

Attention and the multiple stages of multisensory integration: A review of audiovisual studies

Thomas Koelewijn^{a,b,*}, Adelbert Bronkhorst^{a,c}, Jan Theeuwes^a

^a Cognitive Psychology, Vrije Universiteit, Amsterdam, The Netherlands

^b ENT Audiology, VU University medical center, Amsterdam, The Netherlands

^c TNO Human Factors, Soesterberg, The Netherlands

ARTICLE INFO

Article history:

Received 10 December 2009

Received in revised form 23 March 2010

Accepted 27 March 2010

Available online 27 April 2010

Keywords:

Multisensory integration

Crossmodal

Attention

Audiovisual

ABSTRACT

Multisensory integration and crossmodal attention have a large impact on how we perceive the world. Therefore, it is important to know under what circumstances these processes take place and how they affect our performance. So far, no consensus has been reached on whether multisensory integration and crossmodal attention operate independently and whether they represent truly automatic processes. This review describes the constraints under which multisensory integration and crossmodal attention occur and in what brain areas these processes take place. Some studies suggest that multisensory integration and crossmodal attention take place in higher heteromodal brain areas, while others show the involvement of early sensory specific areas. Additionally, the current literature suggests that multisensory integration and attention interact depending on what processing level integration takes place. To shed light on this issue, different frameworks regarding the level at which multisensory interactions takes place are discussed. Finally, this review focuses on the question whether audiovisual interactions and crossmodal attention in particular are automatic processes. Recent studies suggest that this is not always the case. Overall, this review provides evidence for a parallel processing framework suggesting that both multisensory integration and attentional processes take place and can interact at multiple stages in the brain.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

When you are reading a newspaper on a train the sound of loud music to your left or someone talking on the phone to your right can be distracting. You may skip a line, misread a header, or even stop reading when a conversation behind you grasps your attention. Why is it so hard to stay focused on your readings when you hear sounds around you? Why can't you block out these sounds when you know that they are irrelevant? Although distracting when you try to read, these interactions between what we hear and what we see can save your life – for example when the sound of a car coming from your left makes you freeze.

These everyday examples illustrate the strong interactions that exist between our auditory and visual systems. These interactions can occur at the level of 'multisensory integration' (see [Stein & Stanford, 2008](#)), as when a voice and a moving mouth are integrated into a single event (e.g., [McGurk & MacDonald, 1976](#)). Multisensory integration helps us perceive information better, which might be why it is so tempting to look over our newspaper when eavesdrop-

ping on a conversation between two people sitting opposite in the train. Additionally, these interactions can be at an attentional level (see [Driver & Spence, 1998](#)) in which, for example, a sound draws our visual attention to a certain location (e.g., [Spence & Driver, 1997](#)). This might be why it is so hard to focus our attention on the words in the newspaper in front of us when someone is snapping chewing gum next to us.

Early studies on perceptual and attentional processes primarily investigated sensory modalities in isolation. However, in the last two decades or so more research has addressed the interaction between modalities. This allows us to get a full picture of how these processes work in the brain, but also to relate these outcomes to more realistic situations in which auditory and visual events hardly ever occur in isolation. With current technology developments the question of when to expect audiovisual interactions becomes more pressing than ever. For instance, in-car technologies like navigational systems overflow us with audiovisual information. The impact of sounds on our driving ability, which is primarily a visual task, has become a hot research topic (see [Ho & Spence, 2005](#); [Spence & Ho, 2008](#)).

Audiovisual interactions may allow us to focus on relevant information and filter out irrelevant information, or may cause distraction when our attention is captured against our will by audiovisual information that is irrelevant for our task. We speak of

* Corresponding author. ENT Audiology, VU University medical center, Boelelaan 1117, 1081HV Amsterdam, The Netherlands. Tel.: +31 20 44 40900.

E-mail address: t.koelewijn@vumc.nl (T. Koelewijn).

attentional capture when spatial attention is drawn to a location in space against our intentions (Theeuwes, Belopolsky, & Olivers, 2009). For example, even though our goal may be to read a book, our attention may get drawn to the location where a person is making a sound. The question that is central in this review is whether visual attention that is voluntarily directed to a specific spatial location can be drawn away automatically from that location towards the location where a sound is coming from. Even though previous studies have shown that attentional capture can occur between the different modalities (e.g., Spence & Driver, 1997), the question remains whether a localizable sound captures visual spatial attention (cross-modal capture) under all circumstances. Recent studies have shown that in some circumstances audiovisual interactions like crossmodal capture do not occur (Koelewijn, Bronkhorst, & Theeuwes, 2009b; Santangelo & Spence, 2007) while other studies have shown that in most circumstances irrelevant sounds do affect our visual system (e.g., Koelewijn, Bronkhorst, & Theeuwes, 2009a; Mazza, Turatto, Rossi, & Umiltà, 2007; McDonald, Teder-Salejarvi, & Hillyard, 2000; Spence & Driver, 1997; van der Lubbe & Postma, 2005; Ward, 1994). This review addresses the question under what circumstances crossmodal capture occurs. Additionally, recent research has shown that multisensory integration and (crossmodal) attention interact at certain brain levels (e.g., Fairhall & Macaluso, 2009; Mozolic, Hugenschmidt, Peiffer & Laurienti, 2008; Talsma, Doty, & Woldorff, 2005; Talsma, Doty, & Woldorff, 2007; Talsma & Woldorff, 2005). This review also addresses the levels at which these interactions may occur.

In addition to vision and audition, multimodal interactions are also known to occur between taste, smell, and touch senses (e.g., see Driver & Spence, 1998; Stein & Stanford, 2008). So far, most research has been directed at the interactions between our visual, auditory, and somatosensory systems and has been focused on interactions at an attentional level or at a multisensory integration level. This review will focus on studies mainly discussing interactions between the visual and auditory modality, although sometimes a reference will be made to somatosensory studies to illustrate that effects apply more generally.

Although our perceptual systems seem fully integrated, modality specific features tend not to interact, as shown by Alais, Morrone and Burr (2006) for auditory pitch and visual contrast perception. However, there is a form of interaction called synaesthesia where non-overlapping features between modalities do integrate. For example Baron-Cohen, Wyke and Binnie (1987) have shown that some people see colors when hearing numbers which seems to imply some form of multimodal interaction. However, Rouw and Scholte (2007) have shown that the structure of the brain of those people that experience synaesthesia may be different from those that do not experience synaesthesia, suggesting that the occurrence of synaesthesia and its implied multimodal interaction is not a general phenomenon.

This paper reviews studies that investigated audiovisual interactions in the form of multisensory integration and crossmodal attention. Both types of interactions take place at multiple processing levels within the brain. The first section describes the processing levels at which information from the auditory and visual modalities meet and integrate. This is followed by a review of studies that specifically look at attentional capture across the auditory and visual modalities. The section that follows introduces the idea that multisensory integration and crossmodal attention sometimes act independently, and at other times interact. To shed light on this issue, different frameworks regarding the level at which multisensory interactions take place are discussed. The final section focuses on the question of whether audiovisual interactions and crossmodal attention are automatic processes. The literature shows that crossmodal attention does not always meet the criteria for automaticity. One possibility is that these findings can be explained in terms of parallel processing. Both behavioral and electrophysiological studies will be

discussed to provide a full picture of the current status on this topic. The present paper presents a broad overview of studies regarding audiovisual integration and attention.

2. Multisensory integration

We need multisensory integration in order to recognize different types of sensory input as belonging to the same object. Multisensory integration helps to reduce noise within our perceptual system by combining information from different sensory modalities (see Stein, Stanford, Wallace, & Jiang, 2004). Less noisy input allows for an easy separation of events from background noise and division between successive events. For example a sound can boost the detectability of visual events (see Noesselt, Bergmann, Hake, Heinze, & Fendrich, 2008). Even though some multimodal behavioral effects and illusions resulting from multisensory integration were reported as early as the 1960's and 1970's (e.g., Hershenson, 1962; McGurk & MacDonald, 1976), research on multisensory integration has skyrocketed in the last two decades. Psychophysical studies have demonstrated that the notion that sensory information is processed within each modality separately in a feedforward fashion is incorrect (see Driver & Spence, 2000). In addition, animal physiology (see Stein & Stanford, 2008), human electrophysiology (Talsma et al., 2007) and human imaging studies (Calvert, Campbell, & Brammer, 2000) have provided evidence that multisensory integration is not restricted to higher multisensory (heteromodal) brain areas (see Macaluso & Driver, 2005). This section discusses under what circumstances and where in the brain multisensory integration takes place. First, some multisensory illusions and effects will be discussed to illustrate the strength of multisensory integration.

2.1. Multisensory integration effects and illusions

Although multisensory integration is the process that binds information from different modalities, most of the time you are not aware of its occurrence. Still, there are some multisensory integration effects or illusions of which we can become consciously aware. Ventriloquism (Thurlow & Jack, 1973) is a well-known example. In this illusion, the voice of the puppeteer seems to project from the mouth of the puppet itself. This attribution of voices to congruent sources is generally beneficial and results in improved perception under noisy circumstances (Sumbly & Pollack, 1954).

Ventriloquism is most commonly demonstrated in the shift of sound toward the location of a visual event. In the puppet illusion, sound is shifted toward a congruent source, but Slutsky and Recanzone (2001) demonstrated that ventriloquism also occurs with simple auditory and visual onsets that have no semantic value. The same study showed that there are spatial and temporal constraints to the ventriloquism effect. This means that these events should take place not too far apart in space and preferably should co-occur in time. Temporal and spatial restrictions generally apply to multisensory integration and will be discussed in the next section. The ventriloquism effect suggests that the visual system is dominant over the auditory system when it comes to spatial localization. However, other illusions that are discussed below demonstrate that this is not always the case.

Ventriloquism can also pull sensory events together in terms of time, such that the perceived temporal proximity of two successive visual events is affected by auditory input. For example, in Morein-Zamir, Soto-Faraco, and Kingstone (2003) participants performed a temporal order judgment task on the onsets of two LEDs. When a sound was presented before the first onset and after the second onset, compared to a neutral condition in which the sound coincided with the LED onsets, the participants' performance benefitted. It seemed as if the visual onset was pulled in time towards the auditory onsets, which made temporal order judgment of the visual events easier.

Ventriloquism and temporal ventriloquism show that one modality can bias another modality in the spatial and temporal domain. These effects suggest that the auditory modality is dominant in the temporal domain (Morein-Zamir et al., 2003) and the visual modality is dominant in the spatial domain (Slutsky and Recanzone, 2001).

Multisensory integration does not only result in a spatial or temporal bias but can also create illusory effects. Shams, Kamitani, and Shimojo (2000) showed that when a single visual flash is accompanied by multiple short auditory events in the form of beeps, the visual event is perceived as multiple flashes. In a follow-up study, Shams, Kamitani, and Shimojo (2002) showed that this illusion only occurs when two beeps are presented within a time window of 100 ms before or after the onset of the flash, which is a characteristic temporal constraint for multisensory integration.

Temporal ventriloquism shows that sound biases visual temporal perception (Morein-Zamir et al., 2003). However, sound can also boost the detectability of a visual event (e.g., Frassinetti, Bolognini, & Ladavas, 2002) and this boost in visual detectability or salience can affect temporal search (Vroomen & de Gelder, 2000). Vroomen and de Gelder (2000) have shown that sound can enhance visual temporal search. In this study, participants had to detect a visual target that was presented within a rapid serial stream of distractors. At the onset of each visual event within the stream a low pitch tone was presented, except for one condition in which a high pitch tone was presented together with the target. Under the latter conditions performance of the participants improved. The authors named this effect the 'freezing effect' because some participants reported that the target seemed to stay on screen longer than the distractors, as if the target image froze for a while.

Multisensory integration is not only helpful in separating successive events, but can also enhance visual spatial search. This has been demonstrated by Van der Burg, Olivers, Bronkhorst, and Theeuwes (2008). In this study, participants had to search for a vertical or horizontal target line segment in-between diagonal line distractors. Both target and distractors changed color (red or green) randomly over time but when the target changed color it was the only element changing color at that moment. The performance on this task resembled that from other serial search tasks showing an increase of search time when the number of distractors was increased. However, when a short sound (a pip) was presented at the onset of the color change of the target, the visual target popped out from the display as evidenced by search functions that were basically flat (i.e., no effect on search time of the number of distractors in the display). Van der Burg et al. (2008) furthermore showed that search performance was optimal when the pip was temporally aligned with the change of the visual target, and decreased when it was presented either earlier or later in time. In a follow-up study in which the time course of the processes underlying the 'pip and pop' effect was investigated, Van der Burg, Talsma, Olivers, Hickey, & Theeuwes (submitted) demonstrated that this effect can be explained in terms of multisensory integration. They measured event-related potentials (ERPs) for stimuli which behaviorally induced the pip and pop effect and found a series of perceptual and attentional effects: First was an early multisensory response (50 ms post-stimulus), which was followed by a contralateral positivity (80–120 ms) suggesting a saliency boost of the multimodal event and an enhanced N2pc reflecting the application of attention to the target location (e.g., Hickey, McDonald, & Theeuwes, 2006). A large sustained posterior contralateral negativity component was also identified, reflecting encoding and maintenance of the target in visual short-term memory (e.g., Klaver, Talsma, Wijers, Heinze, & Mulder, 1999; Vogel & Machizawa, 2004), alongside an enlarged P3 component, reflecting updating in working memory (e.g., Nieuwenhuis, Aston-Jones, & Cohen, 2005). Overall these results indicate that the pip and pop effect can be explained by early multisensory integration, which boosts target saliency and captures attention.

To conclude, these experiments illustrate the strength of multisensory integration by showing that one modality can bias the other (e.g., Morein-Zamir et al., 2003; Slutsky & Recanzone, 2001), enhance the other (Van der Burg et al., 2008), or create strong illusory effects (e.g., Shams et al., 2000). Additionally, these studies show that these illusions or interactions only occur under particular temporal and spatial constraints.

2.2. Temporal and spatial constraints

Our perceptual system seems to effortlessly integrate co-occurring information from different modalities (Ernst & Bühlhoff, 2004). However, for multisensory integration to take place, it is often required that both events occur close in time and space (Bolognini, Frassinetti, Serino, & Ladavas, 2005; Frassinetti et al., 2002). Frassinetti et al. (2002) found an enhancement of the perceptual sensitivity for luminance detection by means of sound. By systematically varying the spatial and temporal proximity of the visual and auditory events, they showed that this enhancement only takes place when both visual and auditory events co-occur in time and space. A strong multisensory integration effect is obtained when the time window between the onsets of auditory and visual events is less than 100 ms (Meredith, Nemitz, & Stein, 1987). In their study Meredith et al. (1987) measured cells of the superior colliculus in the cat's brain. Their results show a clear decline of this integration effect when the time windows became progressively larger than 100 ms. A further increase in temporal disparity between an auditory and visual event could even cause these cells to become inhibited.

This narrow 100 ms time window is a distinct feature of multisensory integration, which sets it apart from attentional effects that can operate at much larger intervals. Moreover, multisensory integration has also a clearly different time window than the time window within which effects occur that are related to advance preparation and warning, also known as foreperiod effects. The foreperiod is the time interval between two successive events. When participants need to respond to the second event the first event can act as a warning cue. Even when this cue is neutral with respect to the target location (or other features) the foreperiod allows for the perceptual system to come in a preparatory state. This preparatory state enables faster responses to the target irrespective of other processes like multisensory integration or attentional effects (see e.g., Los & Schut, 2008; Niemi & Näätänen, 1981). While multisensory integration is at its maximum when events co-occur in time (Meredith et al., 1987), these preparatory effects are known to become larger when the foreperiod becomes larger (Niemi & Näätänen, 1981). Therefore, most studies on multisensory integration control for these foreperiod effects by including different time intervals (e.g., Shams et al., 2002; Van der Burg et al., 2008). In this way, effects that are due to true multisensory integration can be distinguished from preparatory (alerting) effects.

Studies using near-threshold stimuli presented in the central visual field have shown that the location of the auditory cue is not always relevant for crossmodal integration to occur (Lippert, Logothetis, & Kayser, 2007; Noesselt et al., 2008; Stein, London, Wilkinson, & Price, 1996). For example, a study by Stein et al. (1996) showed that an auditory stimulus enhances perceived visual intensities. When both visual and auditory events were presented at peripheral locations, spatial proximity was essential for multisensory integration to occur. These enhancements were strongest at the lowest visual stimulus intensities. However, no spatial proximity was needed for these enhancements to occur at the centre of fixation. These results indicate that spatial constraints for multisensory integration only hold for peripheral visual events and not for central visual events. Although a study by Odegaard, Arie, and Marks (2003) suggests that the results shown by Stein et al. (1996) are based on a response bias rather than on multisensory integration, more recent

studies (Lippert et al., 2007; Noesselt et al., 2008) also support the idea that spatial constraints are not always apparent. For instance, Lippert et al. (2007) showed that a sound which is only temporally informative is sufficient to improve the detection of a centrally presented visual event. The participants perceived a low contrast target as being brighter when additional temporal information was provided by a sound that was presented from a different location than the visual event. This sensation was accompanied by a shift in the detection threshold. A recent study by Noesselt et al. (2008) showed that spatial alignment of an auditory event is also not necessary for crossmodal integration to occur during a central visual spatial discrimination task. Together, these results suggest that in order for multisensory integration to occur within the central visual field only temporal proximity between the auditory and visual events is necessary. For multisensory integration to occur in the peripheral visual field, both temporal and spatial proximity seem to be important. According to Stein et al. (1996), these results are a strong indication that multisensory integration occurs in many areas in the brain and some are likely involved in functions that do not require spatial information.

2.3. Neural correlates of multisensory integration

Since the late 1960's electrophysiological research within the animal brain has discovered neurons that respond to input from more than one modality. These *heteromodal* regions showed up in a number of brain areas (see Calvert & Thesen, 2004), including the superior temporal sulcus (Barraclough, Xiao, Baker, Oram, & Perrett, 2005; Benevento, Fallon, Davis, & Rezak, 1977; Bruce, Desimone, & Gross, 1981), the ventral and lateral intraparietal areas (Lewis & Van Essen, 2000; Linden, Grunewald, & Andersen, 1999), and sub-cortical areas like the superior colliculus (Meredith & Stein, 1996; Meredith et al., 1987; Wallace, Meredith, & Stein, 1998). Fig. 1 provides an anatomical overview of these areas. Note that areas like the ventral and lateral intraparietal areas are the monkey homologue of human brain areas.

Many studies have demonstrated that the superior temporal sulcus (STS) is involved in audiovisual integration (see Hein & Knight, 2008). A study by Barraclough et al. (2005) showed, for example, that

matching sights and sounds of actions, such as in ripping a sheet of paper, integrates in STS. Ghazanfar, Maier, Hoffman, and Logothetis (2005) showed that STS was involved in speech processing when monkeys observed dynamic faces and voices of other monkeys. Consistent with these findings, also in humans STS becomes active when processing multisensory speech information (e.g., Senkowski, Saint-Amour, Gruber, & Foxe, 2008).

The superior colliculus (SC) has a strong topographic organization and is known to be involved in saccadic eye movements. Although it receives input from the visual cortex together with many other cortical areas, the neurons within the superior colliculus also respond to somatosensory and auditory input (Meredith et al., 1987). The receptive fields of these different modalities overlap. Therefore, a sound or visual event presented at the same location will activate the same neuron (Meredith & Stein, 1996). Bimodal stimulation within the same receptive field will result in a super-additive neuronal response (Wallace et al., 1998). Not only spatial but also temporal proximity of multisensory input results in stronger neural activity (Meredith et al., 1987).

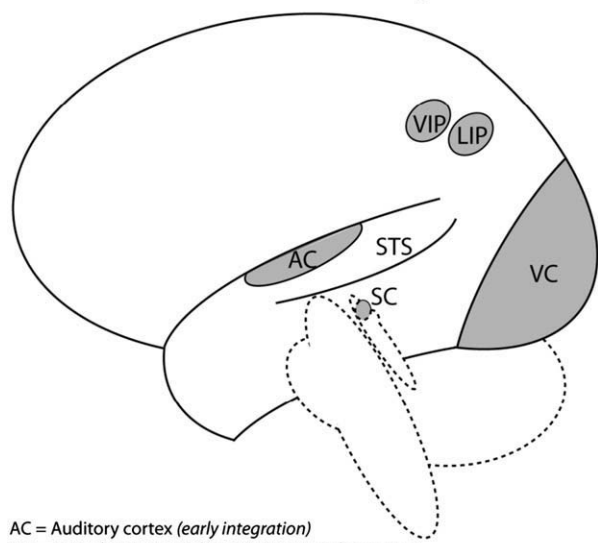
Areas that were long considered part of the unimodal visual cortex, like the lateral intraparietal area, are now known to also receive auditory input (Linden et al., 1999). Lateral intraparietal area neurons become active during the onset of a visual or auditory event and remain active during a delayed saccade response (see Colby & Goldberg, 1999). The neuronal activation in lateral intraparietal area is independent of whether or not an actual saccade is made towards the location of the event (Colby, Duhamel, & Goldberg, 1996). Because of this, the lateral intraparietal area is thought to be involved in visuospatial attention (Colby et al., 1996).

In addition to these heteromodal areas, supposedly unimodal areas like the primary visual cortex also respond to input from other modalities (e.g., Foxe, Morocz, Murray, Higgins, Javitt, & Schroeder, 2000; Martuzzi et al., 2007; Romei, Murray, Merabet, & Thut, 2007; Shams, Iwaki, Chawla, & Bhattacharya, 2005). For instance Shams, Kamitani, Thompson, and Shimojo (2001) showed that the illusory flash effect evokes almost similar event related potentials as the physical flashes do. This suggests that visual perceptual mechanisms can be affected by sound. A follow up study (Shams et al., 2005) confirmed that the sound that causes the illusory flash illusion affects occipital areas known for their unimodal visual processing (see also Mishra, Martinez, Sejnowski, & Hillyard, 2007). Latency differences in auditory and visual information processing may underlie these effects. The speed of cortical responses to auditory stimuli (15–20 ms) (Liegeois-Chauvel, Musolino, Badier, Marquis, & Chauvel, 1994) may allow auditory processes to influence feedforward visual processing (60–90 ms) (see Martinez et al., 1999; Romei et al., 2007). In addition, the primary auditory cortex can also be affected by visual (Romei et al., 2007) or somatosensory information (see Foxe et al., 2000; Ghazanfar and Schroeder, 2006).

2.4. Discussion

Multisensory integration takes place across multiple levels, including sub-cortical areas like the superior colliculus, early cortical areas like the primary visual and auditory cortices, and higher cortical areas like the superior temporal sulcus and intraparietal areas. Different types of illusions illustrate the occurrence of multisensory integration at various levels. For instance, the illusory flash illusion (Shams et al., 2000), the pip and pop effect (Van der Burg et al., 2008), or the freezing effect (Vroomen & de Gelder, 2000) in which, auditory temporal information is used to boost or create illusory visual onsets, seem to take place in the primary visual areas (Shams et al., 2005). Illusions like the McGurk effect (McGurk & MacDonald, 1976) most probably occur at higher cortical areas because of the more complex character of the information.

Brain areas involved in audiovisual integration



AC = Auditory cortex (*early integration*)
 LIP = Lateral intraparietal area (*visuospatial attention*)
 SC = Superior colliculus (*spatial integration*)
 STS = Superior temporal sulcus (*late integration*)
 VC = Visual cortex (*early temporal integration*)
 VIP = Ventral intraparietal area (*late integration*)

Fig. 1. Brain areas and their functional correlate that are involved in audiovisual integration.

The fact that multisensory integration can occur in a number of different brain areas at different processing stages raises the possibility for interactions with attention at different levels. Therefore, the idea of late integration stating that unimodal attention affects the individual sensory input and integrates at a heteromodal level seems incomplete. Many studies show early pre-attentive integration in primary sensory areas (e.g., Shams et al., 2001; Shams et al., 2005), which suggests multisensory integration at multiple levels. In the next section we discuss crossmodal spatial attention, the neural correlates of crossmodal attention, and how crossmodal attention interacts with multisensory integration.

3. Crossmodal attention

Selective attention is the mechanism that allows us to focus on important input while ignoring unimportant events. Attention can be directed to a location in space, to a moment in time or to non-spatial features such as the color of a visual stimulus or the pitch of a sound. It is possible to direct attention to the auditory domain while ignoring the visual and vice versa. The attentional processing can occur in a bottom-up (exogenous) manner, for instance when a salient event pops-out from its background. In this case, an object gets selected even though the observer was not planning to select it. In other cases, attentional processing operates in a top-down (endogenous) manner in which the observer voluntarily controls what is attended and what not.

By directing our attention to a particular moment in time we are able to anticipate an upcoming event (e.g., Coull & Nobre, 1998; Kingstone, 1992). In a study by Coull and Nobre (1998), participants had to detect a target onset as fast as possible. The time interval before the onset could be long (1500 ms) or short (300 ms). At the beginning of a trial, an endogenous cue was presented indicating the upcoming interval duration with 80% validity. Results showed behavioral costs for invalidly cued intervals but only for the short interval. The reason that no cueing was shown at long intervals probably had to do with the fact that omission of the short interval target guaranteed a long interval target. Therefore, participants could reorient temporal attention to the long interval, which did not result in costs. Dalton and Lavie (2006) show that attention can also be captured (exogenously) by a color singleton in a temporal search task. In this study, participants had to search for targets that were slightly larger or smaller than the distractors presented in a rapid serial visual presentation. When the distractor before or after the target was colored red, participants responded slower to the target. Similar temporal capture effects were shown in the auditory domain. Dalton and Lavie (2004) presented sound sequences with targets that differed from the distractors in frequency, intensity, or duration. A singleton distractor sound that was easily discriminated from other events in the sequence was either present (before, after, or at the same time as the target) or absent. Participants had to detect or discriminate between targets and were instructed to ignore the distracter. The results showed facilitation of search when the singleton coincided with the target. Dalton and Spence (2007) extended these results by showing that these auditory singletons have a similar facilitatory effect on serial visual search. Overall these results show that temporal attention can be affected both in a bottom-up (e.g., Dalton & Lavie, 2006) and top-down (e.g., Coull & Nobre, 1998) fashion, both for unimodal and crossmodal presentation.

Attention can also be directed to specific visual features like shape or color (see Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990), and to specific auditory features like pitch or amplitude (e.g., Dalton & Lavie, 2004; Zatorre, Mondor, & Evans, 1999). For example, Treisman (1988) shows that knowing that the target would have a unique color of shape reduces search time by 100 ms. Overall, there are many features we can direct our attention to. One feature that is shared by both auditory and visual events is their location in space.

By directing attention to a location in space we are able to respond more quickly and more accurately to events occurring at that location (Posner, 1980). We can direct our spatial attention in an overt manner by making eye movements (Theeuwes, Kramer, Hahn, & Irwin, 1998) or in a covert manner without making eye movements (Theeuwes, 1994). This review focuses on covert attentional selection processes. Covert attention can be voluntarily deployed, under what is known as *endogenous* control, or can be involuntary deployed, as when attention is *exogenously* captured. Directing endogenous attention has metaphorically been compared to the movement of a spotlight to a particular location illuminating that area (Posner, Snyder, & Davidson, 1980). If a target object occurs within the spotlight, one is able to respond faster and more accurate than when it occurs outside the spotlight (Broadbent, 1982; Posner et al., 1980). Endogenous attention can for instance be directed by means of a centrally presented arrow that points with a high probability towards a target location in the periphery. By using an 80% valid endogenous arrow cue pointing to one of two peripheral locations, Posner (1980) showed that people respond faster to a target occurring at a validly cued location compared to an invalidly cued location. Exogenous capture of attention can be evoked by the presentation of sudden salient events like a visual onset. When these exogenous cues happen to occur at the location where the target is going to appear reaction times to the target are faster and more accurate than when targets appear at uncued locations. These benefits are shown even when cue validity is at chance level and no reliable prediction of the location of the upcoming target is provided (e.g., Jonides, 1981; Yantis & Jonides, 1984). Endogenous and exogenous cueing effects are reported in both the visual (Posner, 1980) and auditory modality (Spence & Driver, 1994). This section will primarily focus on spatial attention and crossmodal spatial attention in particular.

3.1. Crossmodal spatial attention

The effect of attentional capture on visual perception is elegantly demonstrated by a phenomenon called 'illusory line motion' or the 'shooting line illusion' (Hikosaka, Miyauchi, & Shimojo, 1993). In this illusion, a line that is physically presented at once is perceived as being drawn from one side to the other. Illusory line motion occurs when prior to the presentation of the line attention is captured by means of an exogenous visual cue to one of the ends of the line. This cue location is then perceived as the starting location from which the line is illusorily drawn. Shimojo, Miyauchi and Hikosaka (1997) show that the shooting line illusion is not restricted to visual cueing. Both auditory and somatosensory cues presented at one of the far ends of the line also create the illusory motion sensation. This illusion illustrates that exogenous capture of attention does not only occur within modalities but can also occur across modalities (e.g., Bernstein & Edelman, 1971; McDonald et al., 2000; Simon & Craft, 1970; Spence & Driver, 1997; Ward, 1994). Note that illusory line motion does not occur when attention is focused by means of an endogenous cue (Christie & Klein, 2005; Schmidt, 2000).

Although early studies already found evidence of crossmodal attention (Bernstein & Edelman, 1971; Simon & Craft, 1970; Ward, 1994), they did not control for eye movements, which means that they could not rule out overt rather than covert orienting of attention. In addition participants in these studies had to respond to the left target by pressing a button with their left hand and to the right target by pressing a button with their right hand. However, the cue presented prior to the target was also presented at a left or right location and could therefore prime the response hand in addition to capturing attention. This made it hard to differentiate between attention and response priming effects.

In a seminal study, Spence and Driver (1997) investigated crossmodal attention while controlling for both eye movements and response priming. Participants were required to maintain ocular

fixation and this was monitored by an eye tracker. Response priming was controlled for by using an *orthogonal cueing task* in which participants made an elevation judgment regarding auditory or visual targets presented to the upper or lower visual hemifield on the left or right of fixation. At 100, 200, or 700 ms prior to the onset of the visual target an auditory or visual cue was presented along the horizontal meridian to the left or right side of fixation. In this way, Spence and Driver (1997) decoupled the response dimension from the cueing dimension.

The results of Spence and Driver (1997) showed unimodal visual (visual cue and visual target) and auditory cueing effects (auditory cue and auditory target). In addition, a crossmodal auditory cueing effect on visual target discrimination was shown. These results were only found for cue–target intervals of 100 and 200 ms and not for 700 ms. Crossmodal cueing of an auditory target by a visual cue was notably absent over all cue–target intervals. Spence, McDonald, and Driver (2004) attributed this absence to the higher spatial resolution of the visual perceptual system relative to the auditory system. The idea is that a visual cue focuses spatial attention to a relative small area in between the upper and lower target locations. Because the attentional focus does not include either the upper or lower target location, it does not result in a cueing effect. On the other hand, an auditory cue draws attention to a much larger area in a more diffuse manner, and as such cues both the upper and lower target locations.

McDonald, Ward, and colleagues (McDonald et al., 2000; Ward, McDonald, & Lin, 2000) showed cueing across both modalities using a different paradigm that involved a go/no-go task. To rule out response priming, participants responded with the same button regardless of whether the target was presented at either the left or right side. They had to refrain from responding when the target appeared in the centre. Because this task involved no elevation judgment, cues and targets were presented at the same location. Therefore, the design was not sensitive to differences in spatial resolution between the auditory and visual domains.

The studies by Spence and Driver (1997) and McDonald, Ward, and colleagues (McDonald et al., 2000; Ward et al., 2000) clearly demonstrated that auditory input can affect visual spatial attention and vice versa. As shown in Table 1, similar crossmodal attentional effects occur between the somatosensory and visual modalities (e.g., Kennett, Eimer, Spence, & Driver, 2001; Spence, Nicholls, Gillespie, & Driver, 1998), and between the somatosensory and auditory modalities (e.g., Spence et al., 1998). However, there is little consensus among these studies regarding the level at which crossmodal capture takes place. The asymmetry shown by Spence and Driver (1997)

suggests an interaction at an early unimodal stage. This might explain why there was only capture of visual attention by sound and not vice versa. However, later studies (McDonald et al., 2000; Ward et al., 2000) attributed this asymmetry in crossmodal cueing to some particularities of the Spence and Driver (1997) paradigm. Presumably, a difference in spatial resolution of the visual compared to the auditory perceptual system is associated with a corresponding difference in the size of the spatial area that is attended (Spence et al., 2004). By using another paradigm McDonald et al. (2000) and Ward et al. (2000) indeed showed that there was crossmodal cueing in both directions (from audition to vision and vice versa). Such symmetry in cross-modal cueing suggests that crossmodal capture occurs at an amodal level. To shed more light, on which level crossmodal attention takes place, it is important to consider the neural correlates underlying crossmodal attention. More specifically, it will be discussed how crossmodal attention and multisensory integration affect one another.

3.2. Neural correlates of crossmodal spatial attention

In an ERP study, McDonald and Ward (2000) showed that auditory capture of visual attention is represented by an ERP effect they termed the negative difference. Participants had to respond to a visual target that was preceded by a spatially valid or invalid auditory cue. The negative difference is calculated by subtracting the ERPs to visual targets on the invalid trials from those of the valid trials. By means of this subtraction, all evoked potentials that are constant over both valid and invalid cueing conditions are filtered out. This results in a negative difference potential that only reflects effects of spatial attention. At short cue target intervals (100–300 ms) this negative difference potential was largest over the occipital cortex contralateral to the target location. This lateralization in the occipital cortex suggests modulation of the early visual cortex by means of spatial attention. Similar negative difference effects were also shown for visual cues and auditory targets (McDonald, Teder-Salejarvi, Heraldez, & Hillyard, 2001) and tactile cues and visual targets (Kennett et al., 2001). In a follow-up study McDonald, Teder-Salejarvi, Di Russo and Hillyard (2003) investigated the neural correlates causing these negative difference effects and their time course. They found early activation in the superior temporal sulcus and gyrus (120–140 ms), then in the fusiform gyrus of the ventral occipito-temporal cortex (150–170 ms), followed by activity in the peri-sylvian cortex of the inferior parietal lobe (200–300 ms). The superior temporal sulcus is known as a site where multisensory information meets and integrates (for a review, see Stein & Meredith, 1993). Neurons of the fusiform gyrus of the ventral occipito-temporal cortex are known to respond to different kinds of visual stimuli (e.g., Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1991; Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999) and this activation can be modulated by attention (e.g., Corbetta et al., 1991; Hopfinger, Buonocore, & Mangun, 2000). McDonald et al. (2003) suggested that the activity in the peri-sylvian cortex of the inferior parietal lobe reflects enhanced perceptual processes based on attentional control rather than crossmodal attention itself (see McDonald et al. (2003); McDonald, Teder-Salejarvi, Heraldez, et al., 2001).

Studies performing functional magnetic resonance imaging (fMRI) investigating the brain areas involved in crossmodal attention (Degerman, Rinne, Pekkola, Autti, Jääskeläinen, Sams et al., 2007; Weissman, Warner, & Woldorff, 2004) show activation in both heteromodal and early sensory brain areas. An fMRI study by Weissman et al. (2004) looked at neural mechanisms that might reduce crossmodal distractions. In this study, participants had to identify a visual or auditory letter (i.e., written or spoken) that co-occurred with an irrelevant congruent or incongruent letter in the other modality. The results showed an increase in activation in the early sensory areas, dorsolateral prefrontal cortex, and in the anterior cingulate cortex. Incongruence between a visual target and auditory distractor resulted in additional activity in the visual cortex but had no

Table 1

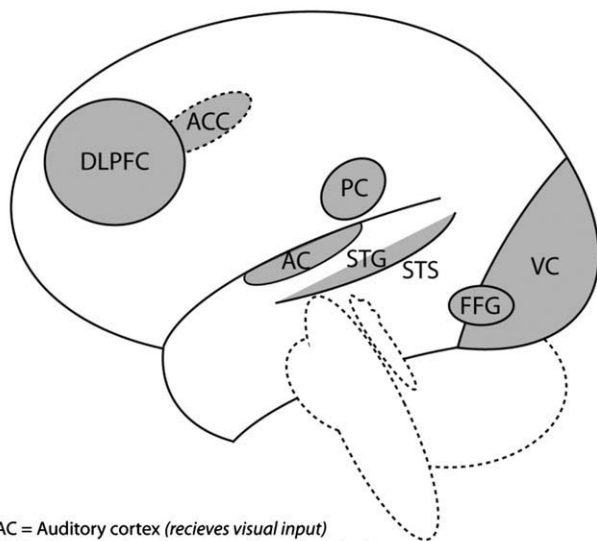
A selection of important results from representative crossmodal cueing studies.

	Method	Cue (SOA ms)	Target (Experiment)	Effect size ms * $p < .05$
Spence and Driver, 1997	Orthogonal cueing task	Auditory (100)	Auditory (1)	24*
			Visual (1)	33*
		Visual (100)	Visual (2)	51*
			Auditory (2)	-3
Ward et al., 2000	Go/no-go	Auditory (100)	Auditory	39*
			Visual	11
		Visual (100)	Visual	5
			Auditory	30*
McDonald et al., 2000	Go/no-go	Auditory (100–300)	Visual (1)	20*
Spence et al., 1998	Cueing task	Auditory (150)	Tactile (1)	10*
			Tactile (2)	18*
		Visual (200)	Auditory (3)	15*
			Tactile (200)	Visual (3)
Kennett et al., 2001	Cueing task	Tactile (200)	Visual	28*

effect in the auditory cortex. Incongruence between an auditory target and a visual distractor resulted in additional activity in the auditory cortex but did not affect the visual cortex. Increased activation in the dorsolateral prefrontal cortex suggests increased biasing in goal-relevant attention during incongruent trials. The anterior cingulate cortex becomes active when conflicting events take place (Carter et al., 1998). Overall, these findings show a role for unimodal and multimodal processing levels when it comes to minimizing effects of distracting stimuli. A study by Degerman et al. (2007) investigated whether audiovisual attention activates similar brain areas as do visual and auditory attention alone. During this experiment visual events (blue or red circle) presented on a central display and auditory events (high and low pitch) presented through headphones occurred simultaneously. Participants attended to the visual event, the auditory event, or both. Results show for all conditions show activation in frontal, temporal, parietal, and occipital cortical regions. Occipital visual regions showed modulation during the visual and auditory task, and temporal auditory regions showed also modulation during the visual and auditory task. Overall, these results suggest that top-down control of attention by attending to one modality can affect early sensory areas of the other modality such that the crossmodal distracting effect is minimized (Weissman et al., 2004). However, when these auditory and visual events have task relevant features that are non-conflicting like color and pitch (Degerman et al., 2007), attentional modulation in both sensory areas were shown. Interestingly, as mentioned above, Alais et al. (2006) show related behavioral results in the form of separate attentional resources for modality-specific features like auditory pitch and visual contrast.

Overall, these studies demonstrate that crossmodal attention affects sensory processing at an early unimodal stage as shown by the activation in the early sensory areas (McDonald et al., 2000). Additionally, modulation of heteromodal areas was found (McDonald et al., 2003), which suggests effects of crossmodal attention at multiple stages of sensory processing (see Fig. 2).

Brain areas involved in audiovisual attention



AC = Auditory cortex (receives visual input)
 ACC = Anterior cingulate cortex (error detection)
 DLPFC = Dorsolateral prefrontal cortex (goal-relevant attention)
 FFG = Fusiform gyrus (modulated by attention)
 PC = Perisylvian cortex (perceptual enhancement)
 STG = Superior temporal gyrus (auditory processing)
 STS = Superior temporal sulcus (multisensory integration)
 VC = Visual cortex (receives auditory input)

Fig. 2. Brain areas and their functional correlate that are involved in audiovisual attention.

3.3. Interaction between multisensory integration and attention

An important question is whether multisensory integration and crossmodal attention interact. The ventriloquism effect – which is known to result from multisensory integration – has been shown to occur preattentively and independently of both voluntary and involuntary spatial attention shifts (Vroomen, Bertelson, & de Gelder, 2001a,b). McDonald, Teder-Salejari, and Ward (2001) argue that multisensory integration and crossmodal attention are different processes with separate neural mechanisms. Consistent with this idea are the differences in temporal constraints under which multisensory integration and attention take place. Multisensory integration is optimal when events co-occur in time (see Meredith et al., 1987), while attention needs some time to engage (see Woodman & Luck, 1999) before it affects other processes. However, Macaluso and Driver (2001) argue that such distinction cannot be made since there are also multisensory cells that still show integration effects for asynchronies up to 600 ms (see Calvert & Thesen, 2004; Wallace, Meredith, & Stein, 1992). This is enough time for engagement of crossmodal attention to occur and would suggest that multisensory integration and attention are based at least partly on similar underlying processes. Spence et al. (2004) argue that crossmodal spatial cueing effects could be explained in terms of spatial attention, multisensory integration, or both (Spence & Driver, 1997; Spence & McGlone, 2001). As long as there are no conclusive studies on this topic, these possibilities seem to remain open. Additionally, there is a controversy about the stage at which multisensory integration takes place. This could be an early pre-attentive stage, which might suggest that multisensory integration drives attention (Vroomen et al., 2001b). Other studies suggest late integration by showing that attention is needed for multisensory integration to occur (e.g., Busse, Roberts, Crist, Weissman, & Woldorff, 2005; Talsma & Woldorff, 2005). A third option might be that multisensory integration occurs at multiple stages in a more parallel fashion (Calvert & Thesen, 2004).

The late integration framework (Fig. 1a) states that unimodal attention affects the individual sensory input and integrates them at a late stage into a single percept. Thus auditory and visual events are first individually enhanced by means of unimodal attention before integrating at a higher heteromodal level. As a consequence, attention is needed for multisensory integration to occur. Some experimental results are consistent with this idea. For instance, Talsma and Woldorff (2005) showed multisensory integration effects in the form of enhanced frontal positivity 100 ms after stimulation. This effect was only present for visually attended stimuli (see also Talsma et al., 2007). These results suggest that there is no multisensory integration without attention.

The early integration framework (Fig. 1b) states that multisensory integration occurs at an early sensory level and at a later stage amodal attention is captured. Therefore, this framework suggests that multisensory integration is independent of attention. Even though independent, bimodal cues can still capture attention at a higher heteromodal level. For example, the idea of early integration is in line with the pip and pop effect (Van der Burg et al., 2008) and the ventriloquism effect (Vroomen et al., 2001b) both of which seem to occur at a pre-attentive stage. As mentioned above, quickly processed auditory information projecting from auditory to visual cortical areas seems able to influence bottom-up visual processing in a way that enhances co-occurring visual information (see Romei et al., 2007). This enhancement by multisensory integration at a pre-attentive stage can lead to attentional capture in a situation where the individual events would not capture attention (Santangelo & Spence, 2007).

The parallel integration framework (Fig. 1c) as proposed by Calvert and Thesen (2004) suggests that multisensory integration takes place at multiple stages. Between these stages there is dynamic modulation, meaning that multisensory integration occurs at an early or late stage

depending on the resources available. Studies of multisensory integration as discussed in the previous section have shown that early or late integration is highly task dependent. There may be qualitative and quantitative differences in these types of multisensory integration. Although parallel integration was originally used to explain different forms of multisensory integration, it might apply to multisensory interaction in general. It is conceivable that similar resources used for multisensory integration are also used for attentional processes (see Meredith et al., 1987, Fig. 3).

Therefore, the parallel integration framework might explain the interaction between attention and multisensory integration. For example, near-threshold events might need attentional resources for integration to occur. If that is the case, integration can only occur at those stages that are sensitive to top-down influences. Also such integration may occur relatively late in time because it takes time for top-down control to have an effect (van Zoest, Donk, & Theeuwes, 2004). However, supra-threshold events may integrate automatically (without attention) at an early stage of processing. Even though this early integration may occur automatically, top-down attention could still affect late integration. This idea is consistent with a recent fMRI study by Fairhall and Macaluso (2009), who showed that spatial attention can affect multisensory integration in cortical and subcortical areas. In this study, participants attended to a visual stream of speaking lips that was either congruent or incongruent with an auditory speech stream. Results showed increased activation in associative regions, visual cortex, and subcortical areas for attended congruent conditions. In other words, these results show involvement of heteromodal brain areas and early sensory areas like the primary visual cortex. The authors concluded that multisensory integration and attention interact in a way that affects an extensive network of brain areas. Even though this framework provides a way to

understand the interaction between top-down attention and multisensory integration, it should be realized that when focused on a single audio-visual event the largest benefit of multisensory integration will be seen with stimuli presented near threshold. For additional reading about multisensory anatomical connections, see Cappe, Rouiller and Barone (2009).

Audiovisual events that integrate at an early stage are known to become more salient than the individual events. These bimodal events are known to draw attention (see Van der Burg et al., submitted). However, bimodal events do not show a super-additive effect at an attentional level (e.g., Koelewijn et al., 2009b; Santangelo, Van der Lubbe, Belardinelli, & Postma, 2008, 2006). This was shown in a study by Santangelo et al. (2006) where participants performed an orthogonal cueing task in which the visual target was preceded by a unimodal visual or auditory cue, or by a bimodal audiovisual cue. The results showed cueing effects that were comparable in size for all three conditions. However, in the bimodal condition the auditory and visual cues were presented at the same time and at the same location, which allows for multisensory integration. Therefore, the results indicate that multisensory integration is not reflected by a stronger cueing effect (see for similar results, Koelewijn et al., 2009b). In a follow up study, Santangelo et al. (2008) used EEG to test whether multisensory integration takes place between bimodal cues. As in the earlier study, the behavioral results showed no additional effect of bimodal cueing compared to unimodal cueing. However, ERPs did show a super-additive effect for bimodal stimuli, indicating multisensory integration. These results thus confirm that multisensory integration is not reflected at an attentional level in the form of larger cueing effect for bimodal cues compared to unimodal cues. See Spence and Santangelo (2009), for a more elaborate review on this topic.

To summarize, some studies show that attention is needed for multisensory integration to occur (e.g., Talsma et al., 2005; Talsma et al., 2007), while others show that multisensory integration occurs independent of attention (Vroomen et al., 2001a,b). Macaluso and Driver (2001) suggest that similar areas or even similar cells in subcortical areas or primary sensory cortices are responsible for both multisensory integration and crossmodal attention. Also heteromodal areas like the superior temporal sulcus are known to play a role in both multisensory integration (Benevento et al., 1977; Bruce et al., 1981) and crossmodal attention (McDonald et al., 2003). Although both multisensory integration and attentional processes take place in similar brain areas they do not necessarily interact.

3.4. Discussion

So far, the literature shows that multimodal interactions like multisensory integration can take place in early unimodal and late heteromodal areas (see Calvert & Thesen, 2004). Crossmodal spatial attention can also take place at an early unimodal stage (McDonald et al., 2000) and at a later heteromodal stage (McDonald et al., 2003). Multisensory integration and attention can interact with one another in a way that we see stronger multisensory integration at an attended location (e.g., Talsma et al., 2005).

Earlier, it was discussed that there are temporal and spatial constraints for the occurrence of multisensory integrations. A further question is whether these constraints are not only a necessary but also a sufficient condition. In other words: to what degree are multisensory interactions automatic? Note that these constraints do not apply in all cases. The complexity of the task in, for example, audiovisual speech perception can reduce the dependence on spatial and temporal overlap (see Jones & Jarick, 2006). One important criterion a process has to meet in order to be called automatic is the *intentionality* criterion (e.g., Jonides, 1981; Jonides, Naveh-Benjamin, & Palmer, 1985; Posner, 1978; Yantis & Jonides, 1990). This criterion states that an automatic process is not affected by voluntary control. For example, voluntarily or top-down directing of attention to a

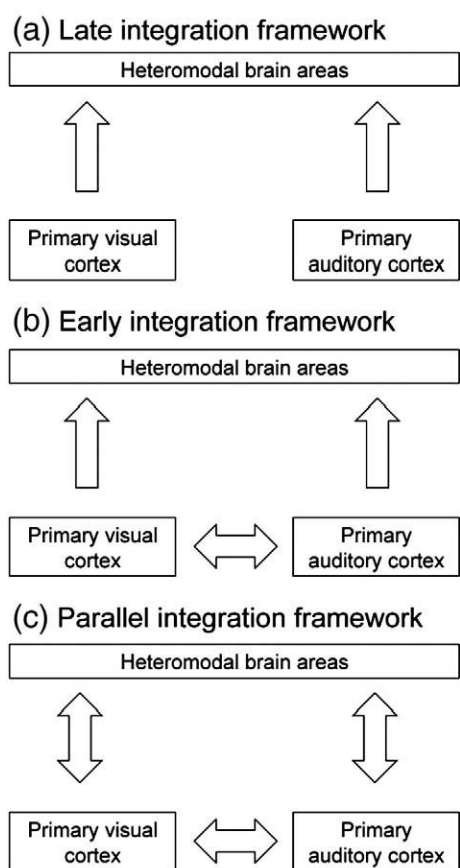


Fig. 3. Schematic representation of; a) a late integration framework, b) an early integration framework, and c) a parallel integration framework.

certain location should not affect multisensory integration. However, as mentioned above, multiple studies show that multisensory integration is indeed modulated by attention (e.g., Fairhall & Macaluso, 2009; Talsma & Woldorff, 2005). This suggests that multisensory integration in general is not an automatic process. Because there is evidence that early multisensory integration takes place without requiring attentional resources (e.g., Van der Burg et al., 2008; Vroomen & de Gelder, 2004), it might be more correct to define early and late integration as different processes of which early integration is automatic and late integration not. In the next section, we discuss whether crossmodal attentional capture is an automatic process.

4. Automaticity of crossmodal attention

The final issue addressed in this review is whether crossmodal attention is an automatic process. In other words, does auditory capture of visual spatial attention always occur? Jonides (1981) stated that a cognitive process occurs in an automatic fashion if it satisfies the *load-insensitivity* criterion, which states that automatic processes are insensitive to the load of current task demands. For attentional capture, this means that the occurrence of capture should not be affected by the presence of other competing events in the display. In addition, the intentionality criterion – already mentioned above – states that an automatic process is resistant to suppression and is insensitive to an observer's top-down control. For capture, this implies that irrespective of the goals of the observers capture should occur.

4.1. Intentionality criterion

Several studies have tested whether attentional capture meets the intentionality criterion within the visual modality (e.g., Jonides, 1981; Müller & Rabbitt, 1989; Theeuwes, 1991; Yantis & Jonides, 1990). Theeuwes (1991) investigated the interaction between endogenous and exogenous visual attention within a single paradigm. In this paradigm, a target letter was presented among three distractor letters all presented equidistantly on an imaginary circle. Prior to the target a nonpredictive exogenous cue in the form of a visual onset was presented near one of the possible target locations. In addition, an endogenous cue in the form of a central arrow was displayed at fixation indicating the upcoming target location with 100% validity. When the endogenous cue was presented after the exogenous cue, attention was drawn to the location of the exogenous cue. However, when the endogenous cue was presented prior to the exogenous cue, no exogenous cueing effect was observed (for similar results, see Yantis & Jonides, 1990). These results suggest that visual exogenous attention is not a fully automatic process and can be affected by top-down control of attention.

The idea that visual capture is not a fully automatic process raises the question whether auditory capture of visual attention is automatic. The fact that capture within the visual modality can be affected by top-down control of attention does not necessarily mean that the same holds for auditory capture. Recently, several studies (Koelewijn et al., 2009a; Mazza et al., 2007; Santangelo, Belardinelli, & Spence, 2007; Santangelo & Spence, 2007; van der Lubbe & Postma, 2005) addressed this issue. In a study by van der Lubbe and Postma (2005), participants performed a variation on the orthogonal cueing task used by Spence and Driver (1997). In this task, participants had to indicate whether a target in the form of an arrowhead pointed up or down. The target was presented on the left or right of fixation on LED grids. An exogenous auditory or visual cue was presented 200 ms prior to the onset of the target at one of the target location. In addition, one second prior to the onset of the target an arrow was presented on a centrally positioned LED grid, which indicated the target side with 100% validity. Both a unimodal visual and a crossmodal auditory

cueing effects were observed. In contrast to the unimodal results of Theeuwes (1991), the results of van der Lubbe and Postma (2005) showed that visual and auditory onsets capture visual attention even when visual attention is endogenously focused.

Mazza et al. (2007) used a task similar to the orthogonal crossmodal cueing and replicated the crossmodal auditory and unimodal visual and auditory cueing effects. In their second experiment, they blocked the target side. Therefore, participants knew where the target would appear which allowed them to endogenously focus their attention to the target location. The results show no unimodal visual or auditory cueing effects. However, crossmodal cueing in the form of auditory capture of visual attention was still observed. Note that Mazza et al. (2007) did not find unimodal cueing during focused attention, which is in line with the results of Theeuwes (1991).

In a recent study (Koelewijn et al., 2009a) we tested how focused visual attention affects auditory capture by differentiating between attentional costs and benefits. In this study participants performed an orthogonal cueing task in which a visual elevation judgment had to be made. The visual target was preceded by an auditory cue that was spatially congruent (valid condition), incongruent (invalid condition), or spatially uninformative (neutral condition). When the RTs to validly cued targets are faster than those in the neutral cue condition, one speaks of performance benefits (Posner, 1980). When RTs to invalidly cued targets are slower than those in the neutral condition, one speaks of performance costs. The results of this study showed that the crossmodal auditory cueing effect as observed by Spence and Driver (1997) consists of both RT costs and benefits. However, when visual attention was focused prior to the presentation of the exogenous auditory cue by means of a 100% valid arrowhead, only costs were observed meaning that attention was still captured towards the invalid target location.

So far, all studies show that focused visual attention does not affect crossmodal cueing (Koelewijn et al., 2009a; Mazza et al., 2007; van der Lubbe & Postma, 2005). Although no attentional benefits are found when attention is focused prior to the presentation of a valid exogenous cue (Koelewijn et al., 2009a), auditory capture still occurs towards an invalid target location and therefore away from the initial focus of attention. In a recent study, Santangelo and Spence (2007) used an orthogonal cueing paradigm in which elevation judgments of visual targets were made. Instead of using an additional endogenous cue, they used a centrally presented task that required subjects to focus attention on the centre of the display. In this task, participants had to respond to a target embedded in a stream of letters presented in the form of a rapid serial visual presentation (RSVP). In the high-load condition, a target digit was presented centrally in 67% of the trials. In the 33% remaining trials peripheral targets were presented for the elevation judgment task. In the no-load condition, no RSVP stream was presented and therefore participants only had to respond to peripheral targets. In all trials, a peripheral exogenous cue was presented that could be valid or invalid. The exogenous cue was visual, auditory, or bimodal (visual and auditory) and was presented prior to the onset of the target. The result for the no-load condition showed auditory, visual and bimodal cueing effects. Importantly, in the high-load condition only a bimodal cueing effect was observed. These results suggest that focusing visual attention at central fixation suppresses unimodal and crossmodal cueing. In other words, no visual or auditory capture of visual attention will occur during focused visual attention.

To summarize, most studies showed no top-down control on auditory capture (Koelewijn et al., 2009a; Mazza et al., 2007; van der Lubbe & Postma, 2005). Focusing attention on an upcoming target location prior to the presentation of the auditory exogenous cue did not affect attentional capture by this cue. However, when visual attention is centrally focused by means of an additional task, no auditory capture is observed.

4.2. Load-insensitivity criterion

The results by Santangelo and Spence (2007) suggest that endogenous attention focused by means of the additional task is able to suppress auditory capture. However, as the authors remark in their review on this topic (Santangelo & Spence, 2008), the RSVP stream used in the additional task also increases perceptual load. Therefore, the authors argue that attentional capture by peripheral onsets may not occur during circumstances of high perceptual load. These results are in line with the load theory as proposed by Lavie and colleagues (see Lavie, 1995; Lavie, Hirst, de Fockert, & Viding, 2004). This theory states that a high perceptual load should reduce distractor (or irrelevant cue) interference. The results by Santangelo and Spence (2007) show that auditory capture does not meet the load-insensitivity criterion while it may still meet the intentionality criterion. This may explain why other studies do not find an effect of top-down control on auditory capture (Koelewijn et al., 2009a; Mazza et al., 2007; van der Lubbe & Postma, 2005).

In a recent study, we tested whether bottom-up competition by a single visual event could affect auditory capture (Koelewijn et al., 2009b). In this study, participants performed an orthogonal cueing task that only required elevation judgments of visual targets. Prior to the presentation of the target, both a peripheral visual and an auditory cue were presented at the same or at opposite locations. In the first experiment, the validity of both the visual and auditory cue was 50% implying that they were presented at chance level and therefore were both pure exogenous cues. The results showed both auditory and visual cueing effects that did not interact but influenced response times in an additive manner. This suggests that a single visual event is not able to affect auditory capture. In the second experiment, the validity of the visual cue was raised to 80% while the validity of the auditory cue remained at chance level. This time only a visual cueing effect remained and the auditory cueing effect disappeared.

These results demonstrate that auditory capture does not occur when a competing and predictive visual event is presented. Note that these predictive visual cues do not only affect auditory capture in a pure bottom-up fashion because of their onset and temporal vicinity, but also top-down because of their high validity. To conclude, these studies (Koelewijn et al., 2009a; Santangelo & Spence, 2007) imply that auditory capture is not an automatic process. For more elaborate reading on this topic, see Spence (2010).

4.3. Discussion

Several studies indicate that auditory capture meets the intentionality criterion (Koelewijn et al., 2009a; Mazza et al., 2007; van der Lubbe & Postma, 2005). Additionally, when auditory capture competes with a purely exogenous visual cue the load-insensitivity criterion seems to be met as well (Koelewijn et al., 2009b). However, Santangelo and Spence (2007) show that when participants have to perform an additional task, no crossmodal capture is observed. The authors suggest that crossmodal capture is affected by high perceptual load. However, an alternative explanation is also possible.

Although the studies by Koelewijn et al. (2009b) and Santangelo and Spence (2007) used different means to focus attention, there are striking similarities in the way both a predictive visual onset and an additional RSVP task can affect visual attention. The onset of the visual peripheral cue as used by Koelewijn et al. (2009b) captures visual attentional resources in a bottom-up fashion. However, when the peripheral cues were made highly valid they added top-down control in addition to the bottom-up capturing effect. Thus, neither purely endogenous (see Koelewijn et al., 2009a) nor purely exogenous cues (Koelewijn et al., 2009b) seem to be able to suppress crossmodal auditory capture. Instead, suppression may only be possible when a combination of both these bottom-up and top-down processes occurs. The RSVP stream used by Santangelo and Spence (2007)

might have affected crossmodal capture the same way. An RSVP stream will capture exogenous attention by means of the onsets of the individual events. Additionally, the fact that 67% of the targets appeared in the central RSVP stream probably caused endogenous focusing of attention. Thus, in order to suppress crossmodal capture, endogenous attention needs some additional bottom-up activity, either in the form of perceptual load (Santangelo, Ho & Spence, 2008) or a peripheral onset (Koelewijn et al., 2009b). The reason for this can be explained by means of the parallel integration framework.

The parallel integration framework of Calvert and Thesen (2004) (see Fig. 1c) proposes that a sound can influence visual processes at an early stage. This would mean that sound can also affect visual attention at an early unimodal level (see Spence & Driver, 1997). Additionally, sound can influence attention at a heteromodal level. If crossmodal capture affects attention at both early and late processing stages in parallel this might explain why interference by a visual event on only one of these levels is not sufficient in suppressing crossmodal capture. Let us assume that endogenous focusing of attention by means of a highly valid cue is able to suppress crossmodal capture at a late heteromodal stage. In this case, a sound is still able to capture visual attention at an early unimodal stage. On the other hand, if we assume that exogenous capture of attention by a visual onset is able to suppress crossmodal capture at an early unimodal stage, sound is still able to capture visual attention at a late heteromodal stage. Only when both stages are affected in parallel by a visual cue that both draws on exogenous and endogenous attentional resources no crossmodal capture is observed. Although this hypothesis is speculative and might be oversimplified, the studies discussed in this review seem to point in this direction.

5. Conclusions

When auditory and visual events are presented at roughly the same time and location they tend to integrate. Note that temporal proximity seems to be a prerequisite for integration while spatial proximity is not always necessary (Van der Burg et al., 2008). This integration can lead to an increased saliency and can draw attention in cases in which individual stimuli would be less effective (Santangelo & Spence, 2007). This multisensory integration can take place in heteromodal brain areas but also in primary sensory areas in a parallel fashion. Multisensory integration is not a pure automatic process since it can be affected by attention. However, these attentional effects on multisensory integration are primarily shown by studies in which late integration takes place at heteromodal brain areas (e.g., Busse et al., 2005; Fairhall & Macaluso, 2009; Talsma & Woldorff, 2005). Early integration as shown by other studies does not seem sensitive to spatial attention (e.g., Van der Burg et al., 2008; Vroomen & de Gelder, 2004). Therefore, late and early integration should be considered as independent processes that take place in parallel (see Calvert & Thesen, 2004).

When events do not co-occur in time or space and one of the events is salient enough, this event can still affect attention in the other modality. This crossmodal attentional capture seems to affect visual attention both at an early stage in the form of a bottom-up process and at a late stage in the form of top-down process. Both processes can occur in parallel in a way similar to what happens in multisensory integration. The results so far suggest that in order to suppress crossmodal auditory capture, presenting a visual event that either competes for bottom-up or top-down attentional resources is not sufficient (Koelewijn et al., 2009a,b). Only when both processes are affected at the same time by a competing event is auditory capture entirely extinguished (Koelewijn et al., 2009a; Santangelo and Spence, 2007; Santangelo et al., 2007).

Based on the studies discussed in this review, we may conclude that audiovisual interactions are not pure automatic processes and therefore do not occur under all circumstances. However,

multisensory illusions show that when these interactions do occur they can have a strong impact. As already mentioned in the Introduction, there is an increase in development and use of multisensory displays like for instance navigation systems. We need to beware of the consequences these applications can have on our everyday functioning (for more elaborate reading on this topic see Ho & Spence, 2008). For example, although audiovisual events or multisensory events in general are well suited as warning signals, when giving too many false alarms they can become distracting.

Acknowledgements

This research was financially supported by the Dutch Technology Foundation STW (07079).

References

- Alais, D., Morrone, C., & Burr, D. (2006). Separate attentional resources for vision and audition. *Proceedings of the Royal Society B-Biological Sciences*, 273(1592), 1339–1345.
- Baron-Cohen, S., Wyke, M. A., & Binns, C. (1987). Hearing words and seeing colors – an experimental investigation of a case of synesthesia. *Perception*, 16(6), 761–767.
- Barraclough, N. E., Xiao, D. K., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*, 17(3), 377–391.
- Benevento, L. A., Fallon, J., Davis, B. J., & Rezak, M. (1977). Auditory–visual interaction in single cells in cortex of superior temporal sulcus and orbital frontal cortex of macaque monkey. *Experimental Neurology*, 57(3), 849–872.
- Bernstein, I. H., & Edelman, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, 87, 242–247.
- Bolognini, N., Frassinetti, F., Serino, A., & Ladavas, E. (2005). “Acoustical vision” of below threshold stimuli: Interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, 160(3), 273–282.
- Broadbent, D. E. (1982). Task combination and selective intake of information. *Acta Psychologica*, 50(3), 253–290.
- Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *Journal of Neurophysiology*, 46(2), 369–384.
- Busse, L., Roberts, K. C., Crist, R. E., Weissman, D. H., & Woldorff, M. G. (2005). The spread of attention across modalities and space in a multisensory object. *Proceedings of the National Academy of Sciences of the United States of America*, 102(51), 18751–18756.
- Calvert, G. A., & Thesen, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*, 98(1–3), 191–205.
- Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, 10(11), 649–657.
- Cappe, C., Rouiller, E. M., & Barone, P. (2009). Multisensory anatomical pathways. *Hearing Research*, 285, 28–36.
- Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science*, 280(5364), 747–749.
- Christie, J., & Klein, R. M. (2005). Does attention cause illusory line motion? *Perception & Psychophysics*, 67(6), 1032–1043.
- Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, 22, 319–349.
- Colby, C. L., Duhamel, J. R., & Goldberg, M. E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of Neurophysiology*, 76(5), 2841–2852.
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, 248(4962), 1556–1559.
- Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1991). Selective and divided attention during visual discriminations of shape, color, and speed – functional-anatomy by positron emission tomography. *Journal of Neuroscience*, 11(8), 2383–2402.
- Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience*, 18(18), 7426–7435.
- Dalton, P., & Lavie, N. (2004). Auditory attentional capture: Effects of singleton distractor sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 30(1), 180–193.
- Dalton, P., & Lavie, N. (2006). Temporal attentional capture: Effects of irrelevant singletons on rapid serial visual search. *Psychonomic Bulletin & Review*, 13(5), 881–885.
- Dalton, P., & Spence, C. (2007). Attentional capture in serial audiovisual search tasks. *Perception & Psychophysics*, 69(3), 422–438.
- Degerman, A., Rinne, T., Pekkola, J., Autti, T., Jääskeläinen, I. P., Sams, M., et al. (2007). Human brain activity associated with audiovisual perception and attention. *NeuroImage*, 34, 1683–1691.
- Driver, J., & Spence, C. (1998). Cross-modal links in spatial attention. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 353(1373), 1319–1331.
- Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, 10(20), R731–R735.
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4), 162–169.
- Fairhall, S. L., & Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience*, 29, 1247–1257.
- Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2000). Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cognitive Brain Research*, 10(1–2), 77–83.
- Frassinetti, F., Bolognini, N., & Ladavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, 147, 332–343.
- Ghazanfar, A. A., & Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences*, 10(6), 278–285.
- Ghazanfar, A. A., Maier, J. X., Hoffman, K. L., & Logothetis, N. K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *Journal of Neuroscience*, 25(20), 5004–5012.
- Hein, G., & Knight, R. T. (2008). Superior temporal sulcus—It’s my area: Or is it? *Journal of Cognitive Neuroscience*, 20(12), 2125–2136.
- Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal of Experimental Psychology*, 63, 289–293.
- Hickey, C., McDonald, J. J., & Theeuwes, J. (2006). Electrophysiological evidence of the capture of visual attention. *Journal of Cognitive Neuroscience*, 18(4), 604–613.
- Hikosaka, O., Miyauchi, S., & Shimojo, S. (1993). Focal visual attention produces illusory temporal order and motion sensation. *Vision Research*, 33(9), 1219–1240.
- Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver’s visual attention. *Journal of Experimental Psychology-Applied*, 11(3), 157–174.
- Ho, C., & Spence, C. (2008). *The multisensory driver: Implications for ergonomic car interface design*. Hampshire: Ashgate Publishing Limited.
- Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience*, 3(3), 284–291.
- Ishai, A., Ungerleider, L. G., Martin, A., Schouten, H. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16), 9379–9384.
- Jones, J. A., & Jarick, M. (2006). Multisensory integration of speech signals: The relationship between space and time. *Experimental Brain Research*, 174(3), 588–594.
- Jonides, J. (1981). Voluntary vs. automatic control over the mind’s eye’s movements. In J. B. Long, & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 187–203). Hillsdale, NJ: Erlbaum.
- Jonides, J., Naveh-Benjamin, M., & Palmer, J. (1985). Assessing automaticity. *Acta Psychologica*, 60, 157–171.
- Kennett, S., Eimer, M., Spence, C., & Driver, J. (2001). Tactile-visual links in exogenous spatial attention under different postures: Convergent evidence from psychophysics and ERPs. *Journal of Cognitive Neuroscience*, 13(4), 462–478.
- Kingstone, A. (1992). Combining expectancies. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, 44(1), 69–104.
- Klaver, P., Talsma, D., Wijers, A. A., Heinze, H. J., & Mulder, G. (1999). An event-related brain potential correlate of visual short-term memory. *Neuroreport*, 10(10), 2001–2005.
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2009). Auditory and visual capture during focused visual attention. *Journal of Experimental Psychology-Human Perception and Performance*, 35(5), 1303–1315.
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2009). Competition between auditory and visual spatial cues during visual task performance. *Experimental Brain Research*, 195(4), 593–602.
- Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology-Human Perception and Performance*, 21(3), 451–468.
- Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology-General*, 133(3), 339–354.
- Lewis, J. W., & Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *Journal of Comparative Neurology*, 428(1), 112–137.
- Liegeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked-potentials recorded from the auditory-cortex in man – Evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology*, 92(3), 204–214.
- Linden, J. F., Grunewald, A., & Andersen, R. A. (1999). Responses to auditory stimuli in macaque lateral intraparietal area II. Behavioral modulation. *Journal of Neurophysiology*, 82(1), 343–358.
- Lippert, M., Logothetis, N. K., & Kayser, C. (2007). Improvement of visual contrast detection by a simultaneous sound. *Brain Research*, 1173, 102–109.
- Los, S. A., & Schut, M. L. J. (2008). The effective time course of preparation. *Cognitive Psychology*, 57(1), 20–55.
- Macaluso, E., & Driver, J. (2001). Spatial attention and crossmodal interactions between vision and touch. *Neuropsychologia*, 39(12), 1304–1316.
- Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: A window onto functional integration in the human brain. *Trends in Cognitive Sciences*, 28(5), 264–271.
- Martinez, A., Anillo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., et al. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, 2(4), 364–369.
- Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J. P., Maeder, P. P., Clarke, S., et al. (2007). Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cerebral Cortex*, 17(7), 1672–1679.

- Mazza, V., Turatto, M., Rossi, M., & Umiltà, C. (2007). How automatic are audiovisual links in exogenous spatial attention? *Neuropsychologia*, 45(3), 514–522.
- McDonald, J. J., & Ward, L. M. (2000). Involuntary listening aids seeing: Evidence from human electrophysiology. *Psychological Science*, 11(2), 167–171.
- McDonald, J. J., Teder-Salejari, W. A., & Hillyard, S. A. (2000). Involuntary orienting of sound improves visual perception. *Nature*, 407, 906–908.
- McDonald, J. J., Teder-Salejari, W. A., Heraldez, D., & Hillyard, S. A. (2001). Electrophysiological evidence for the “missing link” in crossmodal attention. *Canadian Journal of Experimental Psychology*, 55(2), 141–149.
- McDonald, J. J., Teder-Salejari, W. A., & Ward, L. M. (2001). Multisensory integration and crossmodal attention effects in the human brain. *Science*, 292(5523), 1791a–1791.
- McDonald, J. J., Teder-Salejari, W. A., Di Russo, F., & Hillyard, S. A. (2003). Neural substrates of perceptual enhancement by cross-modal spatial attention. *Journal of Cognitive Neuroscience*, 15(1), 10–19.
- McCurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748.
- Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology*, 75(5), 1843–1857.
- Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons 1. Temporal factors. *Journal of Neuroscience*, 7(10), 3215–3229.
- Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience*, 27(15), 4120–4131.
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, 17(1), 154–163.
- Mozolic, J. L., Hugenschmidt, C. E., Peiffer, A. M., & Laurienti, P. J. (2008). Modality-specific selective attention attenuates multisensory integration. *Experimental Brain Research*, 184, 39–52.
- Müller, H. J., & Rabbitt, P. M. A. (1989). Reflexive and voluntary orienting of visual-attention – time course of activation and resistance to interruption. *Journal of Experimental Psychology-Human Perception and Performance*, 15(2), 315–330.
- Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction-time. *Psychological Bulletin*, 89(1), 133–162.
- Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, 131(4), 510–532.
- Noesselt, T., Bergmann, D., Hake, M., Heinze, H. J., & Fendrich, R. (2008). Sound increases the saliency of visual events. *Brain Research*, 1220, 157–163.
- Odegaard, E. C., Arieh, Y., & Marks, L. (2003). Cross-modal enhancement of perceived brightness: Sensory interaction versus response bias. *Perception & Psychophysics*, 65, 123–132.
- Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3–25.
- Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology-General*, 109(2), 160–174.
- Romei, V., Murray, M. M., Merabet, L. B., & Thut, G. (2007). Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: Implications for multisensory interactions. *Journal of Neuroscience*, 27(43), 11465–11472.
- Rouw, R., & Scholte, H. S. (2007). Increased structural connectivity in grapheme-color synesthesia. *Nature Neuroscience*, 10(6), 792–797.
- Santangelo, V., & Spence, C. (2007). Multisensory cues capture spatial attention regardless of perceptual load. *Journal of Experimental Psychology-Human Perception and Performance*, 33(6), 1311–1321.
- Santangelo, V., & Spence, C. (2008). Is the exogenous orienting of spatial attention truly automatic? Evidence from unimodal and multisensory studies. *Consciousness and Cognition*, 17(3), 989–1015.
- Santangelo, V., Van der Lubbe, R. H. J., Belardinelli, M. O., & Postma, A. (2006). Spatial attention triggered by unimodal, crossmodal, and bimodal exogenous cues: A comparison of reflexive orienting mechanisms. *Experimental Brain Research*, 173(1), 40–48.
- Santangelo, V., Belardinelli, M. O., & Spence, C. (2007). The suppression of reflexive visual and auditory orienting when attention is otherwise engaged. *Journal of Experimental Psychology-Human Perception and Performance*, 33(1), 137–148.
- Santangelo, V., Ho, C., & Spence, C. (2008). Capturing spatial attention with multisensory cues. *Psychonomic Bulletin & Review*, 15(2), 398–403.
- Santangelo, V., Van der Lubbe, R. H. J., Belardinelli, M. O., & Postma, A. (2008). Multisensory integration affects ERP components elicited by exogenous cues. *Experimental Brain Research*, 185(2), 269–277.
- Schmidt, W. C. (2000). Endogenous attention and illusory line motion reexamined. *Journal of Experimental Psychology-Human Perception and Performance*, 26(3), 980–996.
- Senkowski, D., Saint-Amour, D., Gruber, T., & Foxe, J. J. (2008). Look who’s talking: The deployment of visuo-spatial attention during multisensory speech processing under noisy environmental conditions. *NeuroImage*, 43(2), 379–387.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions – What you see is what you hear. *Nature*, 408(6814), 788–788.
- Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport*, 12(17), 3849–3852.
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14(1), 147–152.
- Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005). Early modulation of visual cortex by sound: An MEG study. *Neuroscience Letters*, 378(2), 76–81.
- Shimojo, S., Miyauchi, S., & Hikosaka, O. (1997). Visual motion sensation yielded by non-visually driven attention. *Vision Research*, 37(12), 1575–1580.
- Simon, J. R., & Craft, J. L. (1970). Effects of an irrelevant auditory stimulus on visual choice reaction time. *Journal of Experimental Psychology*, 86, 272–274.
- Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, 12(1), 7–10.
- Spence, C. (2010). Crossmodal spatial attention. *Annals of the New York Academy of Sciences*, 1191 (issue The Year of Cognitive Neuroscience 2010), 182–200.
- Spence, C., & Driver, J. (1994). Covert spatial orienting in audition – Exogenous and endogenous mechanisms. *Journal of Experimental Psychology-Human Perception and Performance*, 20(3), 555–574.
- Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, 59(1), 1–22.
- Spence, C., & Ho, C. (2008). Multisensory interface design for drivers: Past, present and future. *Ergonomics*, 51(1), 65–70.
- Spence, C., & McGlone, F. P. (2001). Reflexive spatial orienting of tactile attention. *Experimental Brain Research*, 141(3), 324–330.
- Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. *Hearing Research*, 258(1–2), 134–142.
- Spence, C., Nicholls, M. E. R., Gillespie, N., & Driver, J. (1998). Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision. *Perception & Psychophysics*, 60(4), 544–557.
- Spence, C., McDonald, J., & Driver, J. (2004). Exogenous spatial-cuing studies of human cross-modal attention and multisensory integration. In C. Spence, & J. Driver (Eds.), *Crossmodal Space and Crossmodal Attention* (pp. 277–320). Oxford: Oxford University Press.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Cambridge: MIT.
- Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, 9(4), 255–266.
- Stein, B. E., London, N., Wilkinson, L. K., & Price, D. D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, 8(6), 497–506.
- Stein, B. E., Stanford, T. R., Wallace, J. W. V., & Jiang, W. (2004). Crossmodal spatial interactions in subcortical and cortical circuits. In C. Spence, & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 25–50). Oxford: Oxford university press.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 26(2), 212–215.
- Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, 17(7), 1098–1114.
- Talsma, D., Doty, T., & Woldorff, M. (2005). Audiovisual integration and selective attention: Is attending to both modalities a prerequisite for optimal integration? *Journal of Cognitive Neuroscience*, 83–83.
- Talsma, D., Doty, T. J., & Woldorff, M. G. (2007). Selective attention and audiovisual integration: Is attending to both modalities a prerequisite for early integration? *Cerebral Cortex*, 17(3), 679–690.
- Theeuwes, J. (1991). Exogenous and endogenous control of attention – The effect of visual onsets and offsets. *Perception & Psychophysics*, 49(1), 83–90.
- Theeuwes, J. (1994). Stimulus-driven capture and attentional set – Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology-Human Perception and Performance*, 20(4), 799–806.
- Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, 9(5), 379–385.
- Theeuwes, J., Belopolsky, A., & Olivers, C. N. (2009). Interaction between working memory, attention and eye movements. *Acta Psychologica*, 132(2), 106–114.
- Thurlow, W. R., & Jack, C. E. (1973). Certain determinants of ventriloquism effect. *Perceptual and Motor Skills*, 36(3), 1171–1184.
- Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology*, 40 A(2), 201–237.
- Van der Burg, E., Talsma, D., Olivers, C. N. L., Hickey, C., & Theeuwes, J. (submitted). Early multisensory interactions affect the competition among multiple visual objects.
- Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology-Human Perception and Performance*, 34(5), 1053–1065.
- van der Lubbe, R. H. J., & Postma, A. (2005). Interruption from irrelevant auditory and visual onsets even when attention is in a focused state. *Experimental Brain Research*, 164(4), 464–471.
- van Zoest, W., Donk, M., & Theeuwes, J. (2004). The role of stimulus-driven and goal-driven control in saccadic visual selection. *Journal of Experimental Psychology-Human Perception and Performance*, 30(4), 746–759.
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428(6984), 748–751.
- Vroomen, J., & de Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology-Human Perception and Performance*, 26(5), 1583–1590.
- Vroomen, J., & de Gelder, B. (2004). Perceptual effects of cross-modal stimulation: The cases of ventriloquism and the freezing phenomenon. In G. Calvert, C. Spence, & B. E. Stein (Eds.), *Handbook of Multisensory Processes* (pp. 141–150). Cambridge, MA: MIT.
- Vroomen, J., Bertelson, P., & de Gelder, B. (2001). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychologica*, 108(1), 21–33.
- Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, 63(4), 651–659.
- Wallace, M. T., Meredith, M. A., & Stein, B. E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research*, 91(3), 484–488.

- Wallace, M. T., Meredith, M. A., & Stein, B. E. (1998). Multisensory integration in the superior colliculus of the alert cat. *Journal of Neurophysiology*, *80*(2), 1006–1010.
- Ward, L. M. (1994). Supramodal and modality-specific mechanisms for stimulus-driven shifts of auditory and visual attention. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale*, *48*(2), 242–259.
- Ward, L. M., McDonald, J. J., & Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Perception & Psychophysics*, *62*(6), 1258–1264.
- Weissman, D. H., Warner, L. M., & Woldorff, M. (2004). The neural mechanisms for minimizing cross-modal distraction. *The Journal of Neuroscience*, *24*(48), 10941–10949.
- Woodman, G. F., & Luck, S. J. (1999). Electrophysiological measurement of rapid shifts of attention during visual search. *Nature*, *400*(6747), 867–869.
- Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention – Evidence from visual-search. *Journal of Experimental Psychology-Human Perception and Performance*, *10*(5), 601–621.
- Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention – voluntary versus automatic allocation. *Journal of Experimental Psychology-Human Perception and Performance*, *16*(1), 121–134.
- Zatorre, R. J., Mondor, T. A., & Evans, A. C. (1999). Auditory attention to space and frequency activates similar cerebral systems. *Neuroimage*, *10*(5), 544–554.