

Audiovisual Attention in Space

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

Beoordelingscommissie: Dr. Artem Belopolsky
Prof. Dr. Marc Ernst
Prof. Dr. Armin Kohlrausch
Prof. Dr. John van Opstal
Prof. Dr. Charles Spence
Dr. Durk Talsma

Paranimfen: Clayton Hickey
Michiel de Voogt

Cover design: Maarten Verweij (www.sinds1416.nl)

Druk: Ipskamp Drukkers (www.ipskampdrukkers.nl)



vrije Universiteit *amsterdam*




Nederlandse Organisatie voor Wetenschappelijk Onderzoek

VRIJE UNIVERSITEIT

Audiovisual Attention in Space

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Psychologie en Pedagogiek
op dinsdag 3 november 2009 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Thomas Koelewijn

geboren te Arnhem

promotor: prof.dr. A.W. Bronkhorst
copromotor: prof.dr. J.L. Theeuwes

Table of Contents

<i>Overview</i>	p 7
Chapter 1: <i>Introduction</i>	p 11
Chapter 2: <i>Auditory and Visual Capture During Focused Visual Attention (Koelewijn, Bronkhorst, & Theeuwes, in press)</i>	p 43
Chapter 3: <i>Competition between auditory and visual spatial cues during visual task performance (Koelewijn, Bronkhorst, & Theeuwes, 2009)</i>	p 77
Chapter 4: <i>Multisensory integration is more than directing attention (Koelewijn, Bronkhorst, & Theeuwes, in preparation)</i>	p 101
Chapter 5: <i>Priming T2 in a visual and auditory attentional blink task (Koelewijn, Van der Burg, Bronkhorst, & Theeuwes, 2008)</i>	p 113
<i>Samenvatting in het Nederlands</i>	p 141
<i>References</i>	p 146
<i>Dankwoord</i>	p 158
<i>Curriculum Vitae</i>	p 159
<i>Publications</i>	p 160

Overview

Theoretical framework

When performing a visual task such as reading a book our attention can be drawn to the location of an unexpected sound. In such a situation the irrelevant sound stream captures our attention and distracts us from our task. The processes underlying this phenomenon can be described in terms of *crossmodal attention*. Additionally, visual and auditory events like moving lips and speech can integrate into a single coherent event. This process is called *multisensory integration*. Both multisensory integration and crossmodal attention have a large influence on how we perceive the world. In these modern times, we are bombarded by audiovisual information also in situations in which we cannot afford to be distracted and make errors, such as when we are driving a car. In other situations, we may experience a considerable benefit when visual and auditory information coincide, such as when listening to a talker whose lips we can see. Therefore, it is important to understand the processes underlying multisensory integration and crossmodal attention and how these affect our performance.

The literature regarding multisensory integration and crossmodal attention is reviewed in chapter 1. The first section of this introduction describes the temporal and spatial constraints required for multisensory integration to occur and the regions of the brain where it may take place. The studies discussed show that depending on the content of audiovisual information, multisensory integration occurs at different brain levels. This section is followed by a review of studies that specifically look at crossmodal attention between the auditory and visual modalities. This topic is addressed by studies described in chapter 2 and chapter 3 of this thesis. Attention to a task in one modality can be drawn by information coming through another modality and vice versa suggesting that crossmodal attention can take place in brain areas that are not specific for the processing of either auditory or visual information. Additionally, some studies show the involvement of early sensory specific areas. Chapter 1 introduces the idea that multisensory integration and crossmodal attention sometimes act independently but

at other times interact. This issue is the main focus of the study described in chapter 4. To shed light on this issue, different theories regarding the level at which multisensory interactions takes place are discussed. The final section of the introduction focuses on the question whether audiovisual interactions and crossmodal attention in particular are *automatic* processes and therefore not affected by cognitive load. This question is addressed in the studies described in chapter 2 and chapter 3. The literature reviewed in chapter 1 shows that when a sound distractor is presented concurrently with a visual stimulus it will not draw visual attention. Therefore, crossmodal attention is considered not to be a fully automatic process.

Auditory capture

When visual attention is drawn to the location of a sound the processing of a successive visual event at that location is sped up, ultimately resulting in a fast response. This process is called *auditory capture* and it represents one form of crossmodal attention. In the studies presented in chapter 2 and 3 auditory capture was shown by means of a cueing task. In this task participants had to detect a white dot that was presented to the left or right side of a central fixation point, slightly above or below the horizontal midline. Irrespective of its location, participants had to indicate whether this dot appeared above or below the horizontal midline. Shortly before the presentation of the dot a sound was presented on either the left or right side of the screen. In half the trials, this sound and dot were presented at the same side and the other half at opposite sides. During the task participants did not know on which side the visual target would appear and were told to ignore the auditory cue. Participants were faster in responding to the visual target when a sound was presented at the same side and slower when a sound was presented at the opposite side. This difference in performance time is called a crossmodal cueing effect and is a measure for auditory capture.

The question as to whether auditory capture disappears when we know on which side this dot is going to appear is described in chapter 2. In this study participants performed a task similar to the one described above. However, this

time before the sound and target were presented, an arrow was shown in the middle of the screen. This arrow validly indicated the side at which the dot would appear. Although participants did not know whether the dot would appear above or below the horizontal midline they could already focus their visual attention on the correct side. In this task participants showed a cueing effect for sounds even though the arrow validly indicated the location of the dot. Therefore, we conclude that people are not able to suppress auditory capture, even when they know at what location relevant visual information will be presented.

The question whether auditory capture disappears when a visual event is presented at the same time as a sound is investigated in chapter 3. Visual events are known to capture attention. In this study we tested whether auditory capture would survive when a visual and auditory event compete for attentional resources. Participants performed a cueing task in which the dot was preceded by both a visual and auditory cue. When the two cues were presented at chance level, visual as well as auditory capture were observed. However, when the validity of the visual cue was increased to 80% only visual capture occurred and no auditory capture. We conclude from these results that a competing predictive visual event is able to suppress auditory capture. Therefore, auditory capture is not considered to be a truly automatic process. This means that we now know that there are ways to minimize the distracting effects of sounds.

Multisensory integration

It has been suggested that benefits of multisensory integration are due to a more efficient allocation of spatial attention. Other studies suggested that multisensory integration and spatial attention act independently. In chapter 4 we investigated whether audiovisual integration can alter the perception of a visual event in conditions in which there is no role for spatial attention. We presented a visual target at fixation together with a spatially diffuse auditory cue. In the first experiment we show by means of a staircase procedure that sound lowers the visual contrast detection threshold. In the second experiment participants performed a detection task to rule out criterion shifts. The results show an increased sensitivity

for the detection of a visual event when this event is accompanied by a sound relative to when no sound is present. We conclude that multisensory integration can improve the detectability of a visual event independent of spatial attention.

Attentional blink

An *attentional blink* occurs when people have to report two target items (e.g., words or single characters) presented in rapid succession at the same spatial location. People are accurate in reporting the first target (T1) but often fail to report the second target (T2) when the second target is presented between 200-500 ms after the first one. In chapter 5 experiments are described in which participants performed an attentional blink task containing digits as targets and letters as distractors within the visual and auditory domain. Prior to the rapid serial visual presentation a visual or auditory prime was presented in the form of a digit, which was identical to the T2 on 50 percent of the trials. The results not only showed an attentional blink but also an overall drop in performance on T2 for the trials on which the stream was preceded by an identical prime from the same modality. There was no crossmodal priming suggesting that the observed inhibitory priming effects are modality specific. These findings are assumed to represent a special type of negative priming operating at a low feature level.

Chapter 1

Introduction

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 stop in your tracks.

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 eavesdropping 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 attentional capture when spatial attention is drawn to a location in space against our intentions (Theeuwes, Belopolsky, & Olivers, in press). 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 thesis 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 (crossmodal capture) under all circumstances. Recent studies have shown that in some circumstances audiovisual interactions like crossmodal capture do not occur (Koelewijn, Bronkhorst, & Theeuwes, 2009; Santangelo & Spence, 2007) while other studies have shown that in most circumstances irrelevant sounds do affect our visual system (e.g., Koelewijn, Bronkhorst, & Theeuwes, in press; 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, 2005b). 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, there are also processes that do not interact. For example, a process like negative priming where a repeated event is inhibited does not occur across modalities (see Koelewijn, Van der Burg, Bronkhorst, & Theeuwes, 2008). Additionally, 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 introduction 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 theories 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.

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 our visual percept by lowering our visual detection threshold (see Koelewijn, Bronkhorst, & Theeuwes, in prep). Even though some multimodal behavioral effects and illusions resulting from multisensory integration were reported as early as the 1960s and 1970s (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 & 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. This time window is characteristic for multisensory integration but not for attentional, alerting or preparatory effects (see e.g., Los & Schut, 2008; Niemi & Näätänen, 1981), which are known to operate at longer time intervals. Therefore, these results are a strong indication that the ‘illusory flash effect’ is based on 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 participants had the illusion that the target stayed 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 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 and colleagues (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 salience 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 constrains

Our perceptual system effortlessly integrates 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 and colleagues (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 and colleagues (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, Bergmann, Hake, Heinze, & Fendrich, 2008; Stein, London, Wilkinson, & Price, 1996). For example, a study by Stein and colleagues (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 suggest that spatial constraints for multisensory integration only hold for peripheral visual events and not for central visual events. In line with this, Lippert and colleagues (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 and colleagues (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 and colleagues (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 1960s 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 (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 et al., 1987; Meredith & Stein, 1996; Wallace, Meredith, & Stein, 1998).

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 et al., 2000; Martuzzi et al., 2007; Romei, Murray, Merabet, & Thut, 2007; Shams,

Iwaki, Chawla, & Bhattacharya, 2005). For instance Shams and colleagues (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 & 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.

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., 2005; Shams et al., 2001), which suggests multisensory integration at multiple levels. In the next paragraph we discuss crossmodal spatial attention, the neural correlates of crossmodal attention, and how crossmodal attention interacts with multisensory integration.

3. Crossmodal spatial 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 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 Noble (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 distracters 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. 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.

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 or 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 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. This illusion occurs when prior to the presentation of the line attention is captured by means of a 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 and colleagues (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). 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, & 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 & Ward, 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 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 & Ward, 2000; Ward et al., 2000) clearly demonstrated that auditory input can affect visual spatial attention and vice versa. Similar crossmodal attentional effects have been shown 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 visa versa. However, later studies (McDonald & Ward, 2000; Ward et al., 2000) attributed this asymmetry in crossmodal cueing to some particularities of the Spence and Driver (1997) paradigm and 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 and colleagues (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 perisylvian 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 and colleagues (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 et al., 2007; Weissman, Warner, & Woldorff, 2004) show activation in both heteromodal and early sensory brain areas. An fMRI study by Weissman and colleagues (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 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 and colleagues (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 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 and colleagues (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.

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, 2001b). McDonald and colleagues (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. 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 to modulate multisensory integration (e.g. Busse, Roberts, Crist, Weissman, & Woldorff, 2005; Talsma & Woldorff, 2005b). A third option might be that multisensory integration occurs at multiple stages in a more parallel fashion (Calvert & Thesen, 2004).

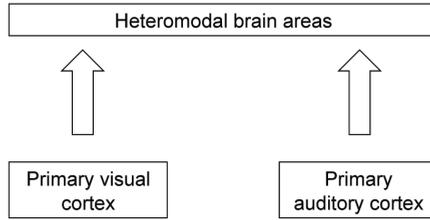
The late integration framework (Figure 1.1a) states that unimodal attention affects the individual sensory input and integrates them at a late stage into a single precept. Thus auditory and visual events are first individually enhanced by means of unimodal attention before integrating at a higher heteromodal 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 (2005b) 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.

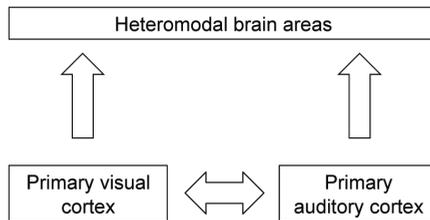
The early integration framework (Figure 1.1b) states that multisensory integration occurs at an early sensory level and at a later stage amodal attention is captured. Therefore, this framework suggests that multisensory integration is independent of attention. Even though independent, bimodal cues can 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, Ho, & Spence, 2008).

The parallel integration framework (Figure 1.1c) 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).

(a) Late integration framework



(b) Early integration framework



(c) Parallel integration framework

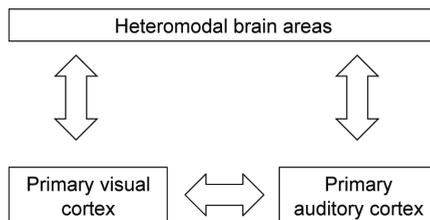


Figure 1.1. Schematic representation of; a) a late integration framework, b) an early integration framework, and c) a parallel integration framework.

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 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 result 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.

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 superadditive effect at an attentional level (e.g., Koelewijn et al., 2009; Santangelo, Van der Lubbe, Belardinelli, & Postma, 2006, 2008). This was shown in a study by Santangelo and colleagues (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., 2009). In a follow up study

Santangelo, Van der Lubbe, and colleagues (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 superadditive 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.

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, 2001b). 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? 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 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, 2005b). 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 crossmodal attentional capture 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; Muller & 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 crossmodal capture. Recently several studies (Koelewijn et al., in press; 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 and colleagues (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 and colleagues (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., in press) we tested how focused visual attention affects crossmodal 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 an 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., in press; 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., in press), 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 crossmodal capture (Koelewijn et al., in press; Mazza et al., 2007; van der Lubbe & Postma, 2005). Focusing attention on an upcoming target location prior to the presentation of the crossmodal exogenous cue did not affect attentional capture by this cue. However, when visual attention is centrally focused by means of an additional task, no crossmodal 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., in press; 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 crossmodal auditory capture (Koelewijn et al., 2009). 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., in press; Santangelo & Spence, 2007) imply that crossmodal capture is not an automatic process.

4.3 Discussion

Several studies indicate that crossmodal capture meets the intentionality criterion (Koelewijn et al., in press; 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., 2009). 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 and colleagues (2009) and Santangelo & 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 and colleagues (2009) 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., in press) nor purely exogenous cues (Koelewijn et al., 2009) 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 et al., 2008) or a peripheral onset (Koelewijn et al., 2009). The reason for this can be explained by means of the parallel integration framework.

The parallel integration framework of Calvert and Thesen (2004) (see Figure 1.1c) 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 a 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, 2005b). 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., 2009, in press). Only when both processes are affected at the same time by a competing event is auditory capture entirely extinguished (Koelewijn et al., in press; Santangelo et al., 2007; Santangelo & Spence, 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 example, although audiovisual events or multisensory events in general are well suited as for instance warning signals, when giving too many false alarms they can become distracting.

Chapter 2

Auditory and visual capture during focused visual attention

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (in press). Auditory and visual capture during focused visual attention. *Journal of Experimental Psychology: Human Perception and Performance*.

Abstract

It is well known that auditory and visual onsets presented at a particular location can capture our visual attention. However, the question whether or not such attentional capture disappears when attention is focused endogenously beforehand has not yet been answered. Moreover, previous studies have not differentiated between capture by onsets presented at a non-target (invalid) location and possible performance benefits occurring when the target location is (validly) cued. In this study we modulated the degree of attentional focus by presenting endogenous cues with varying reliability and by displaying placeholders indicating the precise areas where the target stimuli could occur. By using not only valid and invalid exogenous cues, but also neutral cues, which provide temporal but no spatial information, we found performance benefits as well as costs when attention is not strongly focused. The benefits disappear when the attentional focus is increased. Our results indicate that there is bottom-up capture of visual attention by irrelevant auditory and visual stimuli that cannot be suppressed by top-down attentional control.

Introduction

Covert visual attention can be directed to a specific location in the visual world without making eye movements. This can happen voluntarily by steering attention endogenously to that location or automatically, when attention is exogenously captured. Endogenous attention has been compared metaphorically to a spotlight that casts its light on relevant visual information (Broadbent, 1982; Posner et al., 1980) and that can be directed to a target location by presenting an informative cue prior to the target. An example of such a cue is an arrow displayed at the center of the visual field which points with a high probability (e.g., 80% valid) to a possible target location. Exogenous capturing of attention can for example be evoked by a visual onset occurring at the target location. Exogenous cues can shorten reaction times to targets even when the cues do not reliably predict (with a validity at chance level) the location of the upcoming target (e.g., Jonides, 1981; Yantis & Jonides, 1984). Exogenous cueing is not restricted to the visual modality alone: tactile (Posner, 1978; Spence & McGlone, 2001), and auditory cueing effects (Spence & Driver, 1994) have also been reported. Both endogenous and exogenous cues can cause an object appearing at an attended location to be detected faster and more accurately than an object appearing at an unattended location (Posner, 1980; Posner et al., 1980). However, peripheral exogenous cues capture attention automatically, while central endogenous cues seem to be less obligatory (e.g., Jonides, 1981; Theeuwes, 1991; Yantis & Jonides, 1984).

Both endogenous and exogenous cueing effects can be described by the allocation of attentional resources to a cued location. For endogenous cueing these attentional shifts are controlled in top-down fashion and for exogenous cueing these shifts are enforced by bottom-up processes. While these processes of endogenous and exogenous visual attention have mostly been studied in separate paradigms, there have also been studies that looked at the interactions between both attentional processes (Jonides, 1981; Muller & Rabbitt, 1989; Theeuwes, 1991; Yantis & Jonides, 1990). Theeuwes (1991) investigated the relation between exogenous and endogenous visual attention within a single paradigm. In this study

participants had to identify a target letter among three distractor letters all positioned equidistantly on an imaginary circle. An endogenous cue (a centrally presented arrow) reliably indicated the location of the target. In addition, a non-predictable exogenous visual onset cue was presented near one of the letters. When the central arrowhead was presented after the exogenous cue, attention was drawn to the location of the exogenous cue. But when the central arrowhead was presented prior to the presentation of the exogenous cue, attention was in a focused state and therefore the exogenous cue had no effect. These results show that exogenous capture of attention can cease to exist when attention is endogenously focused on a location in space. In other words, typically, exogenous events do not cause interference when presented outside the focus of attention (for similar results see also, Yantis & Jonides, 1990). The observed suppression of visual exogenous events suggests that visual exogenous attention is not a completely automatic process.

Endogenous and exogenous cueing effects across modalities were demonstrated in two studies by Spence and Driver (1996; 1997). In these studies participants had to perform an orthogonal cueing task in which they had to make an elevation judgment (up vs. down) regarding auditory or visual targets presented to the left or right of fixation. Loudspeakers and LED's, mounted in front of the loudspeakers were used to generate target stimuli. In the first study (Spence & Driver, 1996), participants were presented with an endogenous cue (a central arrowhead) indicating that the target was more likely to appear on the side indicated by the cue. When the cue was valid, participants were faster to make an elevation judgment to that side, regardless of the modality in which the target was presented. In the second study (Spence & Driver, 1997) the target side was cued (exogenously) at chance level by a visual (LED onset) or auditory (pure tone) cue presented at the same eccentricity as the targets. For the crossmodal conditions ('visual cue and auditory target' or 'auditory cue and visual target'), there was only a cueing effect when a visual target was preceded by an auditory cue but not vice versa. For the unimodal conditions ('visual cue and visual target' or 'auditory cue and auditory target') both visual and auditory cueing effects were found. Spence,

McDonald, and Driver (2004) attribute this observed asymmetry to a higher spatial resolution of the visual compared to the auditory perceptual system. Presumably, this difference in resolution is associated with a corresponding difference in the size of the spatial area that is attended. According to Spence and colleagues, “*when testing for visual-upon-auditory effects, the auditory targets were in effect presented too far away in external space (in elevation) from the preceding visual cue for any crossmodal cuing effect to have been observed*” (p. 286). In other crossmodal cueing studies, McDonald and colleagues (McDonald & Ward, 2000; Ward et al., 2000) used a go/no-go task in which cues and targets were presented at the same spatial location. Since they now found a cueing effect for auditory targets preceded by visual cues, their results are consistent with the above explanation of Spence et al. for the earlier observed asymmetry in cross-modal cueing.

The finding of cueing effects across modalities raises the question whether unimodal and crossmodal cueing effects are based on similar processes. One way to investigate this issue is to test whether effects reported in unimodal conditions would also apply to cross-modal conditions. For example, if cross-modal attention is similar to unimodal attention, one would expect similar effects as reported by Theeuwes (1991). In other words, both auditory and visual exogenous events should cease to capture attention when visual attention is in a focused state.

In a recent study, Van der Lubbe and Postma (2005) tested this notion. They used a variation on the orthogonal cueing task. Participants had to perform a discrimination task on targets presented as arrowheads pointing up or down. These targets were displayed on LED grids placed on the left and right side at 28.3° or 19.3° visual angle relative to a centrally positioned LED grid used for fixation. One second prior to the onset of the target, the central grid displayed an arrow pointing to the correct target location (100 % valid) or displayed a non-informative (neutral) cue. An exogenous visual or auditory cue was presented 200 ms prior to the target, indicating the target location at chance level. Unlike Theeuwes (1991), Van der Lubbe and Postma (2005) found that abrupt onsets of both visual and auditory cues captured attention even in conditions in which observers were focused on the endogenously cued location. Note, however, that Van der Lubbe and Postma

(2005) used quite large visual angles between fixation and targets (28.3° or 19.3°), while the angle in the Theeuwes (1991) experiment was only 4.2° . Even though Van der Lubbe and Postma (2005) failed to replicate Theeuwes' original results (1991), they found results for unimodal and crossmodal cueing that are comparable to each other. Therefore, their results are consistent with the idea that unimodal and crossmodal cueing basically operates according to the same underlying mechanisms.

In a recent study by Mazza, Turatto, Rossi, and Umiltà (2007) participants performed an orthogonal crossmodal cueing task similar to the one used by Spence and Driver (1997). In their first experiment the target side was random and results (for a 150 ms cue-target interval) showed both unimodal visual and auditory cueing effects, but only a crossmodal cueing effect when a visual target was preceded by an auditory cue. These results replicate the crossmodal cueing asymmetry shown by Spence and Driver (1997). In their second experiment the target side remained the same during an entire block. Thus, participants knew at which location the target would appear while they were also presented with non-predictive exogenous cues. The results (for the 150 ms cue target interval) show a crossmodal cueing effect for visual targets preceded by valid (512 ms) or invalid (530 ms) auditory cues. Surprisingly, an opposite effect was found when auditory targets were preceded by valid (629 ms) or invalid (586 ms) visual cues. Interestingly, no unimodal cueing effects were observed, this in contrast to the results of their first experiment. Mazza and colleagues (2007) therefore did not find a unimodal cueing effect when attention was in a focused state, basically replicating Theeuwes' (1991) results. In addition they showed that when a visual target is preceded by an auditory cue, the crossmodal cueing effect holds even during visual focused attention. Thus, while these results differ from those of Van der Lubbe and Postma (2005) for the unimodal conditions, both studies agree that exogenous crossmodal cueing cannot be suppressed by focused attention. However, it is not yet clear why results for crossmodal cueing should deviate from the classic visual exogenous cueing effects demonstrated by Theeuwes (1991), and Yantis and Jonides (1990).

A recent study by Santangelo and Spence (2007), used an orthogonal cueing paradigm similar to the one used by Spence and Driver (1997). Their design, however, only contained elevation judgments concerning visual targets. In addition to the elevation judgment task, a second task was introduced to keep the participants' attention focused to the centre of the display. In this additional task, observers had to respond to digits embedded in a stream of letters presented in a rapid serial visual presentation (RSVP). In the high-load condition, a target digit was presented centrally in 67% of the trials and a peripheral target for the elevation judgment task was presented in the remaining 33% of the trials. In the no-load condition no RSVP stream was presented. In all trials a peripheral visual, auditory, or bimodal (visual and auditory) exogenous cue was presented on the left or the right side. With respect to the position of the peripheral targets, these cues could either be valid or invalid. The results for the no-stream condition showed auditory, visual, and bimodal cueing effects. However, in the high-load condition only a bimodal cueing effect was observed. These results indicate that unimodal and crossmodal exogenous cueing effects disappear when, in addition to the cueing task, participants focus their attention on an RSVP stream containing a possible target. Santangelo and Spence (2007) conducted an additional experiment confirming that disengagement from the RSVP stream could be ascribed to multisensory integration instead of the double perceptual input provided by the bimodal cue. In this experiment redundant visual and auditory cues were compared to the bimodal cue. Again, these outcomes only showed a cueing effect for bimodal cues. The authors conclude that unimodal but not bimodal exogenous cueing effects can be suppressed by means of endogenous focused attention.

To summarize; Van der Lubbe and Postma (2005) showed that there are both crossmodal and unimodal exogenous cueing effects when attention is focused endogenously, in contrast to Theeuwes (1991) who found no unimodal cueing effects and Santangelo and Spence (2007) who found suppression of both unimodal and crossmodal exogenous cueing. Mazza and colleagues (2007) show suppression of unimodal but not of crossmodal cueing. It is not clear to what degree these results can be explained by methodological differences. Important factors may be

where and how strongly attention was focused endogenously but these factors do not seem to correlate well with the results. For example, Santangelo and Spence (2007) not only required participants to focus their attention close to fixation (while all other studies used peripheral locations) but also did this in a way that presumably caused stronger attentional focus. Although this might explain why their results for crossmodal cueing deviate from those of the others, it is then difficult to understand why their results for within-modality cueing are in essence the same as those of Theeuwes (1991) and Mazza and colleagues (2007). Another issue is that all studies quantify attentional capture by taking the difference between response times for valid and invalid exogenous cues. It is typically assumed that effects of endogenous and exogenous cueing are due to attention *shifts* to or away from the target (Jonides, 1981; Posner, 1980; Spence & Driver, 1996). However, it is not clear whether this holds for the observed cueing effects found in these and other studies (e.g., McDonald & Ward, 2000; Spence & Driver, 1997). More specifically, one can ask the question whether crossmodal cueing effects are the result of ‘costs’ by attention being drawn away (captured) to an invalid cue location, and/or of ‘benefits’ of attention being captured by a valid cue location. No previous attempts have been made to separate these two components, although it seems evident that a factor such as the strength of the (endogenous) attentional focus will affect them differently. A further issue that complicates comparison of the above studies is that eye movements were not always registered (e.g., Mazza et al., 2007). In particular when participants are instructed to endogenously focus their attention on a location prior to the presentation of a target, one cannot exclude the possibility that they will make eye movements (thus introducing a confounding factor).

In order to shed more light on these issues we have conducted a series of experiments based on the classic orthogonal cueing paradigm introduced by Spence and Driver (1994). We used both exogenous and endogenous cues and we modulated the degree to which observers focused their attention on the target location. This modulation was achieved by changing the validity of the endogenous cue and by using placeholders that indicate the precise spatial region where the

target is about to appear. Furthermore, we included a baseline condition in which the exogenous cue provided temporal, but no (reliable) spatial information (for elaborate explanation on this topic see, Jonides & Mack, 1984). This condition enables us to separate performance costs due to invalid cues from benefits caused by valid cues. A spatially neutral auditory cue was created by simultaneously presenting two uncorrelated noise bursts from two loudspeakers, which in our setup were located to the left and right of the monitor on which the visual stimuli were presented. This causes a broad spatial percept in front of the participant, which extends to the sides beyond both loudspeakers (Blauert, 1997). Note that the use of uncorrelated signals is essential here because two *correlated* signals are perceived as a easily localizable sound in the middle between the two loudspeaker positions (in our case this would be directly in front of the participant) due to summing localization (Blauert, 1997). This principle for creating spatially neutral auditory cues was used in all our experiments. In our final experiment, we also presented a neutral visual cue, which was created by simultaneously displaying cues on both possible target locations. In all experiments except the first one (which did not include endogenous cues), eye movements were monitored to make sure all observed cueing effects could be attributed to covert attention

In our first experiment we validated our paradigm and specific setup by replicating the crossmodal cueing effects found earlier by Spence and Driver (1997) and others. Additionally, we introduced the spatially neutral auditory cue that allowed us to specify the observed cueing effect in terms of ‘costs’ and ‘benefits’. In Experiments 2 and 3 endogenous cueing was added by means of a central arrowhead presented prior to the auditory cue indicating the correct target side in 80% or 100% of the trials, respectively. In Experiment 4 we investigated whether the addition of placeholders, which are assumed to induce an even stronger attentional focus on the position of the target side prior to its appearance, would alter the previously observed cueing effect. Finally, in Experiment 5 we directly compared effects of crossmodal (auditory) and unimodal (visual) cueing in conditions with and without endogenous focus of visual attention.

It should be noted that there is an ongoing discussion concerning the use of arrowheads as endogenous cues (Hommel, Pratt, Colzato, & Godijn, 2001; Santangelo & Spence, 2008). We are aware of the fact that some studies show that arrowheads can also have an exogenous cueing effect when presented at chance level. This is probably due to the fact that overlearned symbols are almost automatically processed (Hommel et al., 2001) and are therefore directing attention partly bottom-up. Possibly, there are also exogenous effects of the arrowheads that we have used but because they were presented at least 650 ms before the target and were always followed by an exogenous location cue, it is unlikely that they have influenced our results.

Experiment 1

The task in this experiment was similar to the orthogonal cueing task used by Spence and Driver (1997). Instead of using LED's, visual stimuli were presented on a computer screen. The loudspeakers that generated the auditory cues were located to the left and right of the computer screen (for a comparable setup see Mondor & Amirault, 1998; Talsma & Woldorff, 2005b). We used only auditory cues in combination with visual targets, and presented the cues at an SOA of 200 ms, because this condition yielded a large cueing effect in earlier studies (e.g., Spence & Driver, 1997). In our paradigm we also used an auditory spatially diffuse cue, which served as a neutral condition, i.e., it did not seem to emanate from a specific direction.

Method

Participants. Twelve students of the Vrije Universiteit Amsterdam (6 male, mean age 21.4 years old, ages between 18 and 28) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing. Participants were informed about the experimental procedure and were naïve as to the purpose of the experiment.

Apparatus and Design. Participants were seated in a dimly lit room at approximately 80 cm distance from a computer screen (CRT, 17 inch, 120 Hz). The experiment was run in E-Prime 1.1 (SP3). To the left and to the right of the screen, a loudspeaker was placed at an angle of 18.3 degree from fixation and both loudspeakers were aligned to the vertical middle of the screen. The experiment consisted of five blocks containing 36 trials each. There was a valid condition (33% of the trials) where the auditory cue and visual target were presented on the same side, an invalid condition (33%) where the auditory cue and visual target were presented on opposite sides, and a neutral condition (33%) where the auditory cue could not be assigned to a specific location in space (but still provided the same temporal information as the other cues). All conditions were presented randomly within blocks; the first block was for practice purposes leaving 48 trials for each condition.

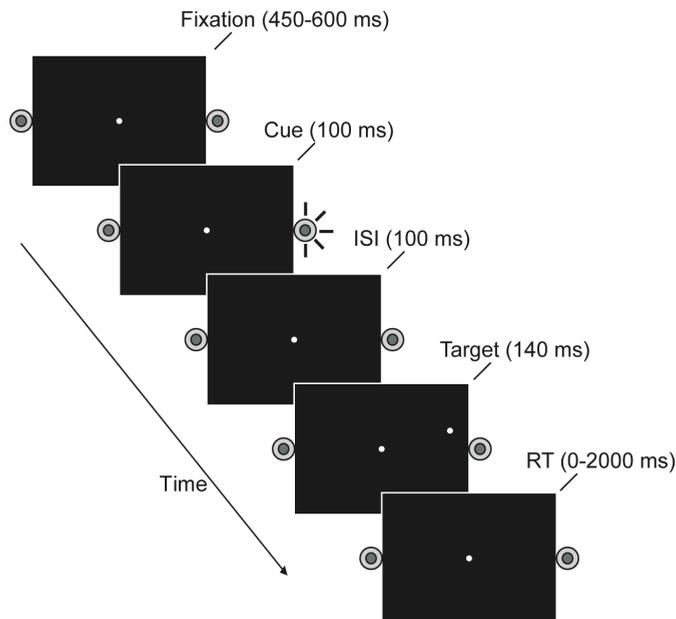


Figure 2.1. Schematic representation of the paradigm used. Participants performed an orthogonal cueing task where they had to discriminate between targets presented above or below the vertical middle of the screen. Targets were presented on the left or right side of the screen and were preceded with an SOA of 200 ms by a non-predictable auditory cue coming out of the left or right loudspeaker.

Procedure and Stimuli. Figure 2.1 gives an example of a typical trial. At the beginning of each trial, a white fixation dot (diameter 0.2°) appeared and stayed on screen until a response was made. Participants were instructed to fixate on this dot during the entire trial and to refrain from making eye movements. After a random delay time of 400 ms to 650 ms, an auditory cue consisting of a white noise burst was presented for 100 ms. This cue came equiprobably from the left or right loudspeaker and was valid or invalid with respect to the target location, or it was neutral and came from both loudspeakers at the same time. In the latter case, two uncorrelated noise bursts were used. The cues presented through a single loudspeaker were boosted by an extra 3 dB, to create the same subjective loudness as the neutral cue that was presented through two loudspeakers. Two hundred ms after the onset of the auditory cue a visual target consisting of a white dot (diameter 0.2°) was presented for 140 ms. There were four possible target locations; two locations were positioned 10.5° to the left of fixation, and two locations 10.5° to the right of fixations. The two locations at each side were positioned above each other, one 2.4° above and the other one 2.4° below the vertical centre of the screen. The target appeared at one of these four locations at chance level. The participants' task was to report in a speeded but accurate fashion whether the target appeared above or below the vertical centre of the screen by pressing the number 8 or number 2 key, on the number pad of a QWERTY keyboard, respectively. Participants responded with both index fingers and were free to choose which finger to use for which button as long as they kept it the same during the experiment. Because of the task's orthogonal design there was no need to balance out for possible motor response effects. Responses had to be made within a time window of 2000 ms after target onset. After the response, participants received feedback when they had made an error – the fixation dot then turned red for 150 ms. After each trial, an interval of 850 ms followed before the next trial started. Following each block participants received feedback in the form of a percentage correct score and a mean reaction time. At the beginning of the experiment, participants were told that the auditory cues would provide no information about the location of the targets and therefore could be ignored.

Results

Premature (< 200 ms) and slow (> 1000 ms) responses (in total 1.3%) were removed from further analysis. For the remaining trials mean reaction times for the correct response trials (92.6%) were calculated for each subject for each condition. Figure 2.2 presents the mean reaction time for each condition (valid 343 ms, invalid 365 ms, and neutral 352 ms) averaged over subjects. The error bars in this figure represent the .95 confidence interval (5.8 ms) for the exogenous cueing main effect, following Loftus and Masson (1994). An overview of the mean reaction times, their standard deviations, and the mean error scores, for each condition and for all experiments is shown in Table 1. An ANOVA on RT with cue validity (valid, invalid, neutral) as a factor revealed a significant effect [$F(2,22) = 15.808$, $MSE = 92.210$, $p < .001$]. Three pairwise 2-tailed t-tests between the cueing conditions were conducted. Valid compared to invalid ($p < .001$), neutral compared to valid ($p = .006$), and neutral compared to invalid conditions ($p = .021$) all differed significantly.

The mean error scores (valid 5.9%, invalid 8.3%, and neutral 8.1%) were also calculated. An ANOVA on error scores with cue validity (valid, invalid, neutral) as a factor showed no reliable effect.

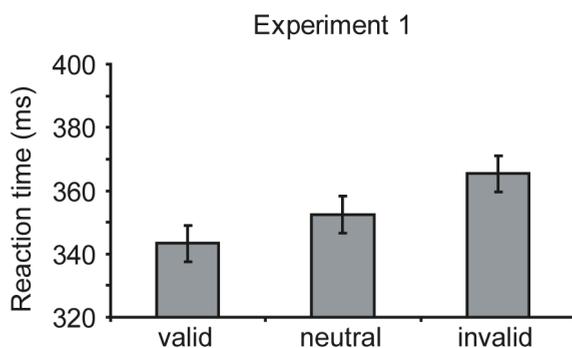


Figure 2.2. Results of Experiment 1 presented as an average reaction time (ms) for each condition (valid, neutral, and invalid). The error bars show the .95 confidence intervals for the exogenous cueing main effect (Loftus & Masson, 1994).

Discussion

The current experiment shows a cueing effect similar to that observed by Spence and Driver (1997). An extension with respect to earlier studies was that it included a neutral condition allowing separation of reaction time costs and benefits. Our results indicate that the typical costs and benefits of crossmodal cueing can be attributed to shifts of spatial attention just as in the classic way of explaining unimodal visual cueing effects (Jonides, 1981; Posner, 1980; Spence & Driver, 1996). The explanation goes as follows: First, directing attention is a fast but not an instantaneous process. In order to direct attention, it has to be disengaged from its old location, shifted to a new location, and then engaged on this new location (Posner et al., 1980). When a target location is validly cued, attention is already directed to the target location prior to the presentation of the target. This results in more attentional resources at the target location allowing for easier processing of the target. The benefit of this is reflected in a faster response relative to the neutral condition. When on the other hand a target location is invalidly cued, attention is directed to a location opposite to the target location. This results in less attentional resources being available on the target location for processing of this target, which results in a higher response time relative to the neutral condition.

The results of Experiment 1 thus show that crossmodal cueing is associated with costs and benefits that can be explained by shifts of spatial attention. As discussed in the introduction, within the visual domain these costs can be suppressed when attention is focused prior to the presentation of the exogenous cue (Mazza et al., 2007; Theeuwes, 1991; Yantis & Jonides, 1990). In order to test whether this same principle holds for crossmodal cueing, we conducted a second experiment in which an endogenous visual cue in the form of a centrally presented arrowhead was shown prior to the exogenous auditory cue. Because this cue was a relatively small centrally presented symbol, and appeared at least 650 ms before the target, we did not expect that this cue exerted an additional exogenous effect (Jonides, 1981). One of the consequences of using an endogenous cue with a longer cue-to-target interval is that it gives participants time to make eye movements towards the target. When eye movements are made the effects are not

solely produced by covert visual attention. In order to control for possible overt orienting of attention eye movements were recorded in all the following experiments.

Experiment 2: Endogenous cue 80% correct

As explained above, an endogenous visual cue presented prior to the exogenous auditory cue was used in this experiment. The cue was an arrowhead presented in the center of the screen that allowed the participants to focus their visual attention on the target location before the onset of the peripheral auditory cue. When an auditory cue does not automatically capture attention, no costs on invalidly exogenous cued trials are expected. However, when the auditory exogenous cue is still able to capture attention even when participants are highly focused on a location in space, costs on invalidly cued trials are expected. If the cueing effects resulting from the presentation of an auditory exogenous event are only due to shifts of spatial attention, then one expects no benefits of the auditory cue when presented at a location at which participants are already focused. Therefore, we expect no benefits on valid trials. To ensure that participants indeed used the endogenous cue it was valid on the majority (80%) of trials.

Method

Participants. Twelve new students from the Vrije Universiteit Amsterdam (2 male; mean age 20.7; ages between 18 and 30) participated in the experiment.

Stimuli and Design. The task was basically identical to the one used in Experiment 1 except that at the start of each trial a visual cue, an arrowhead appointing to the left or right (width 0.5° and height 0.4°), was presented for 600 ms. In 80% of the trials this arrowhead indicated the side where the subsequent target would appear. The combination of two types of endogenous visual cues (valid or invalid) with three possible auditory cues (valid, invalid, and neutral) resulted in six conditions. A total of nine blocks containing 60 trials each were presented during the experiment. All conditions were presented randomly within

blocks and the first block was for practice purposes only. Thus, there were 128 trials for each exogenous cue condition (valid, invalid, and neutral) when the endogenous cue was valid (80%) and 32 trials for each exogenous cue condition when the endogenous cue was invalid (20%).

EOG recoding and analysis. The horizontal and vertical electrooculogram (EOG) were recorded bipolarly by electrodes located on the outer canthi of both eyes, and a pair of electrodes on the supraorbital and infraorbital ridge of the right eye, respectively. Recordings were made at a 500 Hz sampling rate. For detecting eye movements a spike detection algorithm was used (for a full description see Talsma & Woldorff, 2005a). In short, this algorithm uses a sliding time window (sliding with 2-ms steps) set to 100 ms in which the maximum amplitude differences are calculated between all possible time point combinations within the window. The maximum allowed amplitude difference was set to $70\mu\text{V}$. All trials showing larger amplitude differences, from the onset of the visual cue till the offset of the target, were excluded from analysis. An amplitude range of $70\mu\text{V}$ filtered out trials containing eye movements bigger than 5° horizontal angle (Peelen, Heslenfeld, & Theeuwes, 2004), less than half the angle needed to focus on the target location. This range prevented leaving out trials containing small muscle artifacts unrelated to eye movements. To reduce the possible loss of trials based on eye blink artifacts (causing large amplitude differences sometimes hard to distinguish from eye movements), an intertrial interval of 2000 ms instead of 850 ms was used and participants were instructed to blink their eyes during this period.

Results

Trials containing eye movements (8.6%) and premature (< 200 ms) or slow (> 1000 ms) responses (0.6%) were removed from further analysis. For the remaining trials there were on average 95.3% correct responses. For these trials we calculated per subject the mean reaction times for each condition. The mean reaction times for each condition averaged over subjects are plotted in Figure 2.3. The error bars in this figure represent the .95 confidence interval (6.4 ms) for the exogenous cueing main effect. An ANOVA conducted on the RT with visual cueing (valid and

invalid) and auditory cueing (valid, invalid, and neutral) as the within-subject variables showed a main effect of visual cueing [$F(1,11) = 48.749$, $MSE = 408.216$, $p < .001$] and of auditory cueing [$F(2,22) = 30.052$, $MSE = 115.332$, $p < .001$]. In addition the interaction between visual and auditory cueing was reliable [$F(2,22) = 4.781$, $MSE = 69.451$, $p < .05$].

Significant differences in the ‘visual valid’ condition were shown for auditory valid (349 ms) compared to auditory invalid (372 ms; $p < .001$), auditory neutral (357 ms) compared to auditory valid ($p = .004$), and auditory neutral compared to auditory invalid ($p < .001$). For the ‘visual invalid’ condition, both auditory valid (378 ms) compared to auditory invalid (402 ms; $p < .001$), and auditory neutral (398 ms) compared to auditory valid ($p = .003$) showed a significant effect. No effect for auditory neutral compared to auditory invalid (402 ms) was found ($t < 1$).

These results suggest that the observed 2-way interaction is due to the fact that

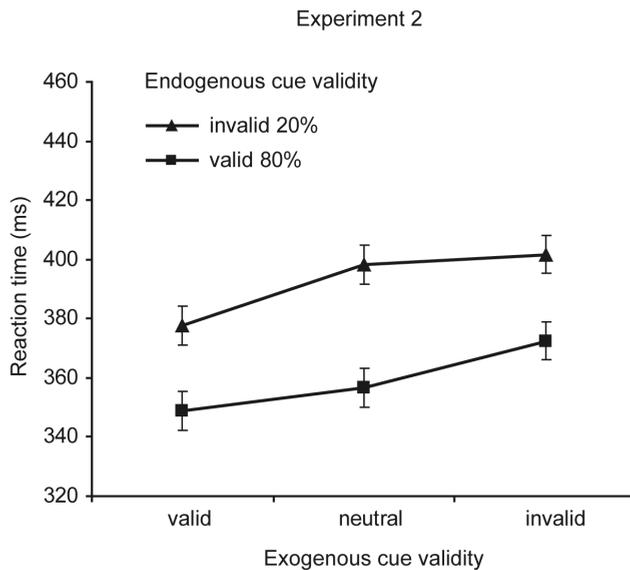


Figure 2.3. Results of Experiment 2 presented as an average reaction time (ms). The endogenous visual cue conditions (valid 80% and invalid 20%) are plotted as separate lines. The exogenous auditory cue conditions (valid, neutral, and invalid) are plotted along the x-axis. The error bars show the .95 confidence intervals for the exogenous cueing main effect.

the effect of the visual cue is different in the neutral condition than in the conditions in which the auditory cue provided location information. To investigate this notion an additional ANOVA with visual cueing (valid and invalid) and auditory cueing (valid and invalid) as within subject variables was conducted. An endogenous visual cueing effect [$F(1,11) = 30.729$, $MSE = 327.475$, $p < .001$] as well as an exogenous auditory cueing effect [$F(1,11) = 50.906$, $MSE = 134.369$, $p < .001$] were found. The two-way interaction between visual and auditory cueing was non-significant ($F < 1$), suggesting that both cueing effects occur independently from one another.

The mean error scores for the valid endogenous cueing condition (auditory cue valid 3.0%, neutral 3.6%, invalid 5.2%) and the invalid endogenous cueing condition (auditory cue valid 6.0%, neutral 8.3%, invalid 8.0%) were also calculated. An ANOVA showed a significant main effect for visual cueing [$F(1,11) = 5.691$, $MSE = .004$, $p < .05$], no effect for auditory cueing [$F(2,22) = 2.014$, $MSE = .001$, $p = .157$], and no interaction effect between visual and auditory cueing ($F < 1$).

Discussion

Even though participants focused their visual attention on a location in space prior to the presentation of the auditory cue, there was still a reliable auditory cueing effect. These effects replicate earlier results of Van der Lubbe and Postma (2005) and are in line with the results found by Mazza and colleagues (2007). Furthermore, a strong endogenous visual cueing effect was found, confirming that the participants used this cue to enhance their performance.

When we compare the results for the condition in which the endogenous visual cue was valid with those of Experiment 1, we see that the exogenous cueing effects are similar. When the locations of the auditory cue and target are congruent (valid cue condition) there is a decrease in RT relative to the neutral cue condition, and when the locations of auditory cue and target are incongruent (invalid cue) there is an increase relative to the neutral condition. Thus, the data suggest that even when participants are focused on the target location they still benefit from an additional

valid auditory cue. There are also RT costs in the invalid auditory cue condition, indicating that the exogenous cue is able to capture attention while attention is focused. In contrast, for conditions in which the endogenous visual cue was invalid, only a valid auditory cue had an effect compared to the neutral cue condition. In other words, when attention is focused on a non-target location, an additional invalid auditory cue (which is congruent with the invalid visual cue) has no further effect on response times. However, there is in that case a large effect of the valid auditory cue, suggesting that the exogenous cue helps participants to disengage their attention from the “wrong” location. Thus the results indicate that sounds facilitate the disengagement of attention from a location, which is beneficial when attention is needed somewhere else.

In summary we find that in both the visual valid and the visual invalid condition, auditory cues from a location opposite to the attentional focus are able to capture attention. In the case of a valid endogenous visual cue this will have a cost, and in the case of an invalid endogenous cue this will yield a benefit. In the valid visual condition we observe an extra benefit of the valid auditory cue indicating that attention was not completely focused by the endogenous cue alone. This finding can possibly be explained by the fact that the endogenous cue was only valid in 80% of the trials. Yantis and Jonides (1990) have shown that the validity of the endogenous cue strongly influences its ability to suppress exogenous cueing. In their study visual exogenous cueing effects were only suppressed when a 100% valid endogenous cue was used (see for similar results, Theeuwes, 1991). However, when the endogenous cue had a 75% validity an exogenous cueing effect was still observed. According to Yantis and Jonides (1990), the uncertainty concerning the validity of the visual cue could have influenced the way in which participants focused their attention. If in the current experiment participants attention was not fully focused because of this uncertainty, the auditory cue could have improved this focus, resulting in better performance (see also, Muller & Rabbitt, 1989). In other words, both the observed costs and benefits can be explained in terms of attention not being completely focused before the location was indicated by the cue. To test this hypothesis a third experiment was performed

in which the endogenous visual cue was valid in 100% of the trials. When a 100% valid endogenous cue is able to fully suppress the capture of visual attention by means of an auditory cue, the exogenous cueing effect should disappear. This would be in line with earlier studies showing a suppression of exogenous cueing when attention is focused endogenously (Santangelo et al., 2007; Theeuwes, 1991; Yantis & Jonides, 1990).

Experiment 3: Endogenous cue 100% correct

In Experiment 3 a 100% valid visual endogenous cue was used, to check whether the crossmodal cueing effects found in Experiment 2 are caused by the fact that the endogenous cue was invalid in a small proportion of the trials. If top-down processes are indeed able to suppress exogenous cueing, as was shown earlier within the visual modality, we expect that this manipulation will cause all exogenous cueing effects to disappear.

Method

Participants. Twelve new students of the Vrije Universiteit Amsterdam (2 male, mean age 20.8, ages between 18 and 28) participated in the experiment.

Design. In Experiment 3 the stimuli and method for eye movement registration were identical to those used in Experiment 2, but endogenous visual cues were used that were valid in 100% of the trials. The design in terms of conditions and amount of trials (48 per condition) was identical to that of Experiment 1.

Results

Trials with eye movements (3.4%), and premature (< 200 ms) or slow (> 1000 ms) responses (0.4%) were removed from further analysis. For the remaining trials mean RTs for the correct response trials (96.2%) were calculated for each subject for each condition. The mean reaction times for each condition averaged over subjects are plotted in Figure 2.4. The error bars in this figure represent the .95 confidence interval (5.9 ms) for the exogenous cueing main effect. A within-

subjects ANOVA showed that auditory cueing (valid 335 ms, invalid 351 ms, neutral 343 ms) had a significant effect on the reaction times [$F(2,22) = 7.836$, $MSE = 97.068$, $p < .005$]. Three Post-hoc pairwise 2-tailed t-tests between the cueing conditions were conducted. Valid compared to invalid ($p = .008$), neutral compared to valid ($p = .039$), and neutral compared to invalid ($p = .046$) were all significant.

A similar ANOVA applied to the error data (valid 2.3%, invalid 4.3%, and neutral 4.7%) showed a significant cue effect [$F(2,22) = 4.256$, $MSE < .001$, $p < .05$]. Post-hoc analysis only shows a significant difference between valid and neutral ($p = .002$), a strong trend for valid compared to invalid ($p = .059$), and no effect for invalid compared to neutral ($t < 1$).

A separate ANOVA for a between-group comparison between Experiment 2 (visual valid) and Experiment 3 showed no significant RT differences [$F(1,22) = 1.257$, $MSE = 2868.708$, $p = .274$].

Discussion

The results of this experiment are similar to those observed in the visual valid condition of Experiment 2. Just as in Experiments 1 and 2, the current results show two effects. Compared to the neutral condition, we see benefits and costs for valid and invalid auditory cues, respectively. This means that the observed cueing effect

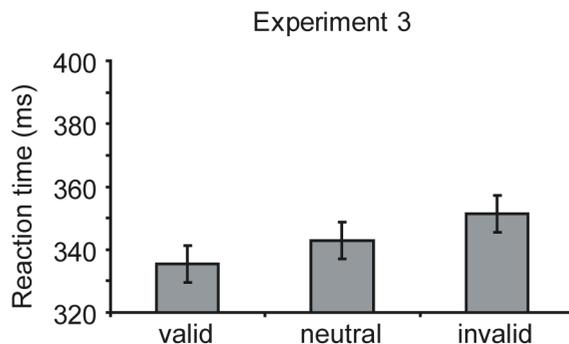


Figure 2.4. Results of Experiment 3 presented as an average reaction time (ms) for each condition (valid, neutral, and invalid). The error bars show the .95 confidence intervals for the exogenous cueing main effect.

is not sensitive to the validity of the endogenous visual cue as suggested earlier. These findings are somewhat unexpected because the 100% valid endogenous cue should have allowed a firmer and more focused attention on the target location before the auditory cue was presented (Yantis & Jonides, 1990). Therefore, if crossmodal cueing is based on shifting attention, we would have expected no benefits because shifting attention in response to the auditory cue is not necessary when attention is already focused on the target location. One possibility is that participants did not make full use of the endogenous cue. This seems unlikely because we already observed a significant effect of endogenous cueing on response times in Experiment 2; also, response times in Experiment 3 are similar to those measured in Experiment 2 for valid endogenous cueing. However, there is still an alternative explanation for the observed costs and benefits. Note that the current paradigm differs with respect to a further aspect in comparison to other studies such as Santangelo and Spence (2007) and Theeuwes (1991).

Theeuwes (1991) showed that cueing effects disappear when there is focused attention in combination with no-onset targets. These no-onset targets in the form of figure-eight premasks (that can turn into letters by removing two of the line segments) were already on screen when the endogenous cue was presented. This allowed participants in response to the endogenous cue to focus their attention tightly on the premasks before the exogenous cue was presented. The same holds for the study by Santangelo and Spence (2007) where an RSVP stream was presented during the entire trial also allowing participants to focus their attention. Both the no-onset targets and the RSVP stream could have functioned as placeholders allowing participants to direct their attention to a predefined location in space. It could be the case that the cueing effect observed in the current experiment is the result of 'not' predefining the specific target location. In other words, the observed cueing effect could reflect merely that attention was not entirely focused on one specific location because no placeholders were present. The observed costs and benefits could then simply be explained in terms of shifts (or fine-tuning) of attention induced by the exogenous auditory cue. In order to test

this notion, a fourth experiment was conducted in which placeholders were used to indicate target positions.

Experiment 4: Introducing placeholders

Experiment 4 was similar to Experiment 3 except that place-holders marking the target locations stayed on the screen during the entire trial. These placeholders enabled the participants to accurately focus their attention on the target location. If spatial uncertainty of the target location is indeed the reason why exogenous cueing effects were still observed in the previous experiments, we expect that this manipulation will cause the effects to disappear.

Method

Participants. Twelve new students of the Vrije Universiteit Amsterdam (3 male, mean age 20.8, ages between 18 and 25) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing.

Stimuli and Design. This experiment was similar to Experiment 3 including the 100% valid endogenous cues in the form of arrowheads and the number of trials per condition (48). Additionally, during the entire trial placeholders were displayed on both sides of the screen indicating the possible target locations. The placeholders were thin gray (9.34 cd/m^2) lined squares with a width and height of 1.3° , that indicated the area in which targets could appear. On each side of the screen two connecting placeholders were shown – one for the targets displayed above the midline of the screen, and the other for the targets below the midline. The horizontal centre of the squares was separated by an angle of 10.5° from the centre of the screen. The placeholders made the task of target discrimination (above or below the vertical middle) easier, because their locations were now predefined. To compensate for this targets were displayed closer to the vertical middle (on average 0.6° above and below) than in the previous experiments and their location was varied by placing the targets randomly within a range of 0.3° from the centre of the placeholder.

Results

Trials containing possible eye movements (5.9%), and premature (< 200 ms) or slow (> 1000 ms) responses (0.5%) were removed from further analysis. For the remaining trials mean RTs for the correct response trials (93.2%) were calculated for each subject for each condition. Mean reaction time for each condition averaged over subjects is plotted in Figure 2.5. The error bars in this figure represent the .95 confidence interval (7.0 ms) for the exogenous cueing main effect. For the analysis a within-subjects ANOVA was used which showed that auditory cueing (valid 362 ms, invalid 373 ms, neutral 364 ms) had a significant effect on the reaction times [$F(2,30) = 4.238$, $MSE = 140.244$, $p < .05$]. Three Post-hoc pairwise 2-tailed t-tests between the cueing conditions were conducted. Valid compared to invalid ($p = .016$), and neutral compared to invalid ($p = .020$) were both significant. Neutral compared to valid ($t < 1$) was non-significant

A similar ANOVA applied to the error data (valid 7.1%, invalid 6.5%, and neutral 6.8%) showed no effect ($F < 1$).

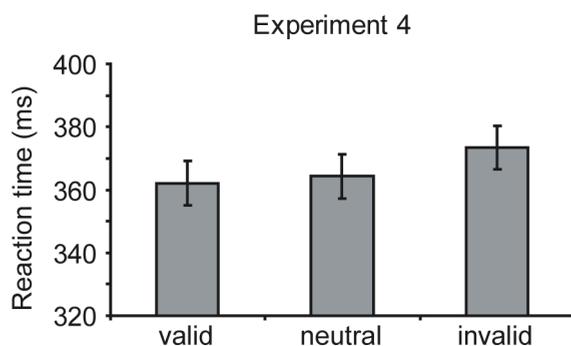


Figure 2.5. Results of Experiment 4 presented as an average reaction time (ms) for each condition (valid, neutral, and invalid). The error bars show the .95 confidence intervals for the exogenous cueing main effect.

Discussion

The results show that the exogenous cueing effect remains, but that it is now solely based on costs: valid cues yield no improvement relative to neutral cues, whereas invalid cues still result in higher reaction times. The results suggest that the effects for valid auditory cues, observed in Experiments 2 and 3 were indeed due to the fact that these cues caused an improved and/or narrowed attentional focus on the location of the impending target. We should also note that the use of placeholders allowed for a smaller vertical angle between the up and down target locations on the left and right side of the screen. This also allowed participants to use a smaller attentional focus in comparison to the previous experiments. Altogether, the results indicate that exogenous auditory cues can still capture attention, even when visual attention is fully focused and the target location is predefined. Apparently, the auditory cue presented at the non-target location causes attention to shift away from the target location causing a small but reliable RT cost.

So far, the results of Experiment 1 to 4 show how exogenous cueing benefits disappear when attention is focused on a valid and predefined target location. For these experiments we used crossmodal auditory cues that were compared to a neutral condition. It is important to compare this with results of visual cues obtained in identical conditions, because previous results on visual cueing are inconclusive. Recall that Mazza and colleagues (2007) showed no cueing effect for unimodal cues when target location is blocked, but they did show crossmodal cueing effects, Santangelo and Spence (2007) found no unimodal and no crossmodal cueing effects on peripheral targets when attention is focused to the centre of the screen, and Van der Lubbe and Postma (2005) found both unimodal and crossmodal cueing effects when visual attention is focused on the correct target location.

Experiment 5: Visual exogenous cues

In this final experiment we wanted to determine whether the results shown in Experiment 4 would change when a visual instead of an auditory exogenous cue was used. In other words, the question is whether knowing at which location a target will appear influences unimodal exogenous cueing in the same way as crossmodal exogenous cueing. In order to test this we used the same paradigm as in Experiment 4, but we made the following changes: endogenous cueing was either 100% valid or absent (a question mark was displayed instead of an arrow), and both visual and auditory exogenous cues were presented. These factors were tested in different blocks in a within-subjects design. The conditions without endogenous cueing were included so that we could verify that exogenous cueing also occurs in the absence of an endogenous cue. We used a modified setup which allowed us to exactly align the auditory cue with the visual targets, to prevent spatial disparities. We did this by displaying the visual stimuli on an acoustically transparent screen and by placing the loudspeakers at the exact target locations (10.5° left and right from the middle).

Method

Participants. Sixteen new students of the Vrije Universiteit Amsterdam (2 male, mean age 20.2 years old, ages between 18 and 25) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing.

Apparatus. In the experiment the visual stimuli were presented on a sound transparent (micro-perforated) screen by means of a projector (Theme Scene HD70, 60Hz). Participants were seated in a dimly lit room approximately 150 cm from the screen. All visual stimuli were rescaled so their retinal images were of identical size as in the previous experiments.

Stimuli and Design. Compared to Experiment 4, two within-subject factors were added to the experiments design. First, we presented either a 100% valid endogenous cue (an arrowhead: < or >) or a neutral cue in the form of a question mark (?) of similar size. Second, the exogenous cue (which could be valid, neutral,

or invalid) was auditory or visual. This visual cue was a thinly lined dark gray circle with a diameter of 3.8° that was flashed for 100 ms at target location (10.5° left or right from the middle). In the neutral condition the circles were flashed simultaneously at both sides. The onset of the exogenous cues occurred 200 ms prior to the onset of the target. The two additional factors were tested in four conditions that were presented in a blocked fashion in the form of four sub-experiments. The following combinations were presented: 1. Auditory exogenous cue (valid, neutral, invalid) - no endogenous cue. 2. Auditory exogenous cue – endogenous cue. 3. Visual exogenous cue (valid, neutral, invalid) - no endogenous cue. 4. Visual exogenous cue – endogenous cue. Each sub experiment contained five blocks each containing 40 trials for each exogenous cueing condition (a total of 24 trials per block). The order in which the four sub experiments were presented to each subject was balanced by means of a Latin square. The session started with one practice block of 24 trials that was identical to the first block of the participants' first sub-experiment.

Results

Trials containing possible eye movements (4.3%), and premature (< 200 ms) or slow (> 1000 ms) responses (0.3%) were removed from further analysis. For the remaining trials mean RTs for the correct response trials (94.6%) were calculated for each subject for each condition. Mean RT per condition is shown in Figure 2.6. The error bars in this figure represent the .95 confidence interval (9.8 ms) for the exogenous cueing main effect. For the analysis a within-subjects ANOVA was conducted containing the factors; exogenous cue modality (visual, auditory), exogenous cue validity (valid, neutral, invalid), endogenous cue presence (cue, no-cue). This ANOVA shows a main effect of exogenous cue modality [$F(1,15) = 90.395$, $MSE = 657.465$, $p < .001$] indicating overall faster RTs on targets preceded by an auditory cue than by a visual cue. There was also a main effect of exogenous cue validity [$F(2,30) = 81.052$, $MSE = 370.603$, $p < .001$], and a main effect of the presence of the endogenous cue [$F(1,15) = 7.941$, $MSE = 1074.622$, $p < .05$] indicating an overall faster performance when the target was preceded by a

valid endogenous cue relative to a no cue condition. Additionally, there was a two-way interaction between exogenous cue modality and exogenous cue validity [$F(2,30) = 17.023$, $MSE = 331.302$, $p < .001$] indicating an overall stronger exogenous cueing effect for visual cues. Also, a two-way interaction was found between endogenous cue presence and exogenous cue validity [$F(2,30) = 4.463$, $MSE = 196.075$, $p < .05$] suggesting a reduced exogenous cueing effect when a valid endogenous cue is present, which is in line with the results from Experiment 4 that showed a reduced cuing effect in comparison to Experiment 1. This reduction is primarily based on the disappearance of benefits when the valid target location is endogenously cued. No interaction between exogenous cue modality and endogenous cue presence was observed ($F < 1$).

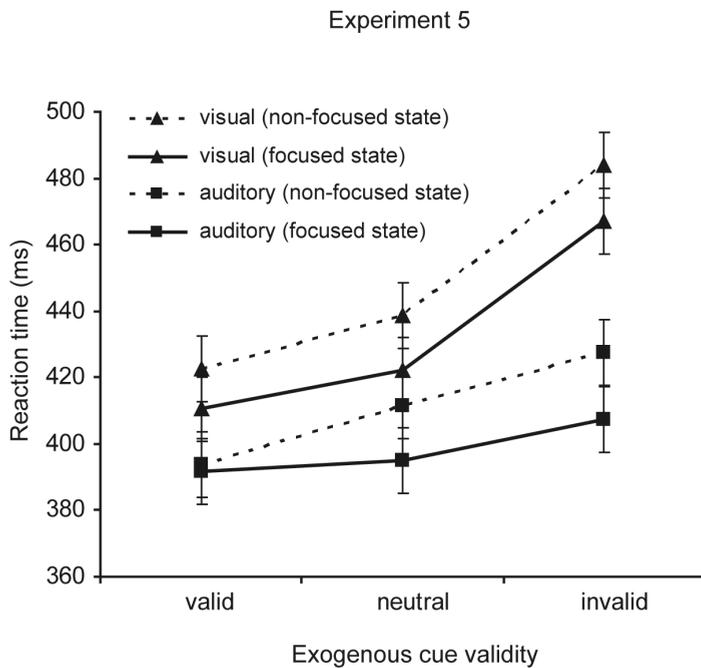


Figure 2.6. Results of Experiment 4 presented as an average reaction time (ms) for each condition. The four combinations for the factors exogenous cue modality (visual and auditory) and endogenous cue presence (present (focused state) and not present (non-focused state)) are plotted as separate lines. The endogenous cue validity (valid, neutral, and invalid) is plotted on the x-axis. The error bars show the .95 confidence intervals for the exogenous cueing main effect.

However, more importantly there was no three-way interaction [$F(2,30) = 1.268$, $MSE = 233.245$, $p = .296$] suggesting that there is no difference in the way that the endogenous cue interacts with visual or auditory exogenous cues.

A similar ANOVA conducted on the error data showed no significant effects but only a trend [$F(2,30) = 2.964$, $MSE = .001$, $p = .067$] for exogenous cue validity (valid 4.9%, neutral 5.6%, and invalid 6.1%). These results indicate that there was no speed accuracy tradeoff.

Experiment	Exogenous cue modality	Endogenous cue condition	Exogenous cue validity			
			valid	neutral	invalid	
1	auditory	no cue	RT	343 (32)	352 (34)	365 (32)
			%	5.85	8.14	8.26
2	auditory	80% valid cue	RT	349 (36)	357 (40)	372 (41)
			%	3.02	3.61	5.21
		20% invalid cue	RT	378 (32)	398 (41)	402 (41)
			%	5.96	8.33	8.04
3	auditory	100% valid cue	RT	335 (31)	343 (36)	351 (43)
			%	2.32	4.71	4.27
4	auditory	100% valid cue	RT	362 (33)	364 (39)	373 (36)
			%	7.10	6.77	6.50
5	auditory	no cue	RT	394 (26)	411 (28)	428 (31)
			%	4.14	5.16	8.56
		100% valid cue	RT	392 (41)	395 (39)	407 (39)
			%	4.32	5.62	5.41
	visual	no cue	RT	423 (30)	439 (32)	484 (37)
			%	5.09	5.38	5.58
		100% valid cue	RT	410 (25)	422 (30)	467 (36)
			%	5.98	6.20	4.74

Table 2.1. Results for Experiments 1 to 5 in the form of mean reaction times (RTs; in milliseconds), their standard deviations (in parentheses), and percentages of errors, for visual target, as a function of exogenous cue modality (auditory and visual), exogenous cue validity (valid, neutral, and invalid), and for the different endogenous cue conditions (between and/or within experiments).

Discussion

The results of this experiment show similar effects for both visual and auditory cues. The results for auditory crossmodal exogenous cues replicated those of Experiment 4 by showing a reduced exogenous cueing effect when visual attention is focused on the correct target location. In addition, the current results show that the spatial disparity between auditory cues and visual targets when presented in all previous experiments had no noticeable influence on the main cueing effect. This is because the new setup used in this experiment, which allowed us to present auditory cues and visual targets at the same location, showed the same crossmodal cueing effects. A similar interaction between endogenous cue presence and exogenous cue validity is present in the visual unimodal exogenous cue condition. Overall these results indicate similar attentional processes involved in unimodal and crossmodal cueing and are in line with the results shown by Van der Lubbe and Postma (2005).

General discussion

This study was conducted for three reasons: First, to investigate whether an endogenous visual cue can suppress exogenous crossmodal cueing. In other words, is exogenous crossmodal cueing an automatic process, or are top-down processes able to influence exogenous crossmodal cueing? As mentioned, the results from previous studies (Mazza et al., 2007; Santangelo et al., 2007; van der Lubbe & Postma, 2005) are inconclusive on this issue. Second, we wanted to examine how both costs and benefits contribute to the crossmodal cueing effect. In order to test this, a neutral exogenous cue was introduced that was spatially uninformative but still provided the same temporal information as the valid and invalid cues. Faster responses on valid conditions compared to neutral would indicate benefits, and slower responses on invalid conditions compared to neutral would reflect costs (Posner, 1980). A neutral baseline condition has not been used in earlier studies investigating endogenous and exogenous crossmodal cueing. Third, we wanted to

test whether an endogenous cue would influence crossmodal and unimodal exogenous cueing in the same way.

The first experiment not only replicated the findings of Spence and Driver (1997) but also showed that crossmodal cueing generates both RT costs and benefits. Similar to notions based on unimodal cueing (Posner, 1980), these results suggest that cueing effects reflect shifts of attention. In the following experiments an endogenous visual cue (a centrally presented arrowhead) was presented prior to the presentation of the exogenous auditory cue. The results from Experiments 2 and 3 showed that when attention was in a focused state, there was still a cueing effect observed consisting of both costs and benefits. Comparison of the results of Experiments 2 and 3 also indicates that the cueing effect is not sensitive to the validity of the endogenous visual cue. However, when attention can be focused on a predefined target location by means of placeholders as in our Experiment 4, only costs are observed and no benefits. For the auditory cueing conditions the results of Experiment 5 replicate those of Experiments 1 and 4. In addition, the results show that visual exogenous cues, at least with the visual task that we have used, in essence have the same effect as auditory cues.

As to the first question of this study, whether an endogenous cue can suppress the exogenous capture of attention by an exogenous crossmodal event, the results from Experiments 2 to 5 provide compelling evidence that the answer should be 'no'. These results show that when attention is in a focused state by means of a centrally presented arrowhead (Experiment 2) which is 100% valid (Experiment 3) and pointing to a predefined target location (Experiments 4 and 5), there is still an effect of the presence of an exogenous auditory and visual (Experiment 5) event. These results extend those of Van der Lubbe and Postma (2005) and Mazza and colleagues (2007) by showing benefits and costs in the form of attentional capture for both valid and invalid cues, respectively, and how this capture strongly depends on the attentional focus prior to the presentation of these cues.

The second question, whether crossmodal cueing effects are based on costs, benefits, or both, could indeed be answered by introducing conditions with a neutral cue, and it appeared that the costs and benefits depend strongly on how

strong spatial attention is focused. As stated earlier, these results are in line with the general view that cueing effects are based on shifts of attention. Remarkably, both Experiments 2 and 3 show benefits when visual attention was focused on the valid target location by means of an endogenous visual cue presented prior to the onset of the exogenous auditory cue. When benefits are indeed based on attentional shifts this should not be expected: when attention is already focused on the correct target location, additional spatial information should not result in an extra performance improvement. If anything, one would have expected additional costs because the auditory cue was presented at an eccentricity that was larger than the eccentricity at which the target was presented. In other words, presenting auditory cues from loudspeakers positioned next to a monitor could have resulted in attention being drawn to the loudspeaker location rather than to the target location. Apparently this was not the case even though this was a concern when designing the display setup for this study. The results of Experiment 4 show that performance improvements due to valid cues disappear when the target locations are predefined by means of placeholders. This indicates that an exogenous auditory cue is only able to influence attentional focus when there is uncertainty with respect to target position. Note that attentional capture – the cost associated with an invalid exogenous cue – is neither affected by the validity of the endogenous cue nor by the spatial uncertainty of the target location.

The third question whether an endogenous visual cue would influence crossmodal and unimodal exogenous cueing the same way was answered by Experiment 5. These results show a reduced exogenous cueing effect in both unimodal and crossmodal conditions, when attention is in a focused state. These results are in line with those of Van der Lubbe and Postma (2005). Although Santangelo and Spence (2007) also showed that focused attention influences crossmodal and unimodal exogenous cueing similarly, their results are completely opposite. Note that Mazza and colleagues (2007) are the only ones who show within one study opposite results for unimodal and crossmodal cueing. When they provided endogenous information by blocking target side, the results showed no unimodal cueing effect. However, the same experiment did show a cueing effect

for the crossmodal condition. However, it must be noted that by blocking trials they have potentially introduced a confound that complicates the interpretation of their findings. As explained by Santangelo and colleagues (2007), a drawback of blocking target side is that *“the presentation of a target on one side on one trial may lead to an exogenous shift of attention toward that side, thus potentially facilitating performance when the target on the next trial also happens to be presented from the same side”* (p.138). Thus, their results might only reflect a differential influence of this trial-to-trial within-modality cueing on the actual cueing effect that was under study.

As already mentioned in the introduction there is no simple way to explain why some studies found suppression of exogenous attention and others did not. Still there are some important differences that could play a role. Most studies adopted the task of Spence and Driver (1997) that uses dots as targets which appeared in the form of onsets. This in contrast to the no-onset letter stimuli used earlier by Theeuwes (1991). It may well be that it is easier for participants to ignore onset cues when the targets are no-onsets instead of onsets. This is consistent with the notion of contingent capture (Folk, Remington, & Johnston, 1992) stating that exogenous capture of attention by for instance an irrelevant cue depends on whether or not the cue shares a relevant feature with the target. Folk and his colleagues (1992) show that onset cues affect onset targets but not targets that are characterized by a color change. In other words, when onset targets are used, onsets become a relevant stimulus feature. As a result, the onset of the cue may automatically draw attentional resources to the cue location, which could explain the cueing effect during focused attention.

Letter stimuli were also used in Santangelo and Spence’s (2007) task but this time in the form of a central RSVP stream on which attention should be focused. This RSVP stream presented letters at a high rate and knowing that a possible target could appear in this stream kept participants endogenously focused to it. However, an RSVP stream also tends to generate a high perceptual load which in turn might drain attentional resources required for the processing of the exogenous cues. A follow-up study by Santangelo and colleagues (2008) using a central

morphing shape (instead of an RSVP stream) to manipulate purely perceptual load confirms this idea by again showing suppression of exogenous visual cueing. Taken together, both endogenous attention and perceptual load could explain the suppression of the unimodal and crossmodal exogenous cueing effect as shown by Santangelo and Spence (2007). In a recent review Santangelo and Spence (2008) discuss whether or not unimodal and crossmodal cueing are automatic processes. They evaluate exogenous cueing by means of the intentionality and load-insensitivity criteria (Jonides, 1981; Posner, 1978; Yantis & Jonides, 1990), stating that voluntary control and perceptual load should not interfere with a process in order for it to be automatic. Santangelo and Spence claim that when an RSVP stream is used to focus attention it is hard to distinguish between possible voluntary endogenous effects of the task (find the target in the stream), and perceptual load effects evoked by the information presented in the RSVP stream. Based on their findings they conclude that the capability of abrupt onsets to capture spatial attention depends on how much attentional resources are available. If one's resources are fully engaged by means of a high perceptual load task such as an RSVP stream there will probably be no attentional capture effects. By contrast, if an endogenous cue is used to voluntarily focus attention, it is likely that there will be enough resources left to process peripheral onsets that are able to capture attention. Therefore, our finding that attentional capture by means of exogenous cues cannot be suppressed when attention is focused in a pure endogenous fashion is not necessarily inconsistent with the views proposed by Santangelo and Spence (2008).

To conclude, the results from these current experiments make clear that crossmodal and unimodal exogenous cueing of a visual target location cannot be suppressed by endogenously focusing visual attention. Even when visual attention is fully focused to a predefined target location, an auditory or visual cue coming from the opposite direction is still able to capture visual attention. When visual attention is not focused prior to the presentation of the auditory cue both costs and benefits are shown.

Chapter 3

Competition between auditory and visual spatial cues during visual task performance

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.

Abstract

There is a debate in the crossmodal cueing literature as to whether capture of visual attention by means of sound is a fully automatic process. Recent studies show that when visual attention is endogenously focused sound still captures attention. The current study investigated whether there is interaction between exogenous auditory and visual capture. Participants performed an orthogonal cueing task in which the visual target was preceded by both a peripheral visual and auditory cue. When both cues were presented at chance level, visual and auditory capture was observed. However, when the validity of the visual cue was increased to 80% only visual capture and no auditory capture was observed. Furthermore, a highly predictive (80% valid) auditory cue was not able to prevent visual capture. These results demonstrate that crossmodal auditory capture does not occur when a competing predictive visual event is presented and is therefore not a fully automatic process.

Introduction

It is well known that our attention may be captured by sudden visual or auditory events even when they are irrelevant for our current task. When hearing a sound or seeing a flash we have the tendency to direct our gaze to the location of the visual or auditory event. This enables us to respond more accurately and more quickly to events that occur at that location (Posner et al., 1980). This bottom-up or exogenous capture of attention can occur in an overt manner by making eye movements (Theeuwes et al., 1998) or in a covert manner without making eye movements (Theeuwes, 1994). Within the visual domain, exogenous capture of covert attention is most often studied by means of a cueing task in which a localizable onset is presented at a valid or invalid target location prior to the presentation of the target. People respond faster and more accurately to validly cued targets than to invalidly cued targets. Importantly, this cueing effect occurs when the cue is valid at chance level, which indicates that it is an automatic process (e.g., Jonides, 1981; Yantis & Jonides, 1984). Previous research shows similar cueing effects in modalities other than vision, such as the tactile (Posner, 1978; Spence & McGlone, 2001) and auditory domains (Spence & Driver, 1994).

Exogenous cueing effects are also known to occur across modalities. In a seminal study by Spence and Driver (1997) participants had to perform a cueing task in which they made an elevation judgment regarding auditory or visual targets presented to the left or right of fixation. For the elevation judgment task, the target was presented either at an “up” location above the vertical meridian or at a “down” location below the vertical meridian. This resulted in a total of four target locations with two (up and down) on each side of fixation. Cues were presented along the vertical meridian between the up and down locations on the left or right side. Therefore, only the side at which the target was presented was cued, but not its exact location. This task has become known as the orthogonal cueing task because the response dimension (up or down) is orthogonal to the cue dimension (left or right). This has the benefit that possible response priming effects are canceled out. In Spence and Driver’s (1997) study, the auditory target stimuli were generated by

loudspeakers, and the visual target stimuli were generated by LED's mounted directly in front of these loudspeakers. The side at which the target could be presented was cued at chance level by either a visual or auditory event. Results showed unimodal cueing effects in both visual and auditory domains and a crossmodal cueing effect when a visual target was preceded by an auditory cue.

An important question to address is whether attentional capture by sound as shown by crossmodal cueing is truly automatic. In order for auditory capture to be considered an automatic process, it should not be affected by top-down control (Jonides, 1981; Posner, 1978; Yantis & Jonides, 1990). It is known that one can direct attention in a top-down (or endogenous) way to a location in space (Broadbent, 1982; Posner et al., 1980). In a typical paradigm, a centrally presented arrow points to the likely target location with a high probability (e.g., 80%). Similar to exogenous cueing, endogenous cueing effects show faster responses to validly cued target locations than to invalidly cued target locations. The question whether auditory capture is sensitive to any top-down settings is addressed by several recent studies (Koelewijn et al., in press; Mazza et al., 2007; Santangelo & Spence, 2007; van der Lubbe & Postma, 2005).

Van der Lubbe and Postma (2005) used a combination of endogenous and exogenous cues to investigate whether top-down control could affect attentional capture. In their study participants performed a variation of the orthogonal cueing task in which they had to indicate whether an arrowhead presented to the left or to the right of fixation was pointing up or down. The exogenous cue consisted of either a visual or auditory onset and was presented 200 ms before the target (i.e., the arrowhead). Eight hundred milliseconds prior to this cue an endogenous central cue consisting of an arrow indicated the target location with 100% validity. Note that in this study only elevation judgments of visual targets were made. The results showed that even though the target location was known in advance, exogenous cueing effects still occurred both in the unimodal condition and in the crossmodal condition. A recent study by Mazza, Turatto, Rossi, and Umiltà (2007) showed similar results. In this study a similar design to that of Spence and Driver (1997) was used with the exception that the side at which the target would be presented

was blocked. Therefore, participants knew where the target would appear and could keep their attention endogenously focused on one of the sides during an entire block. In line with the results of Van der Lubbe and Postma (2005), Mazza and colleagues (2007) found an auditory cueing effect on a visual target even though the target side was known to the participant. Finally, Koelewijn, Bronkhorst, and Theeuwes (in press) also showed that auditory crossmodal cueing effects occur during focused visual attention. In this study participants performed an orthogonal cueing task in which each trial started with an endogenous cue in the form of a centrally presented arrowhead indicating the side at which the target would be presented. In Experiment 4 this arrowhead was 100% valid and placeholders indicated the possible target locations during the entire trial. In addition to the standard spatial cues a spatially neutral auditory cue was introduced. This neutral cue had the same temporal onset as the spatial cues but was spatially diffuse. Compared to this neutral baseline condition the results showed attentional capture by sound in the form of costs when the auditory cue was presented at the invalid location. However, no benefits were found when the target location was validly cued.

So far several studies have demonstrated that top-down control of attention does not affect auditory capture (Koelewijn et al., in press; Mazza et al., 2007; van der Lubbe & Postma, 2005). However, no study has addressed whether auditory capture can be affected by bottom-up processes. The current study investigated auditory capture when at the same time a visual event was presented. From studies on visual search (Desimone & Duncan, 1995) it is known that competing visual events share attentional resources. The more events are shown the less salient individual events become in which case less attention is drawn to each individual event. The present study addressed whether the bottom-up salience of a visual and auditory event is affected when these are presented at the same time.

There are previous studies that have used bimodal exogenous cues to investigate crossmodal integration (e.g., Santangelo et al., 2006; Ward, 1994). Santangelo and colleagues (2006) investigated whether a bimodal audiovisual cue shows a super-additive effect in respect to the unimodal visual and auditory cueing

effects. A super-additive effect for the bimodal cue would indicate crossmodal integration. In an orthogonal cueing task, visual, auditory, or bimodal audiovisual cues were presented prior to the presentation of a visual target. The results show similar effects sizes for the visual, auditory, and bimodal audiovisual cues which indicates that crossmodal integration is not reflected in the cueing effects. Similar results were obtained in later studies (e.g., Santangelo, Van der Lubbe et al., 2008) that show electrophysiological but no behavioral integration effects. In the bimodal condition of the Santangelo and colleagues (2006) study both the visual and the auditory cue were always presented at the same location and never at opposite locations. Therefore, these results do not reveal how each individual cue contributes to the overall bimodal cueing effect. In order to show these individual contributions the validity of each cue should be manipulated individually. Such an experiment was performed by Ward (1994) but his results are inconclusive because the paradigm used failed to show crossmodal auditory capture when unimodal auditory cues were presented. By investigating whether the individual contributions of visual and auditory exogenous cues interact we will learn more about the automaticity of auditory capture.

In Experiment 1 we tested how nonpredictive peripheral auditory and visual cues interact with one another and how they each influence performance in a visual spatial discrimination task. In Experiment 2 and 3 the validity of, respectively, the visual or auditory cue was raised to 80%. We wanted to determine how this top-down bias would influence the competition between the auditory and visual events over attentional resources. We hypothesized that it should be possible to prevent auditory capture of visual attention by presenting competing visual events.

Experiment 1

The paradigm used in this experiment was similar to the orthogonal cueing task used by Spence and Driver (1997). However, only elevation judgments of visual targets were made and both unimodal and bimodal cues were used. We used a within subject design consisting of experimental and control conditions. In the unimodal control conditions both types of cues were presented separately and in the bimodal experimental conditions auditory and visual cues were presented together. In the experimental conditions the visual and auditory cues appeared simultaneously at the same or at opposite locations. The visual cues and targets were presented on a computer screen instead of using LED's. The loudspeakers that generated the auditory cues were located to the left and right of the computer screen.

Method

Participants. Ten students of the Vrije Universiteit Amsterdam (3 male, mean age 21.0, ages between 18 and 26) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing. Participants were informed beforehand about the experimental procedure and were naïve as to the purpose of the experiment.

Apparatus and Design. Participants were seated in a dimly lit room at approximately 80 cm distance from the computer screen (17 inch, 120 Hz). The experiment was run in E-Prime 1.2 (1.2.1.847). The loudspeakers were placed at an angle of 18.3° from fixation and were aligned to the vertical middle of the screen. In the control conditions either a visual or auditory cue was presented that was either valid or invalid with respect to the target location. This resulted in 4 conditions. The cues were presented 150 ms prior to the onset of the target. The control conditions were presented in 9 blocks containing 16 trials each. The first block was for practice purposes only which resulted in 32 trials for each of the control conditions. In the bimodal experimental conditions both a visual and an auditory cue were presented 150 ms prior to the onset of the target. There were 4

conditions in which the validity of the cueing side was manipulated independently for the visual and auditory cues. The experiment consisted of 9 blocks containing 16 trials each. The first block was a practice block so that 32 trials for each of the experimental conditions remained. Participants performed both the control and the experimental conditions in two separate blocks and the order of these blocks was counterbalanced over participants.

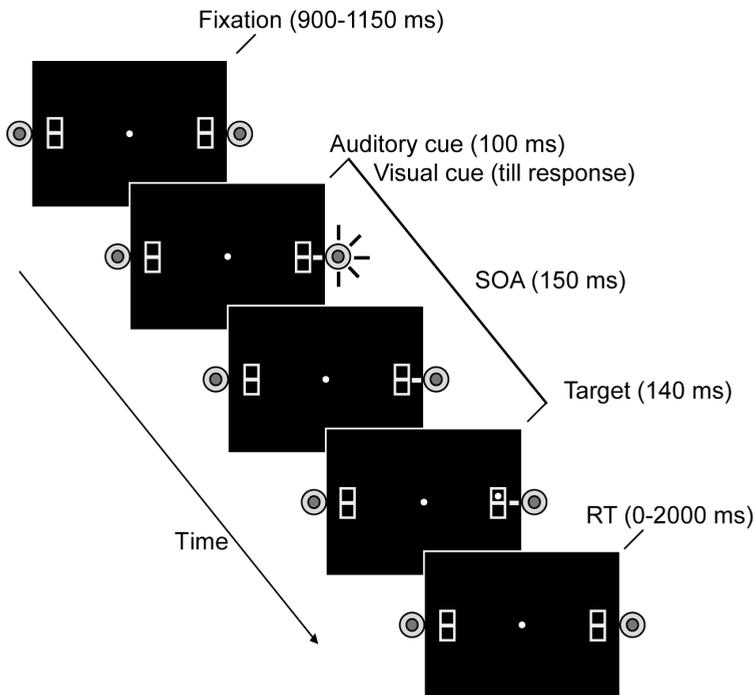


Figure 3.1. Schematic representation of the paradigm used. Participants performed an orthogonal cueing task where they had to discriminate between targets presented in the upper or lower square or the placeholder. Targets were presented on the left or right side of the screen and were preceded independently with an SOA of 150 by a nonpredictive visual or auditory cue. In the example given both the auditory and the visual cue are valid.

Procedure and Stimuli. Figure 3.1 gives an example of a typical trial. At the beginning a white fixation dot (diameter 0.2°) appeared on a black background and stayed on screen until a response was made. Participants were instructed to fixate this dot during the entire trial and to refrain from making eye movements. After a random delay time of 900 ms to 1150 ms the auditory and visual cue were presented. In the control conditions, this was either an auditory or a visual cue. In the experimental conditions, both an auditory and a visual cue were presented. After an SOA of 150 ms the target appeared. The auditory cue consisting of a white noise burst that was presented for 100 ms. This cue was equiprobably presented from the left or right loudspeaker and was valid or invalid with respect to the target location. The visual cue consisted of a dark grey horizontal bar (width 0.6° and height 0.2°) presented at 11.7° left or right of fixation and was valid or invalid with respect to the target location. To prevent attention from being drawn to its offset (Theeuwes, 1991), the visual cue remained on screen until a response was made. The auditory cue was switched off because there is no evidence that auditory offsets are able to draw attention.

The visual target consisted of a white dot (diameter 0.2°) that was presented for 140 ms. During the entire trial placeholders were displayed on both sides of the screen indicating the possible target locations. This allowed participants to focus their attention to a predefined location. The placeholders were thin light grey lined squares with a width and height of 1.3° that indicated the area in which targets could appear. On each side of the screen two connecting placeholders were shown – one for targets displayed above the vertical meridian of the screen, and the other for targets below the vertical meridian. The horizontal centre of the squares was separated by an angle of 10.5° from the centre of the screen. The two target locations at each side were positioned above each other on average 0.6° above and below the vertical meridian of the screen. The exact location was varied by placing the targets randomly within a range of 0.3° from the centre of the placeholder. The target appeared at one of the four locations at chance level.

The participants' task was to report in a speeded but accurate fashion whether the target appeared above or below the vertical meridian of the screen. They did

this using the number pad of a QWERTY keyboard, by pressing the numbers 8 or 2 with their right and left index finger, respectively. Because of the task's orthogonal design there was no need to balance out for possible motor response effects. Responses had to be made within a time window of 2000 ms after target onset. After the response, participants received feedback when they had made an error – the fixation dot then turned red for 150 ms. After each trial, an interval of 850 ms followed before the next trial started. Following each block participants received further feedback in the form of a percentage correct score and a mean reaction time. At the beginning of the experiment, participants were told that both auditory and visual cues would provide no information about the location of the targets and therefore could be ignored.

Results

Control conditions. For each subject the average score and its standard deviation were calculated for each condition over all data. Reaction times above or below average by 2.5 times their standard deviation were considered to be outliers (in total 2.1%) and were removed from further analysis. This method of determining outliers was used for all data analyses in this study. For the remaining trials mean reaction times for the correct response trials (95.8%) were calculated for each condition. An overview of the mean reaction times, their standard deviations, and the mean error scores, for each condition and for all experiments, is shown in Table 1. An ANOVA was performed on RT with cue validity (valid and invalid) and cue modality (auditory and visual) as factors. The outcome revealed a significant effect for cue validity [$F(1,9) = 15.077$, $MSE = 552.058$, $p < .005$] and for cue modality [$F(1,9) = 14.152$, $MSE = 1161.858$, $p < .005$]. In addition, the results show an interaction between cue validity and cue modality [$F(1,9) = 10.560$, $MSE = 228.969$, $p < .05$] indicating a stronger cueing effect for visual cues than for auditory cues. Two pairwise two-tailed t-tests for the separate cue modalities were conducted. These results show a cueing effect for both the visual ($p = .004$) and auditory ($p = .018$) cues.

	Exogenous cue modality		Exogenous cue validity		Cueing effect
			valid	invalid	
Experiment 1	auditory	RT	395 (41)	409 (49)	14
		%	1.59	2.57	
	visual	RT	420 (51)	464 (66)	44
		%	5.35	7.36	
Experiment 2	auditory	RT	361 (56)	378 (60)	17
		%	8.47	9.23	
	visual	RT	362 (47)	425 (63)	63
		%	9.13	14.18	
Experiment 3	auditory	RT	379 (20)	407 (29)	28
		%	2.26	4.20	
	visual	RT	412 (27)	463 (24)	51
		%	3.78	4.35	

Table 3.1. Results of the control condition of Experiment 1, 2, and 3. Shown are reaction times (RTs; in milliseconds), their standard deviations (in parentheses), and percentages of errors, for visual targets as a function of exogenous cue modality (auditory and visual), exogenous cue validity (valid and invalid), together with the cueing effects for each modality.

A similar ANOVA conducted on error data revealed no effect, which indicates that there was no speed accuracy tradeoff.

Experimental conditions. Outliers (in total 2.0%) were removed from further analysis. For the remaining trials mean reaction times for the correct response trials (94.0%) were calculated for each subject for each condition. The mean reaction times for each condition averaged over subjects are plotted in Figure 3.2. The error bars in all figures represent the .95 confidence interval for the auditory cueing main effect, following Loftus and Masson (1994). An ANOVA was conducted on RT with as factors auditory cueing and visual cueing. The results show an effect for auditory cueing [$F(1,9) = 6.424$, $MSE = 343.281$, $p < .05$] and for visual cueing [$F(1,9) = 7.377$, $MSE = 4396.303$, $p < .05$]. No interaction between auditory and visual cueing ($F < 1$) was observed.

A similar ANOVA conducted on error data revealed a trend for visual cueing [$F(1,9) = 4.506$, $MSE = 0.007$, $p = .063$] indicating that participants made more

errors on invalid visual cueing trials (14.2 %) than on valid trials (9.1 %). The error data indicates that there was no speed accuracy tradeoff.

Cueing effects. Two paired sample t-test were conducted to test for possible size differences of auditory and visual cueing effects between the control and experimental conditions. No difference between the unimodal auditory cueing effect and the experimental auditory main cueing effect ($t < 1$), and between the unimodal visual cueing effect and the experimental visual main cueing effect ($t < 1$) were observed.

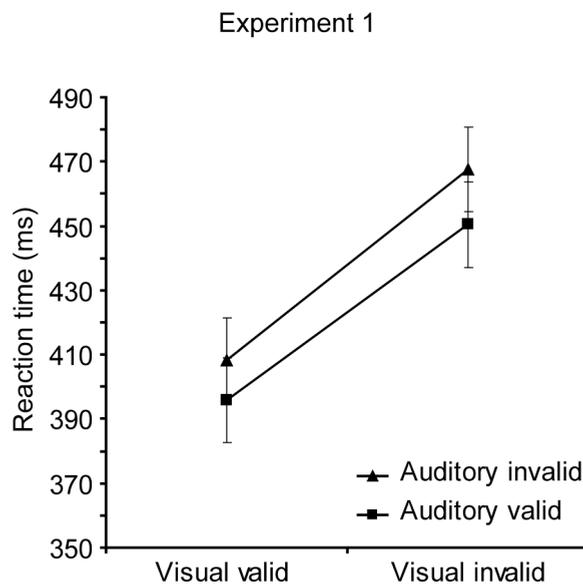


Figure 3.2. Graph with the results of the experimental part of Experiment 1. The graph shows the average reaction time (ms) for all auditory cue (valid and invalid) and visual cue (valid and invalid) validity combinations. The error bars show the .95 confidence intervals for the auditory cueing main effect (Loftus & Masson, 1994).

Discussion

The results show auditory as well as visual exogenous cueing effects in both the unimodal control conditions and the bimodal experimental conditions. In addition, no differences in cueing effect size between the control and experimental conditions were observed. This indicated that the overall bimodal cueing effect is additive and made up of the summation of visual and auditory cueing effects. The current results indicate that auditory capture is not affected by a competing exogenous visual cue. Earlier studies (Koelewijn et al., in press; Mazza et al., 2007; van der Lubbe & Postma, 2005) show that auditory capture is not affected by top-down control. The current results suggest that the same holds for bottom-up competition. Overall these outcomes seem to indicate that auditory capture is an automatic process not affected by other competing processes.

Note, however, that this conclusion is inconsistent with that of Santangelo and Spence (2008) who showed no auditory capture when attention was endogenously focused by an RSVP stream. In their study, participants performed an RSVP task at the center of the display while exogenous cues were presented in the periphery. In other studies (Koelewijn et al., in press; Mazza et al., 2007; van der Lubbe & Postma, 2005) information about the upcoming target location was presented *prior* to the presentation of both cue and target. For example, the central arrowheads in (Koelewijn et al., in press; van der Lubbe & Postma, 2005) pointed towards the target location before the target was presented. Therefore, from trial to trial, participants had to endogenously refocus their attention at the start of each trial, and were required to maintain their attention focused on the target location during a trial. The RSVP stream used by Santangelo and Spence (2008) contained most targets and was therefore inevitably presented at the most valid target location so that participants could keep their attention focused at the centre during the entire block. It could be that it is harder to disengage attention from stimuli presented at the same location as the target than from a target location indicated symbolically by means of an arrowhead. To test this assumption, in Experiment 2 predictive peripheral visual cues were presented at the same time as the exogenous auditory cues.

Experiment 2

Experiment 1 demonstrated that auditory capture still occurs when a nonpredictive visual cue is presented at the same time as the auditory cue. In Experiment 2 we investigated whether making the visual cue predictive would affect capture by the auditory cue. The validity of the visual cue was set at 80% while the validity of the auditory cue remained at chance level. Again we used a within-subject design with bimodal experimental and unimodal control conditions.

Method

Participants. Ten new students of the Vrije Universiteit Amsterdam (4 male, mean age 21.4, ages between 16 and 30) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing. Participants were informed beforehand about the experimental procedure and were naïve as to the purpose of the experiment.

Apparatus and Design. The setup and design were basically identical to the one used in Experiment 1. However, this time the visual cue was valid in 80% of the trials. The control conditions were presented in 9 blocks containing 28 trials each and the first block was for practice purposes. The visual cue was presented in 160 trials; in 128 trials it was valid and in the remaining 32 trials invalid. The auditory cue was presented in 64 trials and was valid in 50% of the trials. The experimental conditions were presented in 9 blocks containing 40 trials each and also here the first block was for practice purposes. The 2 visual valid conditions (visual valid – auditory valid, visual valid – auditory invalid) consisted of 128 trials each, and the 2 visual invalid conditions (visual invalid – auditory valid, visual invalid – auditory invalid) consisted of 32 trials each. Again, participants performed both the control and the experimental conditions in separate blocks and the order of these blocks was counterbalanced over participants.

Procedure and Stimuli. Procedure and stimuli were mostly identical to the one used in the previous experiment. However, at the beginning of the experiment, participants were now told that the visual cues would provide information about

the location of the targets and that they should make use of this information. The auditory cue would provide no information about the location of the targets and therefore could be ignored.

Results

Control conditions. Outliers (in total 2.7%) were removed from further analysis. For the remaining trials mean reaction times for the correct response trials (90.5%) were calculated for each condition. An ANOVA was performed on RT with cue validity (valid and invalid) and cue modality (auditory and visual) as factors. The outcome revealed a significant effect for cue validity [$F(1,9) = 58.023$, $MSE = 270.267$, $p < .001$] and for cue modality [$F(1,9) = 15.432$, $MSE = 367.067$, $p < .005$]. In addition, the results show an interaction between cue validity and cue modality [$F(1,9) = 12.571$, $MSE = 413.511$, $p < .01$] indicating a stronger cueing effect for visual cues than for auditory cues. Two pairwise two-tailed t-tests for the separate cue modalities were conducted. These results show a cueing effect for both the visual ($p < .001$) and auditory ($p = .001$) cues.

A similar ANOVA conducted on error data revealed no effect which indicates that there was no speed accuracy tradeoff.

Experimental conditions. Outliers (in total 2.3%) were removed from further analysis. For the remaining trials mean reaction times for the correct response trials (93.0%) were calculated for each subject for each condition. The mean reaction times for each condition averaged over subjects are plotted Figure 3.3. An ANOVA was conducted on RT with as factors auditory cueing and visual cueing. The results show an effect for visual cueing [$F(1, 9) = 352.334$, $MSE = 129.892$, $p < .001$]. No Auditory cueing effect ($F < 1$) nor an interaction between auditory and visual cueing ($F < 1$) was observed.

A similar ANOVA conducted on error data revealed no effects which indicate that there was no speed accuracy tradeoff.

Experiment 2

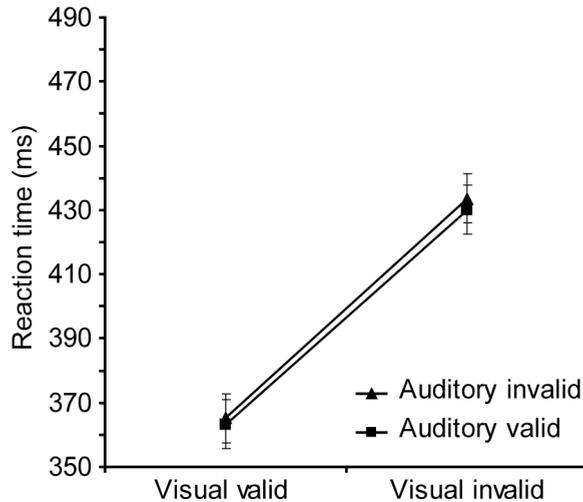


Figure 3.3. Graph with the results of the experimental part of Experiment 2. The graph shows the average reaction time (ms) for all auditory cue (valid and invalid) and visual cue (valid and invalid) validity combinations. The error bars show the .95 confidence intervals for the auditory cueing main effect (Loftus & Masson, 1994).

Discussion

The results of this experiment show no exogenous auditory cueing effect when the auditory cue was presented together with an 80% valid visual cue. When comparing the current results to those of Experiment 1 it appears that the validity of the visual cue is an important factor determining the occurrence auditory capture. Interestingly we do not know whether the same will hold for visual cueing. In other words are we able to prevent visual capture when we are presenting a nonpredictive visual cue at the same time as highly predictive auditory cue. To test whether such symmetry in competition between modalities exist we conducted a third experiment.

Experiment 3

Experiment 2 demonstrated that auditory capture does not occur when a predictive visual cue is presented at the same time as the auditory cue. In Experiment 3 we tested whether a predictive auditory cue could affect visual capture. Therefore, the validity of the auditory cue was set at 80% while the validity of the visual cue now remained at chance level. Again we used a within-subject design with a bimodal experimental and unimodal control condition.

Method

Participants. Ten new students of the Vrije Universiteit Amsterdam (5 male, mean age 23.8, ages between 18 and 44) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing. Participants were informed beforehand about the experimental procedure and were naïve as to the purpose of the experiment.

Apparatus and Design. The setup and design were basically the same as the one used in Experiment 2. However, this time the auditory cue was valid in 80% of the trials while the visual cue was valid in 50% of the trials. Again, participants performed both the control and the experimental conditions in separate blocks and the order of these blocks was counterbalanced over participants.

Procedure and Stimuli. Procedure and stimuli were basically the same as used in the previous experiments. However, at the beginning of the experiment, participants were now told that the auditory cues would provide information about the location of the targets and that the visual cue could be ignored.

Results

Control conditions. Outliers (in total 1.6 %) were removed from further analysis. For the remaining trials mean reaction times for the correct response trials (97.0 %) were calculated for each condition. An ANOVA was performed on RT with cue validity (valid and invalid) and cue modality (auditory and visual) as factors. The outcome revealed a significant effect for cue validity [$F(1,9) = 81.925$, $MSE =$

187.567, $p < .001$] and for cue modality [$F(1,9) = 120.885$, $MSE = 165.289$, $p < .001$]. In addition, the results show an interaction between cue validity and cue modality [$F(1,9) = 13.203$, $MSE = 98.433$, $p < .01$] indicating a stronger cueing effect for visual cues than for auditory cues. Two pairwise two-tailed t-tests for the separate cue modalities were conducted. These results show a cueing effect for both the visual ($p < .001$) and auditory ($p < .001$) cues.

A similar ANOVA conducted on error data revealed no effects which indicate that there was no speed accuracy tradeoff.

Experimental conditions. Outliers (in total 2.3 %) were removed from further analysis. For the remaining trials mean reaction times for the correct response trials (97.5 %) were calculated for each subject for each condition. The mean reaction times for each condition averaged over subjects are plotted Figure 3.4. An ANOVA was conducted on RT with as factors auditory cueing and visual cueing. The results show an effect for auditory cueing [$F(1,9) = 8.153$, $MSE = 158.003$, $p < .05$] and for visual cueing [$F(1,9) = 25.481$, $MSE = 684.069$, $p < .005$]. A trend in the interaction between auditory and visual cueing [$F(1,9) = 5.004$, $MSE = 237.692$, $p = .051$] was observed.

A similar ANOVA conducted on error data revealed an effect for visual cueing [$F(1,9) = 18.116$, $MSE = 0.000$, $p < .001$] indicating that participants made more errors on invalid visual cueing trials (1.7 %) than on valid trials (3.3 %). The error data indicates that there was no speed accuracy tradeoff.

Discussion

The results show auditory as well as visual exogenous cueing effects in both the unimodal control conditions and the bimodal experimental conditions. Additionally, the results show a trend for the interaction between visual and auditory cueing, which is illustrated in Figure 3.4. The current results suggest that a highly predictive auditory event cannot prevent visual capture. Because it was found in Experiment 2 that a predictive visual event is able to prevent auditory capture, we find an asymmetry in how the visual and auditory modalities affect one another. This asymmetry can most likely be attributed to the modality in which the

task was performed. In all three experiments participants performed a visual spatial discrimination task. They had to discriminate between a visual target event that was presented either above or below the vertical meridian. The fact that participants had to respond to visual target events made them possibly more sensitive towards the visual cues as well. This bias towards the detection of visual events might explain why visual capture seems to be less sensitive for crossmodal competition than auditory capture.

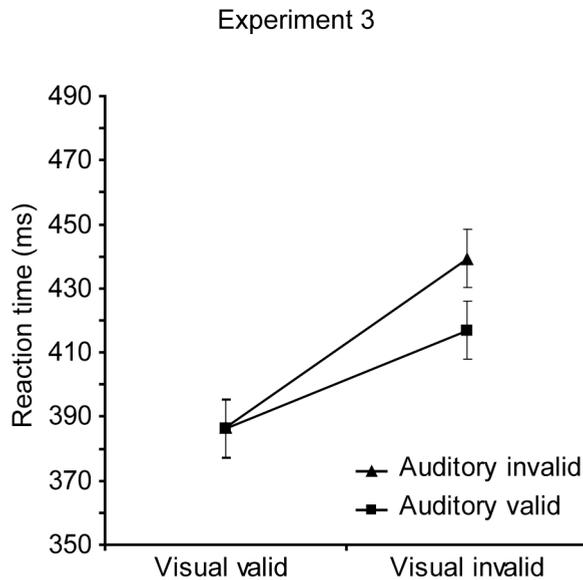


Figure 3.4. Graph with the results of the experimental part of Experiment 3. The graph shows the average reaction time (ms) for all auditory cue (valid and invalid) and visual cue (valid and invalid) validity combinations. The error bars show the .95 confidence intervals for the auditory cueing main effect (Loftus & Masson, 1994).

General discussion

There is debate in the cueing literature as to whether crossmodal auditory attentional capture is fully automatic. On the one hand there are studies that have shown that endogenous focusing of visual attention does not affect auditory capture (Koelewijn et al., in press; Mazza et al., 2007; van der Lubbe & Postma, 2005). On the other hand, Santangelo and Spence (2007) have recently shown that when visual attention is focused by means of an additional task auditory attentional capture ceases to exist. The current study investigated what happens to auditory capture when at the same time a peripheral visual event is presented. In Experiment 1 a nonpredictive peripheral visual cue was presented to test whether bottom-up competition would affect auditory capture. In Experiment 2 a predictive peripheral visual cue was presented to test whether making a cue more endogenous could explain the discrepancy between previous studies. Additionally, in Experiment 3 we tested whether a predictive auditory cue could prevent capture by a nonpredictive visual cue.

The results of the control condition of Experiment 1 replicated the findings of Spence and Driver (1997) by showing cueing effects for both visual and auditory unimodal cues on a visual elevation judgement task. In addition Experiment 1 showed visual and auditory cueing effects when both auditory and visual cues were present in one trial. This means that auditory capture was not affected by the nonpredictive visual event. Whereas in Experiment 1 all cues were nonpredictive, in Experiment 2 the visual cue was 80% valid but the auditory cue remained nonpredictive. Here the results showed no auditory cueing effect when the auditory and visual events co-occurred, while a clear unimodal auditory cueing effect was still observed. These results indicate that capture of visual attention by means of an auditory event does not occur when a predictive peripheral visual event is presented at the same time. In Experiment 3 the auditory cue was 80% valid and the visual cue was nonpredictive. Here the results show both auditory and visual cueing effect. In other words there seems to be an asymmetry in how vision and sound affect one another.

The pattern of results observed in Experiment 2 is similar to those reported by Santangelo and Spence (2007), in that both studies show that auditory capture of visual attention can be prevented. However, in the study by Santangelo and Spence (2007) visual attention was focused by means of an RSVP stream presented at visual fixation. The authors suggest that this RSVP stream induced high perceptual load and that this load prevented auditory capture. In contrast, in the current study visual events were simple onsets presented in the periphery and therefore unlikely to create a high perceptual load. Therefore, it seems that a high perceptual load is not a prerequisite for suppression of auditory capture. Both the current results and the results by Santangelo and Spence (2007) suggest that auditory capture is not a fully automatic process.

It is interesting to compare the current results with those of (Koelewijn et al., in press). In both studies, information about the probable location of the target was provided using top-down information. Also, the auditory cues as well as the placeholders that were used to demarcate where target stimuli could occur were identical. Nevertheless, the outcome of both studies is fundamentally different: whereas auditory capture was still observed when an arrowhead indicated the side where the target would occur with 100% accuracy, no capture occurred when the side was indicated by an 80% predictive peripheral visual cue. One explanation of the difference could be that a cue presented close to the target location results in a smaller attentional focus than a central arrowhead pointing towards a target location. However, the comparison of the results for valid, neutral and invalid cueing in Koelewijn and Colleagues (in press) demonstrates that attentional benefits disappear when a 100% valid arrowhead is used in combination with placeholders. Apparently, in that condition there is already such a narrow focus that the additional (valid) auditory cue does not affect performance anymore. Possibly the use of placeholders of similar size surrounding the target locations may also have contributed to this effect. The fact remains that the endogenous system is not able to prevent for auditory capture towards an invalid target location. This suggests that it is not the size of the attentional focus that matters but whether the attentional system allows for an easy disengagement of attention.

We suggest that it may be harder to disengage attention when the target location is cued by a predictive onset at target location then by a predictive arrowhead pointing towards target location. This may explain the discrepancy between the current results and those of (Koelewijn et al., in press). The current results show that the presence of an 80% valid, localizable visual event prevents auditory capture. This type of events recruits both bottom-up and top-down resources. Top-down because these cues were predictive and bottom-up because there is a peripheral onset which is similar to the pure exogenous cues used in Experiment 1. The importance of this top-down or endogenous factor in suppressing auditory capture is evident when we compare the result of Experiments 1 and 2 where auditory capture is prevented only when the visual cue is predictive. This is so even though endogenous attention alone is not sufficient to suppress auditory capture (Koelewijn et al., in press; Mazza et al., 2007; van der Lubbe & Postma, 2005). Bottom-up or exogenous resources are drawn to the onset of the visual cue, which is presented at target location. As shown by the results of Experiment 2 location of this cue is of high importance. As such, it may be that the concurrent activation of both the endogenous and exogenous attentional system results in the ability to prevent auditory capture.

The results of the current study are consistent with those of Santangelo and Spence (2007) that showed no capture when attention was focused by means of an RSVP stream. Santangelo and Spence (2008) argue that attentional capture by a peripheral onset does not occur under circumstances of high perceptual load. However, the current results show that high perceptual load is not a necessity in preventing attentional capture. It should be noted that in addition to generating high perceptual load the onset of each individual item in the RSVP stream could also draw on bottom-up attentional resources. In a recent study Santangelo and colleagues (2008) tested whether these onsets could explain their results, by replacing the RSVP stream by a central morphing shape. Because of the morphing no abrupt visual onsets occur and therefore the shape should not draw exogenous attention. Also under these circumstances no cueing effect for visual peripheral events was observed. However, just as in the RSVP stream most targets were

presented within this morphing shape. Therefore, the morphing shape was inevitably presented at the most valid target location. The importance of this correspondence between cue and target location is something we also find in the current study.

To conclude, the current results show that auditory capture of visual attention will not occur when a predictive peripheral visual event is presented at the same time. Therefore, it seems that both visual and auditory events compete for the same attentional resources. In order for attention to remain focused on the correct target location a correspondence between cue and target location is of importance. Based on these outcomes we conclude that attentional capture by sound is not a fully automatic process and can be prevented at an attentional level.

Chapter 4

Multisensory integration is more than directing attention

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (in prep). Multisensory integration is more than directing attention.

Abstract

It has been suggested that benefits of multisensory integration are due to a more efficient allocation of spatial attention. Other studies suggested that multisensory integration and spatial attention act independently. The current study investigated whether audiovisual integration can alter the perception of a visual event in conditions in which there is no role for spatial attention. We presented a visual target at fixation together with a spatially diffuse auditory cue. In the first experiment we show by means of a staircase procedure that sound affects the visual contrast detection threshold by lowering it. In the second experiment participants performed a detection task to rule out criterion shifts. The results show an increased sensitivity for the detection of a visual event when this event is accompanied by a sound relative to when no sound is present. We conclude that multisensory integration can improve the detectability of a visual event independent of spatial attention.

Introduction

Spatial attention is the mechanism by which a particular location is selected for further sensory processing. It has been argued that spatial attention may enhance the efficiency of processing (e.g., Posner, 1980), reduce stimulus uncertainty (e.g., Eckstein, Shimozaki, & Abbey, 2002; Palmer, 1994), or reduce interference from unattended locations (e.g., Theeuwes, 1991). Recently Carrasco, Ling and Read (2004) showed that spatial attention can alter the apparent stimulus contrast. These results imply that directing spatial attention results in a greater neuronal sensitivity (i.e., a decreased threshold), changing the strength of the stimulus by increasing its salience. A crucial role for spatial attention has also been implied in audio-visual integration. Indeed, in a recent study describing the so called *pip and pop effect* Van der Burg and colleagues (2008) demonstrated that audiovisual integration may help directing attention to the visual target stimulus. It was argued that the auditory signal boosts the salience of a synchronized visual event, allowing a better and more accurate allocation of spatial attention to the location of the visual event. In subsequent experiments Van der Burg and colleagues (2008) demonstrated that the pip and pop effect occurs in an automatic fashion: when a tone coincided with the presentation of a distractor observers automatically directed spatial attention to the distractor location resulting in a performance decrement. Similarly, in their classic study Spence and Driver (1997) demonstrated that a spatially non-predictive cue in the auditory modality can attract covert visual spatial attention suggesting a clear crossmodal link in attention. Overall these studies suggest that multisensory integration results in a more effective allocation of exogenous (stimulus-driven) attention to the target location.

On the other hand some studies have shown that multisensory integration and shifts of spatial attention are independent. For example the ventriloquism effect, a well-known perceptual consequence of multisensory integration, has been shown to occur preattentively and independently of both voluntary and involuntary spatial attention shifts (Vroomen et al., 2001a, 2001b). Moreover, a recent study by

Koelewijn and colleagues (2009) showed independent attentional effects for a visual and auditory cue presented at the same time in close proximity. In addition, previous studies using near threshold stimuli in the central visual field (Noesselt et al., 2008; Stein et al., 1996) showed that the location of the auditory cue does not have to be relevant for audiovisual integration to occur. This suggests that multisensory integration and exogenous spatial attention can act independently (see for discussion McDonald, Teder-Salejarvi, & Ward, 2001). In other words, multisensory integration by itself could lead to perceptual benefits without a role for spatial attention.

Previous studies that have claimed multisensory integration to be independent of attention (e.g., Noesselt et al., 2008; Stein et al., 1996) were not able to fully exclude attentional effects. For example, in the study of Stein and colleagues (1996) it was shown that an auditory stimulus enhances perceived visual intensities irrespective of the location of the sound. These enhancements were strongest at the lowest visual stimulus intensities but only occurred when the visual stimulus was presented at the centre of fixation. However, the visual stimuli were randomly presented from different non-predictive locations and therefore were able to draw on exogenous attentional processes. Moreover, a recent study by Noesselt and colleagues (2008) showed that spatial alignment of an auditory event is not necessary for audiovisual integration to occur during a visual spatial discrimination task. In their setup sound was presented through a single speaker positioned below the display and the visual target was presented non-predictive above or below the vertical meridian of the screen. Again this design is not able to exclude attentional effects influencing multisensory integration.

The present study examined this classic issue again using a paradigm in which we ensured that spatial attention could not play a role. Throughout the experiment we presented the visual stimulus at fixation. Therefore, there was no spatial uncertainty, nor did attention have to be shifted from one location to the next to respond to the visual stimulus. In addition, to rule out any spatial attentional effects of the auditory cue the auditory cue used in this study was spatially diffuse. This ensures that attention is not drawn to any specific location (Blauert, 1997;

Koelewijn et al., in press). If under these circumstances, the mere presentation of a sound can affect processing then we have evidence that multimodal integration can occur without crossmodal attention contributing to the effect.

In the present study, we presented a central visual event at different contrast ratios with or without an accompanying sound. In the first experiment the detection threshold of the visual target was measured in both a sound and a no-sound condition by means of a staircase procedure. These results showed an effect of sound on the visual detection threshold. In order to ensure that a shift in the detection threshold was not primarily based on criterion shifts a second experiment was conducted using a detection task and using contrast values within the same threshold range. Experiment 2 showed that sound affected visual sensitivity as A-prime (A') was reduced in conditions in which the sound was presented.

Experiment 1

In Experiment 1 we tested whether early audiovisual integration is independent of spatial attention by looking at shifts in the detection threshold. We presented a low contrast visual target either alone or together with a sound. In order to measure the detection threshold in both conditions a psychophysical staircase procedure was used. The sound used only provided temporal but not spatial information.

Method

Participants. Eight students of the Vrije Universiteit Amsterdam (4 male, mean age 21.9, ages between 18 and 29) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing. Participants were informed beforehand about the experimental procedure and were naïve as to the purpose of the experiment.

Apparatus and Design. Participants were seated in a dimly lit room at approximately 80 cm distance from the computer screen (17 inch, 120 Hz). The experiment was run in E-Prime 1.2 (1.2.1.847). The loudspeakers were placed at an angle of 18.3° from fixation and were aligned to the vertical middle of the screen.

The experiment consisted of one block in which four different psychophysical staircases were intermixed randomly. For the staircase direction variable there was an ‘up’ condition where the staircases started below the visual detection threshold, and there was a ‘down’ condition where the staircase started above the visual detection threshold. For the sound variable there was a ‘sound’ condition in which together with the visual target a sound cue was presented, and there was a ‘no-sound’ condition where only the visual target was presented.

Procedure and Stimuli. Each trial started with a random time interval of 400-650 ms after which the target in the form of a capital letter ‘S’ (diameter 0.57°) was presented for 20 ms. Immediately after the offset of the target a mask in the form of a filled light gray circle (diameter 1.65° , 20.754 cd/m^2) appeared for 80 ms. The participants’ task was to report in an unsped and accurate fashion whether or not they observed the target. They did this by pressing the letters ‘J’ for yes or ‘N’ for no on a QWERTY keyboard. Participants were told that the only target that could be presented was the letter ‘S’. Following the response, there was an inter-trial interval of 1000 ms before the next trial started. When the participant responded ‘yes’ the luminance of the next target within that staircase was lowered by 0.302 cd/m^2 . When the participant responded ‘no’ the luminance of the next target within that staircase was raised by 0.302 cd/m^2 . The background to which the target was presented had a luminance of 11.684 cd/m^2 . The two staircases that went up started with a target luminance of 11.684 cd/m^2 and therefore had a Michelson contrast ratio of 0. The two staircases that went down started with a target luminance of 16.219 cd/m^2 and therefore had a Michelson contrast ratio of 0.1625. All results will be presented in terms of Michelson contrast ratios instead of absolute values. When participants reached their tenth staircase reversal the experiment was terminated. In the sound condition 20 ms before the onset of the mask the auditory cue was presented at ± 65 decibels for 100 ms. The auditory cue consisted of two uncorrelated white noise bursts presented at the same time separately through each of the two speakers. This resulted in a auditory cue that was spatially diffuse and therefore creating a broad spatial percept in front of the participant, which extends to the sides beyond both loudspeakers (Blauert, 1997;

Koelewijn et al., in press). Note that the use of uncorrelated signals is essential here because two *correlated* signals are perceived as an easily localizable sound in the middle between the two loudspeaker positions due to summing localization (Blauert, 1997). At the beginning of the experiment, participants were told that the auditory cue provided no information about the appearance of the targets and therefore could be ignored.

Results

The ten staircase reversals for each condition averaged over subjects are plotted in Figure 4.1a. In addition Figure. 4.1b shows the average over these reversals (reversals 1 and 2 were excluded) for each condition. As a dependent variable we used the Michelson contrast ratio, which we calculated from differences in luminance values between the target and its background. The error bars in Fig. 1b

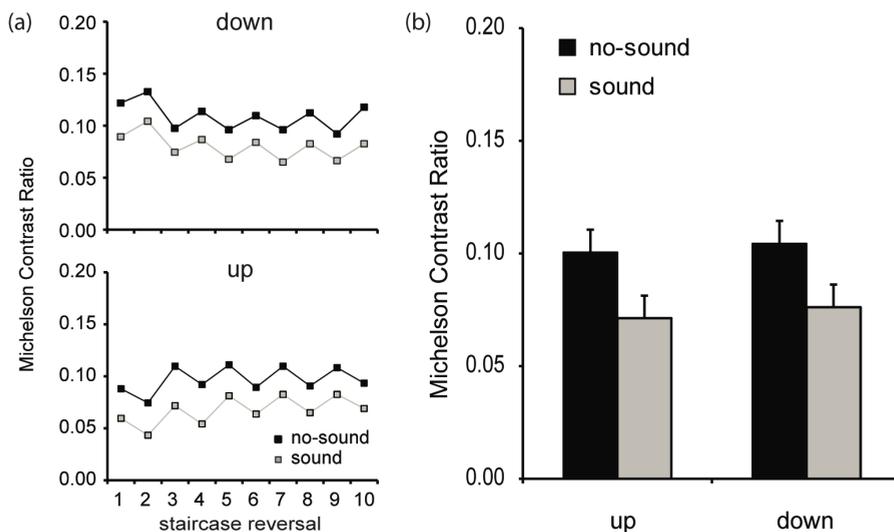


Figure 4.1. (a) Graph with the staircase reversals presented separately for the up and down staircase. The graph shows contrast detection thresholds for targets presented with or without sound, with reversal points presented on the x-axes and Michelson contrast ratios on the y-axes. (b) Average contrast ratios calculated over turning points 2 to 10 for each condition. The error bars show the 0.95 confidence intervals for the sound main effect.

represent the .95 confidence interval for the sound main effect according to Loftus and Masson (1994). An ANOVA was conducted on these averaged reversal scores with as variables staircase direction (up or down) and sound (sound or no-sound). The results show an effect for sound [$F(1,7) = 46.187, MSE = 1.222, p < .001$] and no effect for staircase direction. No interaction between sound and staircase direction ($F < 1$) was observed.

Discussion

The results show a lowering of the visual contrast detection threshold when a sound was presented together with the visual target. This effect is consistent over both the up and down going staircase. These results indicate that audiovisual integration gives a perceptual boost by lowering our detection threshold. These results are in line with earlier studies (Noesselt et al., 2008; Stein et al., 1996) that also show changes in visual detectability by means of sound. However, the current results can also reflect a criterion shift rather than an early multisensory integration effect. In order to exclude this possibility a second experiment was conducted in which participants performed a detection task in which the target was present in half of the trials. This allowed us to calculate hit and false alarm score determining the sensitivity in detecting the visual target in terms of A' (a more robust measure of sensitivity than the typical d' measure)^{4.1}.

Experiment 2

In Experiment 2 we tested whether lowering of the contrast detection threshold is indeed due to differences in detectability rather than criterion shifts. A similar setup was used but this time participants performed a detection task, which allowed us to calculate the A' .

^{4.1} The sensitivity measurements d' and A' require assumptions about their distributions. D -prime is based on a normal distribution. In the case of non-equal variances between the hit and false alarm rate the assumption of a normal distribution is violated and therefore d' cannot be used as a reliable dependent measure. In contrast, A' has no underlying distribution assumptions and thus provides the better sensitivity measure (see e.g., Verde, MacMillan, & Rotello, 2006).

Method

Participants. Eight new students of the Vrije Universiteit Amsterdam (3 male, mean age 21.1, ages between 18 and 24) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing. Participants were informed beforehand about the experimental procedure and were naïve as to the purpose of the experiment.

Apparatus and Design. The apparatus used was identical to Experiment 1. However, this time participants performed a detection task in which they only responded when they observed the target. The targets were shown at five different contrast ratios (0.0325, 0.0650, 0.0975, 0.1300, and 0.1625) distributed around threshold. Again there was a sound and no-sound condition

Procedure and Stimuli. The stimuli used were identical to Experiment 1 as well as the presentation duration of the stimuli. After the target and the mask were presented participants had 2500 ms to respond before the next trial started. Participants only responded when they perceived the target by pressing the spacebar on the keyboard. Within the sound and no-sound conditions in 50 percent of the trials there was no target presented in the other half a target was shown within one of the five contrast ratios. For each individual contrast condition with or without sound, 30 trials were presented. The experiment consisted of 6 blocks containing 100 trials each and within each block all conditions were randomized. Participants first practiced the task until they were confident in performing it.

Results

The A' detectability values for each contrast ratio within the sound and no-sound conditions averaged over subjects are plotted in Fig. 4.2. In an ANOVA we tested the linear trend which revealed an effect for sound [$F(1,7) = 32.412$, $MSE = 0.008$, $p < .005$] contrast ratio [$F(1,7) = 110.391$, $MSE = 0.021$, $p < .001$] and an interaction between sound and contrast ratio [$F(1,7) = 15.701$, $MSE = 0.223$, $p < .01$]. Post-hoc analysis in the form of five Bonferroni corrected 1-tailed paired samples t-tests revealed significant effects of sound for contrast ratios 0.0325 ($p < 0.05$), 0.0650 ($p < 0.005$), and 0.0975 ($p < 0.05$).

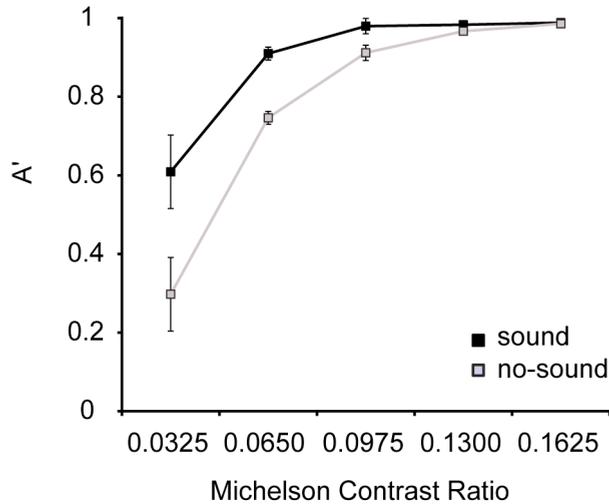


Figure 4.2. A' scores for each of the five contrast ratios for the sound and no-sound conditions. The error bars show the standard error mean for each contrast ratio.

Discussion

The results of Experiment 2 show a lowering of the A' detectability of a visual target when presented together with a spatially diffuse sound. The results confirm that lowering of the contrast detection threshold as shown in Experiment 1 reflects early audiovisual integration. As is clear from Figure 4.2, at relatively low visual contrasts the benefit of the sound is larger than with high visual contrasts. It is clear that at high contrasts, the sensitivity in detecting the target goes to ceiling (A' value of 1).

General discussion

Research on multisensory integration is inconclusive about whether multisensory integration can occur independent of crossmodal spatial attention. In other words, will multisensory integration by itself lead to perceptual benefits or does it always lead to more efficient allocation of attention, which in turn leads to performance benefits? The results of Experiment 1 showed that spatially diffuse sound lowers the detection threshold of visual contrasts within the central visual field. In addition, Experiment 2 showed a change in the sensitivity in detecting a visual target indicating that audio-visual integration can improve target detectability above and beyond any effects due to spatial attention

The current study shows a clear shift in sensitivity in detecting the visual target, which suggests that multisensory integration can occur independent of effects due to spatial attention. Although the current results are in line with those of Stein and colleagues (1996) and Noesselt and colleagues (2008) their results do not fully exclude attentional effects. However, their data suggest independence between multisensory integration and the location of the auditory event in respect to the central visual event. In other words, auditory and visual events do not need to be presented at the same location in order to integrate which is in line with the current results.

Studies looking at the effects of multisensory integration at an attentional level (Koelewijn et al., 2009; Santangelo, Finio et al., 2008; Santangelo et al., 2006; Ward, 1994) also confirm the idea that multisensory integration can occur independent of crossmodal attention. When using bimodal cues in order to draw visual attention, bimodal and unimodal cues show similar effect sizes (Santangelo et al., 2006). In other words, no integration effects show up when looking at bimodal spatial cueing. Even when multisensory integration is observed at a neuronal level these effects tend not to show up at a behavioral attention level (Santangelo, Van der Lubbe et al., 2008).

To conclude, the current results show that the detectability of a visual event improves when presented together with a sound. There was no spatial uncertainty

for the visual event and the sound provided no additional spatial information. Therefore, we conclude that audiovisual integration by itself leads to perceptual benefits independent of spatial attention.

Chapter 5

Priming T2 in a Visual and Auditory Attentional Blink Task

Koelewijn, T., Van der Burg, E., Bronkhorst, A., & Theeuwes, J. (2008). Priming T2 in a Visual and Auditory Attentional Blink Task. *Perception & Psychophysics*, 70(4), 658-666.

Abstract

Participants performed an attentional blink (AB) task containing digits as targets and letters as distractors within the visual and auditory domain. Prior to the rapid serial visual presentation a visual or auditory prime was presented in the form of a digit, which was identical to the second target (T2) on 50 percent of the trials. In addition to the ‘classic’ AB effect there was an overall drop in performance on T2 for the trials on which the stream was preceded by an identical prime from the same modality. There was no cross-modal priming suggesting that the observed inhibitory priming effects are modality specific. The current findings are assumed to represent a special type of negative priming operating at a low feature level.

Introduction

Capacity limitations in our visual system become evident when a vast amount of information needs to be processed within a limited period of time. A classic example of such a capacity limitation is the *attentional blink* (AB) deficit (Broadbent & Broadbent, 1987; Raymond, Shapiro, & Arnell, 1992). An AB occurs when people have to report two target items (e.g., words or single characters) presented among distractors in a *rapid serial visual presentation* (RSVP) in which items are presented in succession at a high rate (e.g., 10 items per second). People are accurate in reporting the *first target* (T1) but often fail to report the *second target* (T2). The AB is most severe when T1 and T2 are presented close in time (200 to 500 ms) but gradually disappears when the time period between the targets becomes longer.

One of the early models explaining the AB effect is the *two-stage* model by Chun & Potter (1995). As implied by its name, this model divides target detection into two stages; in the first stage relevant features of the target are detected, and in the second stage the target is consolidated into *short term memory* (STM). The model states that the AB deficit is based on a capacity limitation that occurs during consolidation of T2 into STM (Jolicoeur & Dell'Acqua, 1998). The AB occurs because resources used during consolidation of T1 are not available at that time when consolidating T2 is necessary. This results in a bottleneck in the transfer of sensory codes to STM.

According to the two-stage model consolidation is necessary for reporting T2. To get a better understanding of the process underlying the AB, it is important to know whether or not prior knowledge that is already consolidated into STM has an influence on the AB. A more direct question would be: What happens to T2 performance when an item identical to T2 is already shown and consolidated into STM prior to the presentation of T2? The classic study by Jacoby and Dallas (1981) shows that identification of an item (e.g., a word) improves as a result of prior exposure to an identical or similar item. This effect is called “repetition

priming” and based on the effect repetition priming has on performance, an improvement of T2 performance it to be expected.

However, Akyürek and Hommel (2005) found an overall performance drop on T2 in an AB task when participant held characters from the same class as T2 in memory. In their paradigm participants had to memorize in each trial a number of characters (letters, digits or symbols) displayed prior to the RSVP stream, which they had to report afterwards. Their performance on the AB task became worse when the STM load became higher and when the STM content became more related to the targets (letters, digits or symbols) in the AB task. Interestingly, the drop in performance did not interact with the AB and was constant over the lag condition. Another study (Nieuwenstein, Johnson, Kanai, & Martens, 2007) showed a similar drop in T2 performance when an STM set contained an item identical to T2 compared to an STM set without identical items. Both studies show that when an item identical to or from the same class as T2 is already consolidated in STM, performance on T2 drops. So, performance on T2 seems to suffer from competition between related items already stored in STM (Akyürek & Hommel, 2005), whereas the AB deficit itself is assumed to be the result of capacity limitations during consolidation of items into STM (Jolicoeur & Dell' Acqua, 1998). Nieuwenstein et al. (2007) explain this effect by a failure in attributing the same item information to both the STM and AB task. They refer to the additional failure in reporting T2 when this item is already coupled to a different task, in this case an STM task, as “cross-talk repetition amnesia”. Both studies conclude that the additional drop in T2 performance does not occur during the consolidation stage but seems to reflect interference in STM and is based on a different process than the AB.

The cross-talk repetition amnesia hypothesized by Nieuwenstein and colleagues (2007) is in line with the *episodic retrieval* model (DeSchepper & Treisman, 1996; Neill & Mathis, 1998; Neill, Valdes, Terry, & Gorfein, 1992) which is used to explain a process called *negative priming* (NP). NP is a label that is broadly used for describing perceptual inhibitory processes (Neill & Mathis, 1998; Rothermund, Wentura, & De Houwer, 2005; Tipper, 1985; Wood &

Milliken, 1998). The classical NP paradigm (Tipper, 1985) consists of prime and probe trials each containing a target and distractor item. When a distractor item that needs to be ignored in the prime trial becomes a target item in a subsequent probe trial, a response to this target tends to be slower. The *inhibition* model of NP (Houghton, Tipper, Weaver, & Shore, 1996; Tipper, 1985; Tipper, Weaver, Bastedo, Cameron, & Brehaut, 1991) explains this effect by inhibition of a to be ignored item. When the inhibited item becomes a target the activation threshold of the item is higher than that of uninhibited targets. The episodic retrieval theory of NP (DeSchepper & Treisman, 1996; Neill & Mathis, 1998; Neill et al., 1992) states that the representation of an item is stored together with an ‘action tag’ indicating what to do with the item. This action tag could involve the coupling to a task as suggested by cross-talk repetition amnesia. In the case of NP a distractor item stored with a no-response tag creates conflict when it becomes a target item to which participants have to respond to.

In the present study participants were presented with a RSVP stream of letters containing two digits as targets. Prior to the RSVP stream a prime was presented that was either identical or not identical to T2 but always belonged to the same class (digits) as the targets. So instead of an additional memory task as was used by Akyürek and Hommel (2005) a single prime was used. The question we addressed was whether a similar performance reduction would show up as was reported in experiments that used an additional memory task or whether a positive enhancement effect would show up due to repetition priming. To anticipate the results, our Experiment 1 shows an inhibitory effect on T2 performance when primed by a physically identical item, which is in line with previous work (Akyürek & Hommel, 2005; Nieuwenstein et al., 2007). Three additional experiments were conducted to investigate whether this effect could be explained in terms of NP.

A second objective of this study was to determine whether this effect on T2 performance is restricted to the visual domain. It is quite feasible based on previous priming (Graf, Shimamura, & Squire, 1985) and NP literature (Buchner, Zabal, & Mayr, 2003) that similar effects exist in auditory and even crossmodal conditions.

To answer this question two additional experiments were conducted using an auditory or a visual serial stream preceded by an auditory prime.

Experiment 1: Visual-visual priming

In Experiment 1, participants were presented with an RSVP stream containing two target digits among distractor letters. Prior to the RSVP stream a prime was presented that was either identical or not identical to the second target (T2). Participants had to report whether the target digits presented within the RSVP stream were odd or even.

Method

Participants. 12 students of the Vrije Universiteit Amsterdam (9 female, mean age 22.3, ages between 19 and 33) took part in the experiment. All had normal or corrected-to-normal vision. Participants were informed beforehand about the experimental procedure and were naïve as to the purpose of the experiment.

Design and Stimuli. This experiment had a 2*4 design; factors were: Prime (prime \neq T2, prime = T2), and Lag (1, 2, 3 or 8). The RSVP stream contained twenty elements. Each element was presented for 16 ms followed by an inter stimulus interval (ISI) of 80 ms. T1 was presented at position 7, 8, or 9 in the RSVP stream and T2 was positioned 1, 2, 3, or 8 lags after T1. All T1 and T2 position combinations occurred equally often in a random order. Prior to the RSVP stream a prime was displayed for 1.5 s, followed by a fixation cross for 200 ms as shown in Figure 5.1. The prime was displayed in all trials and was identical to T2 on 50% of the trials and never identical to T1. For the prime, T1, and T2 digits were used (1-9, 5 excluded) and the distractors were capital letters of the alphabet (letters I and X excluded). All characters were displayed at the center of the screen in dark gray in 48-point Geneva (0.63 cd/m², 1.4° width, 1.6° height) on a gray (9.34 cd/m²) background. During practice all characters were displayed in black to familiarize participants with the task.

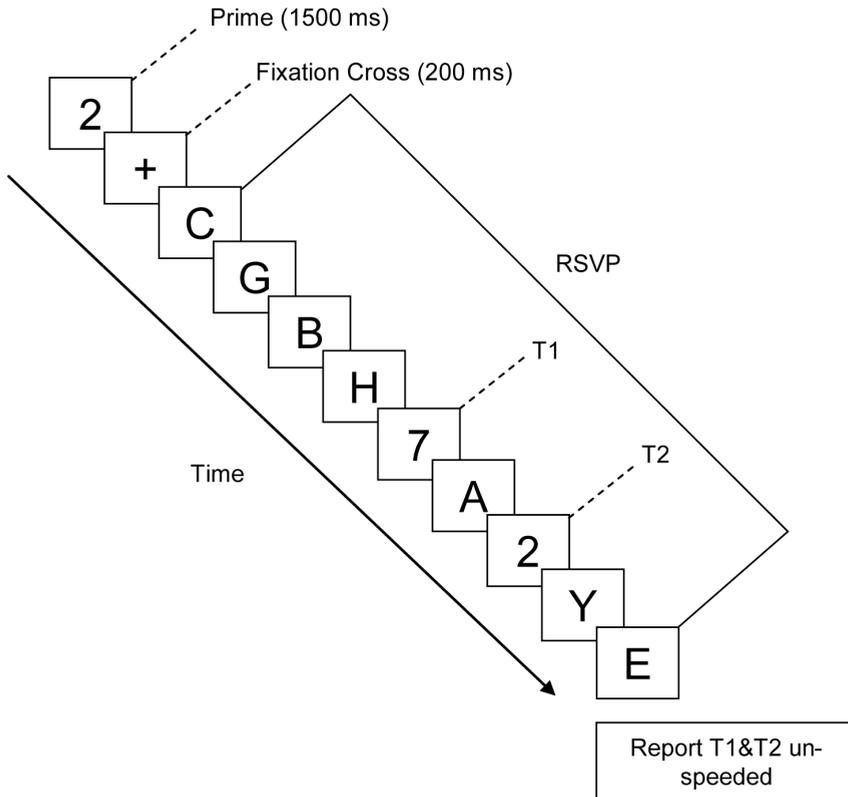


Figure 5.1. Schematic representation of the paradigm used. Participants received a rapid serial visual presentation (RSVP). The task was to identify the two targets (T1 and T2) which were digits and to report un-speeded in order of appearance whether they were odd or even. Prior to the RSVP stream a prime was presented for 1.5 s. In 50% of the trials the prime was identical to T2.

Apparatus and Procedure. Participants were seated in a dimly lit cabin at approximately 80 cm distance from a computer screen (17 inch, 120 Hz). The experiment was run in E-Prime 1.1 (SP3). Task instructions were presented on screen after which the participants started with a practice block of 48 trials. Participants were instructed to look at the prime digit. They were told that it was irrelevant for the task and that they did not need to respond to it. After each block participants received feedback on their overall performance. The experiment

consisted of 6 blocks containing 48 trials each. The participants had to respond un-speeded to T1 and T2 sequentially by pressing the o-key for odd or e-key for even on a QWERTY keyboard.

Results

For all analysis a significance level of $p < .05$ was used, and *MSE* and *p* values were Greenhouse-Geisser adjusted when required. Two separate repeated measurements analyses of variance (ANOVA) were conducted on performance on T1 and on performance on T2 given T1 was correctly identified. Prime and lag were within subject variables. Figure 5.2, shows the mean percentage correct for T2 given that T1 was correctly identified, as a function of lag and prime.

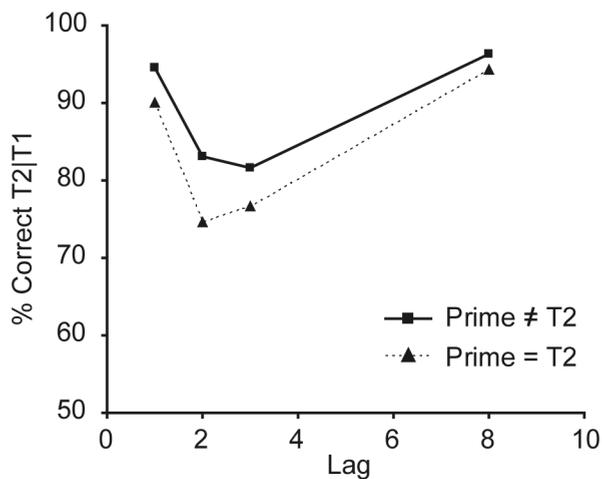


Figure 5.2. Results of Experiment 1 in percentage correct for T2 given T1 correct (T2|T1), as a function of lag and prime.

T1 Accuracy. Overall, performance was at 91% correct. For the ‘prime = T2’ condition the % correct scores for T1 for lags 1, 2, 3, and 8 were 88%, 91%, 92%, and 96%, respectively. The results for the ‘prime ≠ T2’ condition in the same order were 83%, 93%, 92%, and 95%. The main effect of prime failed to reach significance ($F < 1$), performance varied with lag [$F(3, 33) = 10.523, MSE = .006, p < .005$], and the two-way interaction between prime and lag was significant [$F(3, 33) = 3.401, MSE = .003, p < .05$]. The interaction was further analyzed by pairwise t-tests between the prime conditions for each lag (1-3 and 8) which only showed a significant effect for lag 1 ($p < .05$). The main effect of lag was also further analyzed by pairwise t-tests between lags. These results only showed a significant effect for lag 1 relative to lag 2 ($p < .05$).

T2 Accuracy. Performance varied with lag [$F(3, 33) = 16.357, MSE = .023, p < .001$]. There was a significant effect of prime [$F(1, 11) = 12.368, MSE = .005, p < .01$] resulting in an overall drop in performance when T2 was primed (84%) compared to when T2 was not primed (89%). The two-way interaction between prime and lag failed to reach significance [$F(3, 33) = 1.525, MSE = .004, p = .236$].

Discussion

Priming of T2 did not affect T1 performance with exception of the first lag. The drop in performance on T1 for Lag 1 is similar to the effect reported by Akyürek and Hommel (2005), as well as by Potter, Staub and O’Conner (2002). They base this effect on competition between two succeeding targets when selected from a same set of characters (digits). This competition seems to be strongest when T2 is not identical to the prime as shown by the observed interaction. Another explanation for this effect could be that participants remembered T1 and T2 in the incorrect order. At the end of the trial participants had to report whether the two targets (T1 and T2) were odd or even. This had to be done un-speeded and in the correct order. So not only the targets themselves had to be remembered but also the order of appearance. When targets are presented close in time which is the case for lag 1, and both characters are from the same character class, this order judgment can become difficult. This will result in a drop in T1 performance on lag 1. Note

that this effect is not apparent in the scores because only the scores of T2 given T1 correctly identified are shown.

The typical u-shaped drop on T2 performance between lag 1 and 8 shown in Figure 5.2 reflects a standard AB effect (Raymond et al., 1992). The results further show an overall inhibitory effect of prime on T2 performance and no interaction between prime and lag. These results are in line with earlier work (Akyürek & Hommel, 2005; Nieuwenstein et al., 2007) and show that when an item identical to T2 is previously consolidated it becomes harder for participants to correctly identify it.

In this experiment participants were instructed to look at the prime, but were not required to actively maintain it. Unlike in a typical NP paradigm in which participants have to ignore the prime in order to make the appropriate response, in the current experiment observers were asked to attend to the prime but did not have to give an overt response. However, even though we did not instruct participants to do so, we cannot rule out that the participants actively suppressed the prime as a kind of strategy in enhancing their task performance. If participants would have actively suppressed the prime, it would be consistent with the idea that the drop in performance during the AB task is the result of inhibition.

According to the inhibition model, if participants would actively maintain the prime, one would not expect a drop in performance for T2 (since there is no need for inhibition); if anything, on the basis of the results of Tipper (1985) one would expect a performance benefit for T2. Experiment 2 was basically a replication of Experiment 1 but now we ensured that participants actively maintained the prime in STM.

Experiment 2: Memorizing the prime

In this experiment participants were instructed to memorize the prime instead of just attending to it. To make sure that participants followed the instructions, we added a few so-called prime recall trials. In these trials, participants had to report the identity of the prime after presentation of the RSVP stream. Because the prime

is now actively maintained in STM, we expect enhanced T2 performance when the prime and T2 are identical (Tipper, 1985).

Method

The present experiment was identical to Experiment 1, except that participants were instructed to memorize each prime the trial presented. Additional ‘prime recall’ trials were included, constituting 20 % of the trials. These trials were identical to the other trials with the exception that a different response had to be given. Instead of reporting the targets, a number word (e.g., the word ‘eight’) appeared on screen and participants were asked to indicate by pressing ‘J’ for ‘yes’ and ‘N’ for ‘no’ whether the number word displayed was the same or different from the prime digit kept in memory. Twelve new students (6 female, mean age 23.3, ages between 17 and 37) participated in the experiment.

Results

On average participants scored 98% correct on the prime recall trials. This provides a strong indication that participants actively observed the prime and maintained in STM during a trial.

T1 Accuracy. T1 was correctly identified on 87% of the trials. For the ‘prime = T2’ condition T1 performance on lag 1, 2, 3, and 8 was 80%, 90%, 88%, and 92%, respectively. The results for the ‘prime ≠ T2’ condition in the same order were 76%, 88%, 91%, and 93%. There was no significant effect of prime on T1 performance ($F < 1$), T1 performance varied significantly with lag [$F(3,33) = 9.336$, $MSE = .022$, $p < .004$], and there was no significant interaction between prime and lag [$F(3,33) = 1.073$, $MSE = .005$, $p = .363$].

T2 Accuracy. The average scores for each condition are shown in Figure 5.3. T2 given T1 correct varied with lag [$F(3, 33) = 20.381, MSE = .029, p < .001$] indicating an AB effect. There was a significant main effect of prime on T2 performance [$F(1, 11) = 9.308, MSE = .007, p < .05$] resulting in poorer performance when T2 was primed (78%) than when T2 was not primed (83%). The two-way interaction between prime and lag was significant [$F(3, 33) = 4.299, MSE = .008, p < .05$], and further analyzed by two-tailed t-tests for each lag (1,2,3 and 8), which only showed a significant effect for lag 2 ($p < .05$).

Comparing overall T2|T1 performance between Experiment 1 and 2 by means of one-sided independent-samples t-test revealed a significant ($p = .0265$) drop in performance for Experiment 2 relative to Experiment 1.

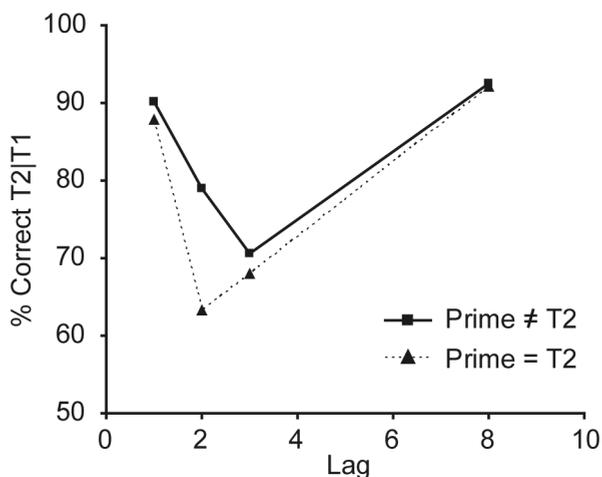


Figure 5.3. Results of Experiment 2 in percentage correct for T2 given T1 correct (T2|T1), as a function of lag and prime.

Discussion

The current experiment replicated and extended the findings of Experiment 1 by showing an inhibitory effect of prime on T2 performance, even when it was ensured that the prime was stored in STM. Thus, contrary to the predictions of the inhibition model of NP (Houghton et al., 1996; Tipper et al., 1991), in the current paradigm actively maintaining the prime does not lead to any facilitation. Note that these results are in line with earlier studies that showed a similar drop of T2 performance when items related to T2 were actively maintained in STM (Akyürek & Hommel, 2005; Nieuwenstein et al., 2007). However, in contrast to Experiment 1, the current results show an interaction between prime and lag, indicating a clear priming effect on lag 2, but not on the other lags.

It is likely that this interaction we obtained in Experiment 2 is due to the fact that the memory task had an overall detrimental effect on performance. Regardless of whether the prime matched T2 or not, performance at lag 2 and 3 was much worse in Experiment 2 than in Experiment 1. It is feasible that in Experiment 2 the effect of the prime was different at the different lags because the overall performance decrement was close to ceiling. Therefore the interaction may be due to the fact that the memory task caused a strong performance decrement, which was only marginally affected by the addition of the prime. Regardless of this interaction, the most important outcome of the current experiment is that there is a main effect of the prime on T2 performance which is opposite to what would be predicted when the prime would not have been inhibited.

All in all, the current findings suggest that the negative effect on T2 performance may be related to NP although not according to the inhibition model. If it is indeed NP then the negative effect does not need to depend on whether the prime is physically identical to the target (for review see Neill & Mathis, 1998). For instance, NP will also occur when using either pictures or words that are semantically related to the target (Tipper, 1985; Yee, 1991). If it is true that the effect observed in Experiment 1 and 2 are based on NP, a similar effect should be observed when using a semantically related prime instead of an identical prime. To

test this assumption a third experiment was conducted in which instead of a digit (e.g., ‘8’) a number word (e.g., ‘eight’) was presented as a prime.

Experiment 3: Semantic priming

In this experiment participants observed a prime in the form of a number word that was either the same number (e.g., prime “eight” target “8”) or not (e.g., prime “eight” target “2”). T1 and T2 were still presented as single digit characters. In contrast to Experiment 1 the prime and T2 were not identical but shared the same semantics. This experiment allowed us to investigate the influence of semantic priming on T2 performance.

Method

The present experiment was identical to Experiment 1 except that a semantic prime was presented as a number word in Dutch (één, twee, drie, vier, zes, zeven, acht, and negen). Sixteen new students (10 female, mean age 20.2, ages between 17 and 31) participated in the experiment.

Results

T1 Accuracy. T1 was correctly identified on 94% of the trials. For the ‘prime = T2’ condition the performance on lag 1, 2, 3, and 8 was 87%, 94%, 97%, and 97%, respectively. The results for the ‘prime ≠ T2’ condition in the same order were 86%, 96%, 98%, and 98%. There was no significant effect of prime on T1 performance [$F(1,15) = 1.259$, $MSE = .001$, $p = .280$]. T1 performance varied significantly across lags [$F(3,45) = 25.833$, $MSE = .007$, $p < .001$]. Further analysis by means of a two-tailed paired sample t-test showed a significant performance drop on lag 1 compared to lag 2 [$t(15) = 4.612$, $p < .001$] and on lag 2 compared to lag 3 [$t(15) = 3.381$, $p < .005$]. No significant two-way interaction between prime and lag was found [$F(3,45) = 1.142$, $MSE = .001$, $p = .338$].

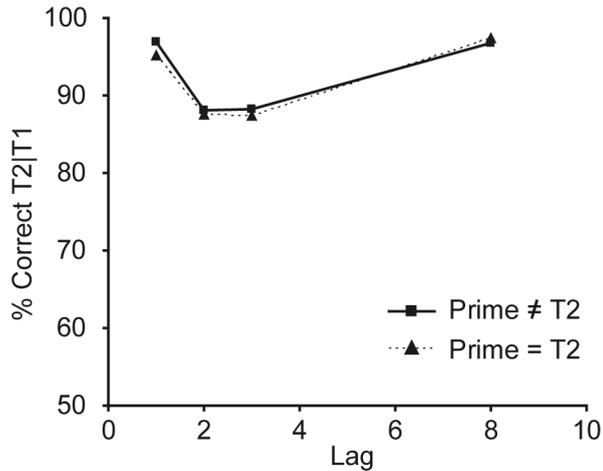


Figure 5.4. Results of Experiment 3 in percentage correct for T2 given T1 correct (T2|T1), as a function of lag and prime.

T2 Accuracy. The average score for the individual conditions are shown in Figure 5.4. The performance on T2 given T1 correct varied with lag [$F(3, 45) = 13.595$, $MSE = .013$, $p < .005$] indicating an AB effect. For T2 there was no significant main effect of prime ($F < 1$) and a no two-way interaction between prime and lag ($F < 1$).

Discussion

When primed semantically no effect on T2 performance was found. The results differ from the results of Experiments 1 and 2 in which performance on a primed T2 dropped relative to the performance on a non-primed T2. For an inhibitory effect to occur (as shown in Experiment 1) it seems necessary that the prime is physically identical to the target. Therefore, these current results are not in line with the classic NP explanation referred to as the inhibition model as developed by Tipper and colleagues (Houghton et al., 1996; Tipper, 1985; Tipper et al., 1991).

However, before excluding NP as an explanation for the performance decrement a close comparison between the NP and current paradigm is needed.

The paradigm here differs from the classic NP paradigm of Tipper (1985). Typically the NP paradigm consists of prime and probe trials each containing a target and distractor item. NP occurs when an ignored distractor item in the prime trial becomes a target in the probe trial. Still, there are also similarities between the two paradigms. First of all, in both paradigms two targets are presented. This means that T1 in the RSVP stream can be seen as the target in the prime trial and T2 as the target in the probe trial. Secondly, in both paradigms distractors are present. Additional to the distractors presented in the RSVP stream surrounding T2, the prime presented prior to the RSVP stream could function as a distractor for T1. Note that in the current experiment targets and distractors were not presented simultaneously, however, a study by Neill and Mathis (1998) showed that this has no consequences for NP. It could be the case that the interaction between the prime and T1, both being from the same character class, is causing the inhibition effect on T2 performance. When T1 and the prime are less related (e.g., from different character classes) there should be less competition between them and therefore no reason for the prime to be inhibited.

Experiment 4: Different character class used for T1

In Experiment 1 and 2 prime, T1, and T2 were all digits. Because they are from the same class of stimuli, the prime can act as a distractor item for T1. To determine whether this is indeed the case, T1 is in this experiment taken from a different character class as the prime and T2. If it is true that the prime is suppressed because it competes with T1 we expect the NP effect to disappear when T1 is of a different character class than that of the prime and T2.

Method

Experiment 4 was similar to Experiment 1 except that instead of digits the symbols ‘#’ and ‘%’ were used for T1. Subjects had to indicate which symbol they had seen by pressing the ‘Z’ key for symbol ‘#’ and the ‘M’ key for symbol ‘%’. For T2 the same digits were shown as in the previous experiments and again subjects had to indicate whether these were odd or even. During a pilot study it became clear that this change made the task much easier and the performance on T1 and T2 may reach ceiling. Therefore, the ISI between the items in the RSVP stream was reduced from 80 ms to 40 ms, which made the task equally difficult as Experiments 1 and 2. Eight new students (6 female, mean age 20.8, ages between 18 and 25) participated in the experiment.

Results

T1 Accuracy. T1 was correctly identified on 95% of the trials. For the ‘prime = T2’ condition the performance on lag 1, 2, 3, and 8 was 97%, 95%, 93%, and 95%, respectively. The results for the ‘prime ≠ T2’ condition in the same order were 97%, 94%, 95%, and 96%. There was no significant effect of prime on T1 performance ($F < 1$). T1 performance varied significantly with lag [$F(3,21) = 4.441$, $MSE = .001$, $p < .05$]. Further analysis by means of a two-tailed paired sample t-test showed a significant performance drop on lag 3 compared to lag 8 [$t(7) = -2.728$, $p < .05$]. No significant two-way interaction for prime x lag was found [$F(3,21) = 1.359$, $MSE = .001$, $p = .287$].

T2 Accuracy. The average scores for the individual conditions are shown in Figure 5.5. T2 given T1 correct varied with lag [$F(3, 21) = 11.133$, $MSE = .006$, $p < .005$]. There was no significant effect of prime on T2 performance ($F < 1$), and the two-way interaction between prime x lag was not significant ($F < 1$).

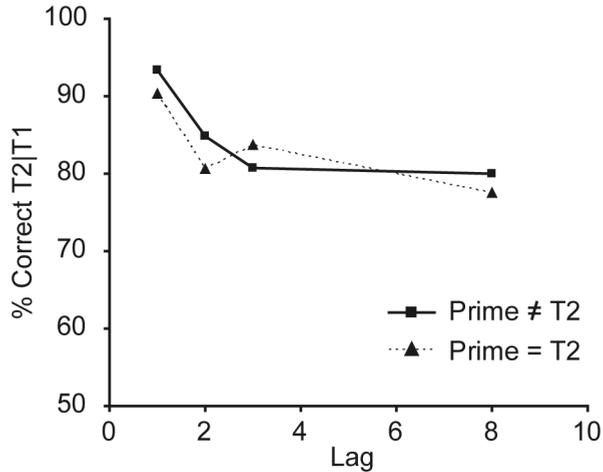


Figure 5.5. Results of Experiment 4 in percentage correct for T2 given T1 correct (T2|T1), as a function of lag and prime.

Discussion

In contrast to what was found in Experiment 1 and 2, priming T2 now had no effect on T2 performance. Changing T1 into a symbol apparently made the prime no longer a distractor item for T1 and therefore no additional inhibitory effect was observed. This is in line with the idea that NP occurs when a distractor item is suppressed by the target item in the prime trial (Houghton et al., 1996; Tipper, 1985; Tipper et al., 1991).

Displaying T1 as a symbol did not affect the AB as shown by the main effect for lag that was still present. However, no typical u-shape, characteristic for the AB, was found. One reason for not finding a classic u-shape is that the u-shape of the AB is based on the time between T1 and T2 and not on the number of lags between the two targets (Martens, Munneke, Smid, & Johnson, 2006). By speeding up the RSVP stream, at lag 8 only 448 ms had passed instead of the 768 ms in Experiment 1 and 2. Obviously, after 448 ms one still expects an AB effect.

A minor point of discussion in Experiment 1 was the drop in T1 performance on lag 1. We suggested that this effect could be explained by the fact that the task requires a correct order judgment of the targets, which becomes difficult when targets are presented as close in time as on lag 1. In the current experiment no correct order judgment was necessary because the participants simply knew that a target symbol was presented before a target digit. Even though the effect of lag remained significant post-hoc analyses revealed no significant drop of T1 performance on lag 1, which is consistent with the above explanation.

A second objective of the current study was to determine whether the additional inhibitory effect on T2 performance is restricted to the visual domain. In the following experiment we investigated if an auditory prime influences the performance on an auditory T2. Such an effect might be expected because Buchner, Zabal, and Mayr (2003) found NP effects in the auditory domain, when participants had to discriminate between sounds of animals or musical instrument.

Experiment 5: Auditory-auditory priming

The additional inhibition effect on T2 performance as shown in Experiment 1 and 2 seems to happen in an early visual processing stage. This raises the question whether this effect is specific for the visual domain. To address this issue, participants were presented a rapid serial auditory presentation (RSAP) comparable to the RSVP stream used in all previous experiments, preceded by an auditory prime. Task and design were similar to Experiment 1.

Method

For this experiment auditory stimuli were used in the form of spoken letters (distractors) and digits (targets and prime). The letters that were used in the RSAP stream as distractors were: B, C, D, F, G, H, J, K, L, M, N, O, R, U, V, W, and X. For the prime and targets the digits 1, 2, 3, and 4 were used. All letters and digits were spoken in Dutch by a male voice and compressed to a duration of 90 ms. The spoken prime digits were identical to the targets and their onset was time locked

1500 ms before the onset of the fixation cross. All vocals were digitally recorded and edited with 16 bit resolution and 44 kHz sampling rate using Cool Edit Pro 2.1 software. During recording voice inflections were kept to a minimum. Amplitudes for all samples were manually normalized and time compression was performed by means of a time stretching routine, which manipulated duration without altering pitch.

To make sure that participants would perceive the edited spoken digits correctly a pilot study was conducted with four subjects. Participants were presented with a RSAP stream similar to the one used in this experiment but which only contained one target digit; no prime was presented prior to the RAP stream. Participants had to give a speeded response by pressing the corresponding number on their keyboard. In this pilot study all digits from 1 to 9 were tested; based on the results digits 1 to 4 were selected as best candidates for the experiment. Nine new students (5 female, mean age 21.5, ages between 18 and 30) participated in the experiment. All had normal hearing and vision, and were paid 7 euros for a single one hour session.

Results

Data of one participant were excluded from further analysis due to not complying to the task instructions. The average scores for the individual conditions are shown in Figure 5.6.

T1 Accuracy. T1 was correctly identified on 84% of the trials. For the ‘prime = T2’ condition the performance on lag 1, 2, 3, and 8 was 83%, 85%, 86%, and 85%, respectively. The results for the ‘prime ≠ T2’ condition in the same order were 81%, 81%, 84%, and 87%. There was no significant effect of prime [$F(1,7) = 2.418, MSE = .002, p = .164$], lag [$F(3,21) = 2.253, MSE = .005, p = .146$], nor was there a two-way interaction between prime and lag ($F < 1$) on T1 performance.

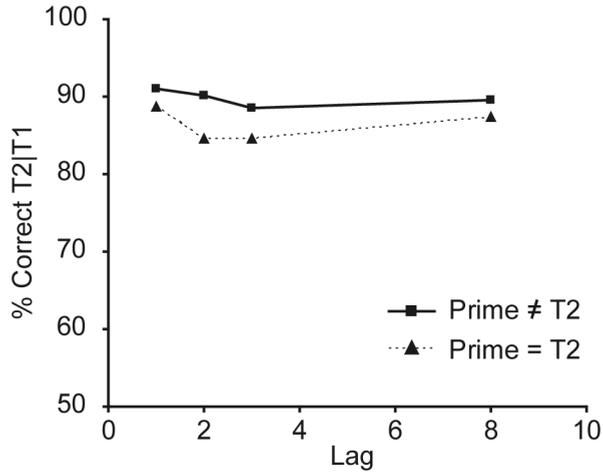


Figure 5.6. Results of Experiment 5 in percentage correct for T2 given T1 correct (T2|T1), as a function of lag and prime.

T2 Accuracy. The average score for the individual conditions are shown in Figure 5.6. T2 given T1 correct did not vary across lags ($F < 1$) indicating that there was no AB effect. There was a significant effect of prime on T2 performance [$F(1, 7) = 8.449, MSE = .002, p < .05$] resulting in an overall drop in performance when T2 was primed (86 %) compared to when T2 was not primed (90 %). The two-way interaction between prime and lag was not significant ($F < 1$).

Discussion

In this experiment no AB effect was found. This is in line with results from earlier studies (e.g., Arnell & Jenkins, 2004; Potter, Chun, Banks, & Muckenhoupt, 1998) which show that auditory ABs can occur, but not when digit targets among letter distractors need to be reported. Nevertheless, we did find an inhibitory effect of the prime on T2 performance suggesting that the NP effect is not specific to the visual

modality. The fact that it occurs in the absence of an AB underlines the idea that NP taps into different resources than the AB.

Experiments 1, 2, and 5 show inhibitory effects on T2 performance in both the visual and auditory domain. It is well known (see for example Arnell & Jolicoeur, 1999; Spence & Driver, 1997) that auditory input can have an effect on processing visual information. Furthermore, Buchner and colleagues (Buchner et al., 2003) have demonstrated NP for a prime and target that are presented in different modalities. A relevant question is therefore whether similar crossmodal inhibitory effects will show up when we use a crossmodal variant of our paradigm.

Experiment 6: Auditory-visual priming

Experiment 6 tests whether the inhibitory effects shown in Experiments 1, 2 (visual), and 3 (auditory) hold in a crossmodal setting, where an RSVP stream is preceded by an auditory prime. When an auditory prime would influence the performance on T2 it could mean that there is interference between the auditory and visual modality at an amodal level. If this effect would yield similar findings as Experiment 1 and 2 it would provide additional information concerning the level at which the observed inhibition takes place.

Method

The present experiment was identical to Experiment 1, except that an auditory prime was presented. The prime was a spoken digit in a male voice in Dutch and was presented by means of a Sennheiser HD 202 headphone. The sound samples used had an average duration of 425 ms (16 bit, 44kHz). The onset of the sound sample was time locked 1000 ms before the onset of the fixation cross. The sound samples were manually normalized for amplitude. Twenty-four new students (10 female, mean age 20.6, ages between 15 and 36) participated in the experiment. All had normal hearing and vision.

Results

T1 Accuracy. T1 was correctly identified on 91% of the trials. For the ‘prime = T2’ condition the performance on lag 1, 2, 3, and 8 was 88%, 94%, 94%, and 97%, respectively. The results for the ‘prime ≠ T2’ condition in the same order were 87%, 94%, 95%, and 97%. There was no significant main effect of prime, $F < 1$. The two-way interaction between prime and lag failed to reach significance, $F < 1$. There was a main effect of lag [$F(3, 69) = 28.474$, $MSE = .004$, $p < .001$]. Further analysis by means of a two-tailed paired sample t-test showed a significant performance drop on lag 1 compared to lag 2 [$t(23) = 4.459$, $p < .001$] and a raise in performance on lag 8 compared to lag 3 [$t(23) = 3.551$, $p < .005$].

T2 Accuracy. The average scores for the individual conditions are shown in Figure 5.7. The performance on T2 given T1 correct varied with lag [$F(3, 69) = 14.534$, $MSE = .009$, $p < .001$] indicating an AB effect. For T2 there was no significant main effect of prime [$F(1, 23) = 2.461$, $MSE = .003$, $p = .130$]. The two-way interaction between prime and lag was not significant [$F(3, 69) = 1,661$, $MSE = .003$, $p = .204$].

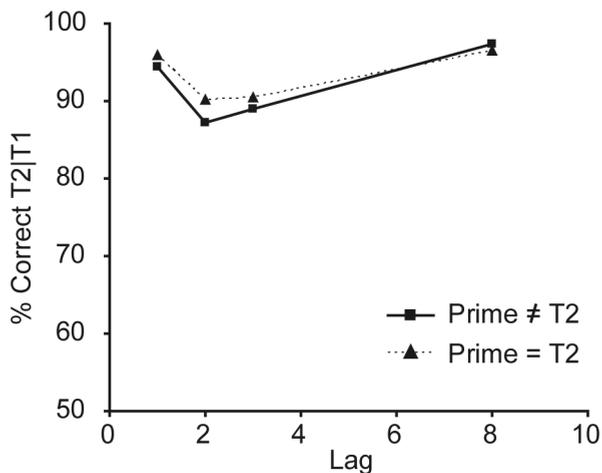


Figure 5.7. Results of Experiment 6 in percentage correct for T2 given T1 correct (T2|T1), as a function of lag and prime.

Discussion

Unlike in Experiment 1 and 2, the current results do not show a significant effect of the prime on T2 performance. A reason for the observed null-result could be that cross-modal priming is simply not possible within this paradigm. Even though earlier reports show cross-modal priming (Graf et al., 1985) and cross-modal NP (Buchner et al., 2003), in these studies, stimuli like words and line drawings were used. In order to find a crossmodal effect, input from different sensory modalities might need to interact on a level where amodal semantic representations are already formed. In line with the results of our Experiment 3 the present study confirms the notion that priming at a semantic level does not cause a negative effect on T2 performance.

General discussion

This study began with the premise that in an AB task, priming T2 with an identical item would have an inhibitory effect on T2 performance. Both Experiment 1 and 2 show this effect and also reveal that this effect adds to the classic AB effect. Similar effects were reported by earlier studies (Akyürek & Hommel, 2005; Nieuwenstein et al., 2007). One of our conjectures was that this effect may be the result of NP. If the inhibitory effect that we reported is indeed due to NP, one would expect that the effect disappears when T1 is from a different character class than the prime and T2. Indeed, one can argue that when there is no competition between prime and T1 anymore then there is no need for inhibition. Our Experiment 4 confirmed this prediction by showing that the inhibitory effect of the prime disappears when T1 is changed. In addition, we wanted to determine whether this inhibitory effect would also occur during auditory unimodal priming and auditory-visual cross-modal priming. Experiment 5 indeed shows an auditory priming effect when T2 is also presented auditorily. However, cross-modal (auditory-visual; Experiment 6) priming had no effect on T2 performance.

The results from Experiments 1, 2, and 5 are consistent with results from previous studies (Akyürek & Hommel, 2005; Nieuwenstein et al., 2007) that showed similar inhibitory effects on T2 performance when it is preceded by an item that is identical to T2 or from a similar character class. The fact that the semantic primes used in Experiments 3 and 6 did not influence T2 performance suggests that the observed inhibition takes place before a semantic level is reached. There are several ways to interpret this uni-modal inhibitory priming effect.

The idea that identical items presented in different tasks during the same trial are harder to retrieve than items used in a single task was named “cross-talk repetition amnesia” by Nieuwenstein and colleagues (2007). In this study an STM set was presented prior to a standard AB task. When an STM item was identical to one of the targets in the AB task, performance for these targets decreased. According to the idea of cross-talk repetition amnesia both the STM item and the targets in the AB task go through a ‘tokenization’ process where binding between episodic features and item representations takes place. When an item is stored with episodic features from different tasks, retrieval of this item causes interference between these different episodic features. This interference can result in errors during retrieval. Our results indicate that it is not necessary to use a whole STM set – also a single prime can evoke such errors. In other words, no cross-talk between tasks is needed in order to observe inhibition on T2 performance, just a single item that does not have to be retrieved will have the same effect. One could argue that looking at the prime can be seen as a separate task, and then the idea of cross-talk repetition amnesia is still valid. In a more general way, our results show that there is an interaction between items perceived prior to an AB task and identical items that are part of the RSVP stream, and that this interference does not have to take place during the retrieval period as cross-talk repetition amnesia suggests but could also take place during consolidation of T2. Still the basic mechanism behind this effect needs to be explained. We reported earlier that the idea of cross-talk repetition amnesia is in line with the episodic retrieval modal of NP.

In our Experiment 2 episodic retrieval still cannot be ruled out as possible explanation. Note that in this experiment the prime was shown for 1.5 seconds and

that the participants sometimes had to report it. In these trials the response to the prime differed from the response to the targets (T1 and T2) in the other trials. For the prime, participants had to indicate whether or not the prime was similar to a test number and for the target whether they were odd or even digits. In line with Neill and Mathis (1998) one could then argue that in this case the episodic retrieval model holds, assuming that the prime and the targets are processed differently. Even though both the prime and targets were stored with a 'respond to' tag, the different type of response ('the same as' compared to 'odd' or 'even') was stored as well. This difference could explain the NP effect that we found. However, Experiment 4 shows no NP effect when T1 is from a different character class than the prime and T2. According to the episodic retrieval model, NP should in this case still occur because the prime and T2 require different responses. This means that the episodic retrieval model can only offer a partial explanation of our results.

According to the "inhibition model", NP seems to operate as a post selective mechanism on a central semantic level and can be influenced by the participants' strategies (see May et al. (1995) for a review). Both auditory and visual, as well as cross-modal priming may occur (Buchner et al., 2003) making it a possible explanation for the inhibitory effects of uni-modal priming observed on T2 performance (Exp. 1, 2, and 5). However, as noted earlier, the results from Experiment 2 are not in line with the idea of the inhibition model of NP, which predicts facilitation effects for actively observed primes. Moreover, the results of Experiment 3 show no effect of semantic priming and the results of Experiment 1 and 2 indicate that participants' strategies have basically no influence. Nevertheless, we have also found evidence that inhibition of the prime underlies the drop in T2 performance that we observe in our experiments. In Experiment 4 we determined whether a prime presented prior to the RSVP stream functioned as a distractor item for T1. When changing the character class of T1 the inhibitory effect disappeared. This finding supports the notion that we are indeed seeing an effect of NP. However, in view of the results of our other experiments we cannot conclude that the inhibition model is able to adequately explain this NP. This

model seems to be based on inhibitory processes on a higher level than those playing a role in the effects observed here.

To summarize, interference found in this study seems to act on a feature extraction level instead of a semantic level as reported in earlier studies. In the current AB task we found uni-modal NP effects on T2 performance in both the auditory and visual modalities, but no cross-modal or semantic influences. For this effect to occur prime and target need to be identical, prime and target have to be presented within the same modality, and the target has to be accompanied by another target from the same character class.

Samenvatting

Audiovisuele aandacht in de ruimte

Theoretisch kader

Tijdens het uitvoeren van een visuele taak, zoals het lezen van een boek, kan onze aandacht getrokken worden naar de locatie van een onverwacht geluid. In zo'n situatie grijpt het irrelevante geluid onze aandacht en leidt ons af van de visuele taak. Het proces dat ten grondslag ligt aan dit fenomeen noemen we *crossmodale aandacht*. Ook kunnen visuele en auditieve gebeurtenissen zoals bewegende lippen en spraak integreren tot één coherente gebeurtenis. Dit proces wordt *multisensorische integratie* genoemd. Zowel multisensorische integratie als crossmodale aandacht hebben grote invloed op hoe we de wereld waarnemen. In deze moderne tijd worden we overspoeld met audiovisuele informatie, ook in situaties waarbij het belangrijk is om gefocuseerd te blijven en geen fouten te maken zoals bijvoorbeeld tijdens het autorijden. In andere situaties kunnen we grote voordelen ervaren wanneer visuele en auditieve informatie samenvallen, bijvoorbeeld wanneer we de mondbewegingen kunnen zien van iemand die spreekt. Het is daarom belangrijk om de onderliggende processen van multisensorische integratie and crossmodale aandacht te begrijpen en te onderzoeken hoe deze processen onze prestaties beïnvloeden.

De literatuur aangaande multisensorische integratie en crossmodale aandacht wordt besproken in Hoofdstuk 1. Het eerste deel van deze introductie beschrijft de temporele (tijd) en spatiële (ruimte) restricties waaronder multisensorische integratie voorkomt en in welke delen van de hersenen multisensorische integratie plaatsvindt. De studies die besproken worden laten zien dat, afhankelijk van het soort en de complexiteit van de audiovisuele informatie, multisensorische integratie plaatsvindt in verschillende hersengebieden. Dit deel wordt gevolgd door een overzicht van studies die specifiek ingaan op crossmodale aandacht tussen de auditieve en visuele modaliteit. Dit is het onderwerp van studies beschreven in Hoofdstuk 2 en Hoofdstuk 3 van dit proefschrift. Aandacht gericht op een taak in één modaliteit kan getrokken worden door informatie afkomstig van een andere

modaliteit en vice versa. Dit suggereert dat crossmodale aandacht kan plaatsvinden in hersengebieden die niet specifiek bedoeld zijn voor het verwerken van of auditieve of visuele informatie. Ook suggereren een aantal studies dat vroege sensorisch specifieke gebieden betrokken zijn bij crossmodale aandacht. Hoofdstuk 1 gaat in op het idee dat multisensorische integratie en crossmodale aandacht soms onafhankelijk zijn maar in andere omstandigheden interacteren. Dit is het thema van de studie beschreven in hoofdstuk 4. Om dit onderwerp beter te belichten worden er verschillende theorieën besproken aangaande het niveau waarop multisensorische interacties plaatsvinden. Het laatste deel van de introductie is toegespitst op de vraag of audiovisuele interacties en crossmodale aandacht in het bijzonder *automatische* processen zijn wat zou suggereren dat ze niet beïnvloed worden door cognitieve belasting. Deze vraag wordt onderzocht in de studies beschreven in Hoofdstukken 2 en 3. De literatuur beschreven in Hoofdstuk 1 laat zien dat wanneer een geluid en een visuele stimulus tegelijkertijd worden aangeboden, de visuele stimulus niet noodzakelijk de aandacht trekt. Gegeven deze bevinding wordt crossmodale aandacht niet beschouwd als een proces dat volledig automatisch verloopt.

Auditory capture

Als visuele aandacht getrokken wordt naar de locatie van een geluid zal het verwerken van opvolgende visuele informatie op die zelfde locatie versneld worden. Dit proces wordt *auditory capture* genoemd en het representeert één vorm van crossmodale aandacht. In de studies gepresenteerd in hoofdstuk 2 en 3 laten we auditory capture zien door middel van een cueing taak. In deze taak dienden proefpersonen een witte stip op te merken die links of rechts van het midden werd aangeboden, iets boven of onder het horizontale midden. Onafhankelijk van de locatie gaven proefpersonen aan of deze stip boven of onder het horizontale midden verscheen. Kort voor het verschijnen van de stip werd er links of rechts van het scherm een geluid aangeboden. In de helft van de gevallen werd het geluid en de stip aan de zelfde kant aangeboden en in de andere helft aan verschillende kanten. Tijdens de test wisten proefpersonen niet waar de stip zou verschijnen en

werden ze gevraagd de geluidscue te negeren. Proefpersonen reageerden sneller op stippen als het geluid aan dezelfde kant werd aangeboden en langzamer als het geluid aan de tegenovergestelde kant werd aangeboden. Dit verschil in reactietijd wordt het crossmodale cueing effect genoemd en is een maat voor auditory capture.

De vraag of auditory capture verdwijnt als we weten aan welke kant de stip gaat verschijnen wordt behandeld in Hoofdstuk 2. In deze studie deden proefpersonen dezelfde taak als hierboven beschreven. Echter, deze keer verscheen er voor het geluid en de stip een pijl in het midden van het scherm. Deze pijl wees de kant aan waar de stip zou verschijnen. Hoewel de proefpersonen niet wisten of de stip boven of onder het horizontale midden ging verschijnen, konden ze al wel hun visuele aandacht richten op de kant waar de stip zou gaan verschijnen. In deze test lieten proefpersonen nog steeds een cueing effect voor geluid zien hoewel de pijl de juiste locatie van de stip aangaf. Daaruit concluderen we dat mensen niet in staat zijn auditory capture te onderdrukken, zelfs als ze weten op welke locatie relevante visuele informatie wordt aangeboden.

De vraag of het auditief trekken van aandacht verdwijnt als er iets visueels wordt gepresenteerd op het zelfde moment als het geluid, wordt onderzocht in Hoofdstuk 3. In het algemeen is bekend dat visuele gebeurtenissen de aandacht trekken. In deze studie onderzochten we of het auditief trekken van aandacht ook optreedt als een visuele en auditieve gebeurtenis moeten concurreren om aandacht. Proefpersonen voerden een cueing taak uit waarin de stip werd voorafgegaan door zowel een visuele als auditieve cue. Wanneer beide cues werden aangeboden op kansniveau werd zowel het visuele als het auditieve trekken van aandacht waargenomen. Echter, wanneer de validiteit van de visuele cue werd verhoogd tot 80%, werd de aandacht alleen nog visueel getrokken. Uit deze resultaten concluderen we dat het trekken van aandacht door een auditieve stimulus kan worden onderdrukt wanneer de visuele gebeurtenis tegelijkertijd met de auditieve gebeurtenis wordt aangeboden en wanneer de visuele gebeurtenis voorspellend is met betrekking tot de locatie waar het target gaat verschijnen. Het auditief grijpen van visuele aandacht wordt daarom niet beschouwd als een volledig automatisch

proces. Dit betekent dat we nu weten hoe we storende effecten van geluid op het visueel presteren kunnen minimaliseren.

Multisensorische integratie

Er wordt gesuggereerd dat voordelen van multisensorische integratie het resultaat zijn van het efficiënter verplaatsen van onze spatiële aandacht. Andere studies daarentegen suggereren dat multisensorische integratie en spatiële aandacht onafhankelijk van elkaar opereren. In hoofdstuk 4 onderzoeken we of multisensorische integratie ons waarneming kan beïnvloeden in condities waarbij spatiële aandacht geen rol speelt. Hiervoor presenteerden we een visuele target in het midden van het visuele veld samen met een niet lokaliseerbaar geluid. Het eerste experiment liet zien dat geluid de waarnemingsdrempel van visueel contrast verlaagt. In het tweede experiment voerden proefpersonen een detectie taak uit om verschuiving van onze waarnemingscriteria als alternatieve verklaring uit te sluiten. De resultaten lieten een verhoogde gevoeligheid voor het detecteren van het visuele target zien wanneer het werd aangeboden met een geluid relatief ten opzichte van geen geluid conditie. Hieruit concluderen we dat multisensorische integratie de waarneming van visuele gebeurtenissen verbetert onafhankelijk van spatiële aandacht.

Attentional blink

Een *Attentional blink* vindt plaats als mensen twee target items (bijv. woorden of symbolen) moeten beschrijven die snel opeenvolgend op de zelfde locatie worden aangeboden. Mensen zijn goed in het identificeren van de eerste target maar zijn vaak niet in staat de tweede target te identificeren wanneer deze kort na de eerste target wordt aangeboden. Hoofdstuk 5 beschrijft experimenten waarin proefpersonen een attentional blink taak uitvoerden die cijfers als targets en letters als distractoren bevatten binnen het visuele en auditieve domein. Voorafgaand aan de sequentieel aangeboden targets en distractoren, werd er auditief of visueel geprimed door middel van een cijfer dat in 50 procent van de gevallen gelijk was aan het tweede target. De resultaten laten niet alleen een attentional blink zien maar

ook een algemene daling van de waarneming van het tweede target in de conditie waar de prime gelijk is aan T2. Er werd geen crossmodale priming waargenomen wat suggereert dat het onderdrukkende priming effect modaliteit specifiek is. Er wordt vanuit gegaan dat deze bevindingen een speciale vorm van negatieve priming representeren, die opereert op een vroeg verwerkingsniveau.

References

- Akyürek, E. G., & Hommel, B. (2005). Short-term memory and the attentional blink: Capacity versus content. *Memory & Cognition*, 33(4), 654-663.
- 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.
- Arnell, K. M., & Jenkins, R. (2004). Revisiting within-modality and cross-modality attentional blinks: Effects of target-distractor similarity. *Perception & Psychophysics*, 66(7), 1147-1161.
- Arnell, K. M., & Jolicoeur, P. (1999). The attentional blink across stimulus modalities: Evidence for central processing limitations. *Journal of Experimental Psychology-Human Perception and Performance*, 25(3), 630-648.
- Baroncohen, S., Wyke, M. A., & Binnie, C. (1987). Hearing words and seeing colors - an experimental investigation of a case of synesthesia. *Perception*, 16(6), 761-767.
- 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., & Edelstein, B. A. (1971). Effects of some variations in auditory input upon visual choice reaction time. *Journal of Experimental Psychology*, 87, 242-247.
- Blauert, J. (1997). *Spatial hearing: Psychophysics of human sound localization*. Cambridge, MA: MIT Press.
- 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.
- Broadbent, D. E., & Broadbent, M. H. P. (1987). From detection to identification - response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, 42(2), 105-113.
- 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.
- Buchner, A., Zabal, A., & Mayr, S. (2003). Auditory, visual, and cross-modal negative priming. *Psychonomic Bulletin & Review*, 10(4), 917-923.
- 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., 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.
- 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.
- Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, 7(3), 308-313.
- 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.
- Chun, M. M., & Potter, M. C. (1995). A 2-stage model for multiple-target detection in rapid serial visual presentation. *Journal of Experimental Psychology-Human Perception and Performance*, 21(1), 109-127.
- 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.
- Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, 22, 319-349.
- 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.
- 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.

- DeSchepper, B., & Treisman, A. (1996). Visual memory for novel shapes: Implicit coding without attention. *Journal of Experimental Psychology-Learning Memory and Cognition*, 22(1), 27-47.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual-attention. *Annual Review of Neuroscience*, 18, 193-222.
- 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.
- Eckstein, M. P., Shimozaki, S. S., & Abbey, C. K. (2002). The footprints of visual attention in the Posner cueing paradigm revealed by classification images. *Journal of Vision*, 2(1), 25-45.
- 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.
- Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology-Human Perception and Performance*, 18(4), 1030-1044.
- 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.
- Graf, P., Shimamura, A. P., & Squire, L. R. (1985). Priming across modalities and priming across category levels - Extending the domain of preserved function in amnesia. *Journal of Experimental Psychology-Learning Memory and Cognition*, 11(2), 386-396.
- 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.
- Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. *Psychological Science*, *12*(5), 360-365.
- Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature Neuroscience*, *3*(3), 284-291.
- Houghton, G., Tipper, S. P., Weaver, B., & Shore, D. I. (1996). Inhibition and interference in selective attention: Some tests of a neural network model. *Visual Cognition*, *3*(2), 119-164.
- 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.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology-General*, *110*(3), 306-340.
- Jolicoeur, P., & Dell' Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology*, *36*(2), 138-202.
- 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., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological Bulletin*, *96*(1), 29-44.
- 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). Competition between auditory and visual spatial cues during visual task performance. *Experimental Brain Research*, *195*(4), 593-602.
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (in prep). Multisensory integration is more than directing attention.

- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (in press). Auditory and visual capture during focused visual attention. *Journal of Experimental Psychology-Human Perception and Performance*.
- Koelewijn, T., Van der Burg, E., Bronkhorst, A., & Theeuwes, J. (2008). Priming T2 in a visual and auditory attentional blink task. *Perception & Psychophysics*, 70(4), 658-666.
- 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.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence-intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1(4), 476-490.
- 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 Neurosciences*, 28(5), 264-271.
- Martens, S., Munneke, J., Smid, H., & Johnson, A. (2006). Quick minds don't blink: Electrophysiological correlates of individual differences in attentional selection. *Journal of Cognitive Neuroscience*, 18(9), 1423-1438.
- Martinez, A., Anllo-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.
- May, C. P., Kane, M. J., & Hasher, L. (1995). Determinants of negative priming. *Psychological Bulletin*, *118*(1), 35-54.
- 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., Teder-Salejarvi, 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.
- McDonald, J. J., Teder-Salejarvi, W. A., Heraldez, D., & Hillyard, S. A. (2001). Electrophysiological evidence for the "missing link" in crossmodal attention. *Canadian Journal of Experimental Psychology*, *55*(2), 141-149.
- McDonald, J. J., Teder-Salejarvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting of sound improves visual perception. *Nature*, *407*, 906-908.
- McDonald, J. J., Teder-Salejarvi, W. A., & Ward, L. M. (2001). Multisensory integration and crossmodal attention effects in the human brain. *Science*, *292*(5523), -.
- McDonald, J. J., & Ward, L. M. (2000). Involuntary listening aids seeing: Evidence from human electrophysiology. *Psychological Science*, *11*(2), 167-171.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748.
- 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.
- Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology*, *75*(5), 1843-1857.
- 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.
- Mondor, T. A., & Amirault, K. J. (1998). Effect of same- and different-modality spatial cues on auditory and visual target identification. *Journal of Experimental Psychology-Human Perception and Performance*, *24*(3), 745-755.
- 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.

- Muller, 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.
- Neill, W. T., & Mathis, K. M. (1998). Transfer-inappropriate processing: Negative priming and related phenomena. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 38, pp. 1-44). San Diego, CA: Academic Press.
- Neill, W. T., Valdes, L. A., Terry, K. M., & Gorfain, D. S. (1992). Persistence of negative priming 2. Evidence for episodic trace retrieval. *Journal of Experimental Psychology-Learning Memory and Cognition*, 18(5), 993-1000.
- 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.
- Nieuwenstein, M. R., Johnson, A., Kanai, R., & Martens, S. (2007). Cross-task repetition amnesia: Impaired recall of RSVP targets held in memory for a secondary task. *Acta Psychologica*, 125(3), 319-333.
- Noesselt, T., Bergmann, D., Hake, M., Heinze, H. J., & Fendrich, R. (2008). Sound increases the saliency of visual events. *Brain Research*, 1220, 157-163.
- Palmer, J. (1994). Set-Size effects in visual-search - the effect of attention is independent of the stimulus for simple tasks. *Vision Research*, 34(13), 1703-1721.
- Peelen, M. V., Heslenfeld, D. J., & Theeuwes, J. (2004). Endogenous and exogenous attention shifts are mediated by the same large-scale neural network. *Neuroimage*, 22(2), 822-830.
- 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.
- Potter, M. C., Chun, M. M., Banks, B. S., & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology-Learning Memory and Cognition*, 24(4), 979-992.
- Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task - an attentional blink. *Journal of Experimental Psychology-Human Perception and Performance*, 18(3), 849-860.

- 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.
- Rothermund, K., Wentura, D., & De Houwer, J. (2005). Retrieval of incidental stimulus-response associations as a source of negative priming. *Journal of Experimental Psychology-Learning Memory and Cognition*, 31(3), 482-495.
- Rouw, R., & Scholte, H. S. (2007). Increased structural connectivity in grapheme-color synesthesia. *Nature Neuroscience*, 10(6), 792-797.
- 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., Finioia, P., Raffone, A., Belardinelli, M. O., & Spence, C. (2008). Perceptual load affects exogenous spatial orienting while working memory load does not. *Experimental Brain Research*, 184(3), 371-382.
- Santangelo, V., Ho, C., & Spence, C. (2008). Capturing spatial attention with multisensory cues. *Psychonomic Bulletin & Review*, 15(2), 398-403.
- 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., 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.
- 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.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions - What you see is what you hear. *Nature*, 408(6814), 788-788.
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, 14(1), 147-152.
- Shams, L., Kamitani, Y., Thompson, S., & Shimojo, S. (2001). Sound alters visual evoked potentials in humans. *Neuroreport*, 12(17), 3849-3852.
- 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., & 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. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology-Human Perception and Performance*, 22(4), 1005-1030.
- 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., 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.
- Spence, C., & McGlone, F. P. (2001). Reflexive spatial orienting of tactile attention. *Experimental Brain Research*, 141(3), 324-330.
- 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.
- 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., & 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., 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*. Oxford: Oxford university press.
- Sumbly, 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., 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.
- Talsma, D., & Woldorff, M. G. (2005a). Methods for the estimation and removal of artifacts and overlap in ERP data. In T. Handy (Ed.), *Event-Related Potentials: A Methods Handbook* (pp. 115-148). Cambridge, MA: MIT Press.
- Talsma, D., & Woldorff, M. G. (2005b). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*(7), 1098-1114.
- 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., Belopolsky, A., & Olivers, C. N. L. (in press). Interaction between working memory, attention and eye movements. *Acta Psychologica*.
- 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.
- Thurlow, W. R., & Jack, C. E. (1973). Certain determinants of ventriloquism effect. *Perceptual and Motor Skills*, *36*(3), 1171-1184.
- Tipper, S. P. (1985). The negative priming effect - inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, *37*(4), 571-590.
- Tipper, S. P., Weaver, B., Bastedo, J., Cameron, S., & Brehaut, J. C. (1991). Inhibitory mechanisms of attention in identification and localization tasks - Time Course and disruption. *Journal of Experimental Psychology-Learning Memory and Cognition*, *17*(4), 681-692.
- 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., 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 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 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.
- Verde, M. F., MacMillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: The accuracy, precision, and robustness of d' , $A(z)$, and A' . *Perception & Psychophysics*, 68(4), 643-654.
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428(6984), 748-751.
- Vroomen, J., Bertelson, P., & de Gelder, B. (2001a). Directing spatial attention towards the illusory location of a ventriloquized sound. *Acta Psychologica*, 108(1), 21-33.
- Vroomen, J., Bertelson, P., & de Gelder, B. (2001b). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, 63(4), 651-659.
- 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.
- 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.
- Wood, T. J., & Milliken, B. (1998). Negative priming without ignoring. *Psychonomic Bulletin & Review*, 5(3), 470-475.
- 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.
- Yee, P. L. (1991). Semantic inhibition of ignored words during a figure classification task. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, 43(1), 127-153.
- 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.

Dankwoord

Hoewel ik dit proefschrift toch echt op een aantal correcties na zelf geschreven heb, was er geen woord op papier gekomen zonder de volgende mensen, die ik hiervoor dan ook hartelijk wil bedanken.

Ik wil graag mijn promotor Adelbert Bronkhorst bedanken voor zijn enthousiasme en steun tijdens het uitvoeren van het project en de wijze les dat onderzoek per definitie leuk is.

Ik wil in het bijzonder ook mijn co-promotor Jan Theeuwes bedanken voor alles wat hij mij geleerd heeft de afgelopen jaren en de wijze les dat onderzoek per definitie een kwestie van doen is.

De leden van de leescommissie voor het kritisch lezen van mijn proefschrift.

Erik van der Burg waarmee ik het project deelde en veel jolijt heb gehad tijdens onze schoolreisjes. Mijn kamergenoten Stefan van der Stigchel, Frank Agter, Yair Pinto, Christel de Vue, Brian White en Clayton Hickey voor hun geduld. Jaap Munneke en Manon Mulckhuysse omdat ze zo'n gezellige inloop kamer hadden. Sander Los (jawel) en Martijn Meeter voor hun hulp bij statistische kwesties. Mijn muzikale collega's Dirk Heslenfeld, Isabel Dombrowe, Artem Belopolsky en Jaap Munneke voor al hun vingerbrekende capriolen. Mieke Donk en Chris Olivers voor hun gemakkelijk verhuis en verbouwingsverhalen tijdens de lunch. Daniel Schreij en Sebastiaan Mathôt voor hun jeugdige enthousiasme voor onderzoek. Wieske van Zoest en Clayton Hickey voor hun gastvrijheid en al het lekkere eten. Richard Godijn voor de gezellige wandelingen naar de trein. Clayton Hickey voor het helpen bij mijn Engels. Mijn stagiair Guido Peek voor zijn goede werk. Barbara Goudriaan voor het op de rails houden van alles. Jarik den Hartog, Rene Nieuwboer en Paul Groot voor al hun hulp bij ongelukken. Wieske van Zoest voor het helpen bij mijn beroepskeuze. Yair Pinto voor alle potjes checkers. Durk Talsma en Mark Nieuwenstein voor al hun kennis over foute rock en gitaren. Leroy Soesman voor zijn goocheltrucs. Marieke van der Hoeven voor de leuke congressen en samenwerking. Ingmar Stel voor al zijn geduld en hulp bij het nog niet helemaal geslaagde rijsimulator project.

Mijn paranimfen Michiel de Voogt en Clayton Hickey voor al het zware werk.

Maarten Verweij en Wietse Bloemzaad voor van alles en nog wat.

Al mijn leuke vrienden en vriendinnen in binnen en buitenland omdat ze bestaan.

Mijn ouders, grootouders, Maaïke en Jasper voor hun steun.

Curriculum Vitae

Thomas Koelewijn was born on April 12, 1979 in Arnhem, the Netherlands. In 2003 he received his MS in experimental psychology on the topic of music perception. He continued his studies with an additional two-year Research Master in Cognitive Neuroscience in Nijmegen. He did an internship at the F.C. Donders Brain imaging centre in Nijmegen. Within the field of perception and action he performed research on the human Mirror Neuron system by means of EEG and MEG recordings. They looked at modulation of primary motor areas by means of correctness and laterality of the observed movement. After his graduation in 2005 he started working as a PhD student on the topic of crossmodal spatial attention in the lab of Jan Theeuwes. His main focus and interest is now on auditory and visual attention both in the spatial and temporal domain, in particular crossmodal interactions.

Publications

Journal articles

- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (in prep). Multisensory integration is more than directing attention.
- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (in press). Auditory and visual capture during focused visual attention. *Journal of Experimental Psychology: Human Perception and Performance*.
- 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.
- Koelewijn, T., Van der Burg, E., Bronkhorst, A., & Theeuwes, J. (2008). Priming T2 in a Visual and Auditory Attentional Blink Task. *Perception & Psychophysics*, 70(4), 658-666.
- Koelewijn, T., van Schie, Hein T., Bekkering, H., Oostenveld, R., & Jensen, O. (2008). Motor-cortical beta oscillations are modulated by correctness of observed action. *Neuroimage*, 40(2), 767-775.
- van Schie, H. T., Koelewijn, T., Jensen, O., Oostenveld, R., Maris, E., & Bekkering, H. (2008). Evidence for fast, low level motor resonance to action observation: An MEG study. *Social Neuroscience*, 3(3&4), 213-228.
- Van der Burg, E., Olivers, C.N.L., Bronkhorst, A.W., Koelewijn, T., & Theeuwes, J. (2007). The absence of an auditory-visual attentional blink is not due to echoic memory. *Perception & Psychophysics*, 69(7), 1230-1241.

Proceedings

- Koelewijn, T., Van Der Hoeven, M., & Bronkhorst, A. (2009). Auditory capture in an auditory and visual cueing task. *NAG/DAGA, 2009*, 1525-1527.

Abstracts

- Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2008). Attentional capture by sound is task dependent. *Perception*, 37(Supplement), 12-12.
- Bronkhorst, A.W., Koelewijn, T., & Van der Burg, E. (2008). Spatiotemporal interactions between audition and vision. *The Journal of the Acoustical Society of America*, 123(5), 3294-3294.
- Bronkhorst, A., Van der Hoeven, M., Theeuwes, J., Van der Burg, E., & Koelewijn, T. (2007). Factors influencing the auditory and cross-modal attentional blink. *The Journal of the Acoustical Society of America*, 120(5), 3128-3128.