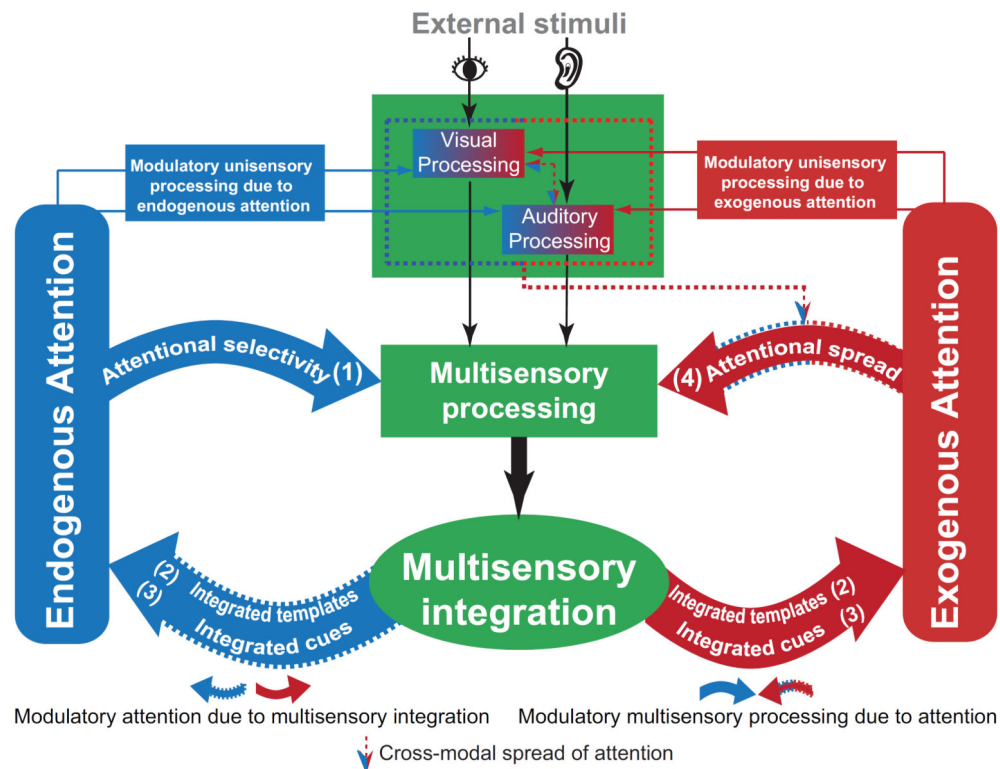


Figure 1.

(a) Multisensory cortical regions (green) that are involved in multisensory integration. SPL = superior parietal lobule; IPS = intraparietal sulcus; STS = superior temporal sulcus; vIPFC = ventrolateral prefrontal cortex; PMC = premotor cortex. (b) Brain areas that are correlated with endogenous and exogenous attention. SPL = superior parietal lobule; IPS = intraparietal sulcus; FEF = frontal eye field; TPJ = temporal-parietal junction; VFC = ventral frontal cortex. Endogenous attention is associated with the dorsal attention network (blue), while exogenous attention is associated with the ventral attention network (red) (Fox et al.,

2006). The dorsal attention network is bilateral. It is involved in voluntary (top-down) orienting and exhibits increases in activity after the presentation of cues that indicate where, when, or to what subjects should direct their attention. The ventral attention network is right lateralized. It is involved in involuntary (stimulus-driven) orienting and exhibits increases in activity after the presentation of salient targets, particularly when they appear in unexpected locations (Chica et al., 2013; Fox et al., 2006). **(c)** A framework for the interactions of multisensory integration with endogenous and exogenous attention. External stimuli from sensory organs can be integrated at multiple multisensory processing levels (Giard and Peronnet, 1999; Talsma and Woldorff, 2005). Multisensory integration is elicited as a consequence of the multiple phases of multisensory processing. Although these multisensory processes are thought to be automatic, attention influences not only unimodal processing but also multisensory processing in both an endogenous and exogenous manner. Endogenous attention can modulate multisensory processing via endogenous attentional selectivity [(1) **Attentional selectivity**]. This modulatory effect determines the extent to which simultaneously presented stimuli from different modalities can be integrated (see Figure 2 & Table 1). Furthermore, the integrated multisensory stimuli can be represented in multisensory templates that are stored in the brain. These multisensory templates exert top-down control over contingent attentional capture [(2) **Integrated templates**]. Due to their increased salience relative to unimodal cues, integrated multisensory cues can influence the exogenous orienting of spatial attention even under quite complex circumstances or can improve visual search efficiency by increasing target sensitivity [(3) **Integrated cues**]. Finally, endogenous attention can spread from one modality to another in an exogenous manner such that the stimuli of the unattended modality come to be “attended” [(4) **Attentional spread**].

Effects of endogenous attentional selectivity on multisensory processing

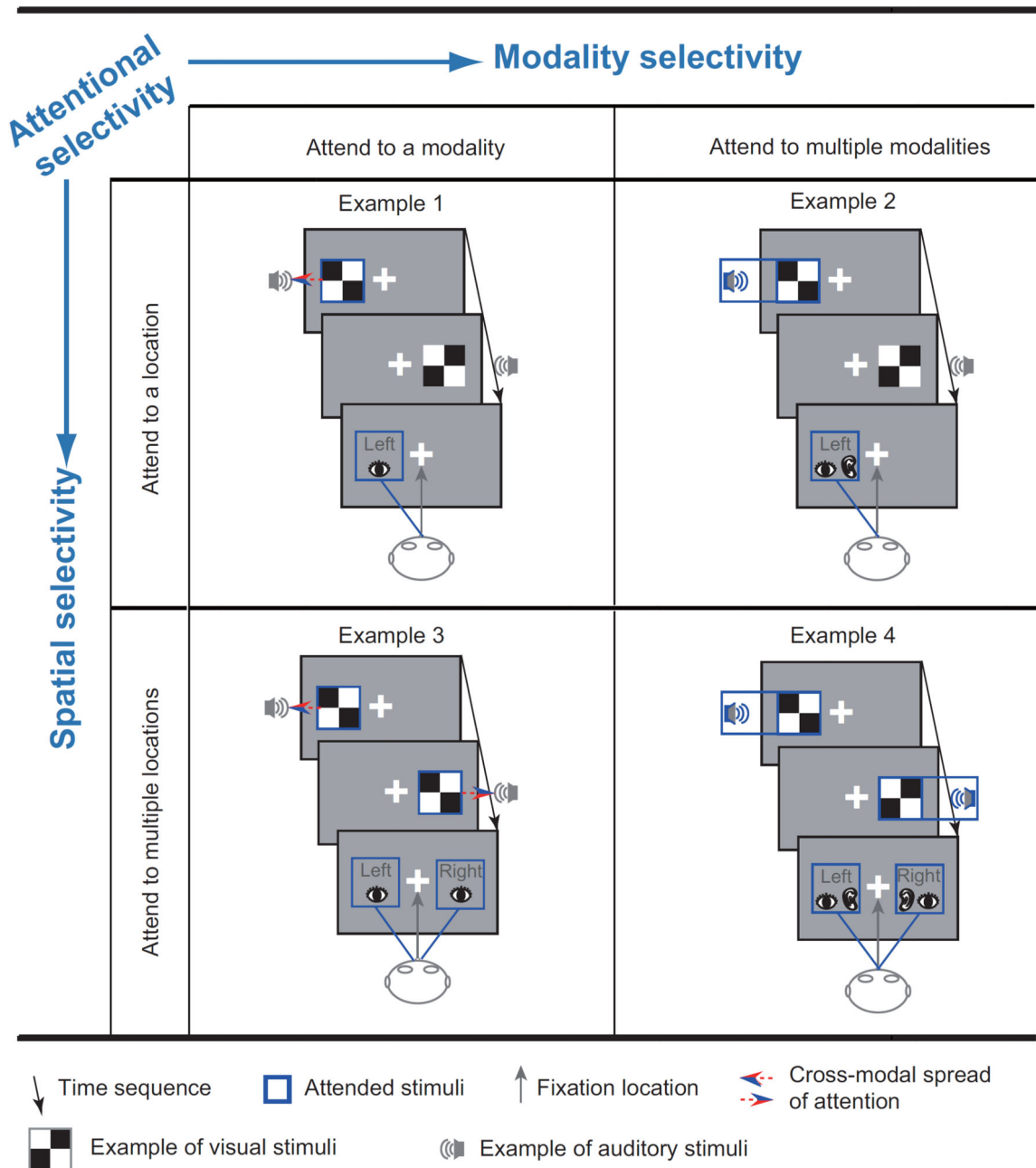


Figure 2. Effects of endogenous attentional selectivity on multisensory processing. Endogenous attention can modulate multisensory performance improvements through spatial or modality selectivity. These two types of attentional selectivity can interact. Here, we list 4 examples of attentional selectivity. **Example 1**, attend to a modality and attend to a location: participants are asked to pay attention to a modality-specific stimuli at a specific location, e.g., they are asked to attend to the left visual stimuli while ignoring all of the auditory stimuli and the right visual stimuli. Consequently, multisensory integration at the attended

Author Manuscript

Author Manuscript

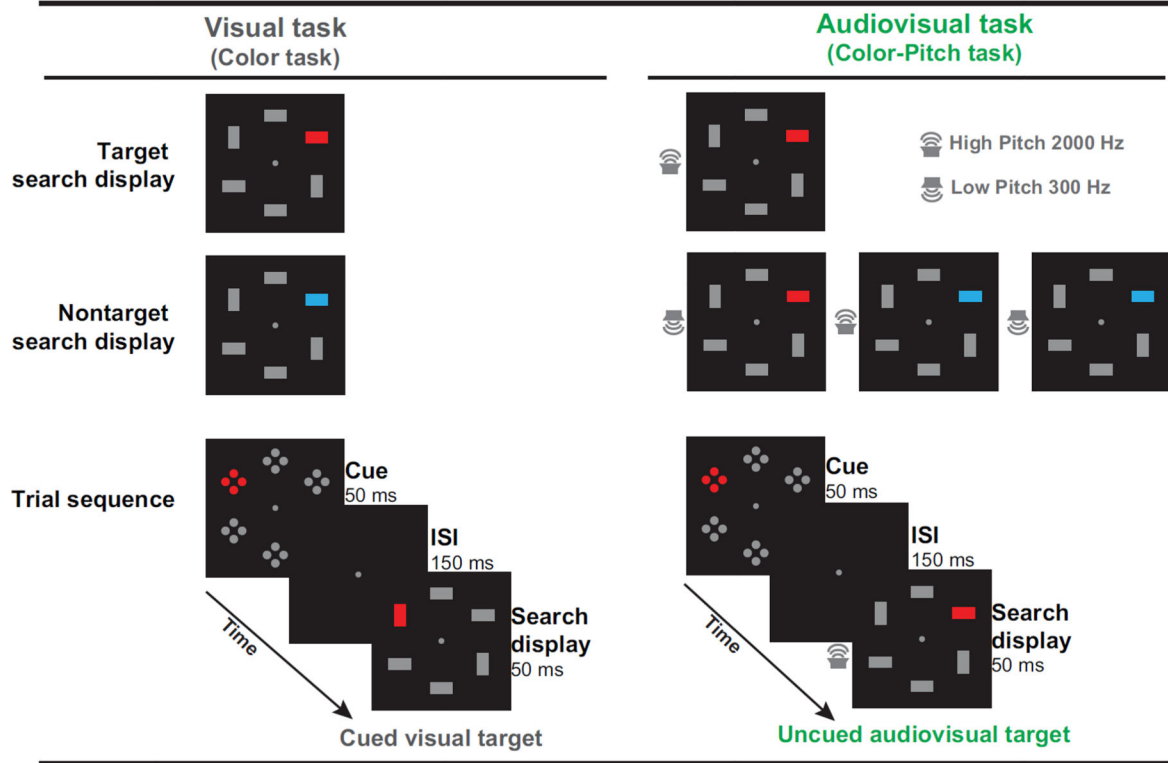
Author Manuscript

Author Manuscript

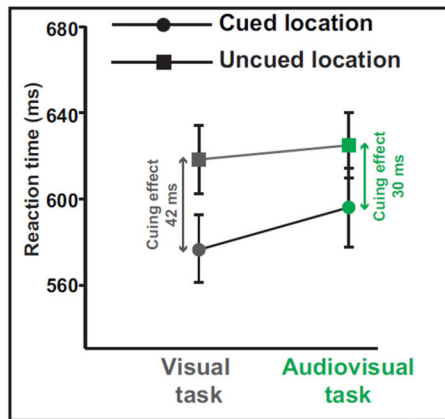
location is more intensive than that at the unattended location (Fairhall and Macaluso, 2009). Although all of the auditory stimuli are ignored, the attention that is directed to the visual stimuli at the attended location can spread to auditory stimuli that are simultaneously presented at the attended location and even at the central location (Busse et al., 2005).

Example 2, attend to multiple modalities and attend to a location: participants are asked to pay attention to stimuli from multiple modalities that are simultaneously presented at a specific location; for example, they are asked to attend to the left visual and auditory stimuli while ignoring all of the stimuli that are presented at the unattended location. Multisensory integration at the attended location has been found to be more intensive than that at the unattended location (Senkowski et al., 2005). **Example 3, attend to a modality and attend to multiple locations:** participants are asked to attend to stimuli in a specific modality at multiple locations; for example, they are asked to attend to visual stimuli while ignoring auditory stimuli regardless of the location of presentation. Consequently, responses to audiovisual stimuli are faster than those to visual stimuli, even though the participants are instructed to ignore the auditory stimuli (Santangelo et al., 2010). **Example 4, attend to multiple modalities and attend to multiple locations:** participants are asked to attend to stimuli in multiple modalities and at multiple locations; for example, they are asked to attend to both visual and auditory stimuli regardless of the location of presentation. Consequently, responses to audiovisual stimuli are faster than those to visual or auditory stimuli (Wu et al., 2012b). Notes: The stimuli illustrated here are only examples and do not depict the actual stimuli used in the previous studies. The tasks and results of the studies that are described in each example are listed in Table 1.

(a) Experiment design



(b) Behavioral results



(c) ERP results

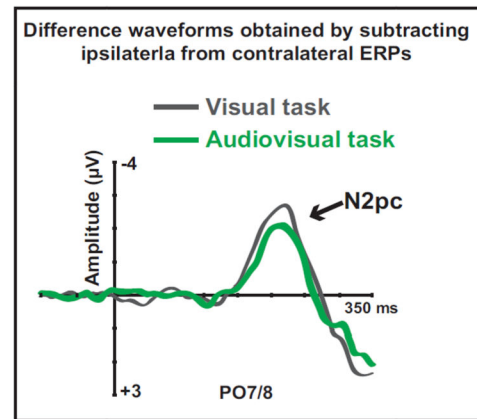


Figure 3.

Multisensory templates exert top-down control on contingent attentional capture. (a) The experiment design of the task and trial sequence. The target and nontarget search displays of the two tasks are illustrated. One is the visual task, in which participants were asked to discriminate whether the red bar was vertical or horizontal. The other is the audiovisual task, in which participants were asked to discriminate between vertical and horizontal red bars when they were accompanied by a high-pitched tone (illustrated here as a 2000 Hz tone). Thus, the blue bars or low pitches (the nontarget search displays) were to be ignored. Each

trial began with the cue array, which was composed of six elements, each consisting of four closely aligned dots. One element was a color singleton that matched the target color (illustrated here as “red”). The red singleton was presented randomly and with equal probability at one of the four lateral locations but never at the top or bottom locations. The visual target (the red vertical or horizontal bar) and the visual nontarget (the blue vertical or horizontal bar) were presented in the same manner as the cue. In the cued trials, the visual target or nontarget was presented at the ipsilateral (same) side as the cued singleton, while in the uncued trials, the visual target or nontarget was presented on the contralateral (opposite) side. **(b)** Behavioral results. Spatial cueing effects, which were calculated by subtracting the reaction time for the cued targets from that for the uncued targets, were found in both the visual and audiovisual tasks. More interestingly, the amplitude of the spatial cueing effect in the visual task was larger than that in the audiovisual task. **(c)** ERP results. The grand average ERP measured at the posterior electrodes PO7/8 contralateral and ipsilateral to the location of a target-color singleton cue. The difference waveforms that were obtained by subtracting the ipsilateral from the contralateral ERPs are illustrated separately for the visual (gray) and audiovisual tasks (green). The N2pc is marked and is an enhanced negativity that emerges approximately 200 ms after the onset of the target-color singleton cue. The results revealed that the amplitude of the N2pc component was larger in the visual task than in the audiovisual task. Adapted with permission from the corresponding author (Matusz and Eimer, 2013). Copyright © 2013 Society for Psychophysiological Research.

Effect of multisensory integration on exogenous attention

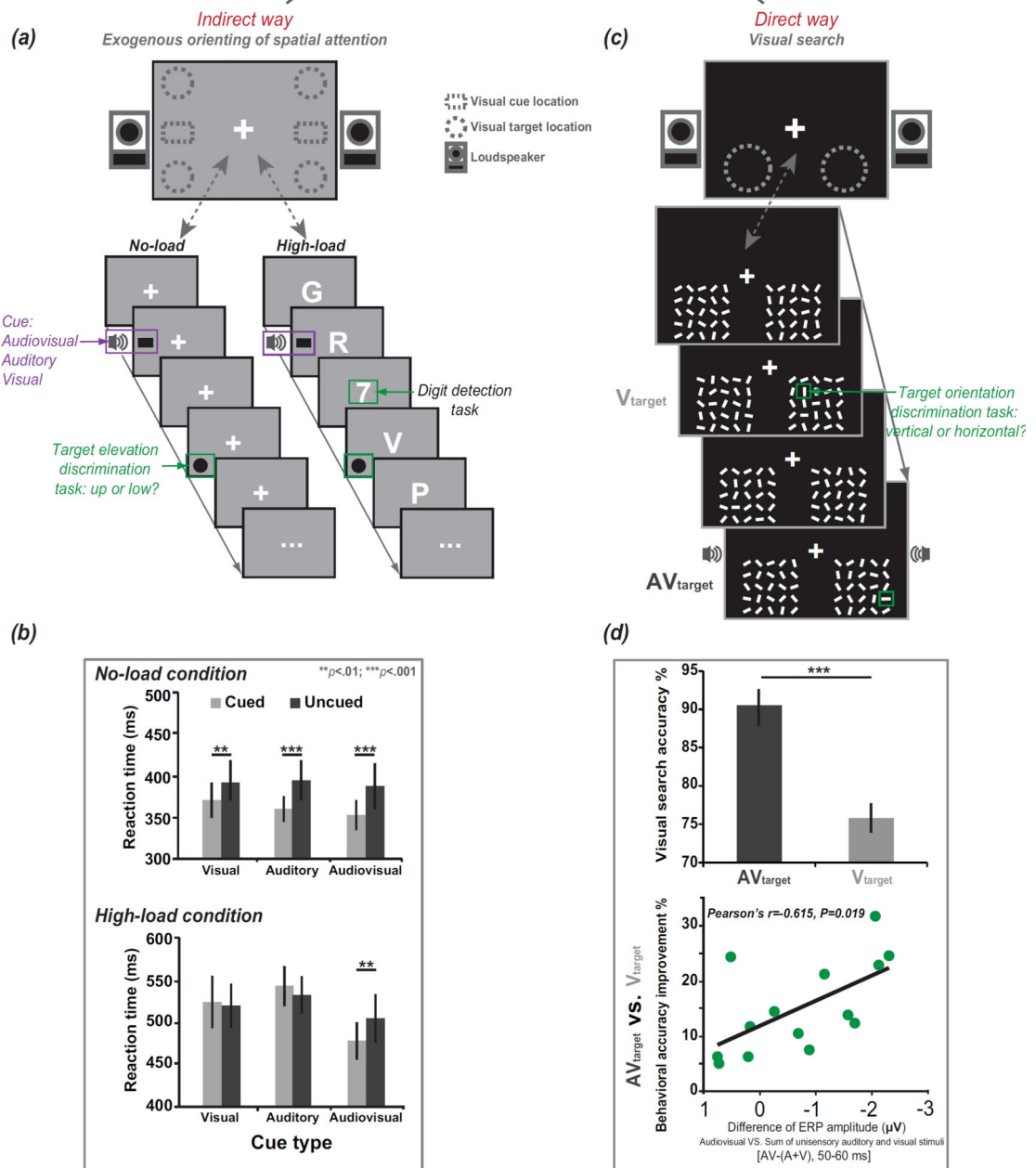


Figure 4. Effects of multisensory integration on exogenous attention. Multisensory integration acts on exogenous attention indirectly or directly. The indirect manner (a) can be observed when the exogenous cueing paradigm and an attentional/perceptual load are applied. As illustrated here, a non-predictive peripheral cue appears; this cue consists of the presentation of auditory stimuli from the two speakers located to the left or right of the monitor, the presentation of visual stimuli within the dashed squares on the monitor, or presentation of audiovisual stimuli. The target is presented in one of the corners of the display (as indicated

by the dashed circles). In the no-load condition (a: left panel), the participants were asked to complete a target elevation discrimination task, i.e., to report whether the target appeared at the top or bottom of the screen. In the high-load condition (a: right panel), the participants were asked to complete not only the target elevation discrimination task but also a center RSVP task in which they were required to detect a digit among distractor letters. **(b)** Consequently, in the no-load condition, all types of cues can capture attention and elicit significant spatial cueing effects; however, in the high-load condition, only the audiovisual cue elicits a significant spatial cueing effect (Santangelo and Spence, 2007). The direct manner **(c)** can be observed when the visual search paradigm is applied. In this paradigm, visual search displays were presented in two dashed circles. The distractor lines changed orientation, and one of them changed into the target line, i.e., the vertical or horizontal line. The participants were asked to discriminate the orientation of the target, i.e., vertical or horizontal. The visual target orientation change was accompanied (AV) or not accompanied (V) by an irrelevant auditory stimulus. **(d)** Consequently (Van der Burg et al., 2011), the responses in the AV condition were found to be more accurate than those in the V condition, and the ERP amplitude elicited in the AV condition differed from the sum of those elicited by the unimodal auditory and visual stimuli (A+V). Further, the value of $[AV-(A+V)]$ that was calculated during the 50–60 ms post-stimulus epoch was significantly correlated ($p<.05$) with the improvements in behavioral accuracy (AV vs. V). Adapted with permission from the corresponding authors (Santangelo and Spence, 2007) [Copyright 2007 by the American Psychological Association] and (Van der Burg et al., 2011) [© 2010 Elsevier Inc. All rights reserved.]

(a) Experiment design



(b) Behavioral and ERP results

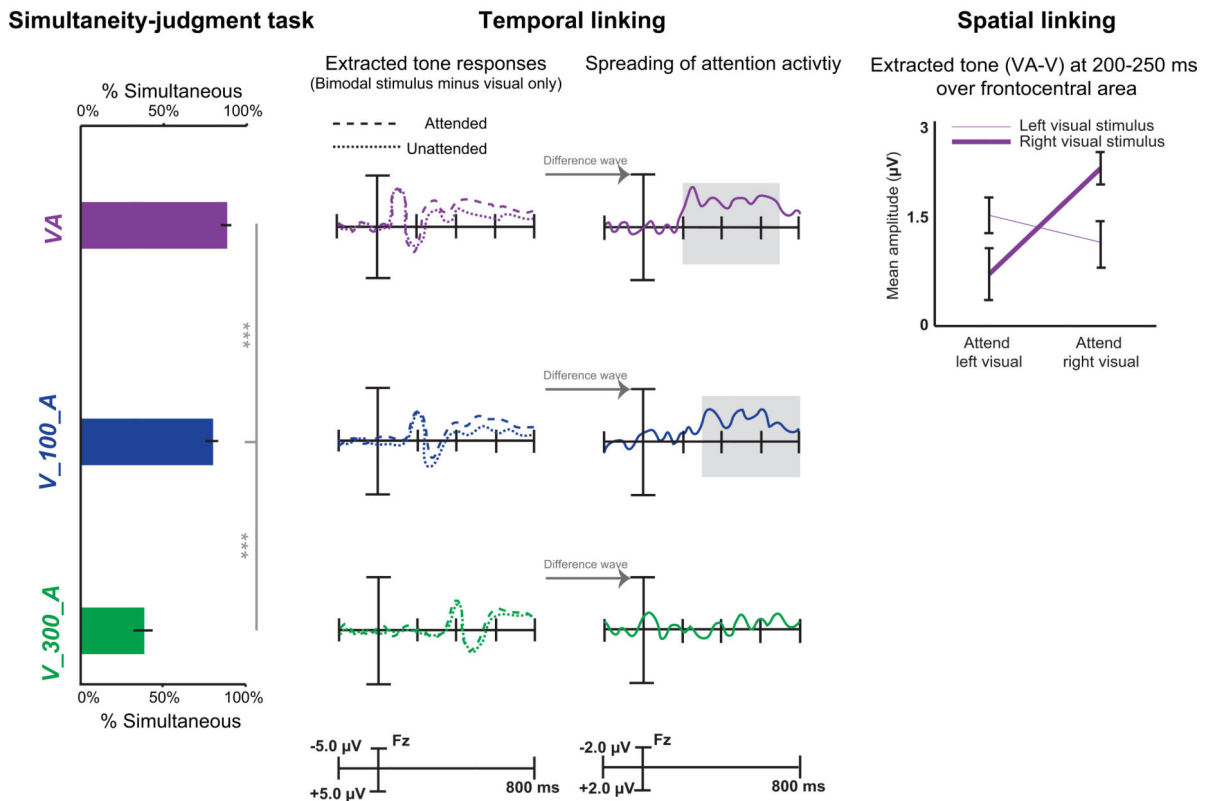


Figure 5. Temporal and spatial constraints on the cross-modal spread of attention. **(a)** Experimental design and types of stimuli. Visual stimuli were presented in the left or right peripheral square, whereas auditory stimuli were presented centrally. The visual target stimulus consisted of a checkerboard containing two dots. The attended locations are marked with the blue circle. The participants were asked to detect the target stimulus that was presented at the attended locations, which is illustrated here as the right side. There were four types of stimuli, including a visual stimulus only (V), a visual stimulus simultaneous with an

auditory tone (VA), a visual stimulus with a tone delayed by 100 ms (V_100_A) and a visual stimulus with a tone delayed by 300 ms (V_300_A). **(b)** The behavioral and ERP results for conditions that correspond to different temporal gaps between the visual and auditory stimulus are illustrated. The behavioral simultaneity-judgment task revealed that the subjects were much more likely to judge the visual and auditory stimuli as occurring simultaneously when the two stimuli were presented simultaneously (VA) or with a temporal gap of 100 ms rather than with a temporal gap of 300 ms. Regarding the ERP results, the tone responses were extracted by subtracting the response to the visual-only stimulus from that for the combination of the visual and auditory stimuli in either the attended or unattended location. Differences in the extracted tone responses between the attended and unattended locations were found over fronto-central areas using a time window of 200–700 ms or 300–800 ms in the VA and V_100_A conditions but not in the V_300_A condition. Furthermore, the contra-laterality of the spreading-of-attention effect was observed only in the VA condition. Specifically, the mean amplitude of the extracted tone (VA-V) response over the fronto-central area during the 200–250 ms time window exhibited the interaction between the attended side and the presented location of the visual stimulus. Adapted with permission from the corresponding author (Donohue et al., 2011) [Copyright © 2011 the authors 0270-6474/11/317982- 09\$15.00/0]

Cross-modal spread of attention in a multisensory object

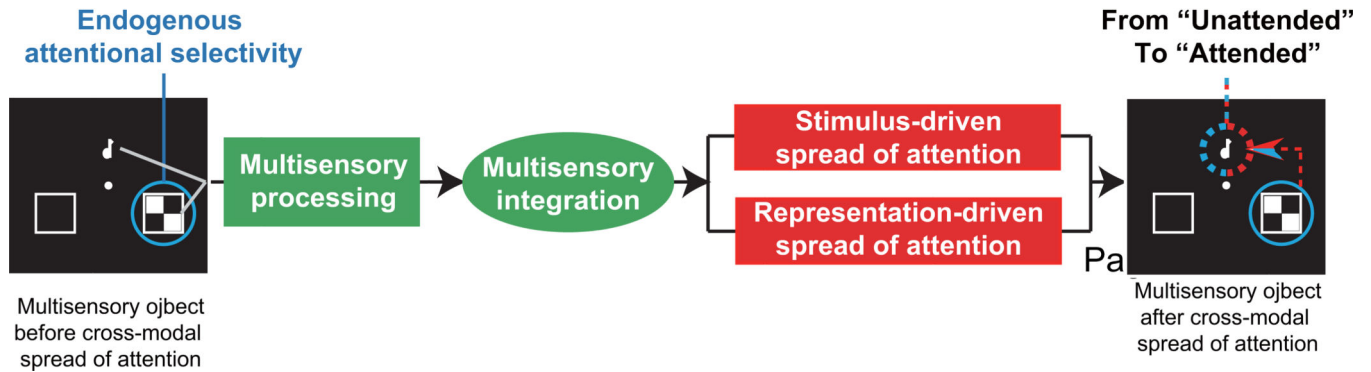


Figure 6.

Processes of the cross-modal spread of attention within a multisensory object. As illustrated here, attention is focused on the visual modality to the right side by endogenous attentional selectivity. When the visual and auditory stimuli are presented simultaneously, they are processed in a multisensory manner. After low- and high-level multisensory processing, these stimuli are integrated into a coherent multisensory object. Within this multisensory object, attention can spread from the attended visual stimuli to the ignored auditory stimuli across modalities and locations, which occurs automatically. Moreover, the process of attentional spread across modalities and space involves dual mechanisms (Fiebelkorn et al., 2010). One mechanism is the stimulus-driven spread of attention, which is affected by spatial or temporal links between the auditory and visual stimuli (Donohue et al., 2011). The other mechanism is the representation-driven spread of attention, which is modulated by congruency when the multisensory stimuli must be checked in terms of matching or congruency (Zimmer et al., 2010a; Zimmer et al., 2010b). After these processing stages, the ignored auditory stimulus acquires attention from the attended visual stimuli. This entire process consists of endogenous attentional selectivity and the exogenous cross-modal spread of attention.

A selection of important results from studies related to the modulation of multisensory integration by spatial- or modality-selective endogenous attention

Table 1

Study	Task	Attended modality	Attended location	Behavioral integration effect	Correlated brain area	Time (ms)
Example 1 (Attend to a modality and attend to a location)						
Fairhall & Macaluso, 2009	Go/no-go	V	L/R	ACC*	Attended $AV_{Con} > Attended AV_{Inc}$ Superior temporal sulcus V1, V2, fusiform gyrus	—
Li et al., 2010	Go/no-go	A	L/R	RT**** ACC****	Centro-medial Right temporal	280-300 300-320
Example 2 (Attend to multiple modalities and attend to a location)						
Talsma & Woldorff, 2005	Go/no-go	A&V	L/R	RT**** ACC****	Attended > Unattended Frontal Centro-medial	100-140 100-140 160-200 320-420 40-60
Senkowski et al., 2005	Go/no-go	A&V	L/R	RT**** ACC****	Medial-frontal	40-60
Example 3 (Attend to a modality and attend to multiple locations)						
Frassinetti & Bolognini, 2002	Go/no-go	V	L&R	Spatial coincidence $d'***$	—	—
Gao et al., 2014	Go/no-go	V	L&R	RT*/ $d'*$	Occipital Fronto-central	180-200 300-320
Example 4 (Attend to multiple modalities and attend to multiple locations)						
Teder-Salejari et al., 2005	Go/no-go	A&V	L&R	RT****	Spatial Con. vs. Inc. Common: Ventral Occipito-temporal Superior temporal Different:	160-220 220-300

Study	Task	Attended modality	Attended location	Behavioral integration effect	Correlated brain area	Time (ms)	
Wu et al., 2012	Go/no-go	A&V	L&R	RT*	Ventral Occipito-temporal	100–400	
					Superior temporal	260–280	
Attend to a modality vs. Attend to multiple modalities							
Talsma et al., 2007	Detection	A&V	Center	RT***	Fronto-central	P50	
					Fronto-central	ACC*	NI
						RT***	420–600
						ACC***	—
Degerman et al., 2007	Go/no-go	A&V	Center	ACC:	<i>AV task > AV task</i>	—	
				A	<i>AV task < A task</i> **	Superior temporal cortex	
Mozolic et al., 2008	Discrimination	A&V	Center	Race modal violation	—	342–426	
					A	Null	
					V	Null	
					A&V	—	
Magnée et al., 2011	Detection	A&V	Center	—	Occipital-temporal	130–210	
					Fronto-central	150–230	
					Fronto-central	150–230	
Modality selectivity vs. Spatial selectivity							
Santangelo et al., 2010	Go/no-go	A/V	L/R	RT:	Fronto-parietal	—	
					Precuneus	A&V>A/V	
						L&R>L/R	
						L&R(A&V-A/V)<L/R(A&V-A/V)	

Notes: A-auditory; V-visual; AV-audiovisual; L-left; R-right; RT-reaction time; ACC-accuracy;

* $p < .05$,

** $p < .01$,

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

 $p < .001$;

Con. and *Inc.* are short for congruent and incongruent conditions, respectively. d' indicates the sensitivity index of signal detection theory; "Attended>Unattended" indicates a greater multisensory integration effect in the attended condition than in the unattended condition; "Null" indicates that no enhancement was found; "Race modal violation" indicates that the cumulative distribution function of the multimodal stimuli was steeper than that predicted by unimodal stimuli. The examples for the 4 combinations of spatial and modality selectivity are illustrated in Figure 2: Example 1, 2, 3 and 4.

Table 2

A selection of important results from studies related to the modulation of the spatial cueing effect by multisensory integration in the exogenous cueing paradigm

Study	Task	Cue modality (SOA/ms)	Target modality (Exp.)	Effect size/ms
Mahoney et al., 2012	Detection task	AV (400)	V	52 ^{**}
		V/A (400)		67 ^{**}
Barrett & Krumbholz, 2012	Temporal order judgment task	AV (200)	V/A	42 ^{***} /25
		V/A (200)		44 ^{***} /25
Matusz & Eimer, 2011	Detection task	AV (200)	V (Exp. 1)	36 ^{***}
		V (200)		29 ^{***}
		AV (200)	V (Exp. 2)	20 ^{***}
		V (200)		11 ^{***}
Santangelo, Ho, et al., 2008	Discrimination	AT (233)	V (Exp. 1 No-load)	35 ^{***}
		A (233)		26 ^{***}
		T (233)		31 ^{***}
		AT (233)	V (Exp. 1 High-load)	35 ^{***}
		A (233)		-5
		T (233)		6
Santangelo & Spence, 2007	Discrimination	AV (233)	V (Exp. 1)	15 ^{***}
		AV (233)	V (Exp. 2 No-load)	31 ^{***}
		A (233)		36 ^{***}
		V (233)		23 ^{**}
		AV (233)	V (Exp. 2 High-load)	27 ^{***}
		A (233)		-10
		V (233)		-3
Santangelo et al., 2006	Discrimination	AV (200/400/600)	V (Exp. 1)	21 [*] /5/-20 [*]
		A (200/400/600)		13 [*] /14 [*] /9 [*]
		V (200/400/600)		16 [*] /19 [*] /4
		AV (200/600)	V (Exp. 2)	15 [*] /2
		A (200/600)		13 [*] /4
		V (200/600)		19 [*] /1/

Notes: A-auditory; V-visual; T-tactile; AT-audiotactile; AV-audiovisual; SOA (i.e., stimulus onset asynchrony) represents the time interval from the cue onset to the target onset. "Effect size/ms" was obtained by subtracting the mean reaction time at the cued location from that at the uncued location. High load indicates paradigms containing dual tasks. The cued location indicates that the target appeared at the same location as the cue, while the uncued location indicates that the target appeared at the opposite location from the cue.

* $p < .05$,

**
 $p < .01$,

 $p < .001$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript